

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

Lexikos 23

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

Lexikos 23

Redakteur

Editor

J.C.M.D. du Plessis

Resensieredakteur

Review Editor

T. Harteveld



African Association for Lexicography

AFRILEX-REEKS 23:2013

AFRILEX SERIES 23:2013



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 © 2013 deur die uitgewer
Alle regte streng voorbehou
Eerste uitgawe 2013

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 978-0-9814434-7-8
ISSN 1684-4904

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

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

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

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

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)
Dr. V. Kukanda (Gaboen/Gabon)
Prof. W. Martin (België en Nederland/Belgium and The Netherlands)
Prof. I.A. Mel'čuk (Kanada/Canada)
Prof. A.M.F.J. Moerdijk (Nederland/The Netherlands)
Prof. M. Schlaefer (Duitsland/Germany)
Dr. J. Tent (Australië/Australia)
Prof. J. Van Keymeulen (België/Belgium)
Prof. P.G.J. van Sterkenburg (Nederland/The Netherlands)
Prof. L.S. Vikør (Noorweë/Norway)
Prof. H.E. Wiegand (Duitsland/Germany)

Redaksiekomitee / Editorial Committee

Dr. M.M. Bagwasi (Botswana)
Dr. H.L. Beyer (Namibië/Namibia)
Prof. W.A.M. Carstens (RSA)
Dr. E. Chabata (Zimbabwe)
Dr. A.R. Chuwa (Tanzanië/Tanzania)
Prof. C.J. Conradie (RSA)
Prof. J.D. Emejulu (Gaboen/Gabon)
Prof. A.E. Feinauer (RSA)
Prof. R. Finlayson (RSA)
Dr. S. Hadebe (Zimbabwe)
Prof. I.M. Kosch (RSA)
Dr. P.A. Louw (RSA)
Mnr. K.J. Mashamaite (RSA)
Prof. P.A. Mavoungou (Gaboen/Gabon)
Dr. J.S. Mdee (Tanzanië/Tanzania)
Prof. B.M. Mini (RSA)
Dr. V.M. Mojela (RSA)
Mnr. M.C. Mphahlele (RSA)
Mnr. M.H. Mpongose (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>J.C.M.D. du Plessis</i>	
'n Woord van AFRILEX	xii
A Few Words from AFRILEX	xiv
<i>V.M. Mojela</i>	
Redaksionele doelstellings	xvi
Editorial Objectives	xvii
Redaktionelle Ziele	xviii
Politique éditoriale	xix

Artikels / Articles

The Afrikaans Orthographic Rules as Guide for Other South African Languages	1
<i>Mariëtta Alberts</i>	
Developing Legal Terminology in African Languages as Aid to the Court Interpreter: A South African Perspective	29
<i>Mariëtta Alberts and Nina Mollema</i>	
The Presentation of Word Formation in General Monolingual Dictionaries	59
<i>Henning Bergenholtz and Rufus H. Gouws</i>	
What is a Lexicographical Database?	77
<i>Henning Bergenholtz and Jesper Skovgård Nielsen</i>	
Le traitement macrostructural de l'emprunt au français dans les dictionnaires locaux au Gabon	88
<i>Edgard Maillard Ella</i>	
Woordeboeke en Dowe gebruikers: huidige probleme en die behoefte aan beter oplossings	113
<i>Hanelle Fourie Blair</i>	

Die leksikografiese aanbieding en behandeling van vaste uitdrukkings <i>Rufus H. Gouws</i>	135
Das lexicographische Beispiel in der Parömiographie. Formen und Funktionen <i>Vida Jesenšek</i>	150
Introducing a New Lexicographical Model: AlphaConceptual+ (and How it Could Be Applied to Dictionaries for Luganda) <i>Deo Kawalya and Gilles-Maurice de Schryver</i>	172
Expectation Levels in Dictionary Consultation and Compilation <i>Inge Kosch</i>	201
New Proposals for the Design of Integrated Online Wine Industry Dictionaries <i>Patrick Leroyer</i>	209
How Dictionary Users Choose Senses in Bilingual Dictionary Entries: An Eye-Tracking Study <i>Robert Lew, Marcin Grzelak and Mateusz Leszkowicz</i>	228
Gabonese French Dictionaries: Survey and Perspectives <i>Paul Achille Mavoungou</i>	255
<i>Birds of a Feather Don't Always Flock Together: User Problems in Identifying Headwords in Online English Learner's Dictionaries</i> <i>Julia Miller</i>	273
A Balanced and Representative Corpus: The Effects of Strict Corpus-based Dictionary Compilation in Sesotho sa Leboa <i>V.M. Mojela</i>	286
Digitizing the Monolingual Lusoga Dictionary: Challenges and Prospects <i>Minah Nabirye and Gilles-Maurice de Schryver</i>	297
Development in Lexicography: From Polyfunctional to Mono-functional Accounting Dictionaries <i>Sandro Nielsen and Pedro A. Fuertes-Olivera</i>	323

IsiXhosa Lexicography: Past, Present and Future <i>Dion Nkomo and Zola Wababa</i>	348
'n Kritiese beskouing van woordeboeke met geamalga- meerde lemmalyste <i>D.J. Prinsloo</i>	371
Old Wisdom: The Highly Relevant Lexicographical Knowledge Obtainable from a Specialized Dictionary from 1774 <i>Sven Tarp</i>	394
Cuban School Dictionaries for First-Language Learners: A Shared Experience <i>Sven Tarp and Leonel Ruiz Miyares</i>	414
Die rol van taalleeronderrig in Afrikaanse woordeboekwerkboeke <i>Michele F. van der Merwe</i>	426
Differences in the Inclusion and Treatment of Terminology in OALD3, OALD4 and OALD8 <i>Marjeta Vrbinc and Alenka Vrbinc</i>	440

Korpusse / Corpora

Moroccorp: tien miljoen woorde uit twee Marokkaans- Nederlandse chatkanale <i>Tom Ruetten en Freek Van de Velde</i>	456
---	-----

Projekte / Projects

Lexicographie et enseignement de langues liées à l'immigration: note sur le contexte sociolinguistique de l'élaboration du dic- tionnaire finnois–lingala–français <i>Kadima Batumona Adi Gilbert</i>	476
Making an Online Dictionary of New Zealand Sign Language <i>Rachel Locker McKee and David McKee</i>	500
Introducing <i>Tlhalosi ya Medi ya Setswana</i> : The Design and Compilation of a Monolingual Setswana Dictionary <i>Thapelo J. Otlogetswe</i>	532

Le dictionnaire de chimie en lingála pour les élèves de Kinshasa <i>Bienvenu Sene Mongaba</i>	548
Leksiko-opnames / Lexicosurveys	
The Role of the Learner's Native Culture in EFL Dictionaries: An Experimental Study <i>Dídac Pujol and Montse Corrius</i>	565
Resensieartikels / Review Articles	
From e-Lexicography to Electronic Lexicography. A Joint Review <i>Valeria Caruso</i>	585
An Analysis of the <i>Oxford Bilingual School Dictionary: Northern Sotho and English</i> (De Schryver 2007) <i>Inge Kosch</i>	611
Resensies / Reviews	
Patrick Hanks. <i>Lexical Analysis: Norms and Exploitations</i> <i>Henri Béjoint</i>	628
Publikasieaankondigings / Publication Announcements	
Voorskrifte aan Skrywers	644
Instructions to Authors	645
Richtlinien für Autoren	646
Instructions aux Auteurs	647

Voorwoord

Lexikos word agtereenvolgens deur proff. E. Taljard, D.J. Prinsloo en R.H. Gouws geredigeer. Prof. Gouws wat vanjaar die redakteur van *Lexikos* sou wees, was vanweë werksdruk nie beskikbaar nie. Alhoewel ek reeds na die publikasie van nommer 20 as redakteur van *Lexikos* uitgetree het, het ek ingestem om *Lexikos* vanjaar weer vir 'n allerlaaste keer te redigeer.

Ek wil graag al die bydraers bedank vir die voorlegging van artikels vir hierdie nommer van *Lexikos*. Baie van hulle het bygedra tot vorige nommers, maar verskeie ander het dit vir die eerste keer gedoen. Hulle almal sal saamstem dat die keurders wat ruimskoots van hulle tyd en ervaring gegee het om die artikels te beoordeel, 'n belangrike rol vervul het met hulle waardevolle opmerkings en voorstelle wat in baie gevalle daartoe gelei het dat dele van artikels heroorweeg en herskryf is. Dit het 'n stimulerende en vernuwende invloed op die gehalte van die artikels gehad wat in hierdie nommer ingesluit is.

Prof. G.-M. de Schryver het uit AFRILEX bedank, maar elkeen kan getuig van die groot bydrae wat hy gelewer het, eers as Raadslid en later as President van AFRILEX. Daarbenewens was hy 'n gereelde medewerker van *Lexikos* waartoe hy vanjaar weer twee medegeskrewe artikels bygedra het. Dr. V.M. Mojela is as President in sy plek verkies. Hy is nie onbekend aan AFRILEX en *Lexikos* nie, aangesien hy vir 'n aansienlike tyd lid van die Raad was en gereeld artikels in *Lexikos* gepubliseer het. *Lexikos* is oortuig dat hy 'n vrugbare termyn as President van AFRILEX sal hê.

Prof. H.M. Batibo het aangedui dat hy uit die Redaksiekomitee wil tree. *Lexikos* wil hom bedank vir sy bydrae deur die jare. Dr. M.M. Bagwasi en prof. T.J. Otlogetswe, albei van Botswana, het ingestem om in die Komitee te dien. Hopelik sal hulle albei 'n aangename verbintenis met *Lexikos* hê.

Na baie jare het ek weer saamgewerk met me. M.H. van der Westhuizen wat die setwerk vaardig en professioneel gedoen het. Soos altyd het me. T. Hartevelt waardevolle tegniese hulp verleen wanneer dit nodig was. Ek wil hulle verseker dat hulle deel aan die voltooiing van hierdie nommer baie waardeer word.

Laastens wil ek hiermee AFRILEX en *Lexikos* alles van die beste vir die toekoms toewens.

J.C.M.D. du Plessis
Redakteur

Foreword

Lexikos is consecutively edited by Profs E. Taljard, D.J. Prinsloo and R.H. Gouws. Prof. Gouws who should have edited *Lexikos* this year, was, owing to pressure of work, unable to do so. Although I had already retired as editor of *Lexikos* after the publication of issue 20, I have agreed to edit *Lexikos* again for the very last time this year.

I would like to thank all the contributors for submitting articles to *Lexikos*. Many of them have contributed to previous issues, but several others have done so for the first time. All of them will agree that the referees, who unstintingly gave of their time and experience to evaluate the articles, fulfilled an important role with their valuable comments and suggestions which in many cases led to rethinking and rewriting parts of articles. This had a stimulating and invigorating influence on the quality of the articles included in this issue.

Prof. G.-M. de Schryver has resigned from AFRILEX, but everyone can attest to the great contribution he made first as Board member and later as President of AFRILEX. In addition to this, he has been a regular collaborator of *Lexikos* to which he has again contributed two co-authored articles this year. Dr V.M. Mojela has been voted President in his stead. He is not unknown to AFRILEX and *Lexikos*, having served on the Board for a considerable time and having regularly contributed articles to *Lexikos*. *Lexikos* is convinced that he will have a fruitful term as President of AFRILEX.

Prof. H.M. Batibo indicated that he wants to retire from the Editorial Committee. *Lexikos* wishes to thank him for his contribution through the years. Dr M.M. Bagwasi and Prof. T.J. Otlogetswe, both from Botswana, have agreed to serve on the Committee. Hopefully both of them will have a pleasant association with *Lexikos*.

After many years, I have again teamed up with Ms M.H. van der Westhuizen who did the typesetting competently and professionally. As always, Ms T. Harteveld gave valuable technical assistance when it was needed. I want to assure them that their part in the completion of this issue is much appreciated.

Lastly I herewith wish AFRILEX and *Lexikos* everything of the best for the future.

J.C.M.D. du Plessis
Editor

'n Woord van AFRILEX

AFRILEX wil dr. Johan du Plessis, een van AFRILEX se drie Erelede, bedank vir sy bereidwilligheid om as redakteur van vanjaar se nommer van *Lexikos* op te tree, alhoewel hy reeds twee jaar gelede uitgetree het. 'n Ander Erelid van AFRILEX, prof. Rufus Gouws, 'n hoogs gerespekteerde leksikograaf en metaleksikograaf, sal *Lexikos* 24 volgende jaar redigeer. AFRILEX wil ook proff. Elsabé Taljard en Danie Prinsloo, twee veteraan-AFRILEX-lede en leksikografiese deskundiges, vooruit bedank wat die redigering van *Lexikos* vanaf 2015 van prof. Gouws sal oorneem. Na dr. Du Plessis se uittrede in 2010, het proff. Taljard en Prinsloo alreeds hulle deskundigheid getoon deur die redigering van twee vorige nommers van *Lexikos* in 2011 en 2012 onderskeidelik. Dit bewys dat *Lexikos* in goeie hande is, en dat sy toekoms verseker is.

As mondstuk vir die leksikografie, nasionaal en internasionaal, het *Lexikos* sedert sy totstandkoming 'n uitstekende platform geskep vir leksikograwe en metaleksikograwe, veral vir AFRILEX-lede, om hulle deskundigheid te toon in vakkundige navorsing. Hierdie tydskrif het baie bygedra tot die ontwikkeling van die leksikografie op die vasteland van Afrika waar daar voorheen min of niks was nie. Vanjaar se nommer van *Lexikos* toon weer vooruitgang wat betref die kwantiteit en die kwaliteit van die gepubliseerde artikels. Soos uitgespreek in stellings deur vorige AFRILEX-presidente, veral prof. Gilles-Maurice de Schryver, het *Lexikos* kwantitatief en kwalitatief van krag tot krag gegaan met die verskyning van elke nommer deur doeltreffend met ander internasionale tydskrifte elders in die wêreld mee te ding. Die groei kan toegeskryf word aan die werk van sy redakteurs, veral dr. Johan du Plessis wat as redakteur van 15 nommers insluitende dié van vanjaar opgetree het. Sy handhawing van gesonde verhoudinge met *Lexikos* se skrywers en aanbieders van referate by AFRILEX-kongresse is een van die hoofredes wat tot die groei van die tydskrif bygedra het. Na elke AFRILEX-kongres het dr. Du Plessis in verbinding gebly met die aanbieders deur hulle uit te nooi om hul referate voor te lê vir moontlike publikasie in *Lexikos*. Dit het baie nuwe leksikograwe en metaleksikograwe wat voorheen nooit enige artikels gepubliseer het nie, aangemoedig om hul referate in te stuur wat bewys dat 'navorsingspotensialiteit in onontgonne en onaangeraakte gebiede lê'. Op hierdie manier het *Lexikos* baie AFRILEX-lede en potensiele metaleksikograwe ontwikkel om welbekende navorsers op die Afrikakontinent te word.

As President van AFRILEX en lid van die personeel van die Noord-Sotho-Nasionale Leksikografie-eenheid, wil ek afsluit deur hierdie geleentheid te gebruik om 'n beroep te doen op AFRILEX-lede, veral ons kollegas in die verskillende Nasionale Lexikografie-eenhede, om voordeel te trek uit die platform geskep deur *Lexikos* vir die bekendstelling van die kennis wat ons opgedoen het en die uitdagings wat ons teëkom in ons daaglikse leksikografiese werk in

die verskillende kantore. Ons moet weet dat *Lexikos* aan elkeen van ons behoort en beskou moet word as 'n platform vir elke leksikograaf om sy/haar navorsingsdeskundigheid te toon. Indien moontlik, behoort elke referaat wat by 'n AFRILEX-kongres aangebied word, voorgelê te word vir moontlike publikasie in hierdie tydskrif.

M.V. Mojela
President: AFRILEX

A Few Words from AFRILEX

AFRILEX wants to thank Dr Johan du Plessis, one of AFRILEX's three Honorary Members, for his willingness to act as editor of this year's issue of *Lexikos*, although he had already retired two years ago. Another Honorary Member of AFRILEX, Prof. Rufus Gouws, a highly respected lexicographer and metalexicographer, will edit *Lexikos* 24 next year. AFRILEX also wants to thank in advance Profs Elsabé Taljard and Danie Prinsloo, two veteran AFRILEX members and lexicographical experts, who will take over the editing of *Lexikos* from Prof. Gouws from 2015 onwards. Profs Taljard and Prinsloo have already demonstrated their expertise in editing two previous issues of *Lexikos* in 2011 and 2012 respectively after Dr Du Plessis's retirement in 2010. This shows that *Lexikos* is in good hands, and that its future is assured.

As a mouthpiece for lexicography, nationally and internationally, *Lexikos* has since its inception created an excellent forum for lexicographers and metalexicographers, especially for AFRILEX members, to demonstrate their expertise in scholarly research. This journal has contributed much to the development of lexicography on the African continent where there was little or nothing previously. This year's issue of *Lexikos* again shows progress with regard to the quantity and quality of the published articles. As expressed in statements by previous AFRILEX presidents, especially Prof. Gilles-Maurice de Schryver, *Lexikos* has grown from strength to strength, quantitatively and qualitatively, with the appearance of each issue, competing effectively with other international journals published elsewhere in the world. This growth can be ascribed to the work of its editors, especially Dr Johan du Plessis who acted as editor for 15 issues, including that of this year. His maintenance of healthy relationships with *Lexikos* authors and presenters of papers at AFRILEX conferences is one of the major reasons which led to the growth of the journal. After every AFRILEX conference, Dr Du Plessis would keep in touch with the presenters, inviting them to submit their papers for possible publication in *Lexikos*. This encouraged many new lexicographers and metalexicographers who had never published any articles before to send in their papers, thereby proving that 'research potentiality lies in unexploited and untapped grounds'. In this way, *Lexikos* developed many AFRILEX members and potential metalexicographers into becoming well-known researchers on the African continent.

As President of AFRILEX and member of staff of the Northern Sotho Lexicography Unit, I want to conclude by using this opportunity to call upon our AFRILEX members, especially our colleagues in the various National Lexicography Units, to take advantage of the forum created by *Lexikos* to make known the knowledge we have gained through and the challenges we are facing in our daily lexicographical work in our various offices. We should know that *Lexikos*

belongs to every one of us, and should be regarded as a forum for every lexicographer to demonstrate his/her research expertise. If possible, every paper presented at an AFRILEX conference should be submitted for possible publication in this journal.

M.V. 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.

Redaktionelle Ziele

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

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

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

Gemäß den Zielsetzungen der AFRILEX-Serie werden:

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

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

Politique éditoriale

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

Les objectifs de la collection AFRILEX sont de :

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

Pour atteindre ces objectifs, la collection AFRILEX

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

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

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

The Afrikaans Orthographic Rules as Guide for Other South African Languages*

Mariëtta Alberts, *Research Unit for Languages and Literature in SA Context, Northwest University, Potchefstroom Campus, Potchefstroom, South Africa (malberts@lantic.net)*

Abstract: The spelling and orthographic rules of a language are very important for compilers of general and technical dictionaries. When compiling a dictionary, the lexicographers and terminographers should adhere to these rules. The word-forming principles of a language form part of these rules, and new terms can only be coined in a given language if the spelling and orthographic rules of the language are followed.

In this article work on the rules of Afrikaans spelling and orthography, by and large, is reported. It is hoped that some of the lessons learned in the process could serve as guidelines for parallel processes of the standardisation of the spelling and orthographies of the other South African languages.

With the establishment of the National Language Body for Afrikaans (NLBA) of the Pan South African Language Board (PanSALB), it was decided that the members of the Taalkommissie (Language Commission) of the Suid-Afrikaanse Akademie vir Wetenskap en Kuns (SAAWK) (South African Academy for Science and the Arts) would become members of the NLBA. These members then became the Technical Committee for Standardisation (TC Standardisation). Since its establishment, the members of the Taalkommissie cum TC for Standardisation (Commission) continued with their work on the standardisation of the spelling and orthographic rules for Afrikaans. Along with work on spelling and orthography, the long-term objectives set by the Commission are, *inter alia*, the conversion and adaptation of the *Afrikaanse Woordelys en Spelreëls* (AWS) to an electronic format to be made available in an electronic version (e-book) and online, as well as a standard grammar for Afrikaans to be compiled for international access on the Internet.

The AWS is a resource compiled by the Commission to assist users of Afrikaans in writing the standard variety of the language. The AWS explains the ground rules of the Afrikaans spelling and orthographic conventions. The basic rules are provided in simplified language.

No language is static, and there are always language changes to be taken into account. The vocabulary and pronunciation of a dynamic language could change over a period of time. These changes should be reflected in the spelling and orthographic system of the language.

The article addresses issues such as the front matter of the AWS, spelling and orthographic principles and rules, and the back matter (i.e. a list of abbreviations, a list of international place

* This article is based on a paper "The Work on the Afrikaans Spelling and Orthography Rules as a Model for the Other Official Languages", presented by the author at the Eighteenth Annual International Conference of the African Association for Lexicography (AFRILEX), which was held at the Nelson Mandela Metropolitan University, Port Elizabeth, South Africa, 2–5 July 2013.

names, transliteration table, etc.). Although the *AWS* is not a dictionary, it also contains a word list. Words which present spelling problems, neologisms and items requiring recognition as part of the Afrikaans vocabulary are some of the categories considered for inclusion in this list.

In 2004 the NLBs of the other nine official African languages started with the revision of the spelling and orthographic rules of these languages. The first editions were published by PanSALB in 2008. No revisions have been compiled since. The *AWS* could serve as example of what could be done for the other African languages and even the Khoe and San languages.

Keywords: ALPHABET, LANGUAGE CHANGE, LEXICOGRAPHER, LEXICOGRAPHY, ORTHOGRAPHY, SIMPLIFIED LANGUAGE, SPELLING, STANDARDISATION, TERMINOGRAPHERS, TERMINOGRAPHY, TRANSLITERATION

Opsomming: Die Afrikaanse spel- en skryfreëls as voorbeeld vir ander Suid-Afrikaanse tale. Die spel- en skryfreëls van 'n taal is baie belangrik vir die samestellers van algemene en vakwoordeboeke. Wanneer 'n woordeboek saamgestel word, moet die leksikografe en terminografe by hierdie reëls hou. Die woordvormingsbeginsels van 'n taal vorm deel van hierdie reëls en nuwe terme kan slegs in 'n gegewe taal geskep word as die spel- en skryfreëls van die taal gevolg word.

In hierdie artikel word van werk aan die Afrikaanse spel- en skryfreëls in breë trekke verslag gedoen. Hopelik kan sommige van die lesse wat in dié proses geleer is as riglyne dien vir parallelle prosesse van die standaardisering van die spel- en skryfreëls van die ander Suid-Afrikaanse tale.

Met die stigting van die Nasionale Taalliggaam vir Afrikaans (NTLA) van die Pan- Suid-Afrikaanse Taalraad (PanSAT), is daar besluit dat die lede van die Taalkommissie van die Suid-Afrikaanse Akademie vir Wetenskap en Kuns (SAAWK) lede sou word van die NTLA. Hierdie lede het toe die Tegnieese Komitee Standaardisering (TK: Standaardisering) geword. Sedert sy totstandkoming het die lede van die Taalkommissie cum TK: Standaardisering (Kommissie) voortgegaan met hul werk aan die standaardisering van die spel- en skryfreëls van Afrikaans. Saam met die werk aan spel- en skryfreëls is die langtermyn doelwitte wat deur die Kommissie gestel is, onder andere om die *Afrikaanse Woordelys en Spelreëls (AWS)* te wysig en aan te pas om dit in 'n elektroniese formaat en aanlyn beskikbaar te stel en om 'n standaardgrammatika vir Afrikaans saam te stel wat internasionaal op die Internet toeganklik kan wees.

Die *AWS* is 'n bron wat deur die Kommissie saamgestel word om gebruikers van Afrikaans te help wanneer hulle die standaardvariëteit van die taal gebruik. Die *AWS* verduidelik die grondreëls van die Afrikaanse spel- en skryfkonvensies. Die basiese reëls is in vereenvoudigde taal geskryf.

Geen taal is staties nie en taalverandering moet gedurig in gedagte gehou word. Die woorde-skat en uitspraak van 'n dinamiese taal kan oor 'n tydperk verander. Hierdie veranderinge moet in die spel- en skryfsisteem van die taal weerspieël word.

Hierdie artikel behandel aspekte soos die voorwerk van die *AWS*, spel- en skryfbeginsels en -reëls, agterwerk (bv. 'n lys afkortings, 'n lys internasionale plekname, transliterasietabelle, ens.). Hoewel die *AWS* nie 'n woordeboek is nie, bevat dit ook 'n woordelys. Woorde wat spelprobleme oplewer, neologismes en gevalle wat erkenning as deel van die Afrikaanse woordeskat verdien, is sommige van die kategorieë wat oorweeg word vir insluiting in hierdie lys.

In 2004 het die NTL'e van die ander nege amptelike Afrikatale begin met die hersiening van die spel- en skryfreëls van hierdie tale. Die eerste uitgawes is deur PanSAT in 2008 gepubliseer.

Geen hersienings is sedertdien saamgestel nie. Die AWS kan as voorbeeld dien van wat vir die ander Afrikatale gedoen kan word.

Trefwoorde: ALFABET, LEKSIKOGRAAF, LEKSIKOGRAFIE, ORTOGRAFIE, SPELLING, STANDAARDISERING, TAALVERANDERING, TERMINOGRAAF, TERMINOGRAFIE, TRANS-LITERASIE, VEREENVOUDIGDE TAAL

1. Introduction

Throughout the ages mankind had the urge to depict their thoughts, ideas, stories and history. Speech had to be reduced to writing. Different cultures developed different writing systems (i.e. cuneiform writing, hieroglyphs, alphabets, etc.). Writing systems had to be standardised to allow for proper communication through encoding and decoding principles and practices.

The English term 'orthography' dates from the 15th century. It comes from the Latin 'orthographia', which derives from the Greek 'orthós' (correct) and 'gráphein' (to write). The term 'orthography' refers to the correct or conventional way of spelling and the study or science of spelling. Orthography describes or defines the set of symbols used in writing a language, and the rules how to use these symbols.

Most significant languages in the modern era are written down, and for most of such languages a standard orthography has developed. The standard orthography is often based on a standard variety of the language and thus exhibiting less dialectal variation than the spoken language (<http://en.wikipedia.org/wiki/orthography>, 2013).

In some cases orthographies are regulated by bodies such as language academies or government agencies. For many languages (including English) there are no such authorities, and orthographies develop through less formal processes (cf. <http://en.wikipedia.org/wiki/orthography>, 2013).

The spelling and orthographic system of a language is just as integral a part of the language as its vocabulary. Language users create or are involved in the process of creating the spelling and orthographic system of their language, and they should be familiar with the approved spelling and orthographic system to enable them to read and write. A standard(ised) spelling and orthographic system should be used in the public domain for instance by educators, language practitioners, the media and government.

Language users are involved in any changes that may occur in the spelling and orthographic system. Since language is dynamic, it is quite natural for the vocabulary, spelling and orthography of a language to change. These changes are mirrored especially in the way that the spelling system of a language adapts according to development and modernisation.

Language modernisation is viewed as a complex and multi-dimensional endeavour whose sole aim is the empowerment of any language for literacy and numeracy for all purposes in formal and informal education (Emenanjo

1998: 44). Usually the basic spelling and orthographic policies and principles remain unchanged, but when the language develops and modernises, this process usually has an effect on the spelling and orthography.

New and unified spelling and orthographic systems should be created for the languages selected for standardisation. Establishing standardised, efficient and practical spelling and orthographic rules is a crucial basis for developing a modern literacy tradition.

In its pristine sense, literacy is concerned with the issues of writing and reading. Writing presupposes a universally accepted writing system, while writing and reading presupposes both a standard(ised) variety and texts in that variety (Emenanjo 1998: 44).

An orthography is to literacy what numeration is to numeracy (Emenanjo 1998: 47).

An efficient and practical spelling and orthographic system is a direct function of enabling languages in educational, cultural, administrative and mass media systems in the countries concerned (Abdulaziz 1989). Standardised spelling and orthographic rules assist with the production of texts in the standard variety of a language. They also assist in the process of developing standardised curricula designed for formal education.

According to Emenanjo (1998: 52), education is the acquisition of permanent numeracy and literacy and the integrated development of the cognitive, affective and psycho-motor domains in humans. The language of achieving these is immaterial, but it is better in the person's mother tongue or the language of the immediate community (cf. Prah 1993, 1995a and 1995b). It is therefore necessary to provide spelling and orthographic rules in the various official languages of South Africa.

In South Africa there is a dire need for reliable orthographic, spelling and editing rules for all the official languages. Documents dealing with spelling rules do exist in all these languages but most of these documents are outdated, need to be revised and modernised urgently and are not as comprehensive as they should be in order to address current needs in the various fields of communication and knowledge dissemination.

Language cannot be used in written communication or documented in dictionaries if the spelling and orthographic rules are not standardised. New terms that are created need a solid linguistic basis with proper word-forming principles as norm. Lexicographers of the various National Lexicography Units (NLUs), other lexicographers and terminographers cannot compile dictionaries for Afrikaans and the other African languages without standardised spelling and orthographic rules. Lexicographers and terminographers adhere to the spelling and orthographic rules of the standard variety of a language when compiling dictionaries. Translators cannot translate, editors cannot verify and edit and publishers cannot publish without standardised spelling and orthographic rules.

Without a proper spelling and orthographic system users of the language would all spell and write in a haphazard way. The Afrikaans language for instance developed from what was regarded as a mere "kitchen language" to a standard language that can be regarded as a language of science and technology. While the Dutch-Afrikaans language developed, there were many attempts to standardise the spoken language into a written variety. Many of these attempts failed, and since there was no authority taking charge of the process, different varieties resulted, i.e. 'Maleierafrikaans', 'Griekwa-afrikaans', 'Oosgrensafrikaans', 'Bolandafrikaans', 'Oranjerivierafrikaans', 'Vrystaatafrikaans', 'Transvaalafrikaans', etc. There was no uniform language, but considerable usage and pronunciation differences and no uniform orthography (cf. McLachlan 2001: 2). It was obvious that everyone wrote the language as he/she deemed fit since there was no 'Standaardafrikaans' variety.

The first *Afrikaanse Woordelys en Spelreëls* was published in 1917 by request of the Suidafrikaanse (sic) Akademie vir Taal, Lettere en Kuns. This publication was the product of the language pioneers known as Die Genootskap van Regte Afrikaners that was established on 14 August 1875 and the Kommissie voor Taal en Letteren of the Suidafrikaanse (sic) Akademie vir Taal, Lettere en Kuns (cf. McLachlan 2001: 2).

There were originally two commissions associated with the Suid-Afrikaanse Akademie vir Wetenskap en Kuns (SAAWK) (South African Academy for Science and the Arts): one to give attention to the development of spelling rules for the new Afrikaans language and another to compile a word list of existing Afrikaans words. The two commissions later merged and the SAAWK established the Taalkommissie (Language Commission) (De Bruyn 2013).

The first legislation that sanctioned the standardisation of Afrikaans spelling and orthographic rules was Act 23 of 1921 that gave the SAAWK statutory status to develop and standardise Afrikaans (Kapp 2009: 53; McLachlan 2001: 1). Since the first 1917 edition, nine revised editions of the *Afrikaanse Woordelys en Spelreëls* (AWS) (Afrikaans Word List and Spelling Rules) were published.

With the establishment of the Pan South African Language Board (PanSALB), that resulted from the recent South African Constitution, its legislation, i.e. Act 59 of 1995 as amended in 1999, gives this organisation the legal authority to establish standardising bodies, i.e. the National Language Bodies (NLBs), for each of the official languages as well as for the Khoe and San languages and the South African Sign Language (Alberts 2005).

The National Language Bodies were established by notice in the *Government Gazette* in terms of section 8(8)b of the Pan South African Language Board Act, Act No. 59 of 1995 (Alberts 2005). PanSALB published Norms and Rules in the *Government Gazette* No. 21907 of 22 December 2000 to ensure that the NLBs are established primarily as language development agencies with special emphasis on standardisation in respect of all official languages, the Khoe and San languages and South African Sign Language (Alberts 2005).

There is a distinct need for a reliable document on orthographic, spelling and editing rules for each of the official languages and the Khoe and San languages in line with what Afrikaans has:

Language cannot be used in written communication or documented in dictionaries if the spelling and orthographic rules are not standardised. New terms that are created need a solid linguistic basis with proper word-forming principles as norm. ... Although the original establishment of Afrikaans as a language in itself was disputed, Afrikaans is now a fully-fledged standard and scientific language. In 1912 the argument was used against the use of Afrikaans at tertiary level, that 'Afrikaans is nie deftig genoeg nie'. (Finlayson 2004)

The legislation on the Pan South African Language Board (PanSALB), Act 59 of 1995 as amended in 1999, made provision for the incorporation of existing standardisation bodies in the newly established National Language Bodies (NLBs). With the establishment of the National Body for Afrikaans (NLBA) of PanSALB it was therefore decided that the members of the Taalkommissie (Language Commission) of the Suid-Afrikaanse Akademie vir Wetenskap en Kuns (SAAWK) would become members of the NLBA. These members then constituted the Technical Committee for Standardisation (TC Standardisation).

Since its establishment the members of the Taalkommissie cum TC for Standardisation (Commission) continued with their work on the standardisation of the spelling and orthographic rules for Afrikaans. The members worked in close collaboration with the SAAWK and PanSALB and two revisions of the *Afrikaanse Woordelys en Spelreëls (AWS)* were published since.

Although there was a unanimous agreement in 2001 that the work of the Commission should be continued without interruption, it transpired in recent years that PanSALB could not afford two meetings per annum as was required. The SAAWK resumed with its funding of the Commission's meetings due to the long-term objectives set by the Commission (e.g. the conversion and adaptation of the *AWS* to an electronic format to be made available as an electronic book (e-book) and online, as well as a request for a standard grammar for Afrikaans to be compiled for international access on the Internet). The SAAWK had no choice but to assist with the financing of the activities since these activities had reached a critical stage of implementation, for which the momentum and continuity could be lost at that stage if the set time frame was to be kept.

An important aspect of the planning was the need for co-opted expertise and the utilisation of experts who have been involved in the compilation of the recently published *AWS* (2009). This aspect was recognised by PanSALB and was also an important consideration for the co-operation between the SAAWK and PanSALB. The SAAWK initially agreed to share the funding of the work of the Commission on a 50/50 basis by taking responsibility for one of the two meetings scheduled per year. In recent years most meetings and activities of the Commission were funded by the SAAWK.

The structure of the Commission needs urgent attention. Over a period of

time seven of the eight members of the Commission who held membership of both the Taalkommissie of the SAAWK and the NLBA of PanSALB decided not to serve another term of office on the Taalkommissie. In 2013 a new Taalkommissie was appointed by the SAAWK and only one member of the original Commission remains on both the Taalkommissie and NLBA. The other members of the TC: Standardisation are therefore no longer serving on the Taalkommissie. This issue is already being addressed by the SAAWK and PanSALB.

Work on the Afrikaans spelling and orthographic rules are discussed in this article to serve as a guide for similar work in the other South African languages. It should be noted that Afrikaans, a language with its roots in Africa, is also recognised as an African language by the Department of Basic Education. However, in the context of this article, the term 'African languages' refers to the nine indigenous Bantu languages of South Africa that received official language status according to Section 6(2) of the Constitution of the Republic of South Africa, 1996 (cf. 2.2).

2. Historical background of the spelling and orthography of Afrikaans and the African languages

2.1 The spelling and orthography of Afrikaans

The Suid-Afrikaanse Akademie vir Wetenskap en Kuns (SAAWK) (South African Academy for Science and the Arts) was established on 2 July 1909. At that stage English and Dutch were the official languages of South Africa. Afrikaans became one of the three official languages in 1925 (cf. Kapp 2009: 54) and it was only in 1983 that Dutch was officially removed as an official South African language.

The Union Parliament of South Africa was not in favour of bilingualism (English and Dutch) or mother-tongue education. There was a language dispute over language in education that lasted since 1906. Although regarded by some as a mere "kitchen language", the Afrikaans language started to develop rapidly alongside Dutch. The Afrikaans speaking community wanted Dutch (read also Afrikaans) to become a compulsory language in mother-tongue education (cf. Kapp 2009: 103-108).

Although Afrikaans was not an official language at the time, preliminary work on the Afrikaans spelling and orthographic rules started in 1875. Since 1911 several requests were made to the SAAWK to standardise the Afrikaans spelling and orthographic rules, but seeing the unsettling relationship between Dutch and Afrikaans speaking groups, these requests were for the time being ignored. The SAAWK, however, took an official decision in 1914 to standardise the spelling and orthographic system of Afrikaans. The Kommissie voor Taal en Letteren (Commission for Language and Literature) (later called the Taalkommissie (Language Commission)) was established to compile the first rules on Afrikaans spelling and orthography (Kapp 2009: 161) and the first version of

the *Afrikaanse Woordelys en Spelreëls* (AWS) was published in 1917 (cf. Kapp 2009: 111) in co-operation with the members of Die Genootskap van Regte Afrikaners (McLachlan 2001: 2).

One of the small steps towards the recognition of Afrikaans as an official language next to English and Dutch was taken in 1917 when copies of the first edition of the AWS were made available to teachers and schools (cf. Kapp 2009: 106, 329).

The legislation regarding the statutes of the SAAWK, Act 33 of 1956, and the legislation on the establishment of the SAAWK, Act 54 of 1956, stipulate that the Afrikaans spelling and orthographic rules of the standard variety of Afrikaans, as standardised by the Taalkommissie of the SAAWK, are enforceable and should be followed by government and general users of Afrikaans. The Taalkommissie should also prove through excellent linguistic example, the directional role of its own publications (e.g. the AWS) and the work of this commission that it deserves to be the standardisation body for Afrikaans (Kapp 2009: 135-136; 323). The legislation was repealed and currently there is no other legislation in place to support the work of the Taalkommissie other than it being part of the TC: Standardisation of the Afrikaans NLB (NLBA) of PanSALB (cf. Introduction).

Several revisions of the AWS were published since 1917 and all these represented the diachronic and synchronic development of the standard variety of the Afrikaans spelling and orthographic system. Since the first publication of the AWS, the Taalkommissie (read also Commission) continued with the standardisation work and endeavoured to eliminate practical spelling and orthographic problems as far as possible and to bring solutions where problems were encountered (cf. McLachlan 2001: 2). Although the Commission's main focus is Standard Afrikaans, it also concentrates on different varieties of Afrikaans (cf. De Bruyn 2013).

The ninth revision of the AWS was published in 2002 and in 2003 the members of the Taalkommissie received the SAAWK C.J. Langenhoven prestigious award. Each member received a medal and certificate.

The 2002 publication was a total reworking of the 1991 publication and all spelling and orthographic rules were rewritten in simplified language. The 2002 revision elaborated on some aspects of the spelling and orthographic system and revisited anomalies in previous revisions. The Commission consulted widely with language practitioners and stakeholders before the publication of the book.

In 2004 PanSALB decided to revise the spelling and orthographic rules of the nine official African languages. Copies of the 2002 version of the AWS were presented to members of the TC Standardisation of all the NLBs at the 2004 consultative meeting of PanSALB which was organised to discuss the revision process.

In August 2006 the Afrikaanse Taal- en Kultuurvereniging (ATKV) awarded an award to the Taalkommissie for the revision and compilation of

the 2002 edition of the *AWS*. The tenth revision of the *AWS* was published in 2009 to coincide with the centenary of the *SAAWK*. The members of the Taalkommissie were rewarded for this effort with the *SAAWK* C.L. Engelbrecht prestigious award. The fact that the Taalkommissie received prestigious awards for the compilation of the 2002 and 2009 editions of the *AWS* is a sure indication that these products met the needs of the target users.

The tenth revision of the *AWS* was published in 2009 to coincide with the centenary of the *SAAWK*. The members of the Commission were rewarded for this effort with the *SAAWK* C.L. Engelbrecht prestigious award.

No language is static and there are always language changes to be taken into consideration. The vocabulary and pronunciation of a dynamic language could change over a period of time. These changes should be reflected in the spelling and orthographic system of the language. The revision of the Afrikaans spelling and orthographic rules is therefore an ongoing process.

2.2 The spelling and orthography of the African languages

Originally the activities of early missionaries transmuted the various indigenous African languages to written languages. The speakers of the languages provided the vocabulary and the missionaries the orthographies and grammar (cf. Ranger 1989).

These missionaries were dealing with foreign languages and they had little knowledge of the history, genesis, or linguistic or political boundaries of the languages. The transmutation and codification of the languages were marred by a number of discrepancies. Each missionary used the orthography which was informed by his linguistic background, and the same words were spelt differently by the different missionaries (cf. Msimang 1998: 169-171).

Several scholars of the indigenous South African languages had since the attempts by the missionaries tried to develop spelling and orthographic rules for these languages. Spelling and orthographic conventions are also embedded in the grammar of these languages (cf. Msimang 1998: 169; Poulos and Louwrens 1994; Poulos and Msimang 1998).

In 1928 the Union government convened an Advisory Committee on Bantu Studies and Research to harmonise the orthographies of the African languages. This was followed by the appointment of the Suto-Pedi-Chuana Subcommittee, to formulate a uniform orthography for the three Sotho languages, i.e. Southern Sotho (Sesotho), Northern Sotho (Sesotho sa Leboa/Sepedi) and Tswana (Setswana). However, the representatives of the three language groups differed so much that the orthographies that were drawn up by this committee could not be accepted (cf. Msimang 1998: 170).

In 1948 the Transvaal Education Department expressed a need for a conference on reducing orthographic differences between the three Sotho languages, and the resolutions taken at the Somerset House Conference formed the basis for the *Sotho Terminology and Orthography No. 1* published in 1951 by

the Transvaal Education Department and in 1957 by the Department of Native Affairs. This Department also published the *Zulu/Xhosa Terminology and Orthography No.1* in 1957 (cf. Msimang 1998: 170).

During the previous dispensation there was an official attempt by the former Language Boards of the former Department of Bantu Education to standardise the vocabulary, spelling and orthography of the various indigenous languages. The products of these attempts by the Language Boards were published by the Department of Bantu Education as booklets on "Terminology and Orthography" and several revisions were published in the indigenous African languages (cf. Departmental Northern Sotho Language Board 1988).

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. According to section 6(1) of the Constitution, the official languages of South Africa are Sepedi, Sesotho, Setswana, Siswati, Tshivenda, Xitsonga, Afrikaans, English, isiNdebele, isiXhosa and isiZulu. Apart from the 11 official languages, the Constitution also recognises other languages such as Khoe and San and the South African Sign Language, which should be promoted and developed. 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 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 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 develop a language policy regarding its 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 and modernisation of the vocabulary, and of the spelling and orthographic rules of the official languages give effect to the constitutional rights of languages regarding the usage and their further development.

The Language Boards were replaced by the National Language Bodies (NLBs) of PanSALB and according to the NLBs the terminology and orthographic lists compiled by the former Language Boards were outdated and had to be revised.

The Strategic Plan of the National Language Bodies that was approved at

the Pan South African Language Board meeting of 19 March 2004 stated that special attention should be paid to the revision of the spelling and orthographic rules of the official African languages (cf. Alberts 2005).

A special consultative meeting was therefore held on 1–2 July 2004 in Pretoria to discuss the revision process with representatives of the Technical Committees (TCs): Standardisation of all the National Language Bodies. PanSALB regarded the NLBs to be well-positioned to deliver products with normative rules on spelling and orthography for each of the official African languages that would enhance communication in South Africa since the NLBs are regarded as the custodians of the languages (cf. Alberts 2006a).

As members of PanSALB structures, i.e. the Technical Committees: Standardisation of the National Language Bodies, the NLBs were the main bodies entrusted with the standardisation of the official languages, the Khoe and San languages and the South African Sign Language. They had to decide on the degree of revision that was needed for the respective languages. Some of the National Language Bodies had already started on their own with the revision of the spelling and orthographic rules of their respective languages before they were officially requested to do so at the consultative meeting of July 2004 (cf. Alberts 2006a).

During the consultative meeting of July 2004 members of the Khoe and San NLB indicated that they felt the need for and were eager to start with the compilation of spelling and orthography rules for these languages (cf. Alberts 2006a). Unfortunately no rules were compiled to date.

The process of the revision of the spelling and orthography rules of the nine official African languages entailed the standardisation as well as the modernisation of the existing rules. PanSALB embarked on this project by officially requesting the National Language Bodies to revise the current spelling and orthographic rules. During this process, the PanSALB structures (e.g. the National Language Bodies (NLBs), the Provincial Language Committees (PLCs), and the National Lexicography Units (NLUs)) collaborated with stakeholders in the various language communities, and with the national and provincial Departments of Arts and Culture, of Education, and of Communication (cf. Alberts 2006a).

The NLBs working on the three languages with cross-border varieties in Lesotho, Botswana and Swaziland, i.e. South Sotho/Sesotho, Setswana and Siswati, attempted to revise the spelling and orthographic rules in conjunction with the language boards or similar structures of these countries. The Swaziland Language Board attended the meetings of the Siswati NLB and the resulting spelling and orthographic booklet is a joint effort by South Africa and Swaziland. The Setswana spelling and orthography revision is also a joint effort by South Africa and Botswana. The same, however, is not true regarding the revised Sesotho spelling and orthographic rules (cf. Alberts 2006). Attempts to harmonise the orthographies of the South African and the Lesotho varieties of South Sotho/Sesotho, which were introduced as far back as 1927 have always

met with resistance from Lesotho although these varieties are almost identical dialects of the same language (cf. Msimang 1998: 167). To date the South African and Lesotho orthography unfortunately still remains different.

The final revised spelling and orthography documents for each of the nine official African languages could be regarded as products of consultation of which the different speakers of the nine official African languages could rightfully claim ownership.

The focus of the revision process was on spelling, orthographic and editing rules. The project teams revisited and revised existing documents on spelling, orthographic and editing rules. The project teams also focused on naming and word-forming principles and on abbreviation and punctuation principles. The Technical Committees charged with standardisation addressed aspects such as:

- **spelling rules** (e.g. hyphen; apostrophe; capital letters (first letter of sentence, nationalities and tribal designations, names of languages, proper nouns, qualifiers of capitalized proper nouns, names of days, toponymic guidelines); word division (parts of speech, demonstrative pronouns, auxiliary formatives, abbreviated forms, numerals, abbreviations));
- **orthographic rules** (e.g. vowels; consonants (e.g. plosives); constituents (e.g. fricatives, liquids, nasals, affricates, semi-vowels); clicks);
- **typography** (the spelling and orthographic rules regarding place names, surnames, street names, month names, day names, names of geographical features, etc.);
- **punctuation norms;**
- **abbreviation norms;**
- **terminology lists** (explaining of the grammar terms being used when dealing with the different rules);
- **word lists** (containing difficult words or words easily or usually misspelt);
- **lists of abbreviations and acronyms** (cf. Alberts 2006b).

Of these issues the top three were regarded as the urgent ones for the first stage of the revision process. In cases where little or no revision was needed, the TCs could devote more time to the latter issues. All these aspects were (and will in future again be) researched, discussed, verified and authenticated by the various NLBs (cf. Alberts 2006b).

The revised spelling and orthographic rules should depict the changes in the relevant languages throughout the past decades. Each National Language Body had tried to capture the dynamic growth of the language concerned.

This process of standardising the languages and revising the current spelling and orthographic rules is supposed to be ongoing. The PanSALB Board decided that the NLBs would in future attempt to keep abreast of

changes in the spelling and orthography by incorporating changes in each revised version of the spelling and orthographic rules (cf. Alberts 2006b).

Before its publication, the first project output of nine draft booklets were disseminated to stakeholders such as language practitioners, journalists, educators, the media, etc. for their input. Consultative meetings were held with these stakeholders and comments were dealt with by their incorporation into the booklets (cf. Alberts 2006b).

In September 2008 PanSALB published the revised versions of the spelling and orthographic rules to serve as easy reference resources to the spelling and orthographic rules of the nine official African languages (cf. PanSALB 2008). PanSALB then expressed its wish that these documents could be used as quick reference tools. These guideline documents on spelling and orthography had been compiled to serve as booklets for use by language users.

PanSALB stipulated that each National Language Body will have to submit a revised version at the end of its five-year term of office. The NLBs of the official African languages started in 2004 with the first revision of the spelling and orthographic rules of these languages, and the first editions were published by PanSALB in 2008 but no revisions were compiled or published since.

Since the revision of the spelling and orthographic rules was regarded as an ongoing process, PanSALB further stipulated that it would welcome comments from users. These comments could assist the NLBs with the future revision of the rules. After the publication of the booklets, PanSALB did receive feedback from several users. Some of the feedback was positive, but most of it extremely negative and seeing that no revisions were compiled since, the valuable feedback was disregarded.

This article focuses on the post-1994 situation regarding the development and/or standardisation of orthographies for the (official) South African languages. The continued work on the Afrikaans orthography and the resulting products can serve as an example of how and what could and should be done regarding the development and/or revision of the spelling and orthographic rules of the other nine official African languages and the Khoe and San languages.

3. Aspects of the work on the AWS

3.1 Background

The *Afrikaanse Woordelys en Spelreëls* (AWS) is a resource compiled by the Taalkommissie (Language Commission) (cf. Pharos 2009) to assist users of Afrikaans in using the standard variety of the language. The AWS explains the ground rules of the Afrikaans spelling and orthographic conventions.

The basic rules are provided in simplified language for easy comprehension and application.

The current Commission, consisting of academics, translators, lexicogra-

phers and terminologists, decided to contact language practitioners and academics before publishing the 2002 and 2009 revisions. The aim was to get the input of the language experts on the changes incorporated in these publications. Valuable comments and contributions were received which were incorporated in these publications.

The Commission meets twice a year for a full week to decide *inter alia* which new words should be incorporated as part of the vocabulary of Standard Afrikaans, how to spell these words, and how to standardise Afrikaans in general (cf. De Bruyn 2013).

3.2 Front matter

The front matter contains an acknowledgement to financial contributors (i.e. financial assistance by Die Dagbreek Trust, Het Jan Marais Nasionale Fonds, the University of Pretoria, and PanSALB) and computer assistance (i.e. the computer expertise from the Centre for Text Technology (CText) of Northwest University).

The front matter also contains a Table of Contents; Preface; Suggestions for Users; Explanations; Modus operandi; and Ground rules of the Afrikaans spelling system.

3.2.1 Table of contents

The table of contents contains acknowledgements; a preface; suggestions for users; explanations; the modus operandi; ground rules of the Afrikaans spelling; the basic spelling and orthographic rules in an alphabetical format; a word list containing examples of the Afrikaans vocabulary; a list of abbreviations; a list of international place names; a transliteration table; colloquial Afrikaans; a terminology list, and an index.

3.2.2 Preface

The preface gives a synopsis of the process regarding the standardisation of the Afrikaans language. It provides a broad outline of the work on the AWS and the influence of language development and change on the spelling and orthographic conventions.

The preface also contains the names of the current members of the Commission who compiled the latest edition of the AWS.

3.2.3 Suggestions for users

The suggestions for users assist the users of the AWS by providing a basic style guide.

3.2.4 Explanations

This section describes the various sections contained in the *AWS*. It explains the basic spelling and orthographic conventions of Afrikaans and provides a list of phonetic signs with their appropriate pronunciation in Afrikaans, e.g. [æ] skêr, bêre; [u:] loer; [e] sê, hê.

The *AWS* is not a dictionary but it contains a word list. This section explains why certain Afrikaans words are included in the word list while others were omitted.

3.2.5 Modus operandi

This section explains how the Taalkommissie functions. The different Commissions responsible for the different revisions and publications of the *AWS* inherited the basic core of the *AWS*. This core was either kept intact or changed according to language change and modernisation.

The Commission had to take different aspects into consideration — it should not only give advice on *how* to spell but also *what* to spell. Obsolete words had to be deleted and neologisms entered.

3.2.6 Ground rules of the Afrikaans spelling system

An outline is provided of the ground rules of the Afrikaans spelling and orthographic system. There are three basic principles which have been kept almost intact since the first *AWS* publication of 1917. The spelling and orthography of the standard variety of Afrikaans are still determined by these ground rules although they were refined throughout the years:

- (a) The current spelling and orthography of words are determined by the tradition of how Standard Afrikaans words are spelt and written.
- (b) The spelling of Afrikaans is based on the phonetic system of Standard Afrikaans.
- (c) Uniformity: words that have related forms should, whenever possible, be spelt the same.

3.3 Spelling and orthographic rules

The previous formats of the spelling and orthographic rules, i.e. editions prior to the 2002 and 2009 publications of the *AWS*, were very difficult to comprehend. The Commissions working on the latest editions decided to rewrite all spelling and orthographic rules in a format that the average grade 10 learner and most users of the Afrikaans language would be able to comprehend. The rules are therefore rewritten in simplified language. Rules are also shorter and

hidden rules gained proper rule status.

The rules of the latest versions of the *AWS* are presented in a user-friendly alphabetical and numerical order. The rules are arranged in an alphabetical sequence to allow for easy access to information (i.e. 1. *Afbreek van woorde*; 2. *Afkappingsteken*; 3. *Afkortings*; 4. *Akuutteken*; 5. *Deelteken*; ... 17. *Wisseling*; 18. *Woorde van vreemde herkoms*). All rules are numbered under the alphabetically arranged category i.e. 1. *Afbreek van woorde: Reël 1.1, Reël 1.2, Reël 1.3*, etc. Several empty lemmas were entered in the 2002 edition of the *AWS*, i.e. *Aanmeakaarskryf*. A cross-referencing system assisted the user in finding information on this lemma, and others that were not provided in the alphabetical arrangement, e.g. *Aanmeakaarskryf* see *Skryfwyse — los en vas*. The 2009 edition of the *AWS* does not contain any empty lemmas.

Orientation paragraphs and some comments supply extra information. Examples of words spelt according to the specific rule are provided. Examples of misspelt words are sometimes also supplied and these are marked with asterisks to indicate that they are wrongly spelt. The reasoning behind providing misspelt words was that users frequently learn more from these wrongly spelt examples than from the correctly spelt examples, i.e.:

Correct	Wrong
art. 21-maatskappy	*art.-21-maatskappy
graad 1-leerling	*graad-1-leerling
km	*km.
10 Mg	*10 M-g/*10M g/*10M.g.

3.3.1 Afbreek van woorde (Division of words)

There might not be enough space in written, typed or printed texts to fit a full word in a given line. Such words should preferably be shifted in full to the next line. Words can, however, be divided according to the rules provided in this section, i.e.

bio-chemie, nooiens-van, voet-stappe (Rule 1.3)
ge-sel-se-ry, tran-sak-sie, vrien-din (Rule 1.4)
bak-ke, bor-rel, mon-nik (Rule 1.10)
beeld-jie, koord-jie, land-jie (Rule 1.14)

3.3.2 Afkappingsteken (Apostrophe)

This section deals with the use of the apostrophe in Afrikaans, e.g.

foto's – foto'tjie, radio's – radio'tjie (Rule 2.1)
hoera's, karba's – karba'tjie (Rule 2.7)
crèche'e or crèches, garage – garage'ie (Rule 2.8)
a's – a'tjie, l'e or l's – l'etjie (Rule 2.11)

gr. 12's, .303's, .22's, 747's (Rule 2.14)
g'n (*geen*), 'k (*ek*), s'n (*syne*), 't (*het*) (Rule 2.15)

3.3.3 Afkortings (Abbreviations)

A representative list of abbreviations and acronyms is provided in the back matter of the AWS. Although the abbreviations of proper names and of political parties are not included in the list, names of some institutions that may no longer exist or that were renamed (i.e. EVKOM, YSKOR, ZARP), are included for historical purposes since users may still find these in documents such as historical or literary texts.

The rules governing Afrikaans abbreviations and acronyms are provided in the 'Afkorting' section, i.e.:

chem. (*chemie; chemies(e)*), elektr. (*elektries(e)*) (Rule 3.1)
bg. (*bogenoemde*), kl. (*kleinletter*) (Rule 3.2)
a.g.v. (*as gevolg van*), o.a. (*onder andere*) (Rule 3.4)
BA or B.A. (*Baccalaureus Artium*) (Rule 3.8)
HOD or H.O.D. (*Hoër Onderwysdiploma*) (Rule 3.9)
SADOU or Sadou (*Suid-Afrikaanse Demokratiese Onderwysersunie*) (Rule 3.10)

3.3.4 Akuutteken (Acute accent)

The acute accent (´) is only placed in Afrikaans on vowels and diphthongs, e.g.

én ... én, óf ... óf; déúr, kóúd, kráái; kóéël, réén (Rule 4.2)
dié, hiér (Rule 4.3)
attaché, cliché, née (Rule 4.4)

3.3.5 Deelteken (Diaeresis)

The diaeresis (¨) is placed in Afrikaans on vowels, but only in certain circumstances and it has different usages in Afrikaans spelling conventions, i.e.:

beëdig, drieërlei, geëis, trofeë; altruïs; reünie (Rule 5.1)
essensieel > essensiële; finansieel > finansiële (Rule 5.2; 5.3)
biëlektries or bi-elektries; poliëtileen or poli-etileen (Rule 5.6)
fee > feë, see > seë; genie > genieë, knie > knieë (Rule 5.10)

3.3.6 Diftonge (Diphthongs)

The spelling of diphthongs is unproblematic in Afrikaans and it is only the diphthong [əi] that is sometimes spelt with *ei* and sometimes with *y*, cf.

brei/bry, lei/ly, steil/styl, wei/wy (Rule 6.1)

3.3.7 Getalle, syfers en simbole (Numbers, figures and symbols)

The rules in this section provide direction on how to write words and groups of words in combination with numbers, figures and symbols, i.e.:

A4-formaat, 7-sitplekvoertuig; 6-erig, 14-jarige, Δ-agtige (Rule 7.1)

16,7 kg, 318,55 kW; 1,5ℓ, R22,75 (Rule 7.4)

0,9 m/*,9 m; R0,25/*R,25; 0,5c/*,5c (Rule 7.5)

2009-11-25 *for* 25 November 2009 (Rule 7.7)

03:15 *for* quarter past three in the morning (Rule 7.8)

5 °C-styging; 50 kg-sakke; 750 ml koeldrank (Rule 7.17)

3.3.8 Gravisteken (Grave accent)

The grave accent (̀) is used in Afrikaans on vowels only, e.g.:

appèl, dè, hè, nè (Rule 8.1)

crèche, pietà, première, vis-à-vis (Rule 8.2)

3.3.9 Hoof- en kleinletters (Upper and lower case letters)

The use of upper and lower case letters relies on convention. The AWS provides guidelines on the usage of upper and lower case letters but specific institutions may have in-house style guides which could influence usage. Rules 9.1–9.26 deal with a variety of issues, i.e.

- The beginning of sentences (Rules 9.1–9.2)
- Proper names (Rule 9.3)
- Form of address (Rules 9.4–9.6)
- Animal and plant names (Rules 9.7–9.8)
- Appellative nouns used as proper nouns (Rule 9.9)
- Proper nouns used as appellative nouns (Rule 9.10)
- Language names and similar issues (Rule 9.11–9.12)
- Geographical names (Rule 9.13–9.15)
- Monetary units (Rule 9.16)
- Geological names (Rules 9.17–9.18)
- Derivations of proper names (Rules 9.19–9.24)
- Fading of proper names (Rule 9.25)
- Abbreviations (Rule 9.26)

3.3.10 Kappie (Circumflex)

The circumflex (ˆ) is used in Afrikaans on vowels only, to indicate that the specific vowel sound should be stretched, e.g.:

sê, hê; wîe; aikôna; brûe (Rule 10.1)
miljoenêr, sekondêr, skêr (Rule 10.2)
crêpe-de-chine, maître d'hôtel, tête-à-tête (Rule 10.3)

3.3.11 Konsonante (Consonants)

The use of consonants is not problematic in Afrikaans. Usage has usually to do with:

- the pronunciation of specific letters, e.g. robvleis [rɔp'fløis]/robbe [rɔ'bə] (Rule 11.1); hibried [hi'brit]/hibriede [hi'bri'də] (Rule 11.2);
- the usage of **-de** or **-te** in past participles (bedekte, geykte, gelapte, verlepte (Rule 11.6) *or* gestyfte/gestyfde; vereiste/vereisde (Rule 11.7) *or* gematigtheid/gematigtheid (Rule 11.8);
- the duplication of consonants between vowels, i.e. vodde; beddens; blokke; landskappe; sinusse (Rules 11.23–11.30).

3.3.12 Koppelteken (Hyphen)

The hyphen has several different usages in Afrikaans spelling conventions (cf. Rules 12.1–12.38), i.e. to assist with the readability of a word (see-eend); to join words in a specific combination to indicate a specific meaning (skriba-kassier); to use as a mark to indicate omission (Moeders- en Vadersdag; taal- en letterkundiges); to indicate a break in a word that cannot be written in full in a given line (ver-gadering).

3.3.13 Meervoudsvorme (Plural forms)

The plural forms of Afrikaans words (cf. Rules 13.1–13.30) are interdependent on issues such as vowels occurring in open and closed syllables, duplication of consonants, the use of the apostrophe, the use of diacritics, proper and appellative nouns, words of foreign heritage, job and title designations, abbreviations, and variants.

3.3.14 Skryfwyse — los en vas (Styling of compounds — disjunctive and conjunctive)

It is often difficult to determine whether Afrikaans words should be written disjunctively or conjunctively. Rules 14.1–14.53 address these issues.

3.3.15 Verkleiningsvorme (Diminutives)

Diminutive forms in Afrikaans end on the following: **-ie**, **-jie**, **-etjie**, **-tjie**, **-kie**

and **-pie**. Rules 15.11–5.13 address the spelling conventions regarding diminutive forms in Afrikaans.

3.3.16 **Vokale (Vowels)**

Several factors influence the usage of vowels in the Afrikaans written format (cf. Rules 16.1–16.23), i.e. open or closed syllables, stress, origin of the word, origin of prefixes and suffixes in combination of the base word, sound environment, tradition, and the fact that the Afrikaans alphabet contains more vowels and diphthongs than the original Roman alphabet on which it was based.

3.3.17 **Wisseling en subkategorieë (Variants and subcategories)**

There are several spelling variation possibilities in Afrikaans, i.e. aalwyn/aalwee; Chinees/Sjinees; restaurant/restourant, etc. Rules 17.1–17.20 deal with variants and its subcategories.

3.3.18 **Woorde van vreemde herkoms (Words of foreign origin)**

Afrikaans, like any other language in contact with other living languages, is influenced by these languages. There are three basic principles governing the inclusion of foreign words in the Afrikaans vocabulary:

- Foreign words transliterated into Afrikaans according to the typical Afrikaans spelling conventions (e.g. atjar, barmitswa, kikoejoe (Rule 18.1));
- Foreign words retaining their original form (e.g. canyon, gentleman, mañana, oeuvre, pizza (Rule 18.2));
- Foreign words that are spelt partly according to the foreign spelling convention, partly according to the Afrikaans spelling conventions (e.g. Fanagalo/Fanagalô/Fanakalo/Fanakalô; shongololo/sjongolôlo/songolôlo/songolôlo; kleza/klêza (cf. Rules 18.3–18.19).

3.4 **Word list**

Although the AWS is not a dictionary, it also contains a word list. The convention for adding words to the word list is to list words giving spelling problems, neologisms, complex words in order to illustrate word-forming processes and words whose existence is recognised in the Afrikaans vocabulary.

The Commission recently gave special attention to the word list by looking at every single word in it to test it against the spelling and orthographic rules. All words used as examples in the rules section were included in the word list.

The rule numbers were added next to these example words to enable users to go back to the rule where more information is given on the specific spelling convention (cf. *katôba of catawba (soort druiwe)* 18.3). The Commission aims to compile an online version of the AWS where rule numbers next to words in the word list would refer the user with the aid of hyperlinks to the relevant rule.

The user receives assistance through the inclusion of typical or difficult Afrikaans words that are usually or easily misspelt (cf. *fotoalbum; kat-en-muis-spelery of kat-en-muisspelery; meisieskool; mes-en-vurk-laai of mes-en-vurk-laai, -e; mans- en dameskoen, -e; netnoumaar; nieregeringsorganisasie, -s; ineens, nogeens, weereens* 14.53; ook al).

The Commission included core terms from various subject areas to give recognition to these terms, e.g. *stochasties of stogasties (wiskunde; statistiek)*. New coinages in Afrikaans are also given in the word list to recognise their existence, e.g. *stratografie (wetenskap van aardlae)*.

Between 1959 and 1963 the SAAWK acted as adviser regarding the translation of the Koran into Afrikaans. A committee consisting of experts of the Middle Eastern languages and cultures and two members of the Taalkommissie of the SAAWK translated the document. Two copies of this translation are available at the SAAWK (cf. Kapp 2009: 382).

Between 1959 and 1963 the SAAWK acted as adviser regarding the translation of the Koran into Afrikaans. A committee consisting of experts of the Middle Eastern languages and cultures and two members of the Taalkommissie of the SAAWK translated the document. Two copies of this translation are available at the SAAWK (cf. Kapp 2009: 382). Since 2006 the current Commission documented the typical core Islamic words taken from the Koran for incorporation in the AWS (cf. Kapp 2009: 382). Several of these Islamic words were incorporated in the word list of the 2009 edition of the AWS. Typical terms relating to other religious denominations were also included.

3.5 Back matter

The back matter of the AWS consists of *inter alia* a list of abbreviations, a list of international geographical names, a transliteration table, etc. The planned online version will also include several other lists, e.g. the periodical table, a reference table on big and small numbers, a list of the monetary systems of the world with the relevant ISO codes, etc. These reference tables and lists are already completed and ready for inclusion.

3.5.1 List of abbreviations

A representative list of abbreviations and acronyms follows directly on the word list. The rules governing the way a word should be abbreviated in Afrikaans are supplied in section 3 of the rules' section.

3.5.2 List of international geographical names with derivations

The South African geographical names are governed by the South African Geographical Names Council. There is, however, no official body governing the spelling and orthographic conventions of international place names in Afrikaans. The Commission therefore added a list with international geographical names with derivations (i.e. adjectives and inhabitants). These names were transliterated according to the Afrikaans spelling conventions.

3.5.3 Transliteration table

Language practitioners frequently encounter problems when Russian and Slavic names or their English versions have to be transliterated into Afrikaans. The transliteration table assists users with these problems.

The Commission also prepared transliteration tables for Arabic and Chinese. These tables will be part of the planned online version of the AWS.

3.5.4 Colloquial Afrikaans

The 2002 version of the AWS contained a list of typical colloquial Afrikaans words. These words were marked with an ° (° = Omgangsafrikaans) to indicate that they are colloquial words and not part of the standard variety of the Afrikaans lexicon. Several users thought that the Commission permitted the usage of these colloquial words in the standard variety. The idea with their inclusion was, however, to assist authors when they have to spell the colloquial variety of Afrikaans. The Commission excluded the representative list of colloquial Afrikaans in the 2009 edition of the AWS, and this publication only contains six guidelines on how to spell colloquial Afrikaans, i.e.:

- Guidelines on spelling of words pronounced in a specific way, e.g. daai° / darie° = daardie; hierie° = hierdie; kinnere° = kinders (Guideline 1).
- Guidelines on words borrowed from English, e.g. ghoef° (swem), geghoef° (geswem); moewies°, moewiese° (tamaai, baie groot) (Guideline 2).
- Guidelines on typical English words frequently used in Afrikaans with Afrikaans pronunciation, e.g. brekfis° (breakfast); orraait° / orrait° (all right) (Guideline 3).
- Guidelines on English words incorporated unchanged in Afrikaans, e.g. budget°, gebudget°; coach°, gecoach°; sexy°, sexiër°, sexyste° (Guideline 4).
- Guidelines on English expressions used partially or fully in Afrikaans, e.g. boots en all° / boots en al°; finish en klaar°; gepoeier en gepaint° (Guideline 5).
- Guidelines on how to use English expressions in Afrikaans text, e.g. *Ons*

*was behoorlik "gecollar en getie" vir die okkasie (not *ge-"collar" en ge-"tie") (Guideline 6).*

3.5.5 Terminology list

A terminology list explaining the meaning of terms used in the AWS was included as back matter. The list is arranged in alphabetical order and contains typical terms that might be problematic, i.e.

aanhefvoorm: *'n Aanhefvoorm is die inleidende woord of woorde wat gerig word tot 'n persoon of persone aan die begin van 'n brief, uitgeskrewe toespraak of dergelyke teks, byvoorbeeld:*

Geagte dr. Smit, Dames en Here, Heil die Leser, Liewe tannie Sarie

schwa: *Die schwa is die vokaal [ə], byvoorbeeld in kamer, lig, peloton en basaar.*

3.5.6 Index

An alphabetically arranged index and a useful quick reference system with running heads assist in the reference process, e.g.

graadwoorde 139-140, 560
leesbaarheidskoppeltteken 75, 78-80, 92

4. Related activities and future developments

4.1 Lara-TK

All data accumulated by the Commission is in computerised format. The specific format was developed by the Centre for Text Technology (CText) of Northwest University, Potchefstroom campus. The computerised format is called "Lara-TK" and it contains all information regarding the word list, rules, etc. Lara-TK can be regarded as a blueprint of the AWS and future publications, whether hard copy or online versions, will be printed from this data base.

4.2 TK Forum

The members of the Taalkommissie of the SAAWK communicate on a daily basis on the TK Forum — this is an ideal way to solve spelling and orthography related problems encountered between the two formal meetings. Decisions taken are ratified at formal meetings. Present as well as previous members have access to the TK Forum. The members (a) address issues such as new words to be added to or deleted from the word list, and problematic words, and (b) share personal information and have quick access to any one or all of the members. The TK Forum was developed by TshwaneDJe Human Language

Technology (cc) by request of Pharos Publishing House and it is for the exclusive use of the Taalkommissie.

4.3 Facebook: Die Afrikaanse Taalkommissie

The Taalkommissie keeps up to date with developments on the social networks and deals on a daily basis with language related matters on its Facebook page "Die Afrikaanse Taalkommissie". The members of the Taalkommissie take turns to address different problematic issues by writing a short paragraph on a specific issue at hand. Members of the public can then react on the relevant issue and receive more assistance. Members of the public could also use this Facebook page to directly consult the Taalkommissie on language related issues.

4.4 e-AWS (electronic AWS)

Although the AWS is already available on the Internet at www.woordelys.co.za (cf. De Bruyn 2013), the Taalkommissie plans to develop an electronic (e-book) and online version of the AWS, i.e. *e-AWS* (cf. SAAWK 2011: 78-82; SAAWK 2012: 9) for international access via the Internet. Users will be able to browse through the word list, abbreviation list and all other lists and will be able to refer to rules with the aid of hyperlinks. Interaction with users would be made available by supplying answers to frequently asked questions, and there would be a link to the Taalkommissie's Facebook page for direct consultation. The Taalkommissie would also be in a position to discuss suggestions for change and obtain comments and proposals from language users. Frequent updates would also be possible in an online version of the AWS.

4.5 e-Grammatika van Afrikaans (e-Grammar for Afrikaans)

The Taalkommissie formed a special subcommittee in close working relationship with various renowned scholars to compile an online grammar for Afrikaans, i.e. *e-Grammatika van Afrikaans* (cf. SAAWK 2011: 78-82; SAAWK 2012: 9). The *e-Grammar* would be descriptive and normative and would function as language portal. The scope would relate to morphology and syntax (morphosyntactic grammar) with a limited grammar (opposed to basic or extensive grammar) and it will focus on the written variety of standard Afrikaans. The *e-Grammar* could later incorporate phonology, school grammar and other varieties of Afrikaans.

The main aim with the *e-Grammar* is to develop Afrikaans linguistics and the target users are Afrikaans mother-tongue students.

The envisaged format is Wiki-based grammar for Afrikaans — a language portal. The language medium would be Afrikaans and it might later incorporate other languages. The time allocated for the development of this *e-Grammar*

is five to ten years.

It is envisaged that the development of the *e-Grammar* would be financed by external bodies. The service to target users, however, would be free of charge.

The Taalkommissie would be the project leader and it would also supply technical assistance and hosting. The Taalkommissie would co-operate with various Afrikaans speaking scholars as well as with the Dutch language portal (Nederlandse "Taalportaal").

5. Co-operation between Afrikaans, the other nine official African languages and the Khoe and San languages

During the consultative meeting organised by PanSALB in July 2004, it was decided to revise the spelling and orthographic rules of the nine official African languages, and the Khoe and San NLBs expressed an interest in the compiling of spelling and orthographic rules for these languages (cf. 2.2).

In 2004 the NLBs of the nine official African languages started with the revision of the spelling and orthographic rules of these languages. The first editions were published by PanSALB in 2008, but no revisions have been compiled since. The AWS could serve as example of what could be done for these African languages, as well as for the Khoe and San languages.

Co-operation between the various National Language Bodies could lead to an approved system of revision. The different orthographic and spelling rules should be streamlined. There could even be a process of harmonisation of spelling and orthographic rules that are similar in different African languages, Khoe and San, and Afrikaans (e.g. capitalisation, abbreviation processes, word-forming principles, etc.).

The history and development of the Afrikaans spelling and orthographic system could serve as an example of what could be done and what should be avoided. The AWS could be used as a basic format to address different spelling and orthographic issues. The Commission is prepared to co-operate with and assist the NLBs with the revision of their spelling and orthography rules and principles.

The revision of the spelling and orthographic rules of a dynamic and living language should concentrate on aspects such as:

- Description of the language (history of its development)
- Description of language change and modernisation
- Basic principles of the language
- A description of the grammar of the language
- The spelling of the distinctive features of aspiration, nasalisation and palatalisation, as well as the non-distinctive ones on vowel and consonant

assimilation and elision (cf. Emenanjo 1998: 44-45)

- Tonal notation
- Pronunciation
- Word-forming principles for the creation of new words or terms (i.e. scientific and technological concepts)
- Word division
- Conjunctive and disjunctive writing
- Loan-words from living languages, transliteration from classical languages, intra-dialectal borrowings
- Punctuation issues
- Spelling of names (pronouns, place names (national and international), etc.)
- Usage and/or elevation of dialects, colloquialisms, slang and jargon (e.g. language usage in popular literature)
- Transcription procedures (i.e. writing of texts in dialects from praise songs, speeches, etc.)
- Principles for the creation of abbreviations and acronyms
- Terminology list (explanation of grammar terms and terms used in the publication on spelling and orthographic rules)
- Alphabetical word list representative of the vocabulary of the language (e.g. words with spelling problems, neologisms, complex words, acknowledgement of the existence of words)

6. Conclusion

Language is dynamic, it develops and changes and its spelling and orthography system should be revised and modernised at regular intervals. The spelling and orthography conventions of a language form an integral part of language renewal and therefore the spelling and orthographic rules of the standard variety of a language should be revised continuously to meet the needs and demands of language users.

The revision of the spelling and orthographic rules of the nine official African languages started in 2004 and the first revisions were published in 2008 by the National Language Bodies of PanSALB. The PanSALB Board requires revised versions to coincide with the five-year term of office of NLB members. Unfortunately there was since 2008 no indication from PanSALB's side that the NLBs of the official African languages will revise the spelling and orthographic

rules of these languages. The Khoe and San NLBs also did not compile spelling and orthography rules for these languages. Only the Commission working on Afrikaans (i.e. Taalkommissie of the SAAWK cum TC: Standardisation, Afrikaans NLB, PanSALB) revises the Afrikaans spelling and orthographic rules on an ongoing basis. The process of the ongoing revision and publication of the Afrikaans spelling and orthographic rules as described in this article could serve as guide for the other official languages. Co-operation between the standardising bodies for Afrikaans and for the official nine African languages, and the Khoe and San languages could be beneficial to the further development of the official South African languages and the Khoe and San languages.

Bibliography

- Abdulaziz, M.H.** 1989. Development of Scientific and Technical Terminology with Special Reference to African Languages. *Kiswahili* 56: 32-49.
- Alberts, M.** 2005. *National Language Bodies (NLBs) and Technical Committees (TCs)*. Paper read at the PanSALB Colloquium on National Language Bodies of 2005-11-11, PanSALB, Pretoria.
- Alberts, M.** 2006a. *Standardisation, Modernisation and Harmonisation — Users' Perspectives*. Paper read at the 11th International Afrilex Conference, 6-7 July 2006, Thohoyandou.
- Alberts, M.** 2006b. *Standardisation, Modernisation and Harmonisation — A Sociolinguistic Perspective*. Wang, Yuli, Yu Wang and Ye Tian (Eds.). 2006. *TSTT' 2006 International Conference on Terminology, Standardization and Technology Transfer: Proceedings*: 182-192. Beijing: Encyclopedia of China Publishing House.
- De Bruyn, E.** 2013. Taalkommissie kom voel polsslag van Afrikaans in Namibië. *Republikein*, 12 Maart 2013.
- Departmental Northern Sotho Language Board.** 1988⁴. *Northern Sotho Terminology and Orthography No. 4/Noord-Sotho terminologie en spelreëls No. 4/Sesotho sa Leboa mareo le mongwalo No. 4*. Pretoria: Government Printer.
- 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 Sport, Recreation, Arts and Culture.** 2005. *Language Policy Framework of the Gauteng Provincial Government*. Johannesburg: Gauteng Provincial Government.
- Emanajo, E.N.** 1998. The Modernization of the Igbo Language and its Implications for Holistic Education. Prah, K.K. (Ed.). 1998: 43-52.
- Finlayson, R.** 2004. *Consultative Meeting: Revision of the Spelling and Orthography Rules of the Official African Languages*. Paper read at the PanSALB Consultative meeting held on 1-2 July 2004, Pretoria.
- 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.
- Kapp, P.** 2009. *Draer van 'n droom*. Hermanus: Hemel & See Boeke.

- McLachlan, J.D.** 2001. AWS se 'gesag'. Van den Berg, L. (Ed.). 2001. *Taalmanak*: 1-3.
- Msimang, C.T.** 1998. The Nature and History of Harmonisation in South African Languages. Prah, K.K. (Ed.). 1998. *Between Distinction and Extinction. The Harmonisation and Standardisation of African Languages*: 165-174. Johannesburg: Witwatersrand University Press.
- Pharos.** 2009. *Afrikaanse woordelys en spelreëls*. Cape Town: Pharos.
- PanSALB.** 2008. *Imithetho yokubhala nobhalomagama LwesiZulu*. Pretoria: PanSALB.
- PanSALB.** 2008. *Imithetho yokutlola nokupeleda isiNdebele*. Pretoria: PanSALB.
- PanSALB.** 2008. *Imitsetfo yekupela nelubhalomagama LweSiswati*. Pretoria: PanSALB.
- PanSALB.** 2008. *Melao ya mongwalo le mopeleto ya Sesotho sa Leboa*. Pretoria: PanSALB.
- PanSALB.** 2008. *Melawana ya mokwalo le mopeleto Setswana*. Pretoria: PanSALB.
- PanSALB.** 2008. *Melawana ya mopeleto le karohanyo ya mantswe mongolong wa Sesotho*. Pretoria: PanSALB.
- PanSALB.** 2008. *Milawu ya mapeleleto na matsalelo ya Xitsonga*. Pretoria: PanSALB.
- PanSALB.** 2008. *Milayo ya kupelelele na kuñwalele kwa Tshivenda*. Pretoria: PanSALB.
- Poulos, G. and L.J. Louwrens.** 1994. *A Linguistic Analysis of Northern Sotho*. Hatfield, Pretoria: Via Afrika.
- Poulos, G. and C.T. Msimang.** 1998. *A Linguistic Analysis of Zulu*. Cape Town: Via Afrika.
- Prah, K.K.** 1993. *Mother Tongue for Scientific and Technological Development in Africa*. Bonn: German Foundation for International Development.
- Prah, K.K.** 1995a. *African Languages for the Mass Education of Africans*. Bonn: German Foundation for International Development.
- Prah, K.K.** 1995b. *Mother Tongue for Scientific and Technological Development in Africa*. Bonn: German Foundation for International Development Education, Science and Documentation Centre.
- Ranger, T.O.** 1989. Missionaries, Migrants and the Manyika: The Invention of Ethnicity in Zimbabwe. Vail, L. (Ed.). 1989. *The Creation of Tribalism in Southern Africa*: 118-151. London/Berkeley/Los Angeles: James Currey.
- SAAWK (Suid-Afrikaanse Akademie vir Wetenskap en Kuns).** 2011. *Notule 1 TK Maart 2011.1. 0GBVH.2011-04-12*. Pretoria: SAAWK.
- SAAWK (Suid-Afrikaanse Akademie vir Wetenskap en Kuns).** 2012. *Notule 1 TK Maart 2012.1. 5GBVH.2012-03-23*. Pretoria: SAAWK.

Internet Reference

<http://www.en.wikipedia.org/wiki/orthography>

Developing Legal Terminology in African Languages as Aid to the Court Interpreter: A South African Perspective*

Mariëtta Alberts, *Research Unit for Languages and Literature in the South African Context, Northwest University, Potchefstroom Campus, Potchefstroom, South Africa (malberts@lantic.net)*

and

Nina Mollema, *Department of Criminal and Procedural Law, University of South Africa, Pretoria Campus, South Africa (mollen@unisa.ac.za)*

Abstract: The need for unambiguous communication in the theoretical and applied fields of human activity, such as the legal profession, is constantly increasing. Terminologists and subject specialists are focused on the provision of unambiguous source- and target-language terms for well-defined concepts. In order to achieve this aim, it is necessary to determine the precise meanings of terms which enable users to comprehend and use them in a universally accepted manner. Legal language is one of the most difficult languages for special purposes since only legal experts versed in the special vocabulary can communicate successfully – laypeople tend to struggle comprehending the language of law, sometimes called legalese. This study deals with legal language, the bridging of communication problems in a legal setting by especially court interpreters, and difficulties these legal linguists may experience in this endeavour. Some word-forming principles are discussed and examples are given of typical multilingual coinages in the legal profession. Various matters relating to terminology usage in a multilingual society receive attention. Some of the principles of the standardisation, harmonisation and the internationalisation of the terminology of the legal profession are discussed. The research also addresses the work of the Centre for Legal Terminology in African Languages (CLTAL) and of other relevant terminology projects, e.g. that of the Centre for Political and Related Terminology in Southern Africa (CEPTSA).

Keywords: CRIMINAL LAW, CRIMINAL PROCEDURAL LAW, CONCEPTUALISATION, COURT INTERPRETER, COURT INTERPRETING, HARMONISATION, INTERNATIONALISATION, HARVESTING PROCESSES, LAW OF EVIDENCE, STANDARDISATION, TERMINOGRAPHY, TERMINOLOGY

* This article is based on a paper presented by Dr Nina Mollema on 14 October 2011 on behalf of CLTAL at the University of the Free State in Bloemfontein which was prepared by her and Dr Mariëtta Alberts. The Conference of 14 to 15 October 2011 was organised by the Department of Justice and dealt with several aspects relating to interpreting. The conference theme was "The value and professionalism of the court interpreter".

Opsomming: Die ontwikkeling van regsterminologie in die Afrikatale as hulp aan die hofvolk: 'n Suid-Afrikaanse perspektief.

Die behoefte aan ondubbelsinnige kommunikasie in die teoretiese en toegepaste velde van menslike aktiwiteit, soos die regsprofessie, neem steeds toe. Terminoloë en vakspesialiste probeer om ondubbelsinnige bron- en doeltaal terme vir goed gedefinieerde begrippe te verskaf. Om dié doel te bereik, is dit noodsaaklik om die presiese betekenisse van terme te bepaal ten einde gebruikers in staat te stel om dit te verstaan en op 'n universeel aanvaarbare wyse te gebruik. Regstaal is een van die moeilikste vaktale aangesien slegs kundige regslui wat die spesiale woordeskat onder die knie het suksesvol kan kommunikeer – leke is geneig om te sukkel om die regstaal te verstaan, soms regsjargon genoem. Hierdie studie handel oor regstaal, die oorbrugging van kommunikasieprobleme in 'n regsmlieu deur veral hofvolke, en probleme wat die regslinguïste in hierdie strewe mag ondervind. 'n Paar woordvormingsbeginsels word bespreek en voorbeelde van tiperende meertalige skeppings in die regspraktyk word ter illustrasie aangebied. Verskeie sake oor terminologiegebruik in 'n meertalige samelewing kry aandag. Sommige van die beginsels van die standaardisering, harmonisering en die internasionalisering van regsterminologie word bespreek. Die navorsing verwys ook na die werk van die Sentrum vir Regsterminologie in Afrikatale (SRTAT) en ook na ander tersaaklike terminologieprojekte, bv. dié van die Sentrum vir Politieke en Verwante Terminologie in Suidelike Afrika (CEPTSA).

Sleutelwoorde: BEWYSREG, HOFTOLK, HARMONISERING, HOFTOLKING, INTERNASIONALISERING, KONSEPTUALISERING, STRAFREG, STRAFPROSESREG, EKSERPEERPROSESSE, STANDAARDISERING, TERMINOGRAFIE, TERMINOLOGIE

1. Introduction

Language plays a central role in allowing a person access to justice. The Constitution of South Africa 1996 endorses this supposition by allowing the accused an opportunity to be tried in a language they understand. Section 35(3)(k) of the Constitution states: "Every accused person has a right to a fair trial, which includes the right to be tried in a language that the accused person understands or, if that is not practicable, to have the proceedings interpreted in that language." In this regard, the court interpreter plays a pivotal role in ensuring the accused or witness full participation in the trial by interpreting information accurately. This includes the interpretation of legal language, which is – for the majority of the country's inhabitants – difficult to comprehend. Any linguistic deficiency will effectively deny the accused or witness the right of access to justice.

Even though there is no comprehensive official professional code governing the service of court interpreting, there are norms that court interpreters have to adhere to. In terms of Rule 68 of the Rules of the Magistrates Courts, every court interpreter has to take an oath swearing or affirming to interpret truly and correctly to the best of their knowledge and ability from the language they are called upon to interpret into an official language of the Republic of South Africa and *vice versa*. This process entails upholding the four main tenets of professional ethics, namely, accuracy, competence, impartiality and confiden-

tiality (cf. NAJIT *Code of Ethics* <http://www.najit.org/certification/profession.php>, accessed July 2013). Interpreting skills are especially grounded in the elements of accuracy and competence. It is generally acknowledged that court interpreting is a complex activity depending on the interpreter's skills in various factors such as the following:

- A large personal vocabulary: A court interpreter with a large vocabulary bank in all levels of the interpreter's working languages recognises incoming words, and understands their meaning and usage better;
- Knowledge of the linguistic context (grammar, spelling and orthography, and terminology related to the subject);
- The situational context (the extra-linguistic features, i.e. cultural differences); and
- Knowledge of the subject under discussion (law, crimes, expert evidence, etc.).

Court interpreters are continually encouraged to improve their vocabulary so that they are able to select from their rich vocabulary only those words and terms which best and concisely explain the source message. However, sometimes these interpreters are at a loss for the appropriate legal or linguistic equivalent, mainly because no such term may exist in the target language (TL) at a given moment. This inadequacy and ways to eliminate it is exactly what this article focuses on. In exploring this relationship between the court interpreter, legal linguistics, vocabulary acquisition and terminology development, aspects such as accessing legal language in a multilingual society and the compilation of specialised dictionaries will also be considered.

2. What is legal language?

Language of law, legal language or legalese – these are all various names for a very complex variety of a language for special purposes (LSP). Legal language is first of all language. However, as language is a vehicle for the transmission of scientific and technical information as well as culture (Alberts 2001: 91), legal language – as a specialist language – presents problems to non-legal persons who regard this variety of language as obscure legal jargon (cf. Alberts 1997). This may hinder a successful communication process, which is of the utmost importance in a court setting.

Legal language falls within the ambit of what Eksteen (1969: 35) calls peripheral language. The users of a given language share a common vocabulary – the core vocabulary. This core is extended when a language user's vocabulary is enhanced. The extended vocabulary further expands when the specialised languages of various working environments, such as the legal pro-

fession, are added. Not everyone in a given speech community's vocabulary develops in the same fashion and therefore the terminology associated with a given working environment, subject area or domain is not known to all the users of the speech community, cf. Diagram 1:

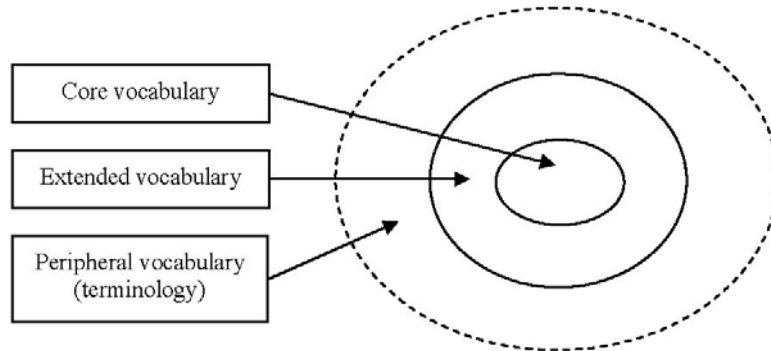


Diagram 1: Core, extended and peripheral vocabulary

According to the diagram on terminology registers (cf. Alberts 1983: 25), legal language falls within the specialised sphere of the humanities since legal language is oriented towards legal professionals who are active in legal institutions such as courts, legal government departments, legislature, universities, the police force, etc. rather than the public at large. Communication is, however, not only between lawyers and legal experts, but also between legal experts and laypeople (non-legal users). And it is here where the problem lies – as the general public does sometimes need to communicate with legal experts, e.g. when they have to appear in court, they do not always understand the legal language since they are not versed in the legal jargon.

Legal language developed from specific legal concepts and as such corresponds with the objects, people and processes it denotes because the terms were formulated through logical thinking and abstraction (cf. Alberts 2001: 90-91). Historical, sociological, political and jurisprudential factors have all contributed to the unique characteristics of legal language (Charrow, Crandall and Charrow 1982: 178-188). Mellinkoff (1963: 3-4) defines legal language as the customary language used by legal professionals. This language is said to include distinctive words, meanings, phrases and modes of expressions. It also includes certain mannerisms of composition not exclusive within the legal profession but prevalent enough to have formed a fixed association. Legal language also contains many Latin terms and phrases which is indicative of elitist and esoteric usage which makes the text sounds prestigious, abstract, formal, authoritative and impersonal. Laypeople without knowledge of the meaning of these Latinisms are excluded from any meaningful participation in courtroom communication.

Many lawyers and legal experts are aware of the fact that the terminology of legal language differs substantially from ordinary usage. However, they tend to attribute the difficulty of understanding legal language to a combination of esoteric vocabulary and the conceptual complexity of the law (cf. Charrow, Crandall and Charrow 1982: 176). Their views of what makes legal language unique differ substantially from what linguistic analysis reveals (cf. Charrow, Crandall and Charrow 1982: 175). Legal experts however recognise that legal language has syntactic constructions that differ from ordinary usage and create ambiguity. They also mention aspects such as Latinisms as well as archaic words and phrases as being problematic. This is evident in the various studies of legal language by lawyers and linguists which have focused almost exclusively on the lexical level (largely on vocabulary) (cf. Charrow, Crandall and Charrow 1982: 175, 176; Alberts 1999, 2001; Cornelius 2012).

Mellinkoff (1963) identifies the following characteristics of legal language:

- **Frequent use of common words with uncommon meanings** (using action *for* lawsuit; party *for* person contracting or litigating; of course *for* as a matter of right; motion *for* formal request for action by court, etc.)
- **Frequent use of Old and Middle English words** once in use but now rare or obsolete (aforesaid, whereas, said, etc.)
- **Frequent use of Latin words and phrases** (in propria persona, amicus curiae, mens rea, prima facie, plaintiff, defendant, status quo, affidavit, alibi, bona fide, quorum, etc.)
- **Use of French words not in the general vocabulary** (lien, easement, tort, etc.)
- **Use of art terms or art jargon** (contributory negligence, fictitious defendant, false pretences, negotiable instrument, eminent domain, etc.)
- **Use of argot** – "in-group communication" or "professional language" (pierce the corporate veil, damages, due care)
- **Frequent use of formal words** (approach the bench; just and equitable; as the Honourable Court pleases)
- **Deliberate use of words and expressions with flexible meanings** (adequate cause, gross malpractice, extreme provocation, informed decision, extraordinary compensation, reasonable man, undue influence)
- **Attempts at extreme precision** ("Know ye that I, _____, of _____, for and in consideration of _____ dollars, to me in hand paid by _____, do by these presents for myself, my heirs, executors, and administrators, remise, release and forever discharge _____, of _____, his heirs, executors, and administrators, of and from any and all manner of action or actions, cause and causes of action, suits, debts, dues, sums of money, accounts, reckon-

ings, bonds, bills, specialities, covenants, contracts, controversies, agreements, promises, trespasses, damages, judgements, executions, claims, and demands whatsoever")

- **Inexactness or indistinctness** ("including but not limited to ..."; "nothing contained herein shall ...")

There are, however, also linguistic reasons for misunderstandings between juristic persons and laypeople. By means of a linguistic analysis of legal language, it is apparent that the following aspects of legal language usage create misunderstandings:

- The usage of **passives** in (past) participial phrases ("to decide all questions of fact *submitted* to you"; "you must not consider ... any statement of counsel *made* during the trial"; "any insinuation *suggested* by a question ... *asked* a witness"; "the reasons *given* for his opinion")
- **Truncated passives** cause problems when they obscure the identity of the agent ("If ... any rule ... *is repeated* or *stated* in varying ways ...")
- **Misplaced phrases**, injected into the middle of clauses ("If, *in these instructions*, any rule, direction or idea is repeated ..."; "a cause which, *in natural and continuous sequence*, produces the injury")
- Phrases beginning with "**as to**" create confusion: ("*As to* any question to which an objection was sustained, you must not speculate *as to* what the answer might have been or *as to* the reason for the objection")
- **Nominalisations** (nouns created from verbs) ("*assumption* of risk"; "... will bar *recovery* of damages"; "failure of *recollection*"; "after a *consideration* of ...")
- **Repetition of entire paragraphs** with slightly different wording and organisation
- Usage of unusual **subordinate clause embedding** ("Whether a discrepancy pertains to a fact of importance or only a trivial detail should be considered in weighing its significance"; "The requirement that affidavits in opposition to summary judgement motions must recite that the material facts relied upon are true is no mere formality.")
- Usage of **unclear anaphora** ("That on or about ____ , plaintiff discovered that the representations made by defendant were false and *he* thereupon elected to rescind the contract, hereinabove referred to, notifying defendant in writing on ____ that he was rescinding the contract.")
- **Wordlists** ("A witness who has special *knowledge, skill, experience, training* or *education* in a particular *science, profession* or *occupation*")

- **Multiple negatives** ("... and innocent *misrecollection* is *not uncommon*") (cf. Charrow, Crandall and Charrow 1982: 175-178; cf. also Cornelius 2012).

From the above linguistic analysis it is clear that legal language is quite distinct from Standard English. One can safely say that legal language developed into a unique variety of English – even more so than other professional languages, i.e. economics, which is more comprehensible to non-experts. As legal language has an important role in regulating society, society has the right to demand more comprehensible legal language in their mother-tongue (cf. Alberts 2001: 90, 91; Cornelius 2012) and especially when involved in a court case.

In this regard the court interpreter plays an important role by not only interpreting court proceedings for witnesses or accused persons into a language they understand, but also making legal language more accessible to them. However, even court interpreters may sometimes not know the appropriate linguistic equivalent for a legal term in the specific TL, mainly because no such term may exist. The absence of an equivalent legal term in the TL is admittedly one of the chief challenges facing interpreters (Matthias and Zaai 2002: 355-356). It is submitted that the development of clear legal terminology, especially in African languages, is an appropriate solution for the formalisation of legal knowledge. As such, legal language can be empowered to be potent media of development. The TLs (i.e. African languages) could also develop into functional languages in the legal domain.

3. Accessibility of legal language in a multilingual society

For legal terminology to become accessible to laypeople, legal experts and court interpreters should simplify legalese when communicating with members of the general public in order to empower them. Laypeople and interpreters should not be afraid to demand further clarification as regards difficult legal concepts from legal experts when necessary, and *vice versa* the experts should cooperate in the process of giving laypeople and interpreters access to these terms (cf. Charrow, Crandall and Charrow 1982: 175-190). Legal language can basically be made more accessible by properly defining legal concepts in understandable words. This is mainly done by terminologists who affix meanings to concepts and their related terms by defining the concepts in a low register or by means of simplified words. Through frequent usage, the concepts are standardised which ensure exact communication. In a typical terminographical environment exact communication is achieved when there is a one-to-one relation between term and concept, and this relationship can be achieved by means of a proper and understandable definition of the concept. This can be illustrated with the terms **arsonist** and **pyromaniac**. Both these terms denote a person who sets fire to an object, but there is a conceptual difference between them:

- An **arsonist** is a *criminal who deliberately sets fire to something, especially a building.*
- A **pyromaniac** is a person who *cannot control the desire to set fire to things, often because of a mental illness.*

Laypersons will understand the legal concepts if they are properly defined – especially if the concepts are defined in a low register or by using simplified words to enhance intelligibility (cf. Alberts 1999, 2010a: 23-24). Similarly, if court interpreters understand the underlying concept, they will be able to interpret it into any target language. It might not be with an appropriate TL equivalent but at least by paraphrasing.

The communication problem that legal language poses is further increased within a multilingual society as the adoption of a multilingual approach to public affairs obviously implies the use of major international and heritage languages such as English and French, but it includes the use of indigenous languages such as Afrikaans and the African languages in the legal sphere as well.

Political and societal changes in the new democratic South Africa have largely transformed the functional role of the indigenous languages. Section 6 of the Constitution further provides for multilingualism and the development of the linguistic heritage. This system of official multilingualism has produced a substantial demand for terminology creation as various business matters (civil service departments, local administrative bodies, courts of law, amongst others) need to be conducted in the different official languages (cf. Government Gazette 2012, 2013).

Multilingualism in South Africa is a sociolinguistic fact to be taken seriously (Alberts 1998: 230). A large proportion of South Africa's inhabitants can only be reached by means of indigenous languages. In a stressful courtroom situation, it is furthermore accepted that legal information is conveyed and assimilated best through the mother-tongue or first language. However, information flow is prevented from being established by factors such as low literacy rates, low levels of proficiency in English and terminologically poorly developed African languages (cf. Fourie 1994: 11-15).

4. Terminological principles regarding conceptualisation and term creation

As previously mentioned, exact communication can only be achieved if the sender of the message and its receiver attach the same meaning to the given message. In the normal communication process, **specialised information** is **encoded** to be conveyed/transmitted from a **sender** (S1) as communication source to a **receiver** (R1) as recipient of information who will **decode** the **message** and then react on the stimulus received, i.e.:

S1 → specialised information → R1

Ambiguity in the specialised information may give rise to confusion and distortion of the communication process. One way to ascertain the exact meaning of a message conveyed through the medium of language is to document and standardise the terminology of languages for special purposes (LSP). When everyone in a specific language group working in a similar working environment understands the same message conveyed by a specific concept denoted by a specific linguistic label, that is the term, one can consider the term to be standardised. This is precisely what terminology is all about. It is the task of the terminologist to make sure that basic terminological principles, cultural differences and language attitudes of the professional group (i.e. legal experts) are taken into account when denoting concepts and coining terms.

As background to vocabulary development and the acquisition of related terminology, a brief explanation of terminological theory is required. According to Sager (1990: 4), terminology is concerned with "the study and use of the systems of symbols and linguistic signs employed for human communication in specialised areas of knowledge and activities". Terminology is "a representation of an equally coherent, but possibly differently structured system of concepts" (Sager 1990: 114). A term is created when various linguistic labels are used to describe or name a specific object or concept. It therefore refers to a definite concept which is clearly defined within specific parameters. A term is as such the linguistic representation of a mental construct. There is a special interrelationship between the symbol, the concept (that is its mental representation in one's brain) and the various linguistic labels used in different languages to describe the object and concept (cf. Diagram 2). If this does not exist, a misunderstanding or miscommunication will result (Sager 1990: 57).



Concept:

Definition: That branch of government which is charged with the making of authoritative and enforceable rules (laws) for a society.

SL term: eng: legislative authority

TL term equivalents:

afr: *wetgewende gesag*

zul: *umkhandlu wesishayamthetho*

nso: *lekgotlatheramelao*

(cf. CEPTSA n.d.)

Diagram 2: Conceptualisation and term creation

According to Alberts (1990: 101), "one of the aims of practical terminological work (terminography) is to order the conceptual systems of subject fields". In terminography, attention is given to the dictionary user's (i.e. interpreter's) needs for specific terms for specialised communication in languages for special purposes, e.g. in certain subject areas, subdivisions or domains (cf. Alberts 1990: 109). This basically entails developing the required terminology.

There are three basic dimensions governing terminology development (cf. Sager 1990: 14-128), namely:

- a cognitive dimension, which relates the linguistic forms to their conceptual content, i.e. the referents in the real world (cf. Diagram 2);
- a linguistic dimension, which examines the existing and potential forms of the representation of terminologies (specialised domains as well as orthographies of SLs and TLs); and
- a communicative dimension that looks at the use of terminologies, and especially at standardisation processes (cf. also Alberts 2010b: 19).

The communicative dimension involves (cf. Alberts 2010b: 46-56):

- **language development** – i.e. terminology development in a variety of subject areas and domains);
- **information retrieval** – reference sources such as terminology lists, dictionaries for special purposes, databases, Internet;
- **coining of terms** – various term creation processes;
- **standardisation** – the frequent usage of terms within specialised environments leads to standardisation of the terms denoting the underlying concepts;
- **cultural differences** – misunderstanding owing to different conceptual environments;
- **consultation** – terminography should be handled through consultation with subject specialists, linguists, language practitioners;
- **dissemination** – terminology should be disseminated to be used in the communication process and to ensure the standardisation of the terms within a subject area or domain.

It is argued that terminography is therefore primarily prescriptive since terms and their underlying concepts need to be standardised to ensure exact communication (cf. Alberts 1990: 109; Alberts 2001: 78). Standardisation involves that terminology must be used correctly and consistently by the author/sender of the information and be known and understood by the reader/recipient of the information. Only if both participants in a speech act know the special ref-

erence of a term, can exact communication and knowledge transfer succeed, e.g. between subject specialists (e.g. lawyers), and between subject specialists and laypeople (cf. Alberts 2010b: 46-56). In the courtroom milieu, court interpreters facilitate this process while making use of terminology in both SLs and TLs.

The participation of a court interpreter as a facilitator in a legal communication process means that the communication activity consists of three, rather than two participants. The court interpreter's role is thus two-fold, first as a receptor of information and secondly as a source of information. As a receptor (R1) he receives the original input from the original speaker, in the source language (SL); and as S2 he relays the SL input to the listener who is now the second receptor (R2) in the listener's TL, i.e.:

S1 (SL) → specialised information → R1 (SL)/S2 (TL) → specialised information → R2 (TL)

This input is usually an utterance made up of words, phrases, sentences and specialised terms (sometimes accompanied by extra-linguistic features, i.e. tone or visible gestures such as facial expressions and other body movements, which will not be discussed here). The interpreter therefore needs background information on the topic under discussion so as to identify any possible lack of knowledge of, or unfamiliarity with legal terms before court proceedings. Accordingly, the court interpreter may either consult a legal dictionary, or if such a term does not exist, attempt to define the term himself. Whatever the case may be, it is clear that there is an urgent need for the development of legal terminology in the African languages as an aid to the court interpreter.

4.1 Terminology development

With the development of new terminology, aspects to be taken into account are that terms are created according to the word-forming principles and spelling and orthography rules of the standard language. A standard language is a codified and uniform language that has grammar, spelling and orthography rules, dictionaries and adequate terminology for the demands made. The primary aim of term creation is, as already seen, to promote communication in all the domains of the working environment, especially in developing environments such as science and technology. In these fields, terms are documented and systematised in order to create LSP, to compile technical dictionaries/language for special purposes dictionaries, terminology lists or electronic term banks. These reference works allow for terms to be standardised through their frequent usage by subject specialists, language practitioners and laypeople.

4.1.1 Conceptual interrelationship of terms

The various concepts belonging to a specific subject or specialised domain

(such as criminal law) have to be collected in a systematic way because terms stand in a specific conceptual relationship to other terms. It is only by dealing with the various related terms and concepts as a whole that terminologists can ensure the correct naming of concepts. This can be illustrated with the term **common law**:

eng	afr	nso
common law 1 n.	gemene reg	molao wo o sego wa ngwalwa
common law 2 n.(general law)	gemene reg	molao wa bohle nageng
Common law 3 n.	Common Law	Molao wa setlwaedi/ Molaotwaedi
common-law marriage	gemeenregtelike huwelik	lenyalo la molaotlwaedi
common-law wife	gemeenregtelike vrou	mosadi (cf. CLTAL database 2013)

In South Africa, the common law (meaning 1) is based on Roman-Dutch law. The Common-law system (meaning 3) originated from the English legal tradition in contradistinction to Civil law (Roman law). Common law (meaning 2) refers to the law that commonly applies to all inhabitants of the country. Northern Sotho translators of this term originally suggested that the phrase *molao wa mmuso* be used as an equivalent for *common law* (general law) (meaning 2), but it was felt that this phrase refers more to *public law* which governs the relationship between the state and its subjects or between government institutions *inter se*, and not to the relationship between people. As such, the phrase *molao wa bohle nageng* was opted for. Cultural differences also have to be accounted for in translations. In African customary law, the common-law wife is regarded as a marriage partner with full marital status because 'lobola' has been paid for her. She is regarded in a totally different light from a concubine whose relationship may be of a temporary nature.

4.1.2 Methods of term creation

Terms are not created in a haphazard way. There are specific ways of supplying term equivalents in the various languages, such as borrowing, transliteration, total embedding, neologisms, etc. (cf. Alberts 1999). Terminologists, subject specialists, linguists and language practitioners (such as court interpreters) who have to supply TL term equivalents for SL terms have a variety of ways to apply, namely:

- *loan words* (from English, Afrikaans, Dutch, Flemish, German, French, Spanish, Italian, etc. (mainly major modern languages), e.g.:

eng	affidavit	afr	beëdigde verklaring	nso	afidavite
eng	Hansard	afr	Hansard	nso	Hansard, Hansate
eng	abortus	afr	abortus	nso	abortuse

The term equivalent (e.g. Afrikaans or Northern Sotho equivalents) must be congruent with the word-forming principles and spelling conventions of the language. The *ff* combination is not normally used in African languages. According to the rule, a consonant must be followed by a vowel and one should adhere to the rules of the specific language, the term equivalent in Northern Sotho is therefore: *afidavite*.

Also, the term **Hansard** (cf. English/Afrikaans) [the official published report of proceedings in the British parliament; named after Luke Hansard (1752–1828) and his descendants, who compiled and printed the journals of the House of Commons until 1889] was borrowed in Northern Sotho: **Hansard** but the synonym **Hansate** was coined according to Northern Sotho orthography.

Certain terms (i.e. **abortus/abortus/abortuse**) are regarded as being sensitive and in African culture form part of mainly taboo or women's language (called *hlonipa* in the Nguni language group and *tlotta* in the Sotho group of languages). It is not part of the natural language of all the people. Although the idiomatic term *lefolotsana* describes abortus in terms of animals, the word is regarded as obscene when applied to humans and therefore the borrowed term **abortuse** is preferred.

- *transliteration* (the use of the Greek or Latin stem to coin a TL equivalent):

addendum (Latin) neutral of *addendus*:

eng addendum afr toevoeging nso adentamo, tlalletso

bail ME *bail, baille*, from MF *bail*, from *bailler* to give, deliver, from Latin *bajulare* to bear a burden, keep in custody, from *bajulus* porter, load carrier:

eng bail afr borg nso peile

In the last example, the SL (English) was used to coin the Northern Sotho term equivalent **peile**. A "p" was used in the TL instead of a "b" to obtain the correct Northern Sotho pronunciation. Terms that are internationally recognisable, i.e. transliterated or borrowed from major international languages such as English, French, German, Spanish, etc., have the added advantage that experts in the field would be able to recognise and comprehend them.

- *total embedding or adoption of a term in its original form without transliteration*

With this method, the original term – which usually stems from classical languages such as Latin or Greek – is embedded in the TL without changing or adapting any part of the original SL term. The terms are usually explained in the TL to enable users to use it appropriately:

Latin	eng	afr
ad litem	for the case/suit	vir die hofsaak/regsgeding
bona fide	in good faith	te goeder trou
cadit quaestio	the question falls away	die vraag verval
contumelia	insult	smaad, belediging
curia advisari vult	the court wishes to	die hof wil sy beslissing oorweeg
(cur. adv. vult)	consider its decision	
non compos mentis	not in full possession of his faculties/insane	nie by sy volle verstand nie
prima facie	at first sight	met die eerste oogopslag
quasi	as if	asof
sine	without	sonder
sub iudice	pending	nog hangende
summa ratio	decisive reason	deurslaggewende rede
uxor	wife	eggenote, vrou
verbis	with words only	slegs met woorde
vice versa	the other way round	omgekeerd
viva voce	orally	mondeling

(cf. Tweetalige Polisiewoordeboek / Bilingual Police Dictionary 1994: 436-441)

– *extension of meaning* (using existing words but broadening the meaning to encompass the new concept), e.g.:

eng abet	afr aanmoedig	nso go thusana le go thusa bosenyinying
eng abductor	afr ontvoerder	nso motshabisamongwe

Note 1: Cf. international alpha codes: three-letter abbreviations for names of languages: <http://www.loc.gov/standards/ISO639-2/php/code>.

The Northern Sotho equivalent of abductor, namely 'motshabisamongwe' literally means a person who steals another person, from the word *-thopa* (captives in a war). This meaning is foreign to the Sotho group of languages. This concept is known in the Nguni tradition as *ukuthwala*. This traditional term has a positive meaning and is used when a man captures the girl he loves and intends to marry, while the term *ukuthumba* has a negative meaning. Terminologists have to know the exact meaning of each word to avoid ambiguity. In most cases, the subject or the context determines the meaning and therefore determines the choice of a term.

– *neologisms* (the coining of a new term), e.g.:

eng polygraph	afr poligraaf	nso seutoliamaaka
eng adopt	afr aanneem	nso adopta

This word 'adopt' is common in Western culture but according to African culture a child can never be attained by nursing and feeding it; as such, a new word had to be coined.

Again, when designating new concepts by means of terms, the underlying meaning of the concept should be understood before coining a term, cf.

eng hit <i>n.</i> <slang> [a murder carried out as the result of an underworld vendetta or rivalry]
afr huurmoord 1
afr huurmoord 2, sluipmoord
nso polaotlhoyo 1 [a killing committed out of hatred]
polaorongwa 2 [a killing when someone has been sent, or contractually employed to kill another; a killing when a killer is hired to assassinate someone] (polaopakišano, polaonkwa , replaced by: polaokwanelwa)
Note 1: The term hit could have more than one meaning – the context would dictate the term to use.
Note 2: Although the English and Afrikaans terms are general words, the Northern Sotho terms had to be created.

One should also take the orthographical rules of a language into consideration when coining terms. When working in a language with a disjunctive orthography one cannot provide a single equivalent for a given term in the same way as in the case of a language with a conjunctive orthography. The Afrikaans orthography allows for a conjunctive manner of writing and therefore Afrikaans terms tend to be one word. English and some of the orthographies of the African languages prescribe a disjunctive manner of writing and therefore the terms tend to consist of more than one word. Although terms may seem to consist out of more than one word in the African languages, these can still be regarded as terms since the orthographies prescribe disjunctive writing.

eng hit list (disjunctive)
afr huurmoordlys (conjunctive)
nso lenaneo la polaorongwa (disjunctive) [this term could also refer to the plans or tactics that are going to be used]
lenaneoina la polaorongwa (disjunctive) [a list of people to be murdered]

When coining terms (neologisms) for new concepts in any of the South African official languages, it is of the utmost importance to consult with the other languages to see whether the concept is already named. By consulting other languages, it is much easier for the terminologist to coin a new term in the relevant language according to the existing terms in the other language(s).

4.1.3 Specific term-creation considerations

– Culture-specific nature of terms

As seen in some of the examples already provided, culture plays an important role in the translation of terms. Owing to the culture-specific nature of some terms they cannot always be translated successfully by the different culture groups. Although to the layperson it may seem that all concepts are represented in all languages, it is definitely not the case and the speakers/interpreters/translators/terminologists experience great difficulty if the national and cultural backgrounds of the SLs and TLs differ from one another. For example, in a certain case the Northern Sotho exclamation: "Sebatakgomo" (English: a cry for help) by a plaintiff (a woman who was raped) was misinterpreted by a Sesotho court interpreter as "Ke batla kgomo" (English: "I want an ox") and as a result the perpetrator (rapist) was set free. In another case, an accused was wrongfully acquitted based on the erroneous interpretation of an African term literally meaning "arrow", but "gun" in the context of the trial (cf. Mikkelsen 2008: 87).

As such, a person working on a multilingual legal terminology project must not only have a sound background of legal approach and court procedure coupled with a good command of Afrikaans and/or English and the vernacular of the relevant language groups, but also knowledge of indigenous culture. It is important to realise that the terminology list or dictionary will be the channel through which legal and linguistic information will flow. This terminology list or dictionary should also contain extra information on usage, especially for the court interpreter. It should not only be a glossary of legal terms, it should guide the user on language and style usage, on the difference in interpreting and translation of certain concepts, and on misinterpreting, misstatements and misrepresentations. The translated equivalent must also be acceptable to the community:

eng	half-blood <i>n.</i> [the relationship between individuals having only one parent in common]
afr	halwe bloed [met een gemeenskaplike ouer]
nso	morwarre; hlaba

In this example, an equivalent with exactly the same meaning as in English does not exist in Northern Sotho owing to cultural differences. All the children in a Pedi family (Afrikaans: gesin) are brothers and sisters, irrespective of the fact that some of the members of this family may have only one parent in common. The term *hlaba* already exist for this concept, but has a negative connotation as it could be used for an illegitimate child. However, the term *hlaba* is used in fighting language, for example when one contends for a higher position, for instance for chieftainship. In this regard, the context will determine

which term (*morwarre* or *hlaba*) to use.

– Connotation

When supplying or coining term equivalents to SL terms subject specialists, linguists and language communities should be consulted. It is of great importance that consensus should be obtained since subject specialists know the subject or domain and linguists could give authority to the term equivalents. By getting consent from subject specialists as well as linguists the terms will be used by both experts and laypeople in the user's environment. Only terms that are both technically and linguistically sound will penetrate into a language and will be disseminated in the subject field through the language.

Although terminology is by nature abstract and exact, and is therefore to a great extent secure against emotive connotations that can be attached to words, one should be aware not to incorporate potentially offensive and/or sensitive items in a term list or dictionary. This can be illustrated with examples from the social and political subject fields with which the legal domain overlaps to a certain extent, e.g.:

eng **homosexual** *n.* (**general**) [*a person sexually attracted to somebody of the same sex*]

afr homoseksueel

nso morobalalewabongbjagwe (**coinage**)

lehomo (**synonym, loan word**)

lehomosektshuale (**synonym, loan word**)

eng **gay** [*a man who is sexually attracted to other men*]

afr homoseksueel

nso morobalalewabongbjagwebonna (**coinage**)

lekei (**synonym, loan word**)

eng **lesbian** [*a woman who is sexually attracted to other women*]

afr lesbiër

nso morobalalewabongbjagwebosadi (**coinage**)

lelesbiene (**synonym, loan word**)

Note: The *le* class is being used in Northern Sotho for matters unacceptable to the culture; *mo* is being used for things out of the ordinary.

– Harmonisation of terms

Where terms are comparatively similar in the different African languages, the harmonisation of terms should be considered. In terminology harmonisation, the terminologist maintains the basic existing term and also develops it by extending its meaning to match the meaning of a new concept, and adjusting it to the orthography rules of the particular language (cf. Alberts 1998: 235).

Nahir (1984: 308) defines this process as "establishing unified terminologies, mostly technical, by clarifying and defining them, in order to reduce communicative ambiguity". The localised term will be nationally recognisable within the relevant language groups since it would consist of the basic stem with language-specific adjustments, i.e. orthographic conventions (cf. Alberts 1998: 235).

An area where the principles of harmonisation can be applied is in the development of the terminologies of the Nguni and Sotho languages. Since the Nguni and Sotho cultural languages have a sizeable core of common vocabularies, it would be ideal if they also had a common pool of terminologies (Msimang 1996: 20). In this case, terms of modern concepts can generally be shared by the Nguni and Sotho groups. In these cases, it will then be unnecessary to coin a separate term for Ndebele (isiNdebele), Swati (Siswati), Xhosa (isiXhosa) and Zulu (isiZulu) (Nguni group) or for Northern Sotho (Sesotho sa Leboa/Sepedi), Southern Sotho (Sesotho), Tswana (Setswana) (Sotho group). A team of terminologists for all the Nguni languages and another team for all the Sotho languages (consisting of members of the separate languages within the specific language group) can work together when denoting concepts in the related languages, e.g.

bribe *v* (eng): omkoop (afr); -reka (nso); -reka (tsn)

trade *n* (eng): handel (afr); kgwebo (nso, sos, tsn); ukuhweba (zul)

law *n* (eng): wet (afr); molao (nso, sot, tsn); umthetho (xho, zul)

– Internationalisation/globalisation and the South African legal system

When creating new legal concepts, one could also consider adopting existing standardised terms from international and global, sources of law, provided that they are precise denotative equivalents of national concepts (cf. Kierzkowska 1995: 129-140, 1999). This practice entails the internationalisation and globalisation of a legal terminological system's concepts. The easiest way to achieve this is to build a primary system of legal terminology and a secondary system of legal terminology concurrently. This takes place in countries such as Canada and Finland, where legal regulations are passed simultaneously in English and French or in Finnish and Swedish. It is also the case with the European Union which does not need to be internationalised because it is international by definition. However, according to Kierzkowska (1995: 129-140, 1999), the internationalisation of legal terminology within the traditional definition is difficult as international legal systems do not seem to be compatible.

Similarly, the language practitioner working with legal terminology in a multilingual society such as South Africa encounters interesting problems. Although Roman-Dutch law forms the basis of the South African legal system, the legal system is also greatly influenced by inherited legal systems such as English Common law and indigenous law in an African context. Conflicts often occur. For example, according to English Common law, two terms are used to describe the South African criminal offence of defamation, namely libel (for written forms of defamation) and slander (for verbal forms of defamation). No

such distinction is made in South African criminal law.

The various legal systems pose a serious problem when determining the exact meaning of a concept. It is therefore necessary to excerpt legal terms from existing documentation and first of all to define these concepts accurately in the South African context before it is possible to denote them with equivalents in Afrikaans or any of the African languages in South Africa. The idea of internationalisation/globalisation can therefore be implemented within a limited scope only, namely:

– to use terms with Greek and Latin cores such as:

eng democracy **afr** demokrasie **zul** intando yeningi **nso** mmušo ka batho,
temokrasi

In this example the English and Afrikaans terms were transliterated and the Northern Sotho term 'temokrasi' was borrowed.

Localised terms are still spelt according to the orthography of the specific languages (i.e. **afr** 'demokrasie' and **nso** 'temokrasi') but by sharing the same stem legal experts and language practitioners can recognise the terms and with a knowledge of the subject field they can therefore easily derive the meanings attached to the terms. This may lead to better communication worldwide.

In so far as legal systems are in principle compatible, it is possible to standardise the terminology associated within the legal systems. If a concept does not, however, exist in the legal system of a specific language group, it is very difficult to denote it in the language. Problems are for instance encountered when non-existing concepts are imported into the African languages from foreign legal systems – e.g. the concept denoted by the terms **eng** 'high treason' and **afr** 'hoogverraad' for which no Northern Sotho term could be found, because the concept is foreign to the speech community (cf. Alberts 1998: 238). The Northern Sotho term 'sa go uša mmušo' was coined.

Full internationalisation/globalisation of localised or national legal terminologies will, according to Kierzkowska (1995: 129-140, 1999), never be possible. It will, though, certainly be possible in the future to gradually internationalise/globalise systems of legal concepts to be followed by the gradual internationalisation/globalisation of local or national terminologies. Before it becomes universal, terminologists should apply existing internationalisms/globalisms, exchange their local or national terminologies on a worldwide scale and support every effort made towards internationalisation/globalisation and standardisation of legal concepts.

4.2 Terminology harvesting methods

The court interpreter needs to know where to find terms. The terminology needed for communication within a specialised environment such as court interpreting (i.e. criminal law, criminal procedural law and law of evidence) is

not necessarily available in specialised dictionaries or term lists. The court interpreter might need to harvest terms from various sources. It is therefore important to discuss the basic methods of terminology harvesting methods, namely subject-oriented terminography (SOT), translation-oriented terminography (TOT), and linguistic community-oriented terminography (LCOT).

4.2.1 Subject-oriented terminography (SOT)

Subject-oriented terminography (SOT) is the traditional method used for term extraction or term harvesting. In the case of SOT, terms are harvested within a given subject area in terms of the conceptual framework (cf. Alberts 2010a: 72). The point of departure is therefore the conceptual framework of the subject or domain.

SOT underpins primary term creation processes and terms are created first in the SL. During the SOT process, terminologists are highly dependent on the input and collaboration of subject-field specialists and experts of various occupational domains, i.e. the legal profession. They have to liaise with various subject committees and linguists who assist with the supplying of terms and their definitions in the SL (i.e. primary term creation) since they, being knowledgeable experts, are the best equipped. Subject specialists also assist with the supplying of term equivalents in TLs, i.e. secondary term creation (cf. Alberts 2001: 79). Trained terminologists assist with the data capturing and dictionary compilation process.

Example of SOT term extraction:

Subject: Law, more specifically criminal law; primary term: *arrest*, which can be both a noun and a verb. Additional terms can be extracted from the SL term:

SL Terms		Additional Terms
arrest (n)	>	arrest (v)
lawful arrest	>	unlawful arrest
arrestee	>	arrestee

4.2.2 Translation-oriented terminography (TOT)

In the translation-oriented terminography (TOT) process, terms are extracted from available SL and TL texts (cf. ISO 2002; Alberts 2010a: 72-80). These texts are aligned and SL terms (primary term creation) and TL equivalents (secondary term creation) are matched (cf. Diagram 3). TOT term extraction methods are used when SL and TL terms are both available in texts, i.e. when a source document in the SL is translated into a TL. Translators and interpreters could use this extraction method to align SL and TL terms in order to compile term lists, e.g.:

Example of TOT term extraction:

SL Terms		TL Terms
binding authority	>	bindende gesag
the authorities	>	die gesag
local authority	>	plaaslike bestuur

The TOT term harvesting process could be illustrated as follows (cf. Diagram 3):

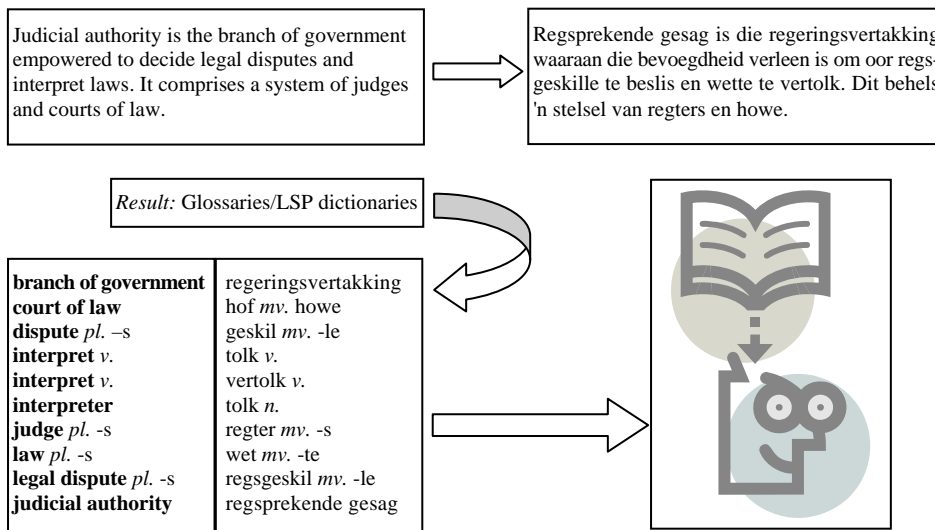


Diagram 3: Translation-oriented terminography (TOT) term harvesting process

4.2.3 Linguistic community-oriented terminography (LCOT)

The linguistic community-oriented terminography (LCOT) process makes use of fieldwork to harvest available terms from linguistic communities (cf. Alberts 2010a: 72). Terminology is gathered through i.e. indigenous knowledge systems (e.g. indigenous law, medicine, herbs, bird and animal names, etc.). Dialects also provide several terms of previously denoted concepts. These terms might be archaic or even obsolete but they could get new life within a conceptual cluster. The value of this method of harvesting is that terms are already available and known in the TL.

The various harvesting processes discussed above support the primary (SL) and secondary term (TL) creation processes. In order to be as comprehensive as possible, translators, interpreters and terminologists should work together to harvest, document, and disseminate these terms. This is done as a first process in order to standardise terminology. However, as emphasised

before, terms will only be standardised when they are frequently used by subject specialists and language practitioners and therefore are able to penetrate into the subject field and languages concerned (i.e. SLs and TLs). One can create the most beautiful terms possible, but if they are not used, they will die a slow death; e.g. the coinage of the isiZulu term *umabonakude* which was disregarded in favour of the English form *ithelevishini*. As such, terms will only penetrate a subject area and language through **continued use**.

4.3 Harvesting methods for the court interpreter

All the terminology harvesting methods described above are of value to the court interpreter. Dictionaries or term lists compiled according to the SOT method would supply terms and would immediately correspond to terminology needs – but this is only relevant in cases where there are special purpose dictionaries available in the given subject field or domain. The terms might even be available in the required SL and TL. In cases where the terminology is not available in the TL, the interpreter still needs to get TL equivalents which might not be standardised or verified by subject specialists, linguists or other language practitioners (i.e. the National Language Bodies (NLBs)).

In the case of terminology acquired through field work (LCOT process), the terms come from the language users themselves and would probably already be in the required TL. These terms would probably also already be standardised through frequent usage, although not documented in dictionaries.

The translation-oriented terminography (TOT) harvesting process will probably work best for court interpreting, where an interpreter would receive documents prior to a hearing. The court interpreter could then study the court documents and with the use of subject-specific dictionaries match SL terms with TL term equivalents. The court interpreter could then be able to compile a term list to prepare for the hearing and to utilise it during the hearing.

In this manner, a court interpreter could compile a more permanent reference resource by also utilising the TOT harvesting method. In this process, the court interpreter has to study SL and TL texts (court documents, legal texts, etc.) to align sentences, and also to identify and match SL terms with TL equivalents. The court interpreter will then extract key SL terms and TL equivalents, record SL and TL terminological information according to a style guide (predetermined, standardised format of a database program), and supply definitions, where possible. It is important though that the court interpreter discuss the SL terms and definitions and TL term equivalents with an established domain-specific terminology committee in order to finalise definitions and term equivalents. The terminology also needs to be authenticated by a language-specific NLB. After updating the terminological database, the information needs to be published and disseminated to fellow language workers, i.e. translators and interpreters and subject specialists, i.e. legal experts.

The TOT harvesting process has certain advantages for court interpreters.

It provides for the continuous recording of terms and neologisms – as the court interpreter creates the terms, he records them. This also leads to the quick and easy retrieval of reliable (verified) terminological information. The new terminology will already be used in court though they may not be standardised according to prescribed terminological processes. However, by frequent usage the terms may become standardised. The continuous recording and usage of terminology further implies continuous support and interaction between translators, interpreters and terminologists. One may also use any language as SL. The TOT process is an easy way to streamline the terminology harvesting process (data collection) and it also facilitates the terminology coordination process. As such, TOT is a reliable method of developing multilingual and polythematic terminology and the South African indigenous languages into functional languages.

On the other hand, the TOT process does require dedication and inputs from various collaborators. There has to be frequent and ongoing harvesting of terminology (and related information) in both SL texts and TL translations. Collaboration can take place via various terminology offices (terminologists, translators and linguists), specific bodies such as the Terminology Coordination Section, National Language Service (NLS), Department of Arts and Culture, the National Lexicography Units (NLU) and the National Language Bodies (NLBs), amongst others.

5. The process of compiling specialised dictionaries: an illustration

A combination of all three practical harvesting methods is recommended when a proper subject-related dictionary is compiled. In this case, the linguistic communities, subject specialists and language practitioners such as translators and interpreters not only collaborate in the supplying of terminology, but they are also the beneficiaries of the subject-related dictionary. A legal dictionary for example would be to the benefit of legal practitioners but also to court interpreters and the members of the community who have to go to court or deal with law enforcement issues.

To make this process more practical, an illustration will be provided by way of the research done by the Centre for Legal Terminology in the African Languages (CLTAL) in the process of compiling a multilingual subject-related dictionary on criminal law, criminal procedural law and law of evidence.

The members of CLTAL have divided the process into four phases (cf. Alberts 2008). The first phase consisted of establishing a team which comprises of subject specialist(s), terminographer(s)/terminologist(s) and linguist(s). This team compiled a basic SL (English) terminology list. For example, the subject specialists from various legal fields extracted terms and their definitions from legal texts, study material, etc. The terminographer or terminologist advised the subject specialist with regards to the extracting of terms. Simplified but

exact definitions and examples were provided to the team. The information supplied in the definitions was deducted and new definitions were compiled using the information gathered from various sources. The terminographer or terminologist advised the subject specialist with regards to the processes concerning the defining of terms. The definitions were again verified by external legal experts.

The second phase entailed the conformation of a SL terminology list with terminographical procedures and standards. The data was consequently computerised (the WordPerfect database which was originally used was converted into a TshwaneTerm database).

The third phase involved the translation of terms, definitions and examples in the TLs. Northern Sotho was chosen as the first African language into which terms would be translated, because in 1987 the then Government of Lebova was the first to draw attention to the need. However, CLTAL serves as an umbrella institution to also develop legal terminology in other African languages. The next language for which funds were acquired is Zulu. Delegates of several other African languages regularly attend the quarterly CLTAL meetings with an aim to add other languages such as Venda, Tswana, Sotho, Swati and Xhosa.

During the translation process, various inputs by collaborators such as language specialists or linguists as well as subject specialists were depended on to accurately translate the terms and definitions. The data was then captured in the TL(s) (i.e. Afrikaans and Northern Sotho). Problematic terms were discussed at quarterly meetings with representatives from different disciplines (e.g. lawyers, magistrates, court interpreters, members of the Justice Training College, the Department of Justice, the SAPS, members of the previous Language Boards, current NLBs (i.e. Afrikaans and Northern Sotho NLBs), NLU (i.e. the Sesotho sa Leboa NLU), Provincial Language Committees (PLCs), the SABC, the Department of Education, the Tshwane Municipality, the SABS, AFRILEX, academics from legal and African language departments, linguists, terminologists, translators, researchers, anthropologists, etc.). All members of CLTAL are voluntary workers (which is an example of the dedication required for the terminology creation process).

The final and fourth phase comprises verification and authentication by NLBs of the relevant TLs. Before this stage, other language specialists and experts in law were appointed for editing and quality assurance purposes. Lastly, the dictionaries are published and disseminated. This dictionary is an example of the SOT process.

6. Terminology from CLTAL and CEPTSA

The terminology/terminography process in South Africa favours English. English is usually regarded as the SL since it is through the medium of English

that most concepts and terms reach the population. Once the English terms are excerpted and supplied with a definition and/or example sentence, they can be supplied with equivalents in the ten indigenous languages.

6.1 Terminological examples from CLTAL

The examples provided are from the SL in English to the TLs in Northern Sotho and Afrikaans. The Afrikaans term equivalents were mostly available since they were already documented in bilingual English/Afrikaans legal dictionaries (cf. Alberts 2012: 15-22).

6.1.1 Neologisms

Some of the Northern Sotho terms were available in general bilingual English/Northern Sotho dictionaries but most Northern Sotho terms did not exist and they had to be created as neologisms.

eng	nso	afr
baby snatcher	seutswalesea	babadief
bush lawyer	moitiraramoloa	bosprokureur
parental kidnapping	tšhabišokamotswadi	ouerontvoering
pilferage	bohotšwana	ontfutseling
pimp	monyakedithobalanong	koppelaar
premeditated crime	bosenyipeakanywa	voorbedagte misdaad
premeditated murder	polaopeakanywa	voorbedagte moord
preponderance	phetano	oorwig
primary evidence	bohlatsehlatse	beste getuienis
polygraph	seutoliamaaka	poligraaf
prime (witness)	sebela (hlatse)	(getuie) afrig
principal	mosenyimogolo	dader
puberty	botšwamahlalagading	puberteit

6.1.2 Homonyms

CLTAL deals with homonyms in the following way:

accessory¹ *n* aiding or being concerned in the commission of a crime [e.g. to a crime]

medepligtige *n* help of betrokke wees by die pleging van 'n misdaad [bv. 'n misdaad]

accessory² *n* additional; subordinately contributive [e.g. in respect of crime]

akcessoor *n* addisioneel; wat bykom [bv. by misdaad]

6.1.3 Synonyms

Terminographical information is supplied in the main entry and the synonym is referred to the main entry, e.g.

custom *n* usual or habitual practice; typical mode of behaviour

gebruik *n* gewone of gebruiklike praktyk; 'n tipiese wyse van optrede of gedrag [gewoonte]

custom gewoonte ► gebruik

6.1.4 Cross-referencing

A cross-referencing system assists the user with related terms, e.g.

supreme court

► *see also* high court

hoogeregshof

► *kyk ook* hoë hof

In the case of related terms such as *particide*, *matricide*, *fratricide*, and *sororicide* the terms are also referred to similar concepts, e.g.

patricide¹ *n* person who murdered his/her own father

patricide² *n* murder of a father by his child

patricide

► *see also* matricide

vadermoordenaar

► *kyk ook* moedermoordenaar

patricide

► *see also* matricide

vadermoord

► *kyk ook* moedermoord

6.1.5 Abbreviations

Information is supplied at the main entry and the abbreviation is referred to the main entry, e.g.

A.D.

► Appellate Division

A

► Appèlafdeling

KB

► Kgorotsheko ya Boipeletšo

Appellate Division A.D. highest court of appeal in the RSA before ... [historic, *see* Supreme Court of Appeal]

6.1.6 The usage of Latin terminology in the CLTAL database

Since Latin terminology forms an integral part of the legal field, these terms were also incorporated into the CLTAL database. In some cases, the Latin terms were translated into Northern Sotho, although they are used as such in English and Afrikaans, e.g.

- Latin terms that are translated into Northern Sotho:
eng prima facie **nso** ponagalomathomong **afr** prima facie

Some Latin terms are, however, embedded as such in the different languages but are defined or explained in the relevant languages:

- Latin terms that are embedded into the various languages:
eng dolus (*intention*) **nso** dolus (*maikemisetso*) **afr** dolus (*opset*)

6.2 Terminological examples from CEPTSA

Subject specialists, i.e. legal experts and language practitioners such as court interpreters should also consult related terminological resources such as political terms. The Centre for Political and Related Terminology in Southern African Languages (CEPTSA) concentrates on political terms. It has compiled a bilingual English/Afrikaans dictionary, i.e. the *Modern Political Dictionary* (MPD) 2002 which was revised as a result of responses to individual enquiries concerning the translation of new terms. Term acquisitions occurring this way up to the end of 2010 have been incorporated into the latest issue of MPD in 2011 (cf. Botha, Le Clus and Venter 2011).

Subsequent to the publication of the bilingual dictionaries, CEPTSA also concentrated on creating the groundwork for the multilingual version. CEPTSA has already defined 1 000 core terms in the field of politics and has translated them into Northern Sotho (nso) and Zulu (zul) with the collaboration of academic colleagues (cf. CEPTSA n.d.). A further 1 500 terms are at present being defined with a view to adding them in the same languages and making the 2 500 terms available as an interim publication in all official African languages (cf. Botha, Le Clus and Venter 2011: 4).

The political and related terminology is also a valuable source for court interpreters since these terms are all defined in the SL and TL. The following SL terms and TL equivalents are examples of the multilingual terminology available in the interim publication (cf. CEPTSA n.d.):

eng	afr	zul	nso
election	verkiesing, eleksie	ukhetho	kgetho
enact	verwetlik	ukumisa umthetho	tihomo ya molao
executive	bestuur	isigungu esiphakeme senhlangano	khuduthamaga
exile n.	balling, banneling	osekudingisweni	molelekwa

forum	forum	inkundla	kgotla, foramo
franchise	stemreg	ilungelo lokuvota	okelo ya go khetha
freedom	vryheid	inkululeko	tokologo
office	amp	ihhovisi	modiro wa ofising
opposition	opposisie	iqembu eliphikisayo	lekoko-kganetšo

CEPTSA also had to deal with homonyms. The fact that all the terms are defined in the SL and TL enables proper interpretation of the concept and its denoting in the TLs:

- executive n. **afr** bestuur, **zul** isigungu esiphakeme senhlangano, **nso** khuduthamaga
- executive n. **afr** uitvoerende beampte, **zul** isikhulu esiphezulu, **nso** lelukuphethiši
- executive n. **afr** uitvoerende gesag, **zul** isigungu esiphakeme sikahulumeni, **nso** khuduthamaga ya mmušo
- executive n. **afr** uitvoerende instelling, **zul** isigungu esiphethe, **nso** khuduthamagakgolo

7. Concluding remarks

This article focused on the background principles, procedures and practice of terminology and terminography for legal interpreting in the African languages. Various terminological principles and processes were discussed. The cognitive, linguistic and communicative dimensions of terminology, word-forming principles, consultation with experts, and planning and compiling of technical term lists were looked at. It was advised that interpreters should use all appropriate methods of term extraction but that TOT would probably be the most useful method of obtaining terms. It was, however, stressed that only by frequent usage terms could penetrate into the subject field and languages concerned and therefore be standardised. The principles of harmonisation and internationalisation/globalisation were discussed in terms of term creation in the legal field. It was suggested that rather than naming every new concept with a totally diverse term, it is more practical to borrow terms from the language of origin to make them locally/nationally and internationally/globally recognisable. Terms can also be coined according to either international languages (through the process of transliteration) or according to the principles of harmonisation (through the process of borrowing). It was further stressed that terminology work should not be done in isolation and that subject specialists (e.g. legal experts), language practitioners and various institutions should be consulted during the process. Further assistance may be obtained from terminology-related centres such as CLTAL and CEPTSA which have as their objectives to make legal and political terminology more accessible to court interpreters, judicial personnel, academics, students and the local indigenous population and by doing so, make it more available and more comprehensible. The multilingual legal and

political terminology would assist the indigenous population in various situations, i.e. enabling users in their first languages when for instance facing court procedures as either the accused, legal representative or court interpreter.

In furtherance of the aim to equip court-interpreting language practitioners with all the basic skills needed to carry out a project through all the stages of the terminological process, the ideal would be to organise a hands-on workshop or multilingual terminology training session where they are encouraged to discover and regard terminology development as a solution to language development. As a more immediate solution, we would like to encourage court interpreters to continue to develop their written and spoken command of their working languages, including any specialist terminology. This is mainly to be done by engaging in continuing education.

Bibliography

- Alberts, M.** 1983. *Aspekte van die rekenariserings van vakwoordeboeke*. M.A. Dissertation. Pretoria: University of Pretoria.
- Alberts, M.** 1990. *'n Bepaling van Afrikaanse vakleksikografiese behoeftes*. D.Litt. et Phil. Thesis. Pretoria: UNISA.
- Alberts, M.** 1997. *Legalese: South Africa's 12th Official Language?* Paper read at the Seminar on Language in Court at the former Venda Parliament (Government Buildings), Thohoyandou, 16–17 May 1997.
- Alberts, M.** 1998. The Harmonisation of Nguni and Sotho: Terminological Perspectives. Prah, K.K. 1998. *Between Distinction and Extinction. The Harmonisation and Standardisation of African Languages*: 229-242. Johannesburg: Witwatersrand University Press.
- Alberts, M.** 1999. *Plain Language in a Multilingual Society*. Paper read at the Plain Language Conference and Workshop, 2 July 1999, Faculty of Law, University of Pretoria.
- Alberts, M.** 2001. *Plain Language in a Multilingual Society*. Viljoen, F. and A. Nienaber (Eds.). 2001. *Plain Legal Language for a New Democracy*. Pretoria: Protea Book House.
- 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.** 2010a. *Electronic Terminology Management. Study Guide TPL 754*. Bloemfontein: University of the Free State.
- Alberts, M.** 2010b. *Theory of Terminography. Study Guide TPL 714*. Bloemfontein: University of the Free State.
- Alberts, M.** 2012. *Sentrum vir Regsterminologie in Afrikatale (SRTAT)*. Paper read at the Prologua meeting, 10 July 2012, UNISA, Florida campus, Florida.
- Botha, S., P. le Clus and A. Venter.** 2011. *Modern Political Dictionary*. Johannesburg: Centre for Political and Related Terminology in Southern African Languages (CEPTSA), Rand Afrikaans University.
- CEPTSA.** n.d. *Multilingual Modern Political Dictionary: 250 Trial Definitions. Meertalige Nuwerwetse Politieke Woordboek: 250 Proefdefinisies. Sesilimi ezahlukene samagama anamuhla apathelene nezepolitiki: Incasiselo yamagama angu-250 angumzamo. Pukuntšu ya dipolelo tše dintši ya sebjalebja le e amanago le mantšu a politiki: Ditlhalošo tše 250 tša mantšu tša boitekelo*. Johannesburg: Centre

- for Political and Related Terminology in Southern African Languages (CEPTSA), Rand Afrikaans University.
- CEPTSA.** 2013. *Database of the Centre for Political and Related Terminology in Southern African Languages*. CEPTSA.
- Charrow, V.R., J.A. Crandall and R.P. Charrow.** 1982. Characteristics and Functions of Legal Language. Kittredge, R. and J. Lehrberger. 1982. *Sublanguage: Studies of Language in Restricted Semantic Domains*: 175-190. Berlin: Walter de Gruyter.
- CLTAL.** 2013. *Database of the Centre for Legal Terminology in the African Languages*. CLTAL.
- Cornelius, E.** 2012. 'n Verkenning van die leksikogrammatikale kompleksiteit van regstaal en die vereenvoudiging daarvan. Paper presented at a meeting of Prolingua, Centurion Office of the Tshwane Municipality, 10 April 2012.
- Department of Justice.** 1996. *Constitution of the Republic of South Africa, 1996. (Act No. 108 of 1996)*.
- Du Toit, I.E.** 1994. *Tweetalige Polisiewoordeboek/Bilingual Police Dictionary*. Pretoria: J.L. van Schaik.
- Eksteen, L.C.** 1969. Die leksikologie van Afrikaans. *Taalfasette* 9: 30-43.
- Fourie, D.J.** 1994. The Rejection of Terminology: Observations from the African Languages. *South African Journal for African Languages* 12(1): 11-15.
- 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.
- Kierzkowska, D.** 1995. Standardisation of Polish Legal Terminology. *Terminology* 2(1): 129-140.
- Kierzkowska, D.** 1999. Polskie Towarzystwo Tłumaczy Ekonomicznych, Prawniczych I Sadowych, Członek Międzynarodowej Federacji Tłumaczy Fit Rada Naczelna. (Fax received on 29 June 1999 from the Polish Society of Economic, Legal and Court Translators (TEPIS)).
- Matthias, C. and N. Zaal.** 2002. Hearing a Faint Echo? Interpreters and Children in Court. *South African Journal on Human Rights* 18(3): 350-371.
- Mellinkoff, D.** 1963. *The Language of the Law*. Boston: Little, Brown & Company.
- Mikkelsen, H.** 2008. Evolving Views of the Court Interpreter's Role: Between Scylla and Charybdis. Valero-Garcés, C. and A. Martin (Eds.). 2008. *Crossing Borders in Community Interpreting: Definitions and Dilemmas*: 81-97. Amsterdam/Philadelphia: John Benjamins.
- Msimang, C.T.** 1996. The Nature and History of Harmonisation. *Proceedings of a LANGTAG Workshop on Issues in Language Development: Focusing on the Future of the Apartheid Language Boards and the Harmonisation and Development of African Languages, 28 March 1996, Unisa, Pretoria*.
- Nahir, M.** 1984. Language Planning Goals: A Classification. *Language Problems and Language Planning* 8(3): 294-327.
- National Association of Judiciary Interpreters and Translators (NAJIT).** *Code of Ethics and Professional Responsibilities*. <http://www.najit.org/certification/profession.php>, accessed July 2103.
- Sager, J.C.** 1990. *A Practical Course in Terminology Processing*. Amsterdam/Philadelphia: John Benjamins.

Internet References

- <http://www.najit.org/certification/profession.php>. Accessed July 2013.
- <http://www.loc.gov/standards/ISO639-2/php/code>. Accessed September 2013.

The Presentation of Word Formation in General Monolingual Dictionaries

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

and

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

Abstract: Word formation in the dictionary belongs, at the latest since the contribution of Mugdan (1984), to the topics frequently discussed by lexicographers. Unfortunately the results have not always been satisfactory. This applies both to the formulation of lexicographic theory and to the lexicographic practice because the lexicographic terms have the linguistic terms as point of departure and questions are put as to how these phenomena should be presented in dictionaries. Instead, one should rather ask which information needs dictionary users experience in which types of user situations and then decide where and how items giving word formation can be presented in order to benefit the envisaged target user of a given dictionary.

The lexicographic practice is also in an unsatisfactory position with regard to items giving word formation because the typical polyfunctional dictionaries have too many items giving word formation for text reception problems and far too few for text production problems or for the real general knowledge needs regarding word formation in the specific language. This paper gives suggestions regarding the theoretical approaches that could lead to a better user-directed lexicographic practice.

Keywords: AFRIKAANS DICTIONARIES, COGNITIVE FUNCTION, COMPLEX FORM, COMPOUND, DERIVATIVE, DICTIONARY FUNCTION, ELECTRONIC DICTIONARIES, TEXT PRODUCTION, TEXT RECEPTION, USER NEEDS, WORD FORMATION

Opsomming: Die aanbieding van woordvorming in algemene verklarende woordeboeke. Woordvorming in woordeboeke is ten minste sedert die bydrae van Mugdan (1984) 'n onderwerp wat gereeld deur leksikograwe bespreek word. Ongelukkig was die resultate hiervan nie altyd bevredigend nie. Dit geld sowel die formulering van leksikografiese teorie as die leksikografiese praktyk want die leksikografiese terme het linguïstiese terme as vertrekpunt en vrae word gestel oor hoe hierdie verskynsels in woordeboeke aangebied moet word. In stede daarvan behoort daar eerder gevra te word watter inligtingsbehoefte gebruikers in watter tipes gebruikersituasies ervaar en dan moet daar besluit word waar en hoe aanduiders van woordvorming aangebied moet word om die beoogde teikenbruiker van die woordeboek ten beste te bevoordeel.

Die leksikografiese praktyk is ook in 'n onbevredigende situasie met betrekking tot aanduiders van woordvorming. Die tipiese polifunksionele woordeboeke het te veel aanduiders wat woordvorming vir teksbegrip bied en veels te min wat op die oplossing van teksproduksiepro-

bleme gerig is of op die werklike algemene kennisbehoefes oor woordvorming in 'n bepaalde taal. Hierdie artikel bied voorstelle oor teoretiese benaderings wat sou kon lei tot 'n beter gebruikersgerigte leksikografiepraktyk.

Slutelwoorde: AFLEIDING, AFRIKAANSE WOORDEBOEKE, ELEKTRONIESE WOORDEBOEKE, GEBRUIKERSBEHOEFES, KOGNITIEWE FUNKSIE, KOMPLEKS, KOMPOSITUM, TEKSBEGRIP, TEKSPRODUKSIE, WOORDEBOEFUNKSIE, WOORDVORMING

1. The function of data regarding word formation in monolingual dictionaries?

The question regarding the function of word formation in monolingual dictionaries constitutes the core of existing lexicographic contributions dealing with this theme. However, in the majority of cases the question is formulated the other way round. The point of departure is linguistic terms like derivative, compound, prefix, affixoid, etc. and the question relates to the way in which these phenomena are presented in existing dictionaries and often also how this could have been done in a better way. This is e.g. the approach followed by most contributors in the recent seminal publication *Wortbildung im elektronischen Wörterbuch*, edited by Annette Klosa (2013a), e.g. contributions by Eichinger (2013), Elsen (2013), Splett (2013) and Klosa (2013c). This does not imply that the user has been completely forgotten. Although mention has in some instances been made of the functions, it was done from the perspective of linguistics that focuses primarily on general knowledge regarding word formation. The following citation illustrates this point:

The presentation of word formation in a dictionary is primarily motivated by the fact that it enables a display of relations and interconnections of words. By means of the inclusion of compounds and derivatives, by means of the lemmatization of affixes and by means of the description of word formation rules in the dictionary grammar the interconnectivity of the vocabulary can be successfully indicated although the alphabetical ordering of head words in the dictionary can only display these relations in an inadequate way. In general the usability and effectiveness of dictionaries should be increased by the inclusion of word formation. The reception of items giving word formation can e.g. lead to an enrichment of the vocabulary especially for learners. (Klosa 2013b)

This is true but also confusing. Someone with e.g. a reception problem looks for the meaning of the word — also if it is a derivative or compound. When the user finds the meaning in a printed dictionary in its alphabetical position or in an electronic dictionary directly without an alphabetical access, his/her problem is solved and he/she can continue reading the text. It could obviously happen that the user forgets the initial problem or moves it to the background in order to explore some more general aspects regarding word formation. This could be due to a general interest in the language or to improve his/her con-

versation in the language or his/her reading and writing skills. In polyfunctional dictionaries all of this is available in one and the same dictionary. But one can also make, especially in the case of electronic dictionaries, monofunctional information tools, i.e. dictionaries, available to the users. This tendency can be seen in the contributions by Bergenholtz (2013) and Ten Hacken (2013).

As earlier indicated, the majority of contributions to this theme do not pose the questions that should, according to our point of view, actually be put by the lexicographer: What kind of problem does the user have? How can a dictionary satisfy his/her information needs? Most contributions, not only in Klosa (2013a) but e.g. almost all in Barz, Schröder and Fix (2000) regard word formation in dictionaries from the perspective of a linguist. As has been argued by Bergenholtz and Tarp (2005) this is motivated by the fact that many lexicographers regard lexicography as a subdiscipline of linguistics and therefore want to bring as much from the field of linguistics into the dictionary. It can be useful but this is not necessarily so. The question should be which theory, linguistic or otherwise, is appropriate to ensure that the selection, method and lexicographic presentation for a given dictionary with a specific genuine purpose can be achieved in an optimal way.

The typical user questionnaires are conceptualised in a similar way. The question is often put, typically to linguists or students of linguistics, which items are rated as the most important or which ones are used the most frequently. Even with 2 200 test persons, as in Töpel (2013), no representative result follows when the test group has not been selected according to a principle of representativeness but have rather voluntarily responded to become test persons following an internet invitation. Töpel (2013) asks which of, among others, word formation, pronunciation, the lemma, meaning, grammar, typical uses or synonyms, etc. are regarded as the most important. Surprisingly enough the result is not that the lemma is important for everyone. In the results obtained from that study the lemma is much less important than meaning or grammar, with pronunciation being the least important. But in these results word formation has also been ranked on the lower end of the popularity scale. What does that say? According to us, nothing because the question should actually have been: Which items are extremely important when you use a dictionary as an aid to solve a reception problem? Or: Which items are extremely important when you learn a language?

2. The presentation of word formation products in current dictionaries

The following paragraphs will focus on the way in which some current dictionaries present word formation products. This discussion does not intend to portray a representative view of the lexicographic treatment of word formation products. However, albeit that examples are only taken from a few randomly selected dictionaries, the presentation found in these dictionaries should not be seen as isolated procedures restricted to these dictionaries or dictionaries of the

respective languages. A similar or comparable presentation can be found in many other dictionaries — of the same and of other languages. Lexicographers dealing with languages with derivatives and compounds as word formation products could do well to embark on innovative presentation procedures to enhance the access of their intended target users to these items, to respond to the needs of these users and to improve the way in which the given dictionary satisfies the identified lexicographic functions.

2.1 Printed dictionaries

Monolingual Afrikaans dictionaries like the *Verklarende Handwoordeboek van die Afrikaanse Taal* (HAT) and *Verklarende Afrikaanse Woordeboek* (VAW) follow comparable ways with regard to the presentation of derivatives and compounds. In both these dictionaries a list of derivatives and compounds is often attached to the article of a lemma representing the first stem of the relevant derivative or compound. The users' guidelines text of VAW gives no explanation of this kind of presentation. HAT does motivate this presentation by indicating that these words are *unexplained headwords*. They are not lemmatised and explained separately because the meaning of the word formation product is self-evident if the different stems are included and explained in their respective alphabetical positions, cf. Gouws (1989).

Attached to the article of the lemma *herstel*² the VAW has the following sequence of words:

~afdeling, ~baar, ~depot, ~krag, ~kuil, ~ler, ~ling.

In the users' guidelines text it is mentioned that the tilde is a place-keeping symbol for the form represented by the lemma. The use of the tilde increases the degree of textual condensation and makes it even more difficult for the user, especially the occasional dictionary user, to have a rapid access to the required form. In this list the words *herstelbaar*, *hersteller* and *herstelling* are derivatives whereas the rest are compounds. The list is presented in an alphabetical ordering without distinguishing between derivatives and compounds. From a linguistic point of view a distinction between these two types of word formation products might have been helpful. For the user looking for the specific word an alphabetical ordering ensures easier access. However, a significant problem lies in the fact that the word *herstel* has no less than eight polysemous senses allocated to it in the specific dictionary article. There is no way the user will know which sense applies in the different complex forms, i.e. derivatives and compounds. Albeit that the primary sense is usually the one used in word formation products, cf. Gouws (1988), this is not always the case. The complex forms included in this list are not also included as lemmata although a number of compounds with *herstel-* as first stem have been included. The user is at a loss when having to decide where to find a given compound or deriva-

tive: attached to the article of the lemma representing the first stem or included in its own alphabetical position as guiding element of a separate article. From a text reception perspective this way of presentation leaves much to be desired. Where the word formation demands the inclusion of a linking morpheme VAW presents the linking morpheme as part of the second component of the complex form, as can be seen in the article of the lemma *kind* that has an *-er-* as linking form:

~eraand, ~erbal, ~erbottel, ~erdrag, ~ergebabbel, ...

to represent *kinderaand*, *kinderbal*, *kinderbottel*, etc. From a linguistic perspective this is unacceptable and from a user perspective the subsequent degree of textual condensation is extremely confusing. This form of textual condensation with the first component of the compound in a remote position makes it difficult for the average user to quickly interpret the compound correctly. If the linking morpheme does not occur in all complex forms the presentation confuses the user even more, cf. the list attached to the article of the lemma *meisie*:

~(s)koshuis, ~agtig, ~span, ~stehuis

as condensed forms of *meisiekoshuis* / *meisieskoshuis*, *meisieagtig*, *meisiespan*, *meisiestehuis*.

Although HAT also makes no distinction between derivatives and compounds and also fails to link the unexplained derivatives and compounds to specific senses of the first stem the listing of unexplained forms is immediately preceded by a repetition of the first stem, e.g. attached to the article of the lemma *doop*:

doop: ~bak, ~bediening, ~boek, ~diens, ...

Where a linking morpheme is needed it is attached to this first stem that stands in the entrance position of the list of unexplained forms as in the list attached to the article of the lemma *meisie*. From both a linguistic and a user perspective this is a more acceptable and satisfying approach:

meisie: ~agtig, ~jare, ~kind, ... ~stem. meisies: ~boek, ~drag, ~gesig, ...

From a text reception perspective the same problems prevail as in VAW but the lesser degree of textual condensation impedes access to the unexplained forms in a lesser way.

In both these dictionaries the decision regarding explained or unexplained is done in an arbitrary way. Gove (1966) remarked that the *self-* in *self-explanatory* should refer to the interpreter of the word and not to the word itself. When planning and compiling a dictionary the *self-* in *self-explanatory* should refer to the user and not to the lexicographer. The application of such an approach will lead to the lemmatisation of many more complex forms.

Nasionale Woordeboek (NW) also includes complex forms in a list attached to the article of a lemma representing the first stem of the complex word. However, the ordering within the list is partially determined by morpho-semantic criteria, cf. the list attached to the article of the lemma *skeer*:

~**skeerder** (by 2); **skeerapparaat**, **-goed**, **-kwas**, **-mes**, ... (by 1), **-geld**, **-hok**, **-kraal**, **-skêr**, ... (by 2).

The markers (*by 1*) and (*by 2*) are semantic markers indicating which sense of the polysemous lemma applies in the specific complex form. The semicolon preceding *skeerapparaat* marks a division between derivatives and compounds. This is valuable linguistic information but it increases the degree of textual condensation. The inclusion of these complex forms is motivated in the users' guidelines text where it is stated that they are included to indicate their existence or to show their spelling. Yet again the distinction between explained and unexplained complex forms is done in a haphazard way because this dictionary also includes compounds and derivatives as guiding elements of fully-fledged articles.

The *Woordeboek van die Afrikaanse Taal* (WAT) also makes a distinction between explained and unexplained complex forms with the former being included as lemmata and the latter as entries in a list attached to the article of the lemma representing the first stem. In this list, a rudimentary treatment is allocated to the complex forms by means of an item giving the part of speech and a marker of the main stress of the word, cf. the list following the article of the lemma *ritme*:

rit'megevoel s.nw., **rit'meloos** b.nw., **rit'mepatroon** s.nw., **rit'mevariasie** s.nw., ...

In this list the word *ritmeloos* is a derivative and the rest are compounds. This is not indicated. The use of a much lesser degree of textual condensation by giving the full forms of the complex words enhances the chances for easy access to these forms.

In VAW, HAT, NW and the WAT the presentation of word formation products assist in text production with the user being able to comprehend something of the system followed in the formation of derivatives and compounds. But the lexicographers eschew the text reception function. Not only will users have problems in determining the relevant sense of the first stem of the unexplained compounds but even worse, no attention is given to the meaning of the second stem in these compounds where different senses could also come to the fore.

Van Dale Groot Woordenboek der Nederlandse Taal includes some complex forms as main lemmata as part of a straight alphabetical macrostructure whereas other complex forms are included as sublemmata being the guiding elements of horizontally ordered niched articles. The word *moeras* is included as

main lemma. Attached to its article a horizontally ordered niched article cluster follows with sublemmata like *moerasachtig*, *moerasaloë*, *moerasandijvie*, *moeras-andoorn*, ... *moerashoenders*. Each niched article contains at least a paraphrase of meaning but in addition also entries like items giving grammatical data or example sentences. The article of the niched lemma *moerashoenders* is followed by the main lemmata *moerashoorn* and *moerasijzer* and then again a niche of horizontally ordered articles with compounds as lemmata. This represents a typical example of multiple niching, cf. Gouws (2005). Although a strict alphabetical ordering is maintained the distinction between vertically and horizontally ordered lemmata remains unclear to the user.

The New Oxford Dictionary of English (NODE) includes compounds as main lemmata whilst derivatives are included within the article of a lemma representing a form from which the derivative was derived. Derivatives are given in a specific article zone and they are preceded by the marker "- DERIVATIVES". The article of the lemma *lemma* has no derivatives but the subsequent article, i.e. that of the lemma *lemmatize* has the noun *lemmatization* as derivative. Derivatives receive a rudimentary treatment, e.g. items giving pronunciation and part of speech. For both text reception and text production purposes the presentation in this dictionary is helpful.

Even when a user is familiar with the system used by a given dictionary with regard to the presentation of derivatives, it is not always that easy to achieve a rapid access to the required form. In its explanatory notes, *Webster's Ninth New Collegiate Dictionary* (W9) explains the inclusion of derivatives as run-on entries, following a "main entry". They are not defined "because their meanings are readily derivable from the meaning of the root word". One of the problems the user experiences with this system is that some derivatives are entered as run-on entries and others as main lemmata. The user does not know where to find a required form. The article of the lemma *incline* contains the run-on entry *incliner*. The user looking for the derivatives *inclinable*, *inclination*, and *inclinational* will not find them in this article. The derivatives *inclinable* and *inclination* are guiding elements of their own articles whereas *inclinational* is presented as a run-on entry in the article of *inclination*. Whilst *inclinable* is presented as lemma and not as run-on entry in the article of *incline*, *includable* is a run-on entry in the article of the lemma *include*, with *inclusion* included as separate lemma.

In W9, main lemmata are typically given an item indicating the date of the earliest recorded use. This is unfortunately not done for run-on entries. It can therefore not be deduced whether the run-on entry has really been derived from the respective main lemma. The dating of derivations given as main lemmata shows the difficulty of guessing which form constitutes a derivation from another form. The following main lemmata have been included (with their dates of earliest recorded use in brackets): *premeditate* (1548), *premeditated* (1590), *premeditation* (15c), *premeditative* (1858). According to these dates the noun *premeditation* is the form from which the verb *premeditate* and the adjective *premedi-*

tated were derived. *Premeditator* is included as run-on entry in the article of *premeditate* but without an item giving its date of earliest recording. Yet again the user has problems in knowing where to find a specific derivative.

The *De Gruyter Wörterbuch Deutsch als Fremdsprache* (WDF) has an article zone identified as accommodating the word family of the word represented by the lemma. A structural indicator precedes this slot and marks it clearly for the user. A system of cross-referencing guides users from articles with different members of the word family as lemmata to the article where the complete family is given. The articles of the lemmata *Diplom*, *Diplomat* and *diplomatisch* have cross-reference items guiding the user to the article of the lemma *Diplomatie*. Here the list of word family members, i.e. *Diplom*, *Diplomatie*, *diplomatisch* is given. It is interesting that the word family list is not given in the article of the base form but rather in the article of a derivative. The system of cross-references allows the user access to all the derivatives. These lists do not always only contain items with the same first stem. The article of the lemma *Kritik* has the following word family list: *Kritiker*, *kritisch*, *kritisieren*, *kritteln*, *gesellschaftskritisch*, *Selbstkritik*, *selbstkritisch* and the article of the lemma *Platz* has the following items in its word family list: *platzieren*, *Arbeitsplatz*, *Campingplatz*, *Flugplatz*, *Parkplatz*, *Schauplatz*, *Sitzplatz*, *Stehplatz*, *Studienplatz*, *Platzkarte*. The inclusion of the different members of a word family as lemmata and the cross-referencing to the article where the full family is found, helps a user to find a given complex form.

From the preceding discussion it is clear that dictionaries have diverse ways of presenting word formation products. One often has the feeling that the needs of the intended target users and the lexicographic functions of the dictionary have not been taken into account sufficiently in the planning of the way in which these word formation products are to be presented. One of the problems for a user consulting an existing printed dictionary in order to find a given derivative or compound is to decide in which article the specific form will be found. Access to word formation products is often impeded by uncertainty as to where the form is entered in the dictionary and what the system of ordering of word formation products is. The typical user of a general dictionary should not have to negotiate the distinction between a base and the derivative forms. He/she merely wants to find the form he/she has encountered in a text and wants to find it as quickly as possible. The average dictionary user is familiar with the alphabetical ordering within dictionaries. Consulting the dictionary to find a specific word formation product should not demand knowledge of an alternative ordering system, e.g. one based on morpho-semantic criteria. Where space restrictions demand that compounds and derivatives should be included in a cumulative list attached to the article of the lemma representing the word that corresponds to the first stem of the compound or derivative the lexicographer should maintain the alphabetical ordering within that list. As many complex forms can be given as regarded necessary by the lexicographer to ensure the best possible presentation in a given dictionary.

Another problem regards the treatment allocated to word formation products. If the inclusion is restricted to their occurrence in a cumulative list the user with a text reception problem will have little assistance. Where text reception is the function a paraphrase of meaning is needed for a complex form. If the dictionary has a cognitive function, data like the date of the first recording of a derivative or compound could be of interest. Additional items could give guidance regarding the history of the complex form, the structure, i.e. the word formation process resulting in the specific complex form, the meaning of the complex form and its frequency of use.

2.2 Electronic dictionaries

In electronic dictionaries, a search for a specific complex form can usually have quicker success because of direct access to the required form, irrespective of it being included as lemma or as derivative/compound in the article of another lemma.

In *ellexico*, an electronic dictionary of German compiled by the Institut für deutsche Sprache, a variety of word formation forms are given. For the word *ernst*, a list is given in which combinations determined by the part of speech of the constituents are ordered alphabetically and for each one its usage frequency in the specific corpus is indicated, e.g. *bierernst* (836), *tiefernst* (1 010), *todernst* (760). Clicking on anyone of these forms takes the user to the specific lemma but also to an alphabetical list of complex forms indicating the alphabetical environment of the specific word.

This dictionary gives access to a wide-ranging selection of complex forms. Clicking on the word *Angst* leads to an article with links to both compounds and derivatives. In the section on compounds, there are alphabetical lists with *Angst-* as first and *-angst* as second stem, e.g. *Angstanfall*, *Angskauf*, *Angsluft* and *Altersangst*, *Bürgerangst*, *Herzangst*. A question can be asked regarding the choice of the form to be indicated as the base. In the article of the word *Liebe*, it is indicated that the verb *lieben* is the base form with the noun *Liebe* (with 211 723 occurrences in the corpus) as a derived form. This might be valuable linguistic guidance. The typical non-linguistic user is not interested in the distinction between base and derivative. In this dictionary, the word *Liebe* can be accessed directly or via the article of the base *lieben*. This is good lexicography. However, the route via *lieben* is not necessarily a quick road. When searching for the verb *lieben* the automatic access is to the noun *Lieben*. On the left of the screen there is a word list with the article stretch to which *Lieben* belongs alphabetically. This list also includes the verb *lieben* and clicking on this word takes the user to the lemma. The choice of *Lieben* as only destination when entering *lieben* as a search word seems to be an arbitrary choice that is not supported by the frequency of use of the respective forms.

The problem indicated in the Afrikaans dictionaries HAT, VAW, NW and WAT that the user does not know which sense of a polysemous word is acti-

vated in the occurrence of the corresponding stem when included as first or second constituent of a compound is cleverly avoided in *The Danish Writing Dictionary*. In this dictionary, the word *portræt* is treated as a polysemous word. The relevant paraphrase of meaning is followed by, among others, a list of complex forms in which the stem corresponding to the word represented by the lemma has the polysemous sense indicated in the specific paraphrase of meaning, e.g.

portræt (= portrait)

1. billede, som gengiver en eller flere personer i form af et maleri, tegning eller fotografi; også om et billede, der kun gengiver ansigtet (= picture reflecting one or more persons by means of a painting, drawing or photo; also a picture giving only the face)

Orddannelser (= word formation)

Nøgenportræt nude (= nude portrait)

portrætalbum (= portrait album)

portrætbillede (= portrait picture)

portrætfoto (= portrait photo)

portrætfotografering (= portrait photographing)

portrætfotografi (= portrait photography)

portrætgalleri (= portrait gallery)

portrætlighed (= portraitness)

portrætmaler (= portrait painter)

portrættere (= to make a portrait)

portrættering (= making a portrait)

portrættør (= maker of a portrait)

selvportræt (= self-portrait)

2. tekst, film, udsendelse eller lignende, som giver et indtryk af en persons eller en gruppe menneskers karakterer (= text, film, television broadcast or something similar that describes the character of a person or group of persons)

Orddannelser (= word formation)

portrætsamtale (= portrait discussion)

portrættere (= to make a portrait)

portrættering (= making a portrait)

portrætsudsendelse sendung (= portrait broadcast)

Here the verb *portrættere* (to make a portrait) is given as a derivative. However, this word is also included as lemma and guiding element of its own article. This article also includes word formation products, including the form *portræt*. The user is not confronted with the distinction between a base and a derived form but the poly-accessibility of this dictionary makes it possible to have rapid access to the desired form but also to link a derived form to its base. Yet again each one of the paraphrases of meaning of the different polysemous senses of this word is allocated an own article zone. This zone also contains the relevant

word formation products and possible base forms:

portrættere

1. gengive en eller flere personer i form af et maleri, tegning eller fotograf (= one or more persons pictured as painting, drawing of photograph)

Orddannelser (= word formation)

portræt (= portrait)

portrættering (= to make a portrait)

portrættør (= maker of a portrait)

2. gengive og give et indtryk af en persons eller en gruppe menneskers karakterer i form af en tekst, film, udsendelse eller lignende; benyttes især i forbindelse med skuespil (= to characterise a person or group of persons, especially in a film or play)

Orddannelser (= word formation)

portræt (= portrait)

portrættering (= to make a portrait)

portrættør (= maker of a portrait)

The article zones for both polysemous senses offer the same word formation products. This will help the user with both text reception and text production needs because the user has no uncertainty with regard to the meaning of the first stem of the word formation products.

2.3 Where can information regarding word formation be found in dictionaries?

Dictionaries can assist their users with regard to word formation products in various ways. In the present discussion, the focus has primarily been on word formation products presented in dictionary articles. Although this is a frequent and important way of conveying word formation information, there are also other ways to do so.

The most typical way remains the lemmatisation of a derivative or compound. Lemmatisation typically leads to a treatment, albeit perhaps of a limited extent, but the user retrieves information that helps him/her to understand and use the complex word correctly. When planning the lemmatisation and treatment of compounds and derivatives, the lexicographer should also negotiate the possibility of variant forms of a given complex. A similar approach that is followed with regard to other variants should also be applied to compounds and derivatives. Both or all the variants could be lemmatised with the necessary cross-references given where needed. In VAW the word *meisie(s)koshuis* indicates that this compound can be used with or without the linking morpheme. No preference is indicated. In a proscriptive dictionary, one specific form could be indicated as the recommended variant. Guidance with regard to word formation can also be given by means of lexicographic comments. In

HAT a comment is added to the article of the word *lief* stating that *lief* is sometimes used as a suffix which is added to nouns, e.g. in *vaderlief* and *kindjief*. Such comments are helpful but they do not allow a systematic retrieval of information. As seen in some of the dictionaries discussed in the preceding sections, word formation items can also be presented in a word formation field or article zone. Separate fields for derivatives and compounds can be included in or attached to an article. Although a word formation product is not lemmatised, its inclusion in such a field or cumulative list can help the user to recognise the word and in some cases even to understand it although text reception is at best accomplished by means of a paraphrase of meaning and not by a mere listing of a so-called self-explanatory complex word. The treatment of word formation is also strengthened by the inclusion as lemmata of relevant terms from the field of word formation where these terms are explained in such a way that the target user of the dictionary can use them to increase his/her understanding of word formation types presented as part of the subject matter of the dictionary.

Lexicographers can also utilise outer texts to present data regarding word formation. One such possibility which is extremely helpful but unfortunately too often neglected is the inclusion of a separate section in the dictionary grammar presented as front or back matter text. This is for example done in the *Malagasy–German Dictionary*. Such an approach integrates grammar and dictionary and brings an added value to the dictionary as language learning instrument. Guidance regarding word formation can also be given within the users' guidelines text, typically presented within the front matter section. Such a discussion makes the user aware of the way in which word formation products are presented and treated in the given dictionary. Dictionaries can also give cross-references or links to relevant internet websites or other dictionary-external sources, e.g. text books or grammars where supplementary discussions of various word formation processes can be found. Such an approach emphasises the relations holding between different reference sources and confirms the fact that dictionaries do not have to be used in isolation but form part of a bigger reference network.

3. Word formation items as a tool to assist dictionary users in different types of situations

3.1 Word formation items as a tool to assist dictionary text reception problems

If you do not understand a derivative or a compound you will not need to know whether it is a derivative or a compound. You do not need to know anything but the meaning. When the required word can be found as lemma in the dictionary the user has performed a successful dictionary consultation. When the word has not been lemmatised, Bergenholtz and Tarp (2005: 581), as

many other lexicographers in various contributions, believe that the dictionary does not offer any help. This is not completely correct; at least not in electronic dictionaries. When one looks for a word that has only been lemmatised in its occurrence with a linking morpheme, e.g. the German *Glaubenssache*, but not *Glaubensache* (without the linking -s-), the electronic dictionary can suggest, as we also know from Google searches: Do you mean *Glaubenssache*? In this case even a printed dictionary can help if the searched word is not given too far from the lemmatised word. But in some other cases only the electronic dictionary can help, e.g. in the case of a derivative where the derivative is only given as item presenting the word formation in the dictionary and not also as lemma. In *The Danish Meaning Dictionary* (2013), the form *ionosfærisk* (= ionospheric) has not been lemmatised. Because it is included as item giving the word formation in the article of the lemma *ionosfære* (= ionosphere) the user retrieves the following dictionary article with the following introduction:

No article *ionosfærisk* can be found. Instead, the following article is shown in which *ionosfærisk* occurs as item giving word formation:

ionosphere noun

the outer part of the atmosphere of about 80 km and more, where the air has been ionised, i.e. the air molecules have been split in ions and electrons.

The knowledgeable user will now realise that the adjective has more or less the same meaning. The user cannot know for sure that he/she can use the explanation found for the noun, but he/she will see that it fits the context of the not understandable word *ionisfærisk*.

However, it does not always work. And the problem is not only whether the lexicographer has presented an item giving word formation or not. It is not relevant to the user with a reception problem whether the following advice had been followed:

Thus, Dutch *handappel* 'eating apple' should be listed under *appel*, not under *hand*, because language users know that the word *handappel* stands for a subset of apple, not of hands, and will look for this word under the heading for *appel*. (Booij 2003: 254)

From a linguistic point of view the argument might be correct, but you cannot be sure that the intended user is thinking in exactly the same way. And much more important is the fact that you should not make any lexicographic choices without considering the intended dictionary functions. If you have a text reception problem, you do not understand the word *handappel*; it is difficult to look up under both the guiding elements *hand* and *appel*. You need the individual lemma *handappel* in order to find help in solving the reception problem. If you do not understand a compound at all and the meaning of the parts of the compound are not really related to the meaning of the compound as a whole, you can only be sure that your assumption or guess is correct, when you find

the word in the lemma list. That means that a really helpful reception dictionary must be a very large one with at least 200 000 lemmata. This is too comprehensive for general printed dictionaries but for electronic dictionaries it only remains a question whether the project has or had enough lexicographers.

3.2 Word formation items as a tool to assist text production problems

All the arguments in favour of including many word formations in order to help solve reception problems are also valid in terms of production problems. But here 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 know that the word formation exists and how it is spelled or whether it is used with a linking morpheme or not. In the latter case you often have variants for which it is not sufficient to only inform about the existence of variants. Instead, the user needs a recommendation for making his choice. At best, this can be done through a simple reference to the recommended word, e.g.:

job noun ...

Word formations:

bijob (= sideline job)

deltidsjob (part time work)

fuldtidsjob (= full time job)

fuldtidjob (= full time job) [this spelling without -s- is unusual and not recommended, use instead fuldtidsjob]

(In this article, the underlined forms constitute links that the user can follow to obtain a more comprehensive treatment of the specific word.)

Word formation items are important if the formation is irregular or somehow not predictable. This is often the case with nouns for people living in a country or a town or adjectives for something belonging to that part of the world. The user knows the name of the town or the country, but not the related noun or adjective, e.g.: People coming from or living in *Copenhagen (København)* are called *københavnner* (with an *-er* as derivation morpheme), people from *Silkeborg* are called *silkeborgenser* (with an *-enser* as derivation morpheme), people from *Århus* are called *århusianer* (with an *-ianer*) as derivation morpheme). In Afrikaans an inhabitant of *Durban* and *Paarl* is called a *Durbaniet* and a *Paarliet* respectively and an inhabitant of *Bloemfontein* is a *Bloemfonteiner* and wine from *Stellenbosch* is *Stellenbosse wyn*. This is not really a problem in a text reception situation, because you can guess what the words *københavnner*, *silkeborgenser*, *århusianer*, *Durbaniet*, *Bloemfonteiner*, etc. mean, but in a text production situation you cannot predict how the adjective derived from Stellenbosch or the noun indicating an inhabitant from *København* must be formed if you do not already know it. Therefore it is crucially important that city and country names should also be included as lemmata in order to be able to enter the relevant

word formation products. This is yet again clear when the word formation is not done in a systematic or predictable way, e.g. the inhabitants of the Danish city of *Ribe* are not *ribener* or *ribenser* (with a *b*), but *ripenser* (with a *p*). Also if *ripenser* is a lemma, it does not help the user, who only knows the name of the town. He/she will not find this lemma, but will need a presentation like the following:

Ribe PROPER NOUN

Word formations ripenser, ripensisk

(*Ribe* is a Danish town in Southern Jutland; a *ripenser* is someone from Ribe, *ripensisk* refers to something from Ribe)

But in principle it also applies to a text production dictionary that it ideally contains all word formation products of a language as lemmata. It is unforgivable to exclude opaque compounds as the earlier mentioned Dutch word *handappel* or the Afrikaans form *broekskeur* (literally to tear pants, meaning "difficult"). It is of little assistance to the user whether the word *handappel* is presented as an item giving word formation at *hand* or at *appel*. If he/she does not know the word and even if he/she finds it as word formation item but does not understand it, he/she cannot use the word. The more lemmata the dictionary contains the better are the chances to help the user in those cases, not so infrequent, where the user is able to form the word but is not sure whether the word exists in the language, i.e. whether it can actually be used in the language. Alternatively, as word formation item, the dubious word can assist the user realising that the word does exist. Admittedly Google nowadays has a comparable function when the user no longer trusts the dictionary.

3.3 Word formation items as a tool to assist knowledge needs

To satisfy a cognitive need, the user would often not consult a dictionary but rather a grammar or a linguistic text book. If a dictionary does have an integrated dictionary grammar, one can find in one information tool both a systematic presentation of the grammar and possible cross-references to the description of the single phenomenon in the dictionary, or explicit cross-references from concrete dictionary articles to the more comprehensive presentation in the dictionary grammar. In a dictionary directed at cognitive needs, the complementary items to individual items giving word formation, especially in the form of comments, play a much more significant role than in communication dictionaries. Furthermore the description of the lemmatised word formation terms also plays an important role. Finally, comprehensive comments regarding linguistically interesting problems are also important, e.g. with regard to the relation between morphological variants. These problems can be illustrated by the following two examples. The first example shows a lemmatised derivation morpheme in the *Malagasy–German Dictionary* (1991). Note in particular here the reference to a dictionary grammar, i.e. a dictionary grammar written in

both languages as part of the dictionary:

- a verbal morpheme which is added to a word stem with two different functions
 1. used for forming action (agisive) verbs, e.g. adaboka (give up) → §1.5.4
 2. used for forming instrumental verbs, e.g. aharatra (shave) → §1.5.6

The second example shows a grammatical comment directed at a variant problem from *The Danish Internet Dictionary*. It is a polyfunctional dictionary with "Knowledge regarding the Danish language" as part of one of its functions. It also is a proscriptive dictionary, i.e. neither a purely descriptive nor a prescriptive dictionary:

massemedium = Massemedium (= mass media)

Medium that reaches the majority of the population of a country or a region, e.g. a news paper, television or radio.

This word formation product is not recommended, rather use → *massemedie*.

Comment: In most cases of two words with the same meaning but with the suffixes *-ium* or *-e*, the variant with the *-ium* is recommended, e.g. the word *gymnasium*. But in words with the stem *medi* and the two variants *medium* and *medie* there is a different proposal. Generally the use of *-medie* is recommended, as also in this case, where a Google search in March 2011 shows 1.980 citations with *massemedium* and 25.900 citations with *massemedie*. In this regard our advice is in line with the language use. But the proposal does not mean that the other variant, i.e. *massemedium*, should be regarded as incorrect. It is merely recommended that the other variant should be used.

It is important that the presentation of word formation products should be done according to a specific and well-defined concept. This is co-determined by whether the envisaged dictionary is compiled for linguists, students of linguistics, linguistically lay people interested in language or e.g. learners of a language. In the above-mentioned Malagasy and Danish examples two articles have been cited from dictionaries compiled for linguistic lay persons. Corresponding articles e.g. for students of linguistics could and should be much more comprehensive. In a similar way articles dealing with linguistic terms regarding word formation should be sufficiently exact and comprehensive.

4. Conclusion

Diversity in lexicography implies different dictionaries for different needs in different situations of use. In specific situations of use specific users consult a dictionary to find a compound or derivative and to retrieve information regarding that word. Current dictionaries display a variety of ways in which

complex forms are presented and treated. In many instances this presentation is determined by linguistic criteria and not necessarily by the needs of the intended target users of the dictionary. The discussion in this paper illustrates the current situation and argues in favour of a presentation and treatment determined by the needs of the users. Compilers of electronic dictionaries should utilise the means at their disposal to give a more comprehensive selection of compounds and derivatives as lemmata and a treatment directed at the functions of the dictionary. In the treatment of word formation products dictionaries should not be seen as isolated products but the use of links and references to other sources should emphasise the role of a dictionary as part of an interactive network of reference sources.

5. References

- Barz, I., M. Schröder and U. Fix (Eds.).** 2000. *Praxis- und Integrationsfelder der Wortbildungsforschung*. Heidelberg: Winter.
- Bergenholtz, H.** 2013. Wortbildungsangaben als Hilfe für den Zugriff auf andere Datentypen und als Hilfe bei kommunikativen und kognitiven Informationsbedürfnissen. Klosa, A. (Ed.). 2013a: 133-156.
- Bergenholtz, H. and S. Tarp.** 2005. Dictionaries and Word Formation. Brown, K. (Ed.). 2005. *Encyclopedia of Language and Linguistics. Volume 3*: 580-583. Second edition. Oxford: Elsevier.
- Booij, G.** 2003. The Codification of Phonological, Morphological, and Syntactic Information. Van Sterkenburg, P. (Ed.). 2003. *A Practical Guide to Lexicography*: 251-259. Terminology and Lexicography Research and Practice 6. Amsterdam/Philadelphia: John Benjamins.
- De Gruyter Wörterbuch Deutsch als Fremdsprache** = Kempcke, G. et al. (Eds.). 2000. *De Gruyter Wörterbuch Deutsch als Fremdsprache*. Berlin/New York: Walter de Gruyter.
- Eichinger, L.M.** 2013. Wortbildung im Wörterbuch. Aus der Sicht eines Grammatikers. Klosa, A. (Ed.). 2013a: 63-86.
- ellexico** = *ellexico* Online Wörterbuch zur deutschen Gegenwartssprache. Compiled by the Institut für deutsche Sprache, Mannheim. Available: <http://www.owid.de/wb/ellexico/start.html> (Accessed 6 April 2013).
- Elsen, H.** 2013. Problemzonen der Wortbildung und der Eintrag im Wörterbuch. Klosa, A. (Ed.). 2013a: 87-103.
- Gouws, R.H.** 1988. Compounds in Dictionaries: A Semantic Perspective. Snell-Hornby, M. (Ed.). 1988. *ZüriLEX '86 Proceedings. Papers Read at the EURALEX International Congress, University of Zürich, 9-14 September 1986*: 91-97. Tübingen: Francke Verlag.
- Gouws, R.H.** 1989. *Leksikografie*. Pretoria/Cape Town: Academica.
- Gouws, R.H.** 2005. Multiple Niching. Gottlieb, H., J.E. Mogensen and A. Zettersten (Eds.). 2005. *Symposium on Lexicography XI. Proceedings of the Eleventh International Symposium on Lexicography May 2-4, 2002 at the University of Copenhagen*: 261-275. Lexicographica. Series Maior 115. Tübingen: Max Niemeyer.
- Gove, P.B.** 1966. Self-explanatory Words. *American Speech* 41: 182-198.
- Klosa, A. (Ed.).** 2013a. *Wortbildung im elektronischen Wörterbuch*. Tübingen: Gunter Narr.
- Klosa, A.** 2013b. Einleitung. Klosa, A. (Ed.). 2013a: 7-11.

- Klosa, A.** 2013c. Wortbildung in elexiko: Gegenwart und Zukunft. Klosa, A. (Ed.). 2013a: 175-196.
- Malagasy–German Dictionary** = Bergenholtz, H. in cooperation with Rajaonarivo, S., R. Ramasomanana and B. Radanielina as well as Richter-Johanningmeier, J., E. Olszowski and V. Zeiss in collaboration with Ranaivoson, H., N. Rasoarimanana, R. Ravololomboahangy and M. Razafiarivony. 1991. *Rakibolana Malagasy–Alema/Madagassisch–Deutsches Wörterbuch*. Antananarivo: Leximal/Moers: aragon.
- Mugdan, J.** 1984. Grammatik im Wörterbuch: Wortbildung. Wiegand, H.E. (Ed.). 1984. *Studien zur neuhochdeutschen Lexikographie IV*: 237-308. Hildesheim/Zürich/New York: Olms.
- Nasionale Woordeboek** = De Villiers, M. et al. (Eds.). 1987. *Nasionale Woordeboek*. Cape Town: Nasou.
- Splett, J.** 2013. Grundlegende Bemerkungen zu einem auf einer pragmatischen Sprachtheorie fußenden Wortfamilienwörterbuch als legitimen Ort einer integrierten Wortbildung. Klosa, A. (Ed.). 2013a: 117-129.
- Ten Hacken, P.** 2013. Wortbildung in elektronischen Lernerwörterbüchern. Klosa, A. (Ed.). 2013a: 157-174.
- The Danish Internet Dictionary.** 2013. = Bergenholtz, H. in collaboration with Agersnap, O., A. Bodilsen, S.W. Dohrn, K.B. Eriksen, H.R. Gudmann, H.H. Jensen, J.H. Jensen, M. Holst, C. Kjølhedede, J. Nguyen and H.A. Pedersen. Database: Almind, R., R.T. Styrk and P. Christensen: *Den Danske Netordbog*. Odense: Ordbogen.com.
- The Danish Meaning Dictionary.** 2013. = Bergenholtz, H. in cooperation with Jensen, H.H. in collaboration with Agersnap, O., A. Bodilsen, S.W. Dohrn, K.B. Eriksen, H.R. Gudmann, J.H. Jensen, M. Holst, C. Kjølhedede, J. Nguyen and H.A. Pedersen. Database: Almind, R., R.T. Styrk and P. Christensen: *Den Danske Betydningsordbog*. Odense: Ordbogen.com.
- The Danish Writing Dictionary.** 2013. = Bergenholtz, H. in cooperation with Eriksen, K.B. in collaboration with Agersnap, O., A. Bodilsen, S.W. Dohrn, H.R. Gudmann, H.H. Jensen, J.H. Jensen, M. Holst, C. Kjølhedede, J. Nguyen and H.A. Pedersen. Database: Almind, R., R.T. Styrk and P. Christensen: *Den Danske Netordbog*. Odense: Ordbogen.com.
- The New Oxford Dictionary of English** = Pearsall, J. (Ed.). 1998. *The New Oxford Dictionary of English*. Oxford: Oxford University Press.
- Töpel, A.** 2013. Die Wortbildungsangaben im Online-Wörterbuch und wie Nutzer sie beurteilen — eine Umfrage zu elexiko. Klosa, A. 2013a: 197-214.
- Van Dale Groot Woordenboek der Nederlandse Taal** = Geerts, G. et al. (Eds.). 1992. *Van Dale Groot Woordenboek der Nederlandse Taal*. Utrecht: Van Dale Lexicographie.
- Verklarende Afrikaanse Woordeboek** = Labuschagne, F.J. et al. (Eds.). 2010. *Verklarende Afrikaanse Woordeboek*. Cape Town: Pharos.
- Verklarende Handwoordeboek van die Afrikaanse Taal** = Odendal, F.F. and R.H. Gouws (Eds.). 2005. *Verklarende Handwoordeboek van die Afrikaanse Taal*. Fifth edition. Cape Town: Pearson Education South Africa.
- Webster's Ninth New Collegiate Dictionary** = Mish, F.C. (Ed.). 1987. *Webster's Ninth New Collegiate Dictionary*. Springfield, Massachusetts: Merriam-Webster.
- Woordeboek van die Afrikaanse Taal** = Botha, W.F. (Ed.-in-Chief). 2009. *Woordeboek van die Afrikaanse Taal, Volume 13*. Stellenbosch: Bureau of the WAT.

What is a Lexicographical Database?

Henning Bergenholtz and Jesper Skovgård Nielsen,
Ordbogen.com. and Lemma.com., Odense, Denmark (jsn@ordbogen.com)

Abstract: Fifty years ago, no lexicographer used a database in the work process. Today, almost all dictionary projects incorporate databases. In our opinion, the optimal lexicographical database should be planned in cooperation between a lexicographer and a database specialist in each specific lexicographical project. Such cooperation will reach the highest level of success if the lexicographer has at least a basic knowledge of the topic presented in this paper: What is a database? This type of knowledge is also needed when the lexicographer describes an ongoing or a finished project. In this article, we provide the description of this type of cooperation, using the most important theoretical terms relevant in the planning of a database. It will be made clear that a lexicographical database is like any other database. The only difference is that an optimal lexicographical database is constructed to fulfil the requirements for a specific lexicographical project.

Keywords: DATABASE, RECORD, DATABASE MANAGEMENT, DESIGNING A DATABASE, USER INTERFACE, LEXICOGRAPHICAL DATABASE, DATABASE SYSTEM, DATABASE STRUCTURE, DATABASE SCHEMA

Opsomming: Wat is 'n leksikografiese databasis? Geen leksikograaf het vyftig jaar gelede 'n databasis in die werksproses gebruik nie. Vandag inkorporeer byna alle woordeboekprojekte databasisse. Na ons mening behoort die beste leksikografiese databasis beplan te word in samewerking tussen 'n leksikograaf en 'n databasisdeskundige vir elke bepaalde leksikografiese projek. So 'n samewerking sal die hoogste suksesvlak bereik as die leksikograaf ten minste 'n basiese kennis van die onderwerp van hierdie artikel het: Wat is 'n databasis? Hierdie soort kennis is ook nodig as die leksikograaf 'n voortgesette of 'n voltooide projek beskryf. In hierdie artikel verskaf ons 'n beskrywing van hierdie soort samewerking deur die belangrikste teoretiese terme te gebruik wat toepaslik is in die beplanning van 'n databasis. Dit sal duidelik gemaak word dat 'n leksikografiese databasis soortgelyk is aan enige ander databasis. Die enigste verskil is dat die beste leksikografiese databasis saamgestel is om aan die behoeftes van 'n spesifieke leksikografiese projek te voldoen.

Sleutelwoorde: DATABASIS, OPTEKENING, DATABASISBESTUUR, ONTWERP VAN 'N DATABASIS, GEBRUIKERSKOPPELVLAKE, LEKSIKOGRAFIESE DATABASIS, DATABASISSTELSEL, DATABASISSTRUKTUUR, DATABASISSKEMA

1. Databases in Lexicography

About 30 years ago, much dictionary work was done without the use of a data-

base. From the selected data, which was typically stored on written cards collected in boxes, one and only one dictionary was produced. At that time, dictionaries were mainly polyfunctional dictionaries containing almost all the selected data. Nowadays, the situation has changed. We still have many polyfunctional dictionaries and only few monofunctional dictionaries, but we will definitely produce more monofunctional information tools in the years to come. To our knowledge, there is no current dictionary project that does not use a database. This is reflected in the lexicographical literature, but in quite a disappointing way as you are never really told what a database is or how the specific database is structured. We will not go into detail with this topic and only take a few, but, as we think, representative examples.

Many contributions about lexicographical databases have in common that you can see that lexicographers use the term database without demonstrating a clear understanding of what a database really is. In reality, many, perhaps most, lexicographical discussions of databases in theoretical contributions are not really informative — sometimes they are even misleading. There are exceptions, e.g. Almind (2005), but they are few in number.

The following quotation gives the impression that the database and the dictionary is the same. This is not said explicitly, but we see no other interpretation:

The Multilingual Dictionary of Lexicographical Terms (MDLT) is an electronic dictionary available on the Internet. The content of this database and the detailed description of the entries serve many purposes. For translators, the system has term equivalents in different languages and related terms, which may help them, make adequate translations from one language into other(s). Beginners can find many interesting facts in the introductory part, which is available in both English and Russian. Transcriptions will help users to pronounce terms correctly. (Krestova and Nürnberg 2013)

If the user interface for the lexicographers is exactly the same as the user interface for the dictionary users, you could say that a database and a dictionary is the same. In reality, the user interface is not the database, but this term is often used as a practical expression for the presentation of the fields from the database. Normally, you do not have the same user interface for lexicographers and users, but you still get the impression that no real differentiation is made between a database, the dictionary planning, the dictionary production and the dictionary:

A data bank consists of information organized into records, each of which is subdivided into data fields. Creating a data bank involves a systematic process that goes from designing the form of the bank desired to implementing it on a computer. This process cannot be improvised, but must follow a number of steps. The planning starts with a definition of the major features expected of the data bank. This stage includes the following operations: a. Identification of needs, usually by means of a needs survey consisting of: identification of target users, delimitation of the needs of each user group [...] Identifying the obstacles

that might arise throughout the process: i.e. time, human resources, budget, psychological attitudes toward the new project or project change. (Cabr  1998: 169)

All of this is certainly important for the lexicographical work in a dictionary project, but not for the construction and the use of the database. The database has no direct relationship to the intended use or user group. This does not mean that the selection of different data fields is without relevance for the intended use, of course not. For example, if you do not have a field for pronunciation, you cannot produce a pronunciation dictionary or a dictionary that also informs you about the way you should pronounce a word. Of course, you have to have different fields for all the different kinds of data that you want to present in the resulting dictionary resp. dictionaries. But you can also have fields in the database that are not used at all in the lexicographical product. In principle, the database has no relationship to any intended use and user group.

The following quotation is not wrong, but misleading:

[W]e use 'database' to refer specifically to the structured collection of material assembled during the analysis process, on the basis of which final dictionary entries will be created. (Atkins and Rundell 2008: 264)

In principle, a database is an empty box created to contain data in the different fields. The database contains data, but it is not a collection of structured data. If the database description in the quotation were correct, a database could not be empty, and you would not have a database until someone puts data into the database. It is understandable that lexicographers are only interested in the data ("material"), but the data and the database should be seen as a box and the things you store in the box. The box exists both with and without content, and the same applies to a database.

2. What is a Database Really?

In this section, we will elaborate on the description of a database by providing definitions of a more technical kind, but still in a way so that non-IT specialists can understand them. We begin by defining a database and a database management system:

Database: A structured collection of values

A *structured collection* could be a table, but other options exist. For lexicographical applications, tables would be the preferred choice. A directory structure that only consists of text files could also be a database. It should be noted that the collection may be empty as a directory or a file can be empty. We shall call the structure itself the *database schema*.

By *values*, we refer to the entities that you could choose to store in the database. These might take the form of strings, numbers, dates, etc. By saving

these values in a structured manner, we allow the database management system to search for given values, retrieve all values of a certain kind or sort a given collection of values. When our values are stored in a table structure, values in the same row are related. This means that if one column consists of idioms and another consists of idiom meanings, we can search for all idioms containing the word "dog" and get the meanings for these idioms.

Database Management System (DBMS): *A software system designed to allow the definition, creation, querying, update, and administration of databases* [Wikipedia (Database), <http://en.wikipedia.org/wiki/Database>, retrieved June 2013]

To clarify this definition: If you were to choose Excel, the files created by Excel (even before the first amount of data is put in and before it is saved) would be the database, and Excel itself would be the DBMS. The structure we enforce on the data is the database schema. Granted, for most applications Excel would be a very poor DBMS, but according to our definition, it is a DBMS. In practice, however, we will choose a DBMS built only to be used as a DBMS.

We are, however, still missing a piece. The DBMS does not provide a way of accessing and modifying data for anyone but IT specialists.

User Interface (UI): *[...] is the space where interaction between humans and machines occurs* [Wikipedia (User Interface), http://en.wikipedia.org/wiki/User_Interface, retrieved June 2013]

The UI can take many forms. For example, it could be the tool that lexicographers use to feed data to the database, or it could be the representation of the data the user sees when accessing a dictionary either on a website or on her Kindle.

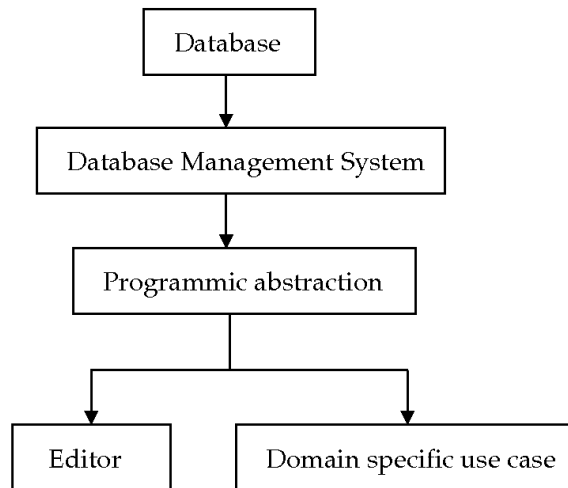
Implications

The three above-mentioned definitions together form what most people call a "database": The database provides useful data through the DBMS to the user in a UI. This more granular definition allows us to talk about the process of creating, modifying and using a database in new ways. A database contains data, and this data can be searched for and presented in different ways. For lexicographical databases, this means that a database is not a dictionary. From one database, you can create different information tools. This can be expressed more generally in the following way: On top of a database, we can build as many UIs as we like, for example one for the lexicographer to create and edit the contents of the database and one or many for the users of the database, each designed to meet a specific use. These examples are all for human utilization, but we can provide an application programming interface (API) for machine exploitation as well.

This means that a lexicographical database could have a UI for the lexicographer for updating the data. It could also have a specific UI for users trying to write a text, another one for users trying to learn more about a word, and lastly an API for a word prediction application such as the spell checker in your word processor.

However, because the database no longer "knows" what it represents, it simply corresponds to values extracted from some domain. Knowledge about what data it represents is still needed. While building the UIs, this knowledge is placed in layers of abstraction expressed in a programming language. In programming languages, we talk about objects (Lemma, Flexion, etc.). That is how we model the data in the database in useful terms.

This leaves us with a layered model of the general perception of what a database is. We call this a database system (DBS):



The most popular type of DBMS for lexicographical applications is used for relational databases. This type of database represents data in much the same way as Excel, that is, as columns of data in different tables (what Excel calls sheets). We can relate rows in different tables by defining "foreign keys", which define the relation between two tables. Some relational databases are targeted for ease of use and thus come with a pre-built UI. These include FileMaker and MS Access. These types of databases also provide a way of building customized UIs. This makes them available to non-IT specialist. As a result, they are a good match for building smaller database-driven applications where speed of development is the primary issue.

Other types of DBMSs are built to support heavy usage. These include MySQL and PostgreSQL. These DBMSs do not come with a UI, which means that more effort needs to be put into the development process. To be a feasible

option for lexicographers, a UI made in some type of programming language is needed. However, these DBMSs are much more capable of handling a large number of users. They also provide advanced ways of finding and joining data. This includes indexes and full text search. As a result, these types of DBMSs are preferable when building an online dictionary and other lexicographical applications.

Another type of database that is gaining popularity at the moment is the NoSQL database. It allows data to be saved and retrieved with lower consistency constraints. This is mainly an advantage for performance issues. This database is, however, limited in its access and storage functionality. This makes it great for storage of large amounts of data in cases where a relational database is not necessary. Examples of NoSQL databases are MongoDB, Cassandra and CouchDB.

3. The Construction and Structure of a Concrete Lexicographical Database

When we claim that a certain database can be used to produce not only one but many dictionaries, it does not mean that there are no restrictions. For instance, the result depends on the number and the content of the fields which the lexicographer has included for the lexicographical working process. For example, you cannot inform about pronunciation if you do not have a pronunciation field, and you cannot inform about the history of a word if you do not have a field for etymology. With this in mind, you can begin the construction of the database even before the lexicographer has a clear conception of the concrete dictionaries.

The fields proposed by the lexicographer do not correspond exactly to the fields made by the database specialist. The specialist may divide the fields that have been proposed by the lexicographer partly into two or more subfields.

In the following, we will describe some of the stages and decisions made in the planning of a database for a number of Spanish monolingual dictionaries. For this project, the lexicographer proposed 28 fields to be part of the database:

1. Lemma
2. Style marker to lemma
3. Sublemma
4. Homonym number
5. Polyseme number
6. Meaning
7. Lexical remark
8. Lexical remark for text production
9. Grammar, word class
10. Grammar, inflection class 1
11. Grammar, inflection class 99
12. Grammar/spelling remark

13. First reference
14. Second reference(s)
15. Collocation(s)
16. Example(s)
17. Word formation(s)
18. Synonym(s)
19. Style marker to synonym
20. Antonym(s)
21. Style marker to antonym
22. Synonym remark
23. Proverb(s)
24. Idiom(s)
25. Idiom meaning
26. Internet link
27. Dictionary grammar
28. Memo field

In practice, field number 26 may be divided in two subfields, one for the real Internet address and one for the shorter name given by the lexicographer to this address. This name will be shown in the dictionary article and used as a link to the real address.

In addition, the lexicographer proposed three kinds of buttons on the UI:

- (a) Buttons to help the lexicographers in their lexicographical work, for example a button to be used for Google searches or one to show the lexicographers the actual dictionary article in its current form.
- (b) Buttons for working with the database such as the commands: find, new lemma, delete an article, delete the data in a field, etc.
- (c) Navigation buttons for going from one of the three different pages of the UI to one of the two other pages (it is necessary to have three pages as the number of fields to be shown is too high to fit into a single screen page).

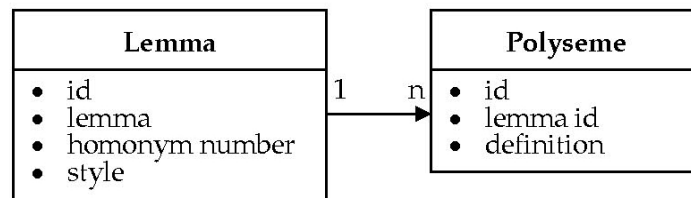
The database discussed above is made for a project with monolingual Spanish dictionaries as the outcome. These dictionaries will be somewhat similar to the six dictionaries in the ongoing Danish dictionary project that consists of one database for six dictionaries, see Bergenholtz (2010, 2012, 2013).

In the following, we will describe how to transform the list of fields into a database structure. In this example, we are working on a very simple dictionary that consists of only lemma, style and meaning (polyseme). This is what the field list looks like:

- Lemma
- Homonym number
- Style

- Definition
- Polyseme number

The first thing to do is to group relevant fields together as they need to go into the same table. We might argue that everything should be put into the same table, but this leaves us with the question of how to handle multiple polysemes. Instead, we place the polysemes into their own table. We give each lemma a unique id. For every polyseme, we add the lemma id which links it to the lemma to which it belongs. We call the lemma id in the polyseme table a *foreign key*. The database schema looks like this:



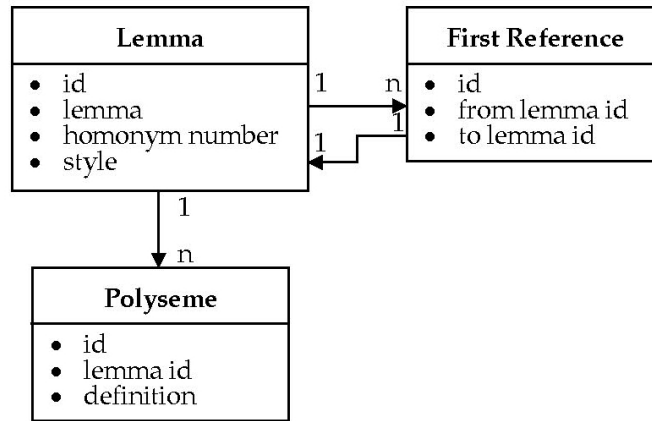
By having a foreign key in the polyseme table, we relate the lemma and the polyseme, in effect creating a one-to-many relationship. This is a technique used in many places in a dictionary. Collocations, synonyms and the like are linked to polysemes in the same manner, that is, by adding polyseme ids to collocations and synonyms. Notice how our use of the polyseme table differs from how we normally talk about polysemes. Normally, a lemma would have zero, two or more polysemes, never one. However, it makes sense to use the polyseme table to store the definition for a record, even if we consider it a lemma record with only one meaning. It should also be noticed how the fields in the database diverge from the fields provided by the lexicographer.

Using Relation Tables

We want to make "first references", in effect linking two lemmata to each other by referring the user from one lemma to another. Now, our field list looks like this:

- Lemma
- Homonym number
- Style
- First reference
- Definition
- Polyseme number

We do this by making a link table:



The field "from lemma id" is the id of the lemma we want to link from. The field "to lemma id" is the id of the lemma we want to link to. In this way, a lemma can link to one or more lemmata. This allows us to make stable connections between lemmata. Had we simply used the lemma and its homonym number to make the link, the connection would not be stable. What if the spelling changed? Or what if the homonym number changed? With this structure, we do not need to worry about that.

We can use the same technique for second references, that is, by linking polysemes. Again, we see the field lists diverge even more. We also see that the way we represent data in the database is widely different from the way it is presented to users of the database via the UI. We cannot show simple ids of the lemmata when we present links between lemmata. For the users of the dictionary, we present links to the articles. For the lexicographers, we present lists of lemmas, which enable them to link lemmata.

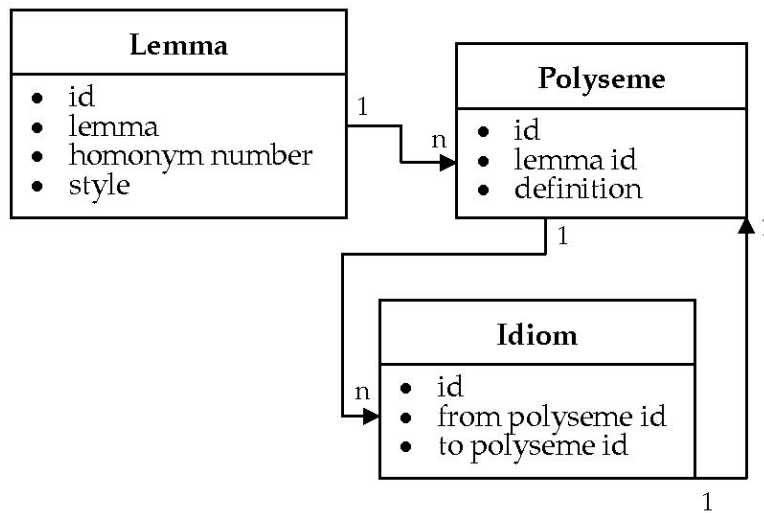
A More Complex Example

If we expand our field list with idioms, we get the following list:

- Lemma
- Homonym number
- Style
- Definition
- Polyseme number
- First reference
- Idiom
- Idiom meaning
- Style

We already know that we want to be able to represent idioms as individual

articles at a later stage. We notice how an idiom can be represented exactly like a lemma with a polyseme. Given that we want to transform idioms into individual articles at a later stage, we can use the lemma and polyseme table to store idioms in the following way:



Firstly, notice that we have made an extra field in the lemma table for the new lemma types called idiom lemma resp. proverb lemma. Because we are storing both lemmata and idioms, we need to know which is which. Secondly, we have made a new table that links one polyseme to another. The "from polyseme id" is the polyseme id which has the idiom. The "to polyseme id" is the polyseme id of the polyseme with the definition of the idiom. Using the lemma id in the polyseme table, we can get the lemma field which is our idiom. This "abuse" of our database solves our problem of letting idioms become individual articles at a later stage while still making us able to link them to polysemes as regular idioms.

This structure also allows us gradually to let the idioms grow into articles and make links directly to them simply by changing their type from idiom to lemma when they become articles themselves. Again, we see how our database has diverged even further from the field list. This is the result of the functional demands we set up for the database. These are hard to express in a field list, and the database expert on the team needs to be aware of these things by asking the right questions to the lexicographer. Furthermore, the UI would also need to display data in a very different manner compared to how it is represented in the database. The abstraction in the database should be hidden from all users of the database.

4. Lexicographical Database

Every database is constructed for use within a narrow or a broad field. If a database is used for its designed function, we can speak about a genuine use of the database; if not, we speak about a non-genuine use. For example, Excel is conceived as a spread sheet processor; this is the genuine use of Excel. But Excel can also be used and is used in dictionary projects; this is a non-genuine use of Excel. Different databases designed to be used in lexicographical work will be different according to the need of the lexicographical project. From these considerations, we can conclude:

Database: A structured collection of values

Lexicographical Database (LDB): A database constructed to contain lexicographical data

References

- Almind, R.** 2005. Designing Internet Dictionaries. *Hermes, Journal of Linguistics* 34: 37-54.
- Atkins, B.T.S. and M. Rundell.** 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.
- Bergenholtz, H.** 2010. Needs-Adapted Data Access and Data Presentation: Bergenholtz, H. 2010. *Doctorado Honoris Causa del Excmo. Sr. D. Henning Bergenholtz*: 41-57. Valladolid.
- Bergenholtz, H.** 2012. Conceptions for Different Types of Language Tools. Ndinga-Koumba-Binza, Hugues Steve and Sonja E. Bosch. (Eds.). 2012. *Language Science and Language Technology in Africa: Festschrift for Justus C. Roux*: 328-339. Stellenbosch: SUN PReSS.
- Bergenholtz, H.** 2013. Wortbildungsangaben als Hilfe für den Zugriff auf andere Datentypen und als Hilfe bei kommunikativen und kognitiven Informationsbedürfnissen. Klosa, Annette (Ed.). 2013. *Wortbildung im elektronischen Wörterbuch*: 133-156. Tübingen: Gunther Narr.
- Cabré, M.T.** 1998. *Terminology. Theory, Methods and Applications*. Amsterdam/Philadelphia: John Benjamins.
- Krestova, S. and P.J. Nürnberg.** 2013. *Multilingual Dictionary of Lexicographical Terms*. (accessed 23 May 2013 http://www.papillon-dictionary.org/static/info_media/42808981.pdf.)

Le traitement macrostructural de l'emprunt au français dans les dictionnaires locaux au Gabon

Edgard Maillard Ella, *Conseil Africain et Malgache pour l'Enseignement Supérieur, (CAMES), Libreville, Gabon*
(ellamaillard@yahoo.fr)

Résumé: Les populations locales empruntent souvent de façon pragmatique aux autres langues les mots dont elles ont besoin dans les leurs. Cela est plus applicable lorsque deux populations de langues différentes se côtoient comme c'est le cas avec français et les langues locales du Gabon. L'emprunt est un processus inconscient et un facteur constitutif de la vie de toutes les langues du monde. C'est pour cela que bien que moins nombreux que les mots locaux, les mots d'emprunt sont très courants dans leur vocabulaire. Certains de ces emprunts sont remplacés par des termes locaux lorsque les populations locales se sont familiarisées avec les concepts ou les choses véhiculés par les mots étrangers. Les uns sont gardés tels quels, et d'autres sont morpho-phonologiquement modifiés afin d'être adaptés aux caractéristiques linguistiques locales. Un dictionnaire est confectionné en fonction des usagers. C'est ainsi que le lexicographe adopte à son tour cette approche pragmatique pour que le public se reconnaisse dans l'ouvrage produit. Pendant le traitement macrostructural quand les mots à inclure comme lemmes sont choisis, il incorpore dans un premier temps les emprunts utilisés par les populations locales. Progressivement, il incorporera dans les éditions revues et corrigées, les termes locaux qui ont remplacé certains emprunts ainsi que les modifications morpho-phonologiques subies par les autres. Dans ce travail, nous voudrions suggérer cette approche pragmatique aux personnes qui confectionnent les dictionnaires au Gabon et qui sont souvent confrontées à l'absence de plusieurs mots modernes dans les langues locales. Ce travail va à l'encontre de la création lexicale qui est connue comme étant une approche puriste et qui s'avère souvent être laborieuse et incertaine. Nous voudrions montrer que l'approche pragmatique qui incorpore les deux approches, constitue une méthodologie qui présente plus d'avantages que une approche puriste. Le dialecte fang ntoumou de Bitam est utilisé ici comme exemple pour étayer ce travail qui est applicable à toutes les autres langues locales du Gabon.

Mots clés: APPROCHE PRAGMATIQUE, APPROCHE PURISTE, CREATION LEXICALE, DICTIONNAIRE, EMPRUNT, FRANÇAIS, LANGUES LOCALES

Abstract: The Macrostructural Treatment of Borrowings from French in Local Gabonese Dictionaries. Local populations often pragmatically borrow words from other languages, which they need in theirs. This applies mostly when two populations with different languages come in contact as is the case with French and the local languages of Gabon. Borrowing is an unconscious process and an essential factor in the life of all the world's languages. This is why, although they are less numerous than local words, borrowed words are very common in their vocabularies. Some of these borrowings are replaced by local terms when local populations become

familiar with the concepts or things referred to by the foreign words. Some are retained as they are and others are morphophonologically modified in order to adapt to local linguistic characteristics. A dictionary is compiled in accordance with the users. This is why the lexicographer in his turn adopts this pragmatic approach so that the users will identify themselves with the finished work. During the macrostructural treatment when words to include as lemmas are chosen, he initially incorporates borrowings used by the local population. Gradually in revised and corrected editions, he will incorporate local terms which have replaced certain borrowings as well as morphophonological changes undergone by others. In this article, we want to suggest this pragmatic approach to persons who compile dictionaries in Gabon and who are often faced with the absence of many modern words in the local languages. This article goes against lexical creation which is known as being a purist approach and which often proves to be laboured and dubious. We want to show that a pragmatic approach, which incorporates both approaches establishes a methodology, which presents more advantages than a purist approach. The dialect Fang Ntumu of Bitam is used here as an example to support this article, which is suitable for all the other local languages of Gabon.

Keywords: BORROWING, DICTIONARY, FRENCH, LEXICAL CREATION, LOCAL LANGUAGES, PRAGMATIC APPROACH, PURISTIC APPROACH

1. Introduction

L'une des nécessités de confectionner des dictionnaires au Gabon aujourd'hui répond au besoin de standardiser les différents lexiques des langues locales. Il y a en effet une volonté de l'État de faire en sorte qu'à l'instar du français, ces langues puissent être également enseignées à l'école¹. Or, dès qu'une langue doit être enseignée, son lexique doit nécessairement s'élargir et se moderniser.

Plusieurs personnes s'engagent actuellement à confectionner des dictionnaires dans ce sens. Il importe donc d'attirer l'attention des uns et des autres sur les risques d'une pratique lexicographique désordonnée. La confection des dictionnaires au Gabon doit se faire selon les règles de l'art afin de produire des ouvrages de bonne qualité. Comme le souligne Gouws (2001: 96), la théorie de la lexicographie n'est pas simplement une théorie pour une théorie, c'est une théorie qui permet d'améliorer la qualité des dictionnaires.

Par ailleurs, il est également nécessaire de souligner l'importance, le rôle et la contribution des locuteurs dans cette entreprise. En effet, d'aucuns semblent souvent oublier que c'est une image d'Epinal, comme nous le faisons souvent assez clairement observer dans nos travaux, que la langue appartient aux locuteurs. Ces derniers ne servent pas que de simples informateurs, mais ce sont bel et bien les principaux acteurs de la langue car ce sont eux qui la façonnent depuis la nuit des temps. La minorité compétente de grammairiens, linguistiques et lexicographes qui pensent qu'ils auraient le pouvoir de statuer sur la langue ne font en réalité que décrire, organiser et standardiser la façon dont les locuteurs pratiquent leurs langues. À ce titre, il est obligatoire de les prendre en compte dans chaque aspect du traitement des langues ainsi que de leur développement. Diop (1979: 345) souligne justement à ce propos que le développement

d'une langue est une entreprise de longue haleine qui est tantôt spontanée, tantôt confiée à des experts, sans que les deux modalités s'excluent.

Le lexicographe² est conscient de ce fait car le dictionnaire est un ouvrage de base qui a une validité sociale acquise et une reconnaissance sociale consacrée. C'est pour cela que celui-ci va vers ses publics, il en est, cf. Kalonji Zezeze (1993: 42). Il publie donc le dictionnaire pour que ce soit un ouvrage dans lequel le public puisse se reconnaître.

C'est à ce propos que dans ce travail, nous nous appuyons sur la façon pragmatique avec laquelle les populations locales empruntent aux autres langues les mots qui manquent dans les leurs. Cette pratique linguistique des populations, destinée à développer et moderniser leurs langues, est le fondement de l'approche pragmatique que nous suggérons dans ce travail. Il s'agit du traitement macrostructural des emprunts au français dans les dictionnaires locaux au Gabon. L'approche pragmatique que nous proposons dans ce travail n'est pas une approche nouvelle. Madiba (2001) l'a suggérée comme modèle pour la modernisation de la terminologie des langues sud africaines. Joseph (1984, cité dans Madiba 2001: 64) l'a recommandée dans la standardisation d'une langue. Jernudd (1977, cité dans Madiba 2001: 64) qui critique l'approche puriste du développement terminologique de l'Hébreu, fait également appel à certaines réalités linguistiques qui sont pragmatiques. Marshad (1984) et Mwanosoko (1990) ont, quant à eux, proposé l'approche pragmatique dans la modernisation du Swahili.

Pour faire ressortir l'intérêt, les objectifs et les enjeux de ce travail, nous allons d'abord commencer par présenter l'approche puriste qui prône la création lexicale dans le développement et la modernisation de la terminologie des langues. Nous allons nous appuyer sur les travaux de Madiba (2001) et nous allons naturellement contextualiser l'approche puriste dans le cas précis du Gabon. C'est dans cette même partie et en nous référant toujours à Madiba (2001) que nous examinerons également les limites de cette approche.

Dans une seconde partie et en nous renvoyant encore une fois encore au travail de Madiba (2001), nous discuterons de l'approche pragmatique qui est en faveur de l'emprunt direct à une autre langue pour le développement et la modernisation de la terminologie d'une langue. En prenant des exemples concrets de l'emprunt au français dans le dialecte fang ntoumou de Bitam, nous allons montrer qu'emprunter au français dans le traitement de la macrostructure des dictionnaires locaux est conforme à la pratique linguistique des populations locales. En sus de présenter plus d'avantages que l'approche puriste, nous ferons aussi observer que l'approche pragmatique de la collecte des données destinées à la macrostructure incorpore l'approche puriste qui, au fil des années, se déroule de façon plus naturelle. Nous verrons par ailleurs que cette approche est l'application d'un aspect du processus lexicographique. L'approche pragmatique aboutira à la production de dictionnaires locaux au Gabon caractérisés par des dispositifs spécifiques correspondant aux principes de la métalxicographie tels qu'énoncés par Wiegand (Smit 1996). Il s'agit de la prévisibilité, de la calculabilité, de l'analysabilité et de la contrôlabilité.

2. L'approche puriste dans le traitement de la macrostructure

2.1 Approche définitoire et les raisons

Nous définirons l'approche puriste, dans le traitement de la macrostructure des dictionnaires africains en général et gabonais en particulier, comme le recours immédiat à la création lexicale en vue d'élaborer les mots n'existant pas dans les langues locales. Cette approche exclut tout emprunt à la langue officielle européenne en particulier. C'est le cas du français au Gabon.

De nombreuses personnes qui confectionnent les dictionnaires dans ce pays sont souvent fortement motivées par le désir de conserver leurs langues et cultures locales en éradiquant et en évitant d'utiliser les mots du français en particulier. Ces mots sont considérés comme des impuretés dont l'usage aura pour conséquence la perte et l'aliénation de l'identité des langues et cultures locales. Cette approche est donc souvent légitimée par des critères idéologiques, politiques, sociaux et émotifs. Elle n'est pas un fait nouveau car elle a été en vigueur France par exemple qui a longtemps considéré l'influence de l'anglais sous cet angle, cf. Weinstein (1989). C'est ainsi que dans le processus de confection de leurs dictionnaires, plusieurs lexicographes gabonais excluent automatiquement dans le choix des items lexicaux à inclure, comme lemmes ou entrées de nos dictionnaires, de nombreux mots français pourtant déjà couramment utilisés, connus et familiers aux populations.

Or, faute des équivalents en langues locales pour le moment, les mots empruntés au français sont naturellement utilisés dans les conversations des populations. C'est ainsi qu'en fang ntoumou de Bitam, il n'est pas rare d'entendre par exemple³:

- *Ma ke a **banque** a ke a gnong **crédit** akala a kousse metoua ya a long nda.*
(Je vais à la **banque** prendre un **crédit** pour acheter une voiture et construire une maison.)
- *Ma ke a kissouane a ke a kousse be **carreaux** ya **ciment** akala e nda ma long.*
(Je vais en ville pour aller acheter des **carreaux** et du **ciment** pour la maison que je construis.)
- *Be ve **nommer** e nome-ngouane wôme **Ministre** e **Conseil** be **Ministres** a ve lote ongwuéghe.*
(On a **nommé** mon beau-fils **Ministre** lors du **Conseil** de **Ministres** qui s'est tenu hier.)

Pourtant aucune impureté ne transparait lorsque les mots français sont si normalement utilisés dans les conversations quotidiennes dans les langues locales du Gabon comme nous le voyons dans cet exemple avec le fang ntoumou de Bitam. Il en sera de même à l'écrit, comme nous venons de le transcrire. Considérer les mots d'une langue étrangère mais connue et familière comme des éléments impurs et des facteurs de dysfonctionnement et d'aliénation de l'iden-

tité des langues et cultures locales du Gabon n'est pas une approche objective. Tout au contraire, celle-ci est contre nature car l'emprunt est un processus inconscient et un facteur constitutif de la vie de toutes les langues du monde. En effet, nous pouvons considérer que ce processus linguistique résulte de la capacité de l'instinct de survie de l'homme à s'adapter à son environnement politique, économique et social sans cesse en mutation en utilisant autant qu'il peut des éléments qui sont à disposition. Dans le cas de la langue, les emprunts font partie de ces éléments.

Pour revenir au cas de la France, nous pensons que l'invasion des mots de l'anglo-américain particulièrement, n'est qu'un refoulement qui a été et est plutôt révélateur d'un conflit d'intérêt entre superpuissances en lutte pour la suprématie mondiale. Dans ce contexte, l'influence de la langue issue d'un autre pays est peut-être considérée comme un signe de faiblesse ou d'infériorité. Cette approche puriste est vaine car le développement de la technologie et la domination de l'anglo-américain dans les sciences et les techniques actuelles confère une suprématie considérable de la langue anglaise à l'échelle planétaire qui n'est pas subie que par la seule langue française. L'anglais est devenu la *lingua franca* du monde contemporain, c'est-à-dire la langue véhiculaire des communications internationales, tant sur le plan commercial que culturel, scientifique, technologique, diplomatique et politique. Pour preuve, de nombreux programmes télévisés en France et à forte audience présentent des titres en anglais en vue d'être plus accrocheurs: Money drop et Secret Story qui sont des programmes télévisés d'une des chaînes publiques françaises, à savoir TF1.

Au Gabon et dans l'arrière-pays, grande fut notre surprise en lisant «Air Force One» comme enseigne d'un snack bar (encore une fois un anglicisme) à Ovan, dans le Département de la Mvoug, province de l'Ogooué-Ivindo. Il est évident que le «Air Force One» d'Ovan n'est pas le seul établissement commercial au Gabon portant un nom anglais. C'est dire combien de fois la pratique langagière qui est inconsciente, obéit naturellement à des besoins prosaïques tels que le business (encore un anglicisme) comme dans le cas que nous venons d'illustrer. En empruntant naturellement à une autre langue un mot qui lui semble plus approprié pour satisfaire un quelconque besoin, l'homme fait que la langue transgresse les considérations émotives, idéologiques et sociopolitiques qui sont contre nature dans le processus de la pratique linguistique.

Dans le cas du Gabon, à l'instar de la plupart des pays d'Afrique noire francophone, l'approche puriste est justifiée par un élan de patriotisme ou de nationalisme peut-être parfois exacerbé. Celui-ci débouche sur une idéologie anti langue française dans le processus de développement des langues locales. Cette approche résulte aussi de la volonté d'ériger ou de rétablir une forme d'indépendance face à l'invasion de la culture, de la politique, du commerce ainsi que de la technologie de l'ancienne puissance colonisatrice occidentale comme le souligne Madiba (2001: 56). Cependant, comme nous avons tenté de le démontrer, l'approche puriste va à l'encontre d'un processus inconscient et naturel de la langue dans l'aspect du développement et de la modernisation de la terminologie.

2.2 Les limites de l'approche puriste

2.2.1 La menace de l'intégrité de l'identité culturelle et de la souveraineté nationale

L'un des arguments ou le seul qui revient régulièrement de la part des personnes qui prônent l'approche puriste dans le traitement de la macrostructure lors de la confection de dictionnaires gabonais, demeure l'aliénation ou purement et simplement la perte de l'intégrité de notre identité culturelle et de notre souveraineté nationale. Nous pouvons dire que cette crainte n'est pas à redouter si nous nous référons à l'histoire des autres langues du monde. C'est le cas de la langue anglaise dont 50% à 60% du vocabulaire est d'origine française et latine, cf. SALIC (*l'Histoire de la langue française*). Cette langue qui est la plus diffusée au monde aujourd'hui n'a pas souffert d'une perte de son identité et celle de la culture de ces locuteurs

C'est aussi le cas de la langue française. L'identité culturelle et de la souveraineté nationale de la France sont intactes quand bien même plusieurs mots proviennent d'autres langues. Henriette Walter et Gérard Walter, deux linguistes lexicologues d'origine française, ont effectué une analyse minutieuse de 70 000 termes puisés dans les dictionnaires *Le Petit Larousse* et *Le Petit Robert*, cf. SALIC (*l'Histoire de la langue française* 23). Cette analyse révèle que 8 088 mots proviennent des emprunts aux langues étrangères, soit 11% du corpus. Ce pourcentage est inférieur à la réalité si d'une part, le latin et le grec sont pris en compte. Et d'autre part, si les lexiques spécialisés d'origine anglaise sont considérés dans cette analyse compte tenu du développement de la technologie et la domination de l'anglo-américain dans les sciences et les techniques actuelles. Cette étude montre par la même occasion que le nombre des langues prêteuses s'élève à plus de 120. Il s'agit entre autres du latin, du grec, de l'anglais, de l'italien, de l'espagnol, du hollandais, de l'arabe et même du swahili, du coréen et de l'iranien. C'est dire que si l'emprunt de mots à une ou plusieurs langues étrangères constituait un facteur d'aliénation et d'intégrité de l'identité culturelle et la souveraineté nationale d'un pays, la France et l'Angleterre seraient aujourd'hui des pays ayant une identité culturelle et une souveraineté nationales peut-être inexistantes. Ce qui est évidemment loin d'être le cas.

C'est dire que l'intégrité de l'identité culturelle et de la souveraineté nationale du Gabon ne pèsent sous aucune menace d'être aliénées si les mots sont empruntés au français pour combler le manque de mots modernes qui existe dans les langues locales. Il n'y aura donc aucun mal à ce que, dans le traitement dictionnaire, ces mots soient inclus comme lemmes ou entrées dans la macrostructure des ouvrages locaux. Il n'en sera pas moins une hérésie scientifique dans cette démarche lexicographique⁴.

2.2.2 Lenteur, difficultés d'adoption et refoulement des termes créés

La confection de dictionnaires constitue une activité laborieuse et de très

longue haleine. Il est connu que Pierre Larousse (1817–1875) est mort bien avant d'avoir vu son dictionnaire terminé, tant on se rend rarement compte du fait que les quantités de travail et de temps à consacrer à la production de ce livre sont énormes. Si l'approche puriste qui consiste à créer les mots qui n'existent pas dans les langues locales du Gabon pour la confection des dictionnaires locaux au Gabon est adoptée, cela signifie qu'il faut s'engager à créer quelques dizaines de milliers de termes dans des domaines autant nombreux que diversifiés. Pour montrer l'ampleur de la tâche à accomplir, il s'agit entre autres des termes provenant de l'économie, de la politique, du social, de l'administration, de la justice, des finances, du commerce, de l'industrie, des sciences exactes et humaines, des techniques, de l'informatique, de l'électronique, de la mécanique, de la médecine, de la télévision, du sport, de la drogue, de la violence, de l'armement, de la culture, de l'histoire, de l'environnement, du cinéma, de la musique, de l'architecture, de l'art, de la danse et du tourisme pour ne citer que ceux-là. Il existe bien d'autres domaines encore.

Cette activité constituerait un énorme travail mais qui serait tout à fait réalisable. Mais à quel moment se terminera-t-il et comment sera-t-il possible de gérer le déferlement de mots qui se créent tous les jours à cause d'une société sans cesse en développement et en mutation? C'est pour cette raison que les dictionnaires de langue tels que ceux produits par les éditions Larousse, se voient aujourd'hui obligés de produire une édition chaque année, ce qui n'était pas le cas il y a de cela quelques décennies, non pas pour des raisons commerciales mais pour fournir aux utilisateurs le lexique le plus actualisé possible. À cela, il faut encore ajouter le fait que les populations locales prendront nécessairement du temps pour adopter les termes créés même si ces derniers proviennent de leurs langues-mères. Pire, elles peuvent tout simplement les rejeter comme le montrent les travaux des auteurs tels que Mwansoko (1990) et Samsom (1988). Ces auteurs montrent clairement dans leurs travaux comment les termes puristes développés dans la mécanique en Swahili ont été ignorés par le public qui leur a préféré les termes empruntés.

Nous pouvons aisément imaginer qu'il sera laborieux en termes de temps et de logistique pour mener à bien les enquêtes de terrain nécessaires. Ce qui est loin d'une affirmation gratuite de notre part comme on nous le fait souvent remarquer.

L'adoption de l'approche puriste n'est donc pas aisée dans un premier temps pour rendre disponible, dans des délais certains, une terminologie moderne des langues locales du Gabon. Dans un deuxième temps, cette approche ne réduira pas dans des délais acceptables l'écart de modernisation qui existe entre ces langues locales et le français afin que ces deux populations de langues soient enseignées de façon équitable.

C'est pour cela que nous, ainsi que de nombreux autres auteurs tels que Madiba (2001: 74), nous montrons relativement sceptiques par rapport à la création lexicale qui est un modèle plausible, mais qui implique un travail de recherche long et laborieux de termes en langue locales pour exprimer les con-

naissances nouvelles et étrangères. Or, l'immense écart de terminologie moderne qu'il y a entre les langues locales africaines et européennes ne permet pas d'accuser encore des retards considérables. Comparaison est raison dans ce cas précis. En effet, il suffit de faire tout simplement comme cela l'a été dans toutes les langues du monde: emprunter à une ou plusieurs autres langues les mots manquants. Il s'agira ensuite de refléter cette mémoire et ce vécu des langues-récepteurs dans la confection des dictionnaires locaux en incorporant ces mots étrangers qui font dorénavant partie du patrimoine linguistico-culturel local comme entrées. Le dictionnaire est un ouvrage extensif, historisant, actuel et ouvert. Comme le dit Kalonji Zezeze (1993: 75) le dictionnaire est à la fois mémoire et vécu. Les populations locales du Gabon montrent la voie en empruntant au français les mots qui manquent dans leurs langues locales.

L'approche puriste, qui rappelons le semble être un processus contre nature dans le mécanisme de développement et de modernisation de la terminologie des langues, n'est pas un modèle adapté au développement et à la modernisation de la terminologie des langues locales du Gabon. Cette approche n'est donc pas également adaptée à la confection de dictionnaires locaux dans le traitement de la macrostructure dans lequel il faudra d'abord emprunter au français les mots qui manquent dans les langues locales.

L'approche qui doit être adoptée pour le développement et la terminologie des langues locales du Gabon ainsi que pour le traitement efficace de la macrostructure des dictionnaires locaux ne devrait pas être puriste ni même anti-puriste également, mais «*apuriste*» pourrait-on dire. Autrement dit, une approche qui soit la plus neutre, mais aussi et surtout qui soit la plus objective possible en tenant compte des réalités du contexte dans lequel la langue est utilisée. Cette approche que nous avons déjà suggérée avant ces lignes se référera dans cette étude en termes d'*approche pragmatique*, cf. Madiba (2001).

3. L'approche pragmatique dans le traitement de la macrostructure des dictionnaires gabonais

3.1 Approche définitoire

Nous allons nous appuyer sur les travaux de certains auteurs afin de tenter de donner une définition de l'approche pragmatique dans le contexte précis du Gabon. Marshad (1984) et Mwansoko (1990) ont proposé les contours de la définition de l'approche pragmatique dans la modernisation du Swahili. Joseph (1984) a également indiqué quelques principes. À partir de ces orientations, nous définissons l'approche pragmatique dans le traitement de la macrostructure de dictionnaires locaux au Gabon, comme l'incorporation des mots empruntés au français dans le choix des mots à inclure comme lemmes ou entrées. Cette démarche n'est pas hasardeuse car, comme l'observe Ndong Ngoua (2007: 29), aucune langue n'est riche en elle-même. Bien au contraire, la langue est fonction du milieu physique, de l'organisation politique et sociale,

du degré de technicité et d'emprunts aux autres langues. En effet, toutes les nations, soucieuses de domestiquer la science et ses applications, ont été confrontées aux mêmes difficultés que connaît le Gabon. Hier, c'était le cas des démocraties européennes telles que la France, l'Allemagne et la Grande Bretagne qui ont considérablement emprunté au latin et au grec. Aujourd'hui encore elles empruntent à l'anglo-américain qui devient la lingua franca. C'est aujourd'hui le cas des nouvelles puissances nucléaires, autrement dit l'Inde, le Pakistan et la Chine Populaire qui ont emprunté et adopté tous les mots internationaux scientifiques nécessaires même si elles sont d'origine occidentale, cf. Ndong Ngoua (2007: 31). Les pays émergents tels que le Gabon doivent également emboîter le pas de cette logique en s'accaparant de tous les mots internationaux scientifiques nécessaires qui sont dans le français pour enrichir et moderniser leurs langues locales.

3.2 Les raisons de l'emprunt au français

3.2.1 Le français et les langues locales, langues partenaires et compagnons de route

Le choix d'emprunter des mots issus du français dans le traitement de la macrostructure des dictionnaires locaux au Gabon n'est pas arbitraire. Comme le français et l'anglais, le français et nos langues locales sont de vieux compagnons de route⁵. Il y a même une relation d'emprunteurs mutuels entre ces langues comme c'est le cas entre le français et l'anglais. En effet, le français du Gabon qui est une variété du français standard en usage dans ce pays, quand bien même que celui-ci ne dispose pas encore d'une reconnaissance officielle, résulte de l'influence des langues locales du Gabon. L'approche qui consiste à emprunter des mots au français pour les intégrer dans la macrostructure des dictionnaires des langues locales du Gabon semble donc se présenter comme un choix d'emblée évident et naturel. En effet, le français et ces langues locales constituent deux populations de langues qui sont en contact depuis près d'un siècle. Même si ces deux populations de langues se distinguent par la famille du point de vue linguistique, il n'en demeure pas moins que le français fait maintenant partie entière du patrimoine linguistico-culturel du Gabon. Que certains linguistes africanistes gabonais le veuillent ou non, il faut nécessairement compter avec le français lorsqu'il est question du développement et de la modernisation des langues locales de ce pays. Le français est la langue de l'administration, de l'enseignement, de la mobilité sociale et du pouvoir du Gabon. Cette langue existe dans le quotidien et l'imaginaire des Gabonais car elle constitue la langue intermédiaire ou langue «hub». C'est-à-dire qu'au Gabon, le français est la langue véhiculaire qui permet la communication entre citoyens de groupes linguistiques différents. De par son omniprésence et son statut, il résulte aujourd'hui que le français est la langue maternelle d'un très grand nombre de Gabonais surtout au sein de la population jeune. Elle est de

ce fait le substrat⁶ de nombreux Gabonais est la langue «*compagnon*» de leurs langues locales. Alors, ne pas tenir compte des mots du français qui sont utilisés, connus et familiers aux populations locales dans le traitement de la macrostructure des dictionnaires locaux au Gabon, ne serait pas une approche objective et naturelle dans le processus de développement et la terminologie des langues.

3.2.2 L'emprunt au français comme la restitution d'un aspect de l'histoire du peuple gabonais dans les dictionnaires

Un dictionnaire n'est pas seulement un ouvrage de référence contenant l'ensemble des mots d'une langue. Il inclut aussi des informations encyclopédiques essentielles qui apportent une connaissance sur notre histoire, une ouverture sur le monde et une réponse à toutes les curiosités des utilisateurs. Les dictionnaires traduisent, véhiculent et restituent tous les aspects de l'histoire des peuples. Les dictionnaires du Gabon doivent traduire et restituer l'étroitesse des relations qui existent entre le Gabon et la France. Ne pas considérer plusieurs mots du français comme lemmes ou entrées dans nos dictionnaires alors qu'ils sont abondamment utilisés, connus et familiers aux populations dans l'usage de leurs langues locales à l'oral, c'est nier une partie de l'histoire du Gabon. En effet, il s'agira donc de réfuter l'existence du français comme langue de la population en contact étroit avec les langues de ce pays. Cela déformerait peut-être pour ainsi dire assez considérablement une partie de l'histoire du Gabon dans laquelle notre pays est étroitement et intimement lié à la France. Nous avons souligné que comme le dit Kalonji Zezeze (1993: 75), le dictionnaire est à la fois mémoire et vécu. De ce point de vue, il nous paraît impossible de ne pas répertorier l'ensemble des faits linguistico-culturels de la société gabonaise dans laquelle le français est inclus et de s'en servir objectivement et naturellement pour développer et moderniser les langues locales de cette société.

3.2.3 L'emprunt au français comme approche systématique du traitement de la macrostructure des dictionnaires gabonais

La langue française est aujourd'hui une langue ouverte sur le monde car elle est faite d'influences mutuelles de diverses langues sur la scène politique, commerciale, économique, scientifique, technologique et sportive et plusieurs autres domaines encore aussi nombreux que divers. C'est donc une langue «*mondiale*» dont les termes que nous utilisons quotidiennement, auxquels les populations locales du Gabon sont accoutumées et qu'elles connaissent bien, nous aideront à moderniser assez facilement nos langues locales. Avoir cette langue dans le patrimoine linguistico-culturel du Gabon est une opportunité pour ce pays pour développer leurs langues et cultures locales. C'est de cette même façon

que les démocraties européennes ont développé leurs langues en s'appuyant sur le latin et le grec dont les pays d'origine, à savoir Rome et la Grèce antique étaient des puissances prépondérantes.

Le français s'est aussi enrichi de nouveaux mots empruntés à l'Italie. Le 16ème siècle fut en effet marqué par la prépondérance de ce pays dans presque tous les domaines en raison de sa richesse économique, sa puissance militaire, son avance technologique et scientifique ainsi que sa suprématie culturelle. Cette influence, mais aussi la proximité⁷ entre la France et l'Italie se sont manifestées par la pénétration de quelque 8 000 mots italiens dans la langue française. Il s'agissait notamment des termes relatifs à la guerre (*canon, alarme, escalade, cartouche*, etc.), à la finance (*banqueroute, crédit, trafic*, etc.), aux mœurs (*courtisan, disgrâce, caresse, escapade*, etc.), à la peinture (*coloris, profil, miniature*, etc.) et à l'architecture (*belvédère, appartement, balcon, chapiteau*, etc.), sans compter les domaines du vêtement, de l'alimentation, de l'équitation et de la musique.

Ignorer les mots du français qui ont pénétré le Gabon et que nous utilisons couramment afin de les inclure dans la macrostructure de nos dictionnaires pour obtenir ceux qui n'existent pas dans nos langues locales, serait à tout point de vue et sans trop grand risque de se tromper, adopter une approche non scientifique et objective. Ce sont plutôt des critères politiques, idéologiques et émotifs qui mettent en avant la sauvegarde de l'identité du Gabon ainsi qu'une volonté d'affirmer totalement l'indépendance vis à vis de la France.

Cependant, s'il est vrai que nous recommandons l'approche pragmatique qui consiste à emprunter au français les mots manquant aux langues locales afin de les inclure comme entrées dans la macrostructure, il est important de montrer comment les usagers vont intégrer ou adopter ces emprunts dans leurs habitudes linguistiques. C'est ainsi que dans les sections qui suivent, nous allons examiner en détail comment s'opèrent les deux phases de l'approche pragmatique dans le traitement de la macrostructure des dictionnaires gabonais. Tout au long de cet examen, nous mettrons en exergue les dispositifs spécifiques de la théorie et de la pratique de la lexicographie qui doivent caractériser les dictionnaires gabonais afin qu'ils soient de bonne qualité et qu'ils contribuent efficacement à l'effort de préservation et de promotion des nos langues locales. Il s'agit de la prévisibilité, de la calculabilité, de l'analysabilité et de la contrôlabilité.

4. La phase de l'emprunt

Cette phase se réfère à l'emprunt car elle consistera tout simplement à incorporer directement dans la macrostructure des dictionnaires de langues locales, les termes français qui n'ont pas officiellement d'équivalents dans ces langues. Il s'agira notamment aussi de conserver les termes modernes qui ont déjà des équivalents en langues locales.

Ces équivalents sont généralement de deux types de termes. Il y a d'une part les emprunts qui ont été substitués ou remplacés par des termes locaux et

d'autre part, ceux qui ont subi des modifications pour se conformer aux structures linguistiques, c'est-à-dire, les structures morphologiques et phonologiques de ces langues. En fang ntoumou de Bitam comme dans presque toutes les langues locales du Gabon, ces deux catégories de mots apparaissent. Nous présentons ci-dessous quelques exemples en fang ntoumou de Bitam.

- Quelques mots substitués ou remplacés par des termes en fang ntoumou de Bitam

akôk (pile électrique), *biyéne* (lunettes), *éloume* (fourchette de table), *mvet* (guitare), *nda mi mbok* (prison), *nda Zama* (église), *nkare fuègne* (pasteur/prêtre), *nseng avion* (aéroport), *nseng ndama* (stade/terrain de football), *ofepeya* (volant de voiture)

- Quelques mots ayant subi des modifications morphologiques et phonologiques

awola (heure, ou de l'anglais «hour»), *dokira* (médecin/titulaire d'un doctorat), *kissine* (cuisine ou de l'anglais «kitchen»), *lama* (lampe ou de l'anglais «lamp»), *metoua* (voiture, de l'anglais «motor»), *nkolo* (montre, de l'anglais «clock»), *sikolo* (école, études ou de l'anglais «school»), *sobo* (savon, de l'anglais «soap»), *tawoula* (serviette de bain, de l'anglais «towel»), *tebele* (table)

Ces mots que nous venons de prendre en exemple ci-dessus illustrent les types de termes locaux existant dans nos langues locales et qui représentent ou désignent des choses et des concepts ne provenant pas de nos sociétés locales. Ce sont à ces deux types de termes que seront donc ajoutés directement les termes empruntés au français qui seront aussi constitués de deux types. Il s'agit d'une part, de ceux que les populations locales utilisent quotidiennement dans leurs langues et qui leur sont par conséquent connus et familiers. Ce sont des mots tels que *ciment*, *brique*, *goudron*, *banque*, *crédit*, *épargne*, *compte bancaire*, *béton*, *Député*, *Président*, *Ministre*, *Sénateur*, *fleur*, *Internet*, *informatique*, *géographie*, *cahier*, *crayon*, *carreau*, *règle*, *équerre*, *climatiseur*, *ordinateur*, *géographie*, *brique*, *salon*, *train*, *avion*, *hôtel*, *gaz*, *compas*, *taxi*, *livre*, *surfer (Internet)*, *s'abonner*, *affecter (être)*, *nommer (être)*, etc.

D'autre part, il s'agit essentiellement des mots du français qui appartiennent au vocabulaire des domaines spécialisés. Toutefois, les populations locales, étant donné leur familiarité avec le français, adoptent souvent facilement et rapidement ces termes dès qu'elles en prennent connaissance.

L'incorporation directe de ces mots empruntés au français, tout en conservant ceux qui en langues locales décrivent les concepts et choses modernes, permettra d'avoir des dictionnaires de nos langues locales dont les macrostructures vont présenter des vocabulaires modernes dans un temps relativement court avec un effet immédiat. En effet, les populations qui utilisent et connaissent déjà ces mots vont les adopter aussitôt. Par ailleurs, les mots qu'elles ne

connaîtront pas, ne leur paraîtront pas totalement étrangers car ces populations connaissent déjà bien la langue française. En dehors de ces aspects, avec ces macrostructures, les lexicographes disposeront des listes de termes dont ils pourront dénombrer et analyser la quantité en fonction de ceux issus directement des emprunts, ceux issus des emprunts mais qui ont été phonologiquement et morphologiquement modifiés et ceux issus des langues locales. Cela conférera à ces dictionnaires gabonais les dispositifs spécifiques essentiels à la confection de bons dictionnaires selon les principes pratiques et théoriques de la métalexigraphie que sont dans cette phase de l'emprunt, la calculabilité et l'analysabilité.

5. La phase de substitution

Dans le traitement de la macrostructure des dictionnaires de nos langues locales, la phase de substitution qui interviendra à court et à moyen termes consistera à remplacer progressivement tous les termes que les lexicographes ont empruntés au français sans savoir qu'ils existaient déjà dans la langue locale traitée. Cette procédure importante se réalisera grâce à l'interaction que les producteurs de dictionnaires doivent entretenir avec les locuteurs. Ils se mettront à la disposition des locuteurs pour recueillir ces mots, en s'assurant néanmoins que ceux-ci sont aussi effectivement reconnus par la communauté.

En effet, comme cela a été analysé par plusieurs lexicologues, l'expression populaire «maintenant, ce mot est dans le dictionnaire» est assez trompeuse. Elle semble bien montrer que les locuteurs, pendant un temps, ont l'intuition que tel mot n'est pas légitime, c'est-à-dire, qu'il «sonne» encore «étranger» et qu'il faut une autorité extérieure pour en déclarer le caractère français. En fait, le processus est inverse.

Les dictionnaires sont en réalité des instruments plus descriptifs que prescriptifs, qui ne font que sanctionner l'usage, quelle que soit la définition donnée à ce terme, et le représenter. Qu'un mot étranger entre dans le dictionnaire ne signifie pas qu'il ait été accepté seulement par une minorité compétente de linguistes, grammairiens et lexicographes qui auraient le pouvoir de statuer sur la langue. C'est une image d'Epinal que la langue appartient aux locuteurs et aucun décret officiel ou théorie ne peut les contraindre à changer leurs usages du tout au tout. C'est ainsi qu'à travers les dictionnaires, les lexicographes ne font que signaler les usages qui sont devenus suffisamment courants. Cette première étape de la phase de substitution interviendra immédiatement dans la deuxième et la troisième édition revue et corrigée si nécessaire du dictionnaire confectionné. Aux dispositifs spécifiques lexicographiques de calculabilité et d'analysabilité de la phase de l'emprunt, va s'ajouter dans cette phase de l'emprunt, le dispositif spécifique de la contrôlabilité. L'informatique facilite aujourd'hui considérablement le travail du lexicographe. Avec les corpus électroniques, il sera facile de calculer, d'analyser et de contrôler le flux des termes locaux qui ont été sanctionnés par les locuteurs pour remplacer ceux qui

ont été empruntés au français. Les dictionnaires de nos langues locales seront ainsi de bonne qualité à partir de la manipulation et de l'actualisation minutieuses de leurs textes. Cela permettra d'offrir régulièrement aux utilisateurs de nouvelles éditions dans des délais devenus de nos jours négligeables grâce à la facilité qu'offre aujourd'hui l'outil informatique dans le traitement de textes.

6.1 La phase de l'actualisation de la création lexicale spontanée et naturelle

Dans l'approche pragmatique du traitement de la macrostructure des dictionnaires gabonais, la phase de l'actualisation de la lexicalisation locale va consister à intégrer régulièrement dans les éditions revues et corrigées, les termes empruntés qui ont été remplacés par les termes locaux. Comme nous l'avons mentionné plus haut, ces mots locaux ne sont pas en réalité nouveaux mais ce sont des extensions de sens qu'on leur a donnés et cela est tout à fait normal dans le fonctionnement de toute langue. Cette actualisation va se dérouler tout au long de l'histoire de notre pays. En nous référant aux propos d'Hagège (1987: 78), nous dirons l'évolution des parlers est aussi nécessaire aux lexiques que l'est à l'homme le renouvellement de l'air pur. Ainsi, au sein du lexique d'une langue, les glissements d'un sens à l'autre sont l'aventure permanente des mots pour accomplir la lourde et importante charge qu'ils ont, de refléter constamment les réalités complexes et changeantes de notre société.

Par l'action dynamique et naturelle des locuteurs dans leur intention de communiquer, ceux-ci attribuent des valeurs aux mots pour que les lexiques de leurs langues locales soient toujours adaptés à leurs besoins. Cela constitue d'ailleurs une des conditions du rayonnement d'une langue et se traduit par l'incorporation des vagues de mots migrants véhiculant de nouvelles choses et nouveaux concepts. Plusieurs de ces mots qui s'assimilent aux structures sémantiques, phonologiques et morphologiques de la langue récipiendaire seront digérés, les autres resteront tels qu'ils ont été empruntés. En fang ntoumou de Bitam par exemple, il s'agit dans le premier cas des termes tels que:

- *akôk* (pierre ou caillou) pour désigner «une pile électrique».
- *amane ngouane* (fin du mois) pour désigner «le salaire mensuel».
- *me ayuèghène* (des repos) pour désigner «les vacances».
- *metoua wome a kouane/a bele okouane* (ma voiture **est malade/a une maladie**), *être malade* ou *avoir une maladie* veut dire aussi *être en panne* ou *avoir une panne* en se référant à une automobile.
- *metoua wome a ne wouane* (ma voiture **est en arrêt de panne**), le mort désigne aussi un arrêt de panne de voiture.
- *E Député wa a fe kousse biè e efefègne énguini a dza* (Notre Député nous a acheté/offert un nouveau **groupe électrogène** au village).

Dans le deuxième cas, il s'agit des mots que nous avons pris en exemple dans la section 4, cf. *awola* (heure/hour), ... *tawoula* (serviette de bain, de l'anglais «towel»), *tebele* (table).

Dans le troisième cas, ce sont les termes que nous avons aussi cités dans cette même section, cf. «*ciment, brique, ... surfer (Internet), s'abonner, affecter (être), nommer (être)*», etc.

Nous allons nous intéresser précisément aux mots locaux qui ont été créés pour remplacer les mots empruntés. Les emprunts sont souvent spontanément et naturellement remplacés à long terme par les termes locaux lorsque les choses et les concepts étrangers exprimés par ces termes sont devenus communément connus au sein de la communauté (cf. Madiba 2001: 69). C'est le cas de certains mots remplacés par des termes en fang ntoumou de Bitam que nous avons présentés plus haut. C'est au vu de cette phase que Diop (1979: 345) rappelle à juste titre que le développement d'une langue est une entreprise de longue haleine qui est tantôt spontanée, tantôt confiée à des experts, sans que les deux modalités s'excluent.

En prenant *akôk* (pierre ou caillou), terme local fang ntoumou de Bitam pour désigner «pile électrique» comme exemple de mots locaux qui ont été créés pour remplacer les mots empruntés, un extrait de traitement bilingue français et fang ntoumou de Bitam se présentera ainsi comme suit:

pile [pil] **1** nf a (= tas) angosse **b** [de pont] yini **c** (électrique) **akôk**...

En plus des dispositifs spécifiques de calculabilité, d'analysabilité et de contrôlabilité que vont acquérir les dictionnaires gabonais dont le traitement de la macrostructure va se réaliser en adoptant l'approche pragmatique plutôt que l'approche puriste, ces dictionnaires seront dotés du dispositif spécifique de prévisibilité. En effet, à partir des faits empiriques les lexicographes et linguistes pourront facilement déterminer les types de structures phonologiques et morphologiques des mots empruntés qui ont du mal à être intégrés dans nos langues, et leur prévoir ainsi une durée de vie limitée. C'est de la même façon que les spécialistes pourront assez aisément spéculer sur les mots empruntés qui seront remplacés par l'extension de sens d'un terme local ou ceux qui ne le seront jamais car ce sont des internationalismes⁸.

6.2 Les raisons de la création lexicale spontanée et naturelle après l'emprunt

Il est important de comprendre pourquoi la création lexicale intervient spontanément et naturellement pour remplacer plusieurs mots qui ont été empruntés. Cela est nécessaire pour que les personnes impliquées dans la confection de dictionnaires de nos langues, surtout celles qui sont favorables à l'approche puriste, comprennent le bien fondé de l'approche pragmatique qui régit de façon naturelle le fonctionnement des langues, et qu'il est fondamental de l'appliquer dans le traitement de la macrostructure de nos dictionnaires. Elles

comprendront aussi et surtout que l'approche puriste que nous tous désirons en réalité car elle vise à donner une coloration et une identité locale à nos langues africaines, intervient spontanément et naturellement après avoir appliqué au préalable une approche pragmatique. Autrement dit, il faut en amont emprunter des mots internationaux scientifiques nécessaires inclus dans les langues occidentales.

6.2.1 Le besoin d'une communication effective entre spécialistes, politiciens et analphabètes de la langue européenne

Plusieurs raisons qui poussent spontanément et naturellement à la création lexicale pour remplacer les mots qui ont été aussi empruntés de la même manière ont été examinées à travers plusieurs travaux, notamment ceux de Marshad (1984), Sager (1989) et Mwansoko (1990).

Comme le souligne Madiba (2001: 70), les spécialistes ne rencontrent aucun problème dans l'usage des termes étrangers du moment où ces derniers possèdent de la matière en ce qui concerne la connaissance dans leur domaine de compétence et aussi une connaissance de la langue d'origine de ces mots. C'est aussi le cas pour les politiciens lors des campagnes électorales par exemple chères à notre pays, ainsi que des personnes en charge de la gestion de notre administration. Cependant, lorsqu'il est souvent question de transmettre cette connaissance en langue locale pour les besoins de communiquer avec un public qui ne maîtrise pas bien la langue européenne, le français dans le cas du Gabon, cela doit être fait en des termes locaux que ce public comprendra facilement. C'est dans ce type de circonstances ou de situations que les locuteurs et dans ce cas précis les spécialistes, administrateurs et politiciens locuteurs de leurs langues ou les personnes assignées à l'interprétation, créent spontanément et naturellement plusieurs mots qui permettront de bien transmettre le message et qui seront maintenant utilisés et acceptés par la communauté. C'est à partir de ce moment que ces mots locaux remplaceront les mots étrangers qui désignent les connaissances qui doivent être transmises.

Le besoin de développer des termes locaux résulte du fait qu'un spécialiste, un administrateur et un politicien ont souvent chacun une grande responsabilité envers sa communauté et doivent pour ce faire être capable de communiquer avec des analphabètes de la langue européenne, langue officielle. C'est face à cette exigence que de manière spontanée et naturelle le génie créateur des locuteurs permet d'élaborer des termes locaux qui remplaceront parfaitement les termes étrangers déjà empruntés ou non. Il ne restera plus alors au lexicographe qu'à les recueillir en vue du traitement de la macrostructure des dictionnaires de nos langues locales. C'est ainsi que dans le Département de la Mvoug par exemple, nous avons observé dans le vocabulaire politique l'usage de «édzikete» (ticket de confiance ou ticket qui donne le pouvoir)⁹ qui désigne le «bulletin de vote» et «ekete» (caisse) pour désigner l'«urne». Ces deux termes locaux font clairement comprendre l'importance et l'intérêt du vote aux popu-

lations locales de cette localité. C'est dans ce même contexte que les personnes en charge de l'enrôlement¹⁰ qui se déroule actuellement dans toute l'étendue du territoire national au Gabon, ont initié un terme local, «e binenga e si» (s'enrouler par terre) car les populations locales connaissent plutôt «enrouler» ou «s'enrouler» et non «enrôler». Il ne serait pas donc étonnant que dans le vocabulaire politique d'Ovan, le terme «e binenga e si» deviennent la forme attestée pour signifier «s'enrôler».

Une création lexicale immédiate demandée aux spécialistes, administrateurs et politiciens ne donnera pas lieu au même résultat car il n'y aura pas le contexte favorable offert par l'interaction avec le public, l'acceptation et l'adoption de ce dernier des termes locaux utilisés à la place des termes étrangers.

6.2.2 Durée de vie limitée de certains mots empruntés

La deuxième raison pour laquelle la création lexicale remplace certains mots qui ont été empruntés, se situe dans la nature même de ces mots. Sager (1989) et Mwansoko (1990) relèvent que plusieurs mots empruntés ont une durée de vie limitée car ils sont dotés d'une nature telle qu'ils ont du mal à être assimilés au système linguistique de la langue récipiendaire. Après un certain temps d'usage, ce type de mots révèlent une existence précaire et sont par conséquent fréquemment remplacés. D'où, les limites de l'approche puriste à l'inverse de l'approche pragmatique, où il n'y aura pas le temps nécessaire pour que ce type de mots se révèlent d'eux-mêmes et soient spontanément et naturellement remplacés au fil du temps par les termes locaux.

6.2.3 L'isolement des emprunts dans les énoncés

La troisième raison qui conduit à la création lexicale après l'emprunt vient du fait que les mots étrangers intègrent souvent une langue de façon isolée et ne couvrent pas alors tout le concept qui gravite autour de lui (Madiba 2001: 72-73). Cette nature individuelle fragilise l'existence des mots empruntés qui doivent en fait se fondre dans un énoncé pour donner des sens ou des idées plus complètes et plus générales qui vont au-delà des sens plus restreints véhiculés par ces mot. La phase de la création lexicale spontanée et naturelle se développe donc en essayant d'intégrer le mot emprunté dans un énoncé, c'est-à-dire, une structure conceptuelle plus élargie que ce mot seul. C'est ainsi que dans la phrase suivante en fang ntoumou de Bitam:

Be ve bere ma e étô. Be ve tuè me Directeur Général.

On m'a **promu(e)**. On m'a **placé(e)/nommé(e)** Directeur Général.

En discutant avec un locuteur natif, il en ressort que le verbe «promouvoir» est complexe pour les populations locales. C'est ainsi que celui s'intégrait avec difficulté dans une phrase en fang ntoumou de Bitam dans laquelle il donne un

sens à un énoncé dans cette langue. Dans ce cas précis, il s'agit d'une idée plus large ou plus complète dans le contexte d'une promotion professionnelle. Il a été mieux compris dans le sens de «élever», «augmenter» ou «faire monter», d'où «à bere» qui a ces sens dans ce dialecte fang du Gabon. L'intégration d'un emprunt dans un énoncé ne constitue donc pas un processus fragmentaire, les locuteurs essaient d'intégrer le concept véhiculé par ce mot dans un ensemble qui reflète les différentes structures énonciatives dans lequel ce mot joue un rôle déterminant dans le sens, cf. Sager (1989: 10).

C'est la raison pour laquelle les locuteurs remplacent spontanément et naturellement le mot étranger par un terme local une fois que le concept qu'il véhicule a été assimilé afin d'obtenir une communication plus effective avec des mots locaux qui sont plus clairs et évidemment accessibles. C'est ainsi que le verbe local «a bere» (faire monter, élever) a remplacé «promouvoir» pour que les fang ntoumou de Bitam comprennent facilement dans leur langue, comme c'est le cas, l'idée et l'acte contenus dans le verbe «promouvoir» dans une phrase ou un énoncé.

6.2.4 Les limites des emprunts dans la production d'autres sens

La dernière raison qui engendre souvent le besoin de recourir spontanément et naturellement à la création lexicale pour remplacer les mots empruntés, tient au fait que ces mots ne possèdent pas la même force productive que les termes locaux que les locuteurs maîtrisent mieux. Autrement dit, les locuteurs ne peuvent pas manipuler aussi bien les termes étrangers confinés à véhiculer un seul sens que leurs propres mots auxquels ils peuvent plus facilement ajouter d'autres nouveaux sens. Cette situation fragilise donc de nombreux mots qui sont empruntés et qui sont remplacés à terme par les mots locaux plus faciles à manipuler par les locuteurs.

7. Les avantages de l'approche pragmatique dans le traitement de la macrostructure des dictionnaires gabonais

7.1 Approche systématique dans le traitement de la macrostructure des dictionnaires locaux au Gabon

L'approche pragmatique que nous avons proposée dans cette étude présente plusieurs avantages. Le premier étant de reconnaître que l'emprunt et la création lexicale spontanée et naturelle de certains mots qui se déroulent à terme constituent deux aspects importants dans le traitement de la macrostructure des dictionnaires gabonais. Ce traitement dictionnaire aura également un impact positif sur le développement et la modernisation du vocabulaire des langues locales du Gabon, compte tenu du rôle et de l'importance du dictionnaire dans les langues. La phase de l'emprunt direct permet d'incorporer tous les mots manquants dans de délais de temps relativement courts et raison-

nables, cf. Madiba (2001: 65).

Par ailleurs, l'emprunt va accroître les chances d'acceptabilité des termes empruntés au français. Ces termes que les populations rencontreront comme lemmes ou entrées des dictionnaires de langues locales leur seront déjà en réalité connus et familiers et seront utilisés. Comme le note Madiba (2001: 72), la plupart des emprunts pénètrent souvent une langue à travers les locuteurs eux-mêmes plutôt que par voie institutionnelle.

L'adoption de l'emprunt se révèle aussi importante pour rendre disponible auprès des populations des communautés ethnolinguistiques résidant dans les zones rurales de nouveaux concepts jusqu'ici inconnus à travers l'introduction des mots qui les représentent dans des dictionnaires qui leur sont destinés. La disponibilité de ces mots ainsi que de leurs significations conduira ces populations à se familiariser avec ces mots étrangers et les concepts qu'ils représentent. C'est à partir de cette familiarité qu'il sera possible à la communauté locale de créer des mots du terroir pour remplacer certains mots empruntés. D'où l'occasion de mentionner encore une fois les limites de l'approche puriste et les avantages de l'approche pragmatique. En effet comme l'indique Fodor (1975: 7), les termes locaux sont trouvés seulement après que les notions que chacun d'eux véhiculent soient devenues suffisamment familières au sein de la communauté et qu'il est aussi devenu clair que ces mots ne sont pas des effets de mode éphémère.

7.2 La possibilité de la création lexicale spontanée et naturelle

Le second avantage de l'approche pragmatique dans le traitement de la macrostructure des dictionnaires gabonais tient au fait qu'il offre la possibilité de réaliser après l'emprunt direct, une création lexicale spontanée et naturelle permettant de trouver des mots locaux qui seront acceptés et attestés par la communauté. C'est le cas des mots tels que akôk (*pile électrique*), biyéne (*lunettes*), éloume (*fourchette de table*), mvét(t) (*guitare*), nda mi mbok (*prison*), nda Zama (*église*), nkare fuègne (*pasteur/prêtre*), nseng avion (*aéroport*), nseng ndama (*stade/terrain de football*) et ofepeya (*volant de voiture, de vélo et de moto*) par exemple. Ces termes ont été créés spontanément par des locuteurs et de façon involontaire et inavouée, ces derniers termes sont acceptés par toute la communauté fang ntoumou de Bitam.

Ces quelques termes locaux que nous venons d'énumérer font rejoindre quelque peu l'approche puriste. À la différence que dans ce cas, cette approche puriste se déroule de façon plus conforme au processus naturel du développement et la modernisation des mots d'une langue. Dans ce «purisme», les locuteurs façonnent de façon naturelle leurs langues. Il nous plaît de signaler que nous souhaitons tous en réalité qu'une grande partie du vocabulaire de nos langues locales en Afrique soient celle du terroir. Cela est source de fierté, et aussi un marqueur et une manifestation importante de notre génie et de notre identité. Paradoxalement, ce résultat n'est possible que si au préalable les mots

étrangers pénètrent ces langues afin que celles-ci les transforment en mots locaux.

8. Conclusion

Dans cet article, nous avons tenté de mettre en place une approche systématique dans le traitement de l'emprunt des mots au français dans la macrostructure des dictionnaires gabonais. Cette approche pragmatique s'est faite sur la base des dispositifs spécifiques des principes théoriques et pratiques de la lexicographie ainsi que sur la base des habitudes linguistiques des populations locales. C'est une image d'Épinal que la langue appartient aux locuteurs et la lexicographie tient compte de ce fait dans le traitement dictionnaire. C'est ainsi que nous avons essayé de montrer que l'approche pragmatique permettra de moderniser de façon systématique et élaborée le vocabulaire des langues locales à travers les dictionnaires et ce conformément aux habitudes des usagers pour qui ces ouvrages sont destinés. Cette approche permettra d'obtenir des ouvrages dans lesquels les populations se reconnaîtront et adopteront.

Les experts que nous sommes, linguistes, lexicographes et grammairiens, n'avons pas pour objectif de prescrire les langues locales en Afrique sans tenir compte des utilisateurs et des usages en cours de ces langues. Même si celles-ci incorporent des mots empruntés aux langues occidentales des anciennes puissances colonisatrices et que cela fait ressurgir les fantômes de la colonisation, il faudrait l'accepter ainsi. La présence de ces mots étrangers dans les langues locales d'Afrique est un processus inconscient et un facteur constitutif de la vie de toutes les langues du monde. Les langues-mères en Afrique n'échappent pas à ce fait et nous les experts ne devons pas faire de ces langues des exceptions. Il sera difficile à la minorité des experts que nous sommes de demander aux populations locales, par rupture immédiate ou progressive, de renoncer à l'utilisation des mots français pour satisfaire nos convictions idéologiques et politiques liées au désir de refouler un passé colonial douloureux. Ce passé a été certes douloureux mais nous ne pouvons pas refaire l'histoire. Nous devons vivre avec notre temps. Une marche en arrière n'est pas envisageable au nom de la recherche d'un pur pour d'identité.

Notre rôle n'est pas de produire des langues locales africaines selon nos idéologies et nos émotions. Nous devons d'une part les décrire telles qu'elles se manifestent et se pratiquent afin de mieux la comprendre et l'étudier. D'autre part, nous avons la mission d'organiser l'usage de la langue dans des ouvrages de base tels que les dictionnaires, les livres de grammaire et de lecture en conformité avec les besoins actuels afin de mieux les préserver et les promouvoir. Aucune langue ne naît ex nihilo par un pur pouvoir d'identité. Le développement d'une langue répond à une nécessité de construire et d'affermir l'État et la nation, de domestiquer la science et la technologie, de défendre et de faire rayonner la culture par des conceptions ou des pratiques qui apparaissent, cf. Ndong Ngoua (2007: 29). Par ailleurs, le traitement des langues locales dans les

types d'ouvrages que nous avons cités plus haut permettra de les standardiser et les rendre conformes à l'enseignement.

Enfin, nous devons nous fonder sur les expériences accumulées sous d'autres cieux et du passé pour comprendre et apprendre comment le français par exemple s'est transformé d'un dialecte en une langue officielle et véhiculaire. Ces références nous permettront de mieux aborder le développement et la modernisation de nos langues africaines car elles ont bien plus de choses en commun que les langues occidentales, c'est-à-dire des universaux. Ne faisons pas de nos langues des objets scientifiques qui ne sont destinées qu'à la recherche fondamentale. Le fondement de toute langue est d'abord le prosaïsme qui résulte de la pratique linguistique tout aussi prosaïque des locuteurs qui la produisent depuis la nuit des temps sans critère scientifique. À force de présenter les langues africaines comme des objets scientifiques plutôt que des instruments pratiques qui servent à communiquer pour des raisons prosaïques, nous experts passons parfois à côté des besoins réels. Nombreux sont les locuteurs qui s'écartent de la scientificité qui est mise en avant leurs langues locales. Cette scientificité de ces langues que nous brandissons fait que les locuteurs pensent que leurs langues ne leur appartiennent plus, mais plutôt aux experts. Une des conséquences de ce caractère scientifique qui est octroyé aux langues africaines c'est que cela conduit les populations locales vers le français et l'anglais qui semblent plus accessibles et car n'étant pas perçus comme des objets scientifiques.

Nous espérons que cet article sera d'un apport considérable pour guider les personnes engagées dans la confection des dictionnaires de langues locales au Gabon ou dans d'autres pays africains francophones en particulier, afin que leurs projets aboutissent à la réalisation des dictionnaires de bonne qualité.

Notes

- Les États Généraux de l'Éducation et de la Formation, Libreville, mars 1983: «Les États généraux recommandent l'introduction de nos langues maternelles dans le système éducatif, et que les moyens en hommes et en matériels soient dégagés pour asseoir ou renforcer les structures de recherche appropriées à la réalisation progressive de cet objectif. Ils insistent pour qu'un conseil technique spécialisé se charge d'initier et de conduire les études en la matière en tenant compte des spécificités propres à notre pays.»

– Les 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: Introduction des langues locales comme activités d'immersion dans l'enseignement a été arrêtée comme une des actions et mesures inscrites dans le projet décennal issu de cette concertation.
- Nous préviendrons ici qu'un reproche, une négation ou une critique acerbe au point de vue pratique et théorique, ne serait justifiable que si l'on occulte au Gabon l'origine historique, la place ainsi que le rôle des dictionnaires et de la lexicographie dans le développement et la modernisation des langues. Les dictionnaires sont confectionnés depuis près de 5000 av. J.-C. Par ailleurs, bien qu'ayant de domaine de recherche et de but différent, la confection de ces

ouvrages était considérée comme une sous-discipline de la linguistique. Mais depuis la fin du 20^{ème} siècle et le début du 21^{ème} siècle, la lexicographie s'est établie comme une discipline autonome grâce aux travaux majeurs de Zgusta (1971), de Wiegand (1989a, 1989b, 1989c, 1992, 1996), de Tarp (1998, 2002), de Bergenholtz et Tarp (1995), de Bergenholtz, Tarp et Wiegand (1999), et de Landau (2001), pour ne citer que ces travaux-là parmi tant d'autres. Les caractéristiques essentielles du dictionnaire ainsi que les principes théoriques et pratiques sur lesquels reposent la confection du dictionnaire ne sont plus nécessaires à démontrer dans la préservation, la standardisation, la promotion et le développement des langues et cultures. C'est à ce titre que le dictionnaire est à la fois mémoire et vécu comme le souligne Kalonji Zezeze (1993: 75). Il était très important de signaler cet aspect dans ce travail. Au Gabon comme dans beaucoup de pays d'Afrique noire francophone, il existe une sorte de contexte préétabli et fort rigide où seul ne semble compter que l'approche de la linguistique africaniste. Les travaux des chercheurs lexicographes gabonais sont soumis à l'évaluation des linguistes africanistes gabonais ou étrangers. Dans ce contexte théorique assez particulier, le lexicographe africaniste en général, et gabonais en particulier semble (étrangement) jouer le rôle "usurpé" d'institution, cf. Kalonji Zezeze (1993: 85). Il lui est quasiment exigé ex abrupto de partir des travaux des chercheurs africanistes gabonais pour élaborer des manuels et des dictionnaires appropriés. Son approche est littéralement balayée d'un revers de la main comme si ses propositions étaient des hérésies. Cependant, l'approche ou les travaux des lexicographes gabonais reposent bel et bien sur les principes de la pratique millénaire et de la théorie actuelle de la lexicographie. La théorie et la pratique de la confection de dictionnaires ne peuvent être ramenées aux seuls aspects, recommandations, théories, méthodes et objectifs particuliers et parcellaires de la linguistique interne. Tout en reconnaissant le bien fondé de la linguistique africaniste, nous pensons que les problèmes linguistiques qui relèvent beaucoup plus de la recherche fondamentale, ne devraient pas occulter les aspects linguistiques pratiques et prosaïques qui sont fondamentaux dans la conception, l'élaboration et la fabrication de dictionnaires locaux au Gabon.

3. Nous signalons que nous n'avons pas présenté les transcriptions phonétiques car le traitement prosodique des langues locales du Gabon n'est pas l'objectif de ce travail. Par ailleurs, nous utilisons l'écriture disjonctive car nous avons observé que cette forme d'écriture sied mieux au traitement dictionnaire. En dehors de cela, elle sera plus accessible aux usagers lambda et ceux du plus bas âge à cause des similarités avec le français, langue à laquelle les populations locales sont accoutumées. Le français est la langue officielle du Gabon. Elle est aussi langue de l'école, cf. Mba Nkoghé (2011: 7).
4. Nous sommes souvent assez surpris et surtout heurtés par la véhémence des observations de nombreux linguistes africanistes qui nous disent que "comparaison n'est pas raison" lorsque nous référons à l'histoire des autres langues pour mieux appréhender le développement des langues locales du Gabon. Quelle est cette hérésie scientifique de si grande ampleur com-mettons-nous dans une approche lexicographique, de vouloir tirer de l'expérience d'autres lieux des enseignements qui peuvent nous être utiles pour mieux appréhender une politique de préservation et de promotion de ces langues? "comparaison n'est pas raison" sont-ils des propos véritablement appropriés au domaine scientifique où "on ne réinvente pas la roue"? Nous ne pouvons pas traiter les langues locales du Gabon comme des cas isolés alors qu'elles ont subi et subissent encore les mêmes situations historiques, politiques, économiques et socioculturelles que les autres langues du monde. Comme dans le cas de l'anglais et du fran-

çais par exemple, il s'agit au Gabon de deux populations de langues différentes, le français et les langues locales qui se côtoient à cause de leurs peuples qui sont en contact. Dans ce cas, il est inévitable qu'il y ait des échanges et des emprunts mutuels dans tous les aspects des activités de l'homme. Il devient nécessaire d'analyser les apports inévitables des uns et des autres. La langue en fait partie. Même si le français gabonais n'est ni scolarisé ni officialisé et ce malgré les appels de nombreux linguistes et lexicographes, il est bel et bien effectif. Il résulte de l'impact des langues locales sur le français standard. Le sens inverse est tout aussi effectif. Les langues locales du Gabon se sont enrichies de mots nouveaux en empruntant au français standard même si cela n'est pas officialisé. Quand bien même au point de vue linguistique le français et les langues locales du Gabon ne sont pas de la même famille, il est nécessaire d'étudier les échanges qu'il y a eu entre ces deux langues. C'est une obligation de voir comment cette interaction est utile pour la scolarisation et l'officialisation du français du Gabon et de celles des langues locales de ce pays. Ce ne sont pas des épiphénomènes comme cela est généralement présenté, de la recherche lexicologique de la linguistique africaine, pour servir d'esquisse à une lexicologie différentielle des langues locales du Gabon et du français. C'est ce dernier aspect qui fait partie de notre appel dans ce travail.

5. Plusieurs linguistes africanistes gabonais en particulier, ponctuent souvent d'"affirmation gratuite" nos propos lorsque nous évoquons ce sujet. Nous en sommes encore une fois surpris et profondément heurtés. Dans les pays d'Afrique noire, on vit un bilinguisme d'état généralisé dû en partie à l'instauration du nouvel ordre social issu du colonialisme, cf. Kalonji Zezeze (1993: 57). La question du français dans la vie nationale du Gabon donne lieu souvent à des débats que nous nous pouvons qualifier de stériles ou relevant de combats d'arrière-garde. Il est vrai que nous nous pouvons nous mettre aujourd'hui à penser ou à rêver de ce qu'aurait pu être nos langues locales si le français ne nous avait pas été imposé depuis un siècle. Malheureusement, nous ne pouvons pas refaire l'histoire. Nous devons vivre avec notre temps et assumer donc pleinement le français comme faisant partie de notre patrimoine linguistico-culturel. Nous affirmons et assumons entièrement ce fait ici car que ce statut soit officiel ou pas la théorie et la pratique lexicographique prend en compte le statut sociopolitique des langues en présence. De nombreux auteurs et personnes qui ne sauraient et pourtant assumer aujourd'hui l'avenir du Gabon si par rupture immédiate ce pays renonçait au français, se cachent derrière la réflexion universitaire et l'aveuglement politique ou idéologique et émotif pour avoir des réponses hypocrites à ce sujet. Cette hypocrisie s'observe aussi chez de nombreux linguistes africanistes d'origine européenne qui, perclus dans leurs rôle et mission avoués ou inavoués de "décolonisateurs", militent pour des langues locales d'Afrique qui soient pures, exemptes de tout élément de la langue de l'ancienne puissance colonisatrice. Nous les comprenons parfaitement car toute suggestion allant dans le sens d'utiliser les mots du français pour développer les langues locales leur donnerait à tort un rôle et une mission de "recolonisateurs".
6. Langue dont on repère les traces dans une autre langue et qui sert de base à cette dernière.
7. Après les guerres d'Italie qui s'étalèrent de 1494 à 1559, Français et Italiens tissèrent des liens étroits et pacifiques. De nombreux Italiens allèrent poursuivre leur carrière à la cour du roi de France et les mariages diplomatiques, comme celui de Catherine de Médicis avec Henri II, amenèrent à la cour des milliers d'intellectuels, d'artistes et de scientifiques italiens.
8. Les internationalismes peuvent être considérés comme les items lexicaux qui sont communs aux vocabulaires de plusieurs langues, cf. Braun (1989: 159). Toutefois, il faut également ajou-

ter que pour qu'un item lexical soit considéré comme un internationalisme, il faudrait qu'il soit utilisé dans plusieurs «langues nationales, qu'il soit internationalement courant et qu'il soit compréhensible sans être traduit, cf. Braun (1989: 159). Sur cette base Braun (1989: 157) considère comme internationalismes, les items lexicaux suivants par exemple: Allemand (baseball, golf, cockpit, boxen, abstrakt, konkret, aktiv), Anglais (baseball, golf, cockpit, boxe, abstract, concret, active) et Français (baseball, golf, cockpit, boxe, abstrait, concret, actif).

9. Ces explications nous ont été fournies par Hidalgo Aboghe Nzome, journaliste en poste à RTG 6 à Makokou, Province de l'Ogooué Ivindo.
10. Il s'agit de Mekeng Edouard, Enseignant du secondaire (Histoire et Géographie), et d'Endama Urcy, Technicien à Gabon-Poste. Ils sont respectivement Représentant de la Majorité et Opérateur dans la Commission d'Enrôlement pour l'Identité Biométrique du Gabon (IBOGA) pour le compte de la Commune d'Ovan.

Bibliographie

- Bergenholtz, H. et S. Tarp (Éds.).** 1995. *Manual of Specialised Lexicography: The Preparation of Specialised Dictionaries*. Amsterdam/Philadelphia: John Benjamins.
- Bergenholtz, H., S. Tarp et H.E. Wiegand.** 1999. Datendistributionsstrukturen, Makro- und Mikrostrukturen in neueren Fachwörterbüchern. *Fachsprachen. Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft/Languages for Special Purposes. An International Handbook of Special-Language and Terminology Research. Volume 2: 1762-1832*. Berlin/New York: Walter de Gruyter.
- Braun, P.** 1989. Internationalisms: Identical Vocabularies in European Languages. Coulmas, F. (Éd.). *Language Adaptation: 158-167*. Cambridge: Cambridge University Press.
- Diop, C.A.** 1979. *Nations nègres et culture. Tome 2*. Paris: Présence Africaine.
- Fodor, I.** 1975. Language Reforms of the Past and in the Developing Countries. Herbert, R.K. (Éd.). 1975. *Patterns in Language, Culture and Society: Sub-Saharan Africa: 3-17*. Columbus: The Ohio State University.
- Gouws, R.H.** 2001. Formation lexicographique: approches et thèmes. Emejulu, J.D. (Éd.). 2001. *Éléments de lexicographie gabonaise. Tome I: 95-134*. New York: Jimacs-Hillman.
- Hagège, C.** 1987. *Le français et les siècles*. Paris: Odile Jacob.
- Hausmann, F.J et al. (Éds.).** 1989-1991. *Wörterbücher. Ein internationales Handbuch zur Lexikographie / Dictionaries. An International Encyclopedia of Lexicography / Dictionnaires. Encyclopédie internationale de lexicographie*. Berlin: Walter de Gruyter.
- Jernudd, B.H.** 1977. Linguistic Sources for Terminological Innovation: Policy and Opinion. Rubin, J., B.H. Jernudd, J. Das Gupta, J.A. Fishman et C.A. Ferguson (Éds.). 1977. *Language Planning Processes: 215-236*. Le Havre: Mouton.
- Joseph, J.** 1984. The Elaboration of an Emerging Standard. *Language Sciences* 6(1): 39-52.
- Kalonji Zezeze, M.T.** 1993. *La lexicographie bilingue en Afrique francophone: L'exemple français-cilubà*. Paris: L'Harmattan.
- Landau, S.I.** 2001. *Dictionaries: The Art and Craft of Lexicography*. New York/Cambridge: Cambridge University Press.
- Madiba, M.R.** 2001. Towards a Model for Terminology Modernisation in the African Languages of South Africa. *Language Matters* 32: 53-77.

- Marshad, H.A.** 1984. *An Approach to Code Elaboration and its Application to Swahili*. Thèse de doctorat non publiée. Urbana-Champaign: University of Illinois.
- Mba Nkoghé.** 2011. La langue de l'école. *Gabon Matin, Quotidien d'informations générales* No. 590: 7.
- Mwansoko, H.J.M.** 1990. *Modernisation of Swahili Technical Terminologies: An Investigation of the Linguistics and Literature Terminologies*. Thèse de doctorat non publiée. York: University of York.
- 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.
- Sager, J.C.** 1989. *Terminological Problems Involved in the Process of Exchange of New Technology between Developing and Developed Countries: Study on Recent Developments in the Relationship between Science, Technology and Society in Different Economic, Social and Cultural Contexts*. Paris: UNESCO.
- SALIC (Service de L'Aménagement Linguistique du Canada).** *Histoire de la langue française*. <http://www.salic.slmc.ca/>.
- Samsom, R.H.** 1988. *Ufundi (wa) magari: mfano wa ukuzaji wa istalahi za Kiswahili nchini Tanzania matumizi wa usanifishaji wake*. M.A. thèse non publiée. Leyde: Université de Leyde.
- Smit, M.** 1996. *Wiegand's Metalexigraphy as a Framework for a Multilingual, Multicultural, Explanatory Music Education Dictionary for South Africa*. Thèse de doctorat non publiée. Stellenbosch: Université de Stellenbosch.
- Tarp, S.** 1998. Leksikografien på egne ben. Fordelingsstrukturer og byggedele i et brugerorienteret perspektiv. *Hermes, Journal of Linguistics* 21: 121-137.
- Tarp, S.** 2002. Translation Dictionaries and Bilingual Dictionaries — Two Different Concepts. *Journal of Translation Studies* 7: 59-84.
- Weinstein, B.** 1989. Francophonie: Purism at the International Level. Jernudd, B.H. et M.J. Shapiro (Éds.). 1989. *The Politics of Language Purism*: 53-79. Berlin: Mouton de Gruyter.
- Wiegand, H.E.** 1989a. Arten von Mikrostrukturen im allgemeinen einsprachigen Wörterbuch. Hausmann, F.J. et al. (Éds.). 1989-1991: 462-501.
- Wiegand, H.E.** 1989b. Aspekte der Makrostruktur im allgemeinen einsprachigen Wörterbuch: Alphabetische Anordnungsformen und ihre Probleme. Hausmann, F.J. et al. (Éds.). 1989-1991: 371-409.
- Wiegand, H.E.** 1989c. Der gegenwärtige Status der Lexikographie und ihr Verhältnis zu anderen Disziplinen. Hausmann, F.J. et al. (Éds.). 1989-1991: 246-280.
- Wiegand, H.E.** 1992. Elements of a Theory towards a So-called Lexicographic Definition. *Lexicographica* 8: 175-289.
- Wiegand, H.E.** 1996. A Theory of Lexicographic Texts. An Overview. *South African Journal of Linguistics* 14(4): 134-149.
- Zgusta, L.** 1971. *Manual of Lexicography*. Le Havre: Mouton.

Woordeboeke en Dowe gebruikers: huidige probleme en die behoefte aan beter oplossings*

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: Gebaretaalleksikografie is 'n afdeling van die leksikografie wat tot dusver 'n grootliks vae, onbelangrike area was. Die ontwikkeling van gebaretaalleksikografie was in die verlede onderhewig aan verskillende vooroordele en beperkinge. Dit was dikwels nie *nodig* geag om 'n tweetalige woordeboek saam te stel waarin dit ook moontlik is om 'n gebaar op te soek nie aangesien tweetalige gebaretaalwoordeboeke grotendeels gemik was op horende persone wat die taal wou of moes leer. Daarby was dit baie moeilik vanweë die aard van gedrukte woordeboeke en die gebrek aan 'n ortografiese vorm van gebaretaal om gebare op 'n manier voor te stel én te rangskik sodat hulle opgesoek kan word sonder om van 'n geskrewe soekterm gebruik te maak. Daar is wel notasiesisteme ontwikkel in 'n poging om hierdie probleem te omseil, maar hierdie sisteme het nooit by gewone gebruikers wyd byval gevind nie. Die ontwikkeling van elektroniese gebaretaalwoordeboeke het egter gesorg vir 'n revolusie in gebaretaalleksikografie. Dit is nie meer ongewoon om elektroniese gebaretaalwoordeboeke te vind waarin nie net 'n woord nie, maar ook 'n gebaar soekbaar is, gewoonlik volgens een of meer van die gebaar se parameters, soos handvorm. Sulke woordeboeke maak dit moontlik vir 'n Dowe gebruiker om direkte toegang tot sy of haar eerste taal (gebaretaal) te bekom sonder enige kennis van die tweede taal (die geskrewe taal). Die artikel verwys ook na 'n nuutvoorgestelde model wat dit vir baie jong Dowe gebruikers moontlik maak om ook via konsep na gebaar of woord te soek.

Sleutelwoorde: GEBARETAAL, LEKSIKOGRAFIE, WOORDEBOEK, GEBAREPARAMETERS, GEBARENOTASIE, ELEKTRONIESE WOORDEBOEKE, DOWE TEIKENGEBRUIKERS, GELETTERDHEID, GEBRUIKERSITUASIE, WOORDEBOEKKULTUUR

Abstract: Dictionaries and Deaf Users — Current Problems and the Need for Better Solutions. Sign language lexicography is a branch of lexicography that has until now been a largely vague, unimportant area. The development of sign language lexicography has in the past been subject to various forms of bias and limitation. Often it was not considered *necessary* to compile a bilingual dictionary in which it is also possible to look up a sign because bilingual sign language dictionaries were mostly aimed at hearing people who wanted to or had to learn the language. In addition to that it was very difficult, owing to the nature of printed dictionaries and the

* Hierdie artikel is 'n uittreksel uit 'n D.Litt.-proefskrif, 'n *Leksikografiese model vir 'n elektroniese tweetalige grondslagfase-woordeboek van Suid-Afrikaanse Gebaretaal en Afrikaans*, wat in Maart 2013 deur die Universiteit Stellenbosch, Stellenbosch, Suid-Afrika, aanvaar is.

lack of a sign language orthography, to present and order signs so that they may be searched without using a written search term. Some notation systems were developed in an effort to circumvent this problem, but these systems never found widespread acceptance with regular users. The development of electronic sign language dictionaries however have caused a revolution in sign language lexicography. It is no longer unusual to find electronic sign language dictionaries in which it is not only possible to search for a word, but also for a sign, usually according to one or more of the sign's parameters, like handshape. Such dictionaries make it possible for a Deaf user to obtain direct access to his or her first language (sign language) without any knowledge of the second language (the written language). The article also points out a newly proposed model for very young Deaf users which allows them to also search for a sign or word via concept.

Keywords: SIGN LANGUAGE, LEXICOGRAPHY, DICTIONARY, SIGN PARAMETERS, SIGN NOTATION, ELECTRONIC DICTIONARIES, DEAF TARGET USERS, LITERACY, USER SITUATION, DICTIONARY CULTURE

Inleiding

Die gemiddelde Dowe¹ gebruiker ervaar verskeie probleme in die gebruik van woordeboeke: die opvoedkundige agtergrond van Doves is gewoonlik (en ook universeel — alhoewel daar altyd uitsonderings is) van uiteenlopende gehalte en die gemiddelde Dowe volwassene bereik die (geskrewe) geletterdheidsvlak van 'n leerder in graad vyf. Woordelyste of woordeboeke wat 'n geskrewe taal en gebaretaal kombineer maar wat slegs toegang in een rigting bied — dus *van* 'n geskrewe woord *na* 'n gebaar — faal die Dowe gebruiker wie se eerste taal gebaretaal is of wat verkies om gebaretaal eerder as 'n geskrewe of gesproke taal te gebruik. Sulke woordelyste en woordeboeke vereis van Dowe gebruikers om deur die *tweede* taal, waarin hulle gewoonlik slegs baie lae geletterdheidsvlakke het, toegang tot hulle *eerste* taal te bekom en daaroor te leer, en dui op 'n oorheersende fokus op horende-Dowe kommunikasie sonder om Dowe-horende kommunikatiewe behoeftes in ag te neem. Dit skep 'n ongebalanseerde en onregverdige gebruikersituasie wat waarskynlik nie in baie ander taalgemeenskappe geduld sou word nie. Tog is dit 'n algemene situasie in die arena van horende-Dowe kommunikasie regoor die wêreld.

Die ideaal sou dus wees om tweetalige *en* tweerigtinggebaretaalwoordeboeke saam te stel wat dit nie net moontlik maak om 'n gebaar deur middel van 'n sekere woord op te soek nie, maar ook om 'n woord deur middel van 'n gebaar op te soek. So 'n woordeboek sou nie slegs baie nuttiger wees vir 'n Dowe gebruiker wat die geskrewe taal enkodeer nie, maar dit sou ook nuttig wees vir die horende gebruiker wat gebaretaal wil dekodeer.

Hierdie artikel sal 'n kort oorsig verskaf van die uitdagings, geskiedenis en tradisies van gebaretaalleksikografie, 'n verduideliking van die notering van gebare, en 'n bespreking van elektroniese gebaretaalwoordeboeke en hulle vermoë om tweetalige, tweerigtingwoordeboeke aan te bied wat baie gebruikersvriendelik is. Dit sal ook verwys na 'n nuwe model wat dit nie net moontlik maak om van woord na gebaar en van gebaar na woord te beweeg nie, maar ook van konsep na gebaar/woord, soos voorgestel deur Fourie (2013).

Gebaretaalleksikografie

Die gebaretaalleksikograaf moet verskeie uitdagings oorkom. Eerstens is daar geen ortografie of skryfwyse vir gebaretale nie. Alhoewel verskeie notasiesisteme reeds ontwikkel is, word geen van hulle algemeen gebruik nie en is die meeste te ingewikkeld en kompleks om gebare in 'n woordeboek weer te gee — in ieder geval vir die algemene gebruiker. Tweedens is daar in byna alle lande slegs 'n paar gebarebronne waaruit die betekenis(se) en grammatikale eienskappe van gebare afgelei kan word en waaruit gebruiksfrekwensies van gebare bepaal kan word. Dertens verskil gebaretale aansienlik van gesproke tale. Aangesien die gebaretaal 'n komplekse struktuur kan hê waar die gesproke taal 'n eenvoudige woord het en ook selfs omgekeerd, is dit dikwels moeilik om 'n woord-vir-gebaar- of gebaar-vir-woord-ekwivalent te gee (Zwitserlood 2010: 444-445). Laastens moet gebaretaalleksikograwe die nalatenskap van 'n tradisie van onvoldoende samestelling van gebaretaalwoordeboeke deur nieleksikograwe oorkom — om hierdie rede het die bestaande gebaretaalwoordeboeke in Suid-Afrika dikwels hewige kritiek ontvang en word dit selde indien ooit deur die Dowe gemeenskap gebruik.

Tradisies in gebaretaalleksikografie

The word [dictionary] suggests authority, scholarship, and precision. (Landau 1984: 5)

Alhoewel gebaretale baie jare — en in sommige gevalle, steeds — as ondergeskikte of mindere tale gesien is, het 'n verskeidenheid eenvoudige gebaretaalwoordeboeke in die verlede verskyn. Hierdie boeke is gewoonlik gerig op (horende) onderwysers en ouers wat nodig het om met hul Dowe leerlinge of kinders te kommunikeer wanneer gesproke kommunikasie faal. Die meeste van hierdie woordeboeke, wêreldwyd sowel as in Suid-Afrika, is tipies saamgestel deur onderwysers, maatskaplike werkers, spraakterapeute, lede van verskeie godsdienstige instansies, e.s.m. Tot ongeveer die middel 1960's is die meeste gebaretale nie eens as volwaardige tale vergelykbaar met gesproke tale gesien nie, maar danksy die publikasie in 1965 van die eerste uitgawe van Stokoe, Casterline en Croneberg se gebaretaalwoordeboek, gebaseer op taalkundige beginsels, het gebaretaalleksikografie stadig momentum begin kry (Schermer 2004: 2). Sodra gebaretaalnavorsing in 'n land behoorlik posvat, is een van die eerste take wat aangepak (moet) word, gewoonlik die samestelling van 'n woordeboek.

Die woordeboeke tot dusver bestaan oor die algemeen uit tweetalige woordelyste wat 'n basiese stel geskrewe woorde uit die gesproke taal bevat tesame met 'n skets of foto van 'n persoon wat die gebaar by elke woord demonstreer. Baie min, indien enige, inligting word oor die gebaar gegee en dit is ook nie gewoon om meer as een gebare-ekwivalent per woord aan te tref nie (bv. Penn et al. 1992–1994). Sulke woordeboeke is noodwendig eenrigting, van die gesproke taal na die gebaretaal, en is volgens die alfabet, temas, of beide, gerangskik.

Namate gebaretaal meer geredelik as natuurlike tale aanvaar is en daar 'n groter behoefte was om hulle aan te leer, was daar ook 'n groter behoefte aan meer woordeboeke met meer gebare en 'n uitvoeriger bewerking van die gebare. Hierdie woordeboeke, dikwels saamgestel deur professionele leksikografe, het steeds dieselfde formaat: hulle heet gewoonlik "Die X Gebaretaalwoordeboek" (waar X staan vir 'n land of streek), hulle is tweetalig en gewoonlik ook eenrigting (met enkele uitsonderings). Die meeste woordeboeke bied steeds een woord vir een gebaar en bevat baie min addisionele inligting oor die gebare (Zwitsers 2010: 445).

Voorbeelde van die mees gesofistikeerde tweerigting- en tweetalige gebaretaalwoordeboeke in boekvorm sluit in Johnston se *Auslan Dictionary* (1989), Brien se *Dictionary of British Sign Language/English* (1992), Radutzky se woordeboek van Italiaanse Gebaretaal (1992) en Malm se Finse Gebaretaalwoordeboek (1998). Hierdie woordeboeke bied die leksikon van 'n gebaretaal deur middel van 'n geordende reeks hoofgebare, wat daaropvolgend in die tweede taal verduidelik word. Dit vereis 'n gedetailleerde analise van die fonemiese eienskappe van gebare (Schermer 2004: 4).

Om werklik reg aan die rykheid en oorvloed van die leksikon van gebaretaal te laat geskied, is dit eintlik nodig om eers eentalige woordeboeke te skep. Tweetalige woordeboeke behoort op die eentalige woordeboekinligting gebaseer te word om die vertaalstruikelblokke tussen die gebaretaal en die gesproke taal van 'n gemeenskap te vermy (Schermer 2004: 4). Ongelukkig is so 'n benadering nie noodwendig prakties of ekonomies haalbaar nie en daarby is dit nie ongewoon vir 'n "klein" taal se eerste woordeboek om 'n tweetalige woordeboek te wees nie.

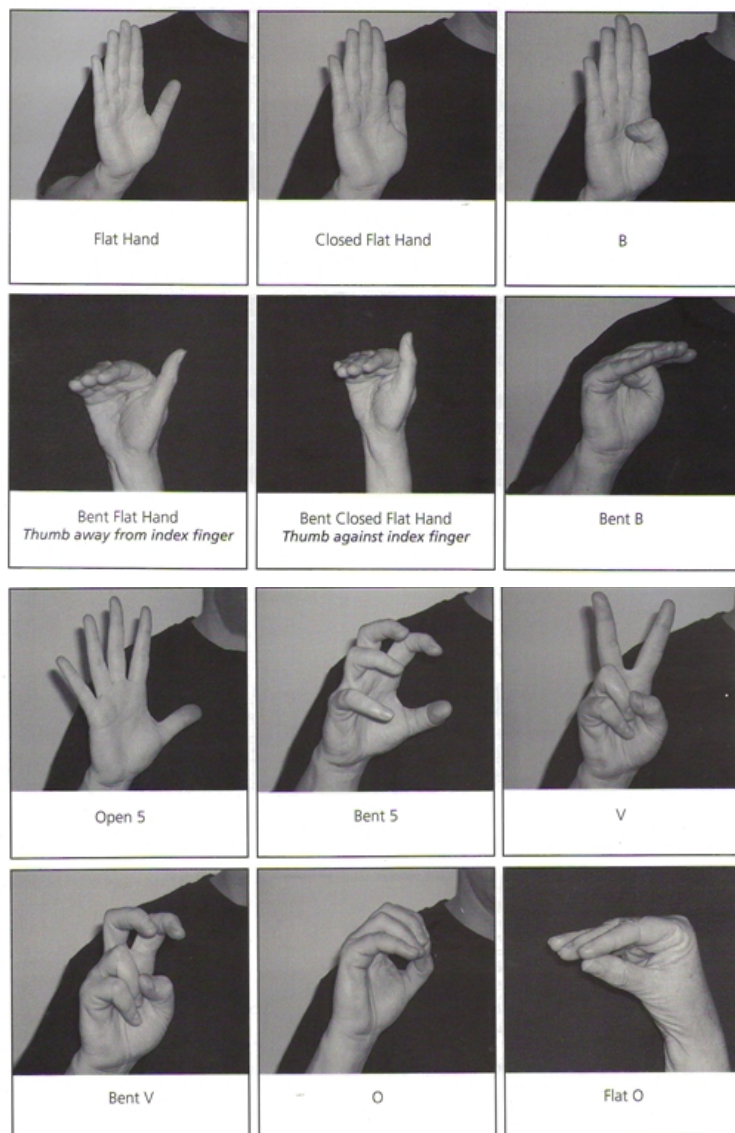
Gebareparameters

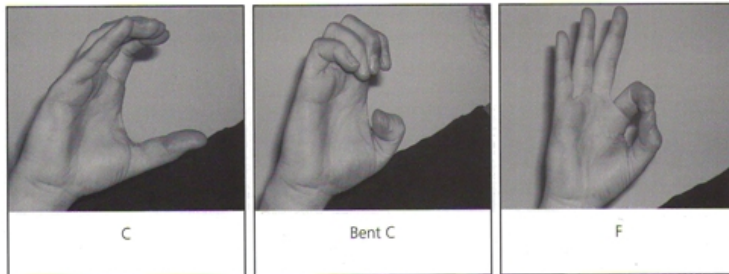
Om die volgende besprekings van die notering van gebare te kan begryp, is dit belangrik om 'n duidelike begrip van die parameters of "boublokke" van gebare te hê. Notasiesisteme beskryf gebare tipies as 'n kombinasie van hulle kleinste kontrastiewe komponente. Die term vir hierdie komponente is "chereme" (Stokoe et al. 1976: xxix), soortgelyk aan die foneme in gesproke tale. Dit is afgelei van die Griekse woord vir "hand". Gebare word volgens handvorm, beweging, plek van artikulasie, ens. geklassifiseer. Die chereme wat belangrik is om individuele gebare te identifiseer en te onderskei, is die volgende gebareparameters:

1. die **handvorm** wat gebruik word om die gebaar te vorm,
2. die **plasing** of **plek** van die gebaar in ruimte,
3. die **oriëntering** van die handpalms, en
4. die tipe **beweging** wat deur die hande uitgevoer word (Mesthrie et al. 2000: 420).

Die bostaande parameters word soos volg deur Howard (2008) vir SASL² uiteengesit en geïllustreer:

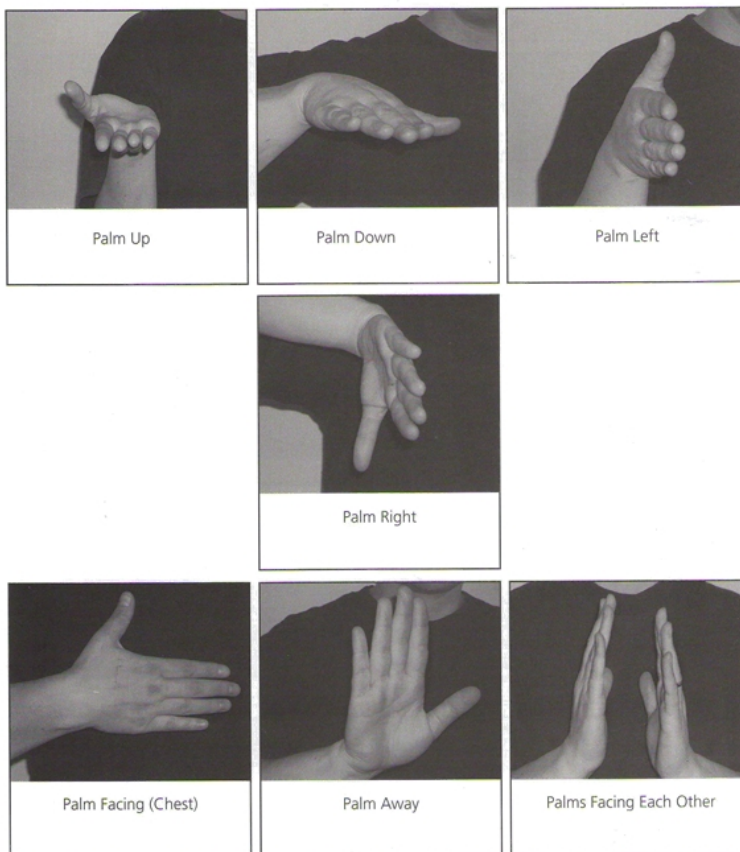
Die **handvorme** in figuur 1 is: Plat Hand, Toe Plat Hand, B (dus die vingerspelvorm van die letter B), Gebuigde Plat Hand, Gebuigde Toe Plat Hand, Gebuigde B, Oop 5, Gebuigde 5 (of 'n klou-handvorm), V, Gebuigde V, O, Plat O, C, Gebuigde C, F, Plat F, 1, 2.





Figuur 1: Handvorms (Howard 2008: 4-6)

Die parameterwaardes ten opsigte van **handpalmoriëntering** in figuur 2 is: Palm Op, Palm Af, Palm Links, Palm Regs, Palm Teenoorstaande (Bors), Palm Weg en Palms Na Mekaar:



Figuur 2: Handpalmoriëntering (Howard 2008: 7)

Die parameters van **Ruimte/Plek** en **Beweging** word soos volg beskryf:

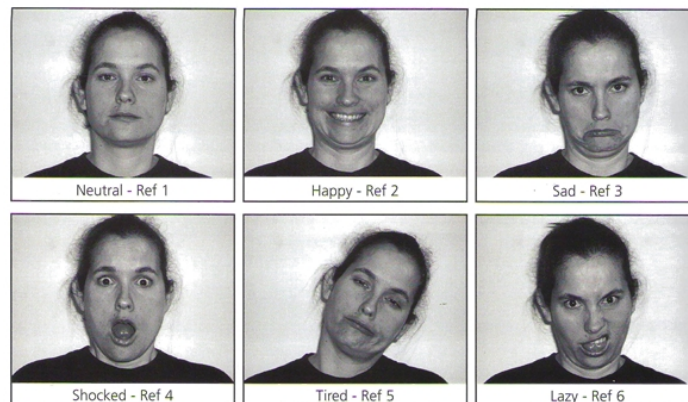
Ruimte/Plek — Die ruimte direk voor jou liggaam word die **neutrale ruimte** genoem. Gebare in die *Finger Talk*-woordeboek bevat inligting oor die presiese plek waar die gebaar uitgevoer word. Ruimte word in gebaretaal op 'n soortgelyke manier gebruik as wat volume in gesproke taal ingespan word: hoe groter die ruimte wat gebruik word, hoe "harder" praat jy, hoe kleiner die ruimte, hoe "sagter" is jou stem.

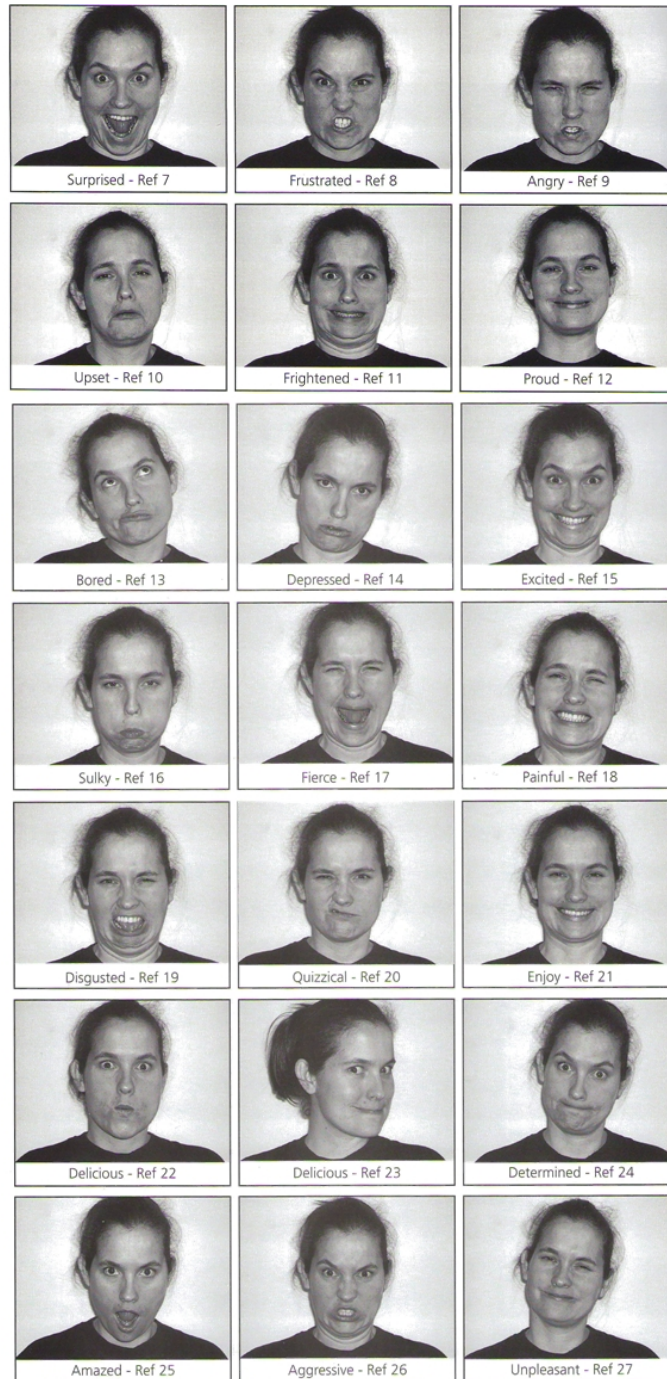
Beweging — dit is die manier waarop die hande tydens die gebaar beweeg word, insluitende die rigting waarin hulle beweeg.

Gebaretaal maak egter nie net gebruik van handgebare om betekenis oor te dra nie, en ander liggaamsdele speel ook 'n rol in taalproduksie. Natuurlike gebaretaal gebruik ook kenmerke wat nie met die hande gevorm word nie (bv. gesigsuitdrukings, lipbewegings, liggaamshouding, oriëntering en beweging van die kop of liggaam) om morfologiese en grammatikale kategorieë aan te dui (Mesthrie et al. 2000: 420). Trouens, verskillende gesigsuitdrukings kan verskillende betekenis aan dieselfde gebaar gee, en die gebrek aan 'n meegaande gesigsuitdrukking kan lei tot 'n ongrammatikale uitdrukking. Dit is dikwels een van die aspekte wat horende gebruikers van 'n gebaretaal die moeilikste aanleer: aangesien horendes meestal selfbewus is weens die oordrewe gesigsuitdrukings van gebaretaal kom hulle uitinge dikwels "stomp" en "dood" voor, soms tot die vermaak van Dowe gespreksgenote en soms tot hulle verwarring.

Die niemanuele eienskappe (NME's of "Non Manual Features", m.a.w. grammatikale elemente wat nie op die hande verskyn nie, soos hierbo beskryf word), is volgens Howard (2008) se beskrywing die volgende:

Die NME's³ in figuur 3 is: Neutraal, Gelukkig, Hartseer, Geskok, Moeg, Lui, Verras, Gefrustreerd, Kwaad, Ontsteld, Bang, Trots, Verveeld, Depressief, Opgewonde, Nukkerig, Heftig, Pynlik, Afkeer, Vraend, Geniet, Smaaklik (lippe rond saamgepers), Smaaklik (lippe plat saamgepers), Vasberade, Verbaas, Aggressief, Onplesierig.





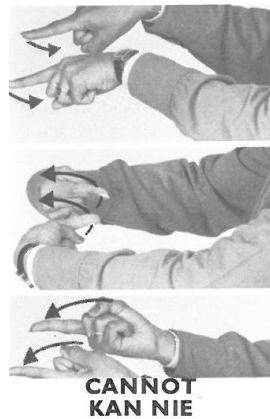
Figuur 3: NME's (Howard 2008: 8-10)

Die notasie van gebare

Woordeboeke van gesproke tale het gewoonlik te make met tale wat wel oor 'n geskrewe vorm beskik — gesproke tale is vanweë hul aard eendimensioneel (spraakklanke word een na die ander uitgespreek) en dit word gereflekteer in die lineêre volgorde van letters in die geskrewe vorm. Sommige woordeboeke bevat ook inligting oor die uitspraak van 'n woord, gewoonlik aan die hand van fonetiese notasie, of gee minstens 'n aanduiding van die lettergrepe en klem vir elke woord.

In teenstelling hiermee beskik gebaretale nie oor enige geskrewe vorm nie, en aangesien spraakklanke nie 'n rol in hierdie tale speel nie, bied die fonetiese skryfwyse ook geen uitkoms nie. Die probleem met gebaretaalnotasie is dat 'n vierdimensionele taal (gebare gebruik driedimensionele ruimte asook 'n sekvensiële of temporele struktuur) moeilik is om in 'n statiese voorstelling weer te gee. Woordeboeksamestellers het hierdie probleem gewoonlik op verskillende wyses hanteer, waarvan een die gebruik is van prentjies/sketse of foto's van 'n persoon wat die gebaar demonstreer. Gebarenotasie word in sommige woordeboeke gebruik, maar gewoonlik saam met die grafiese voorstelling van die gebaar. Danksy meer onlangse tegnologie is die probleem ook omseil deur die gebruik van video-uitbeeldings in elektroniese woordeboeke — elkeen van hierdie metodes het 'n impak op die organisasie van die woordeboek (Zwitsers 2010: 447).

Die meeste woordeboeke, veral die gedrukte woordeboeke, gebruik sketse of foto's om die gebare voor te stel. In die verlede was dit die enigste moontlikheid vir voorstelling en die afbeeldings is gewoonlik so gekies om selfs vir die ongesofistikeerde gebruiker verstaanbaar te wees. Die aktiwiteit van 'n gebaar is egter nie sigbaar in 'n stilbeeld nie, en die aktiwiteit of beweging word gewoonlik daargestel deur pyltjies en ander ekstra simbole wat op die prent aangebring word, of voorgestel deur 'n reeks beelde soos in figuur 4.



Figuur 4: Voorbeeld van 'n reeks beelde met pyltjies om beweging aan te dui (Nieder-Heitmann 1980: 109)

Voorstellings met bygevoegde simbole vir aktiwiteit of beweging is steeds nie baie uitvoerig of gedetailleerd nie, veral nie oor die deel van die gebaar wat nie op die hande plaasvind nie. Daarby neem sulke voorstellings baie ruimte in beslag en is steeds nie maklik om te gebruik nie. Ekstra inligting oor die gebaar se uitvoering is in sommige woordeboeke bygevoeg om 'n oplossing hiervoor te bied (figuur 4 en veral figuur 5 en 7), maar dit dra eintlik slegs by tot die probleem van ruimte — die feit dat afbeeldings soveel ruimte in beslag neem, maak dit feitlik onmoontlik om inligting oor lemmas in die gebaretaal (dus deur middel van afbeeldings) te verskaf. Om hierdie rede bestaan daar geen gedrukte tweetalige woordeboek van 'n gesproke taal na 'n gebaretaal wat inligting oor die lemmas in gebaretaal verskaf nie, en nog minder 'n eentalige gebaretaalwoordeboek. Sommige woordeboeke maak van 'n baie meer minimalistiese metode gebruik en verskaf slegs 'n diagram van die gebaar onder 'n geskrewe soekwoord in 'n alfabetiese lys, sonder enige verdere grammatikale inligting (figuur 7).



Bear (1) *n* a large, heavy mammal with thick, rough fur: *a grizzly bear*.

[Action of a bear scratching itself]

With the arms crossed at the wrist on the chest, scratch the fingers of both *curved hands* up and down near each shoulder with a repeated movement.

Figuur 5: Artikel uit 'n (gedrukte) ASL⁴-woordeboek (Costello 1998) (uit Zwitterlood 2010: 447)

	<p>H S Loc P O NMF Mov</p>	<p>Bear Both hands – C Both sides of head Palms facing each other Ref 2 or Ref 17 Put hands on both sides of head.</p>

Figuur 6: Die artikel vir BEER uit (die gedrukte woordeboek) *Finger Talk* (Howard 2008) — die vier gebareparameters sowel as die nie-manuele eienskappe (NME's) word aangedui

Gegewe die feit dat selfs 'n gebaretaalwoordeboek met relatief min artikels en inskrywings redelik lywig is of selfs verskeie boekdele beslaan, bevat die meeste gebaretaalwoordeboeke redelik min artikels in vergelyking met die

woordeboeke van gesproke en geskrewe tale, en enige ekstra inligting word in die geskrewe taal gegee. Die aantal artikels in 'n woordeboek word egter gesien as 'n aanduiding van die kompleksiteit en gevolglik ook die status van 'n taal (Schermer 2004: 5), met die gevolg dat 'n beperkte aantal artikels die indruk van eenvoud of ondergeskiktheid skep.

bear



Figuur 7: Dié artikel vir die ASL-gebaar BEAR verskyn sonder enige verdere inligting in die gedrukte weergawe van die *Gallaudet Dictionary of American Sign Language*, alhoewel die meegaande CD-ROM wel ietwat meer inligting, soos sinonieme, saam met die video van die gebaar vertoon (Valli 2005: 38)


Die nadeel van gebareafbeeldings is dat hulle holistiese eenhede is wat moeilik is om te rangskik of om deur te soek, in vergelyking met die alfabetiese rangskikking van gesproke tale se woordeboeke. Slegs 'n paar tweetalige woordeboeke bevat afbeeldings van gebare sowel as 'n rangskikking gebaseer op gebarestruktuur, soos die *Preliminary Signing Dictionary of Australian Sign Language* (Johnston 1987), wat naas 'n alfabetiese lys Engels-na-Auslan artikels ook 'n afdeling met Auslan-na-Engels bevat waarin gebare volgens die waardes van die gebareparameters gerangskik is. Vir die meeste gedrukte woordeboeke is hierdie moeite nie gedoen nie en hulle is geneig om steeds tweetalig, eenrigting en alfabeties gerangskik te wees volgens die woorde van die gesproke taal (Zwitsers 2010: 448).

Verskeie notasiesisteme gebaseer op die vorm van gebare is oor die afgelope dekades ontwikkel. Die belangrikste hiervan is:

- die Stokoe-notasiesisteme (Stokoe et al. 1976),
- HamNoSys (Prillwittz et al. 1989), en
- 'n sisteem wat bedoel is om as 'n ortografie van gebaretale te funksioneer, nl. SignWriting (SignWriting for Sign Languages 2011).

Elkeen van hierdie sisteme het stelle simbole vir die verskillende waardes van

die gebareparameters en hierdie simbole word saamgevoeg of gekombineer om die gebarevorm daar te stel (sien figuur 8).

Stokoe-notering	[] √C' √C:
HamNoSys	.. ☞ r 0 × ■ ☞ X [↓ → ☞] +
SignWriting	



Figuur 8: Weergawe van die ASL-gebaar BEAR in die drie notasiesisteme (aangepas uit Zwitserlood 2010: 449)

Waar die Stokoe- en HamNoSys-noterings meestal lineêr gerangskik en redelik abstrak is, is die SignWriting weergawe meer holisties en is die gebare makliker om te herken.

SignWriting is met die eerste oogopslag baie eenvoudiger as die ander twee noterings. Dit is egter ietwat misleidend, aangesien sommige inligting wat eksplisiet gemaak word, deur die ander twee noterings implisiet in die SignWriting weergawe is. In laasgenoemde is inligting oor die oriëntering van die hand altyd teenwoordig in die handvormsimbool. Bv. ☞ dui aan dat die handpalm na die gebaregebruiker gedraai is; ☞ en ☞ dui onderskeidelik aan dat die palm na die kant of weg van die gebaregebruiker gedraai is. 'n Soortgelyke gebruik van kleurskakering dien om rigting van beweging aan te dui, wat nie in hierdie voorbeeld teenwoordig is nie. As resultaat van die kombinasie van handvorm en oriëntering in die handvormsimbole, en beweging en rigting in die bewegingsimbole, het SignWriting 'n veel groter stel simbole as die ander twee sisteme (Zwitserlood 2010: 449).

Waar die SignWriting-notering redelik holistiese voorstellings vorm (alhoewel dit deur die kombinasie van simbole verkry word), is die Stokoe- en HamNoSys-noterings tot 'n sekere mate lineêr gerangskik (vgl. figuur 8). Die meer simultane aard van die gebarestruktuur laat verskillende rangskikkinge van die simbole in elke sisteem toe, bv. handkonfigurasie kom eerste in HamNoSys gevolg deur plek van artikulasie, terwyl die Stokoe-sisteem die plek van artikulasie in die gebaar vóór handkonfigurasie aanbied. Die rangskikking is 'n resultaat van die keuse van die ontwerpers. Danksy die lineêre rangskikking van die simbole vergemaklik die Stokoe-notering en HamNoSys nogtans die rangskikking van gebarelemmas in 'n gedrukte woordeboek. Daar is egter, tot dusver, slegs een woordeboek wat uitsluitlik op die rangskikking van gebare volgens

hul notering staatmaak en glad nie enige afbeeldings van die gebare gebruik nie (sien figuur 10), en dit is Stokoe et al. se *Dictionary of American Sign Language on Linguistic Principles* (1976) (Zwitserlood 2010: 450).

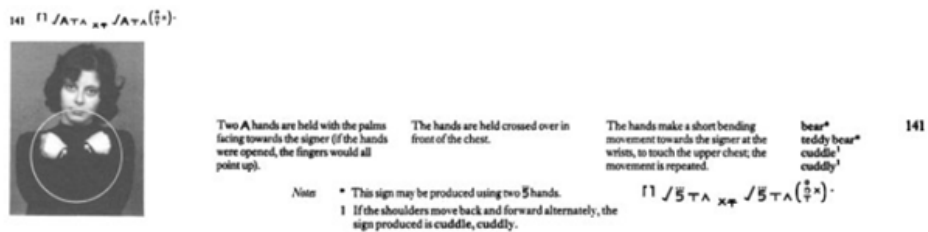
Stokoe-notering		HamNoSys	
[]	hande by of naby bolyf	..	vorm en aksie van linkerhand weerspieël dié van die regterhand
√	voorarm		vingers en duim uitgestrek en reguit
C	gekromde vingers	r	vingers (van die regterhand) wys na links bo
†	arms kruis	0	palm (van die regterhand) na links
v	afwaartse beweging	X	hande of arms gekruis
.	een herhaling	▪ 	(regterhand) aan linkerkant van skouers
x	artikulatore raak	X	kontak met liggaam
		↓	afwaartse beweging
		o	klein beweging
		→	verandering in handvorm (of oriëntering)
		~	gekromde (vinger of vingers)
		+	een herhaling

Figuur 9: Verklaring van die Stokoe- en HamNoSys-noteringsimbole (aangepas uit Zwitserlood 2010: 450)

[] VC† VC_v·
 (imitative of bear hug; arms crossed on chest, dez claw downward) N bear.

Figuur 10: Artikel vir die gebaar vir BEER (Stokoe et al. 1976: 192)

Die *Dictionary of British Sign Language/English* (Brien 1992) gebruik ook 'n aangepaste weergawe van die Stokoe-sisteem om die gebare wat deur die lemmas voorgestel word, te beskryf (sien figuur 11), maar alhoewel die gebare volgens parameterwaardes gerangskik is, volg die rangskikking nie die lineêre rangskikking in die lemmanotering nie, en die gebare is omvattend gerangskik volgens handvorm > plek > oriëntering > beweging, en die spesifieke waardes van hierdie parameters. Hierdie woordeboek verskaf ook foto's van die gebare wat van pyle en ander simbole voorsien is om die aktiwiteit van die gebaar aan te dui (Zwitserlood 2010: 451):



Figuur 11: Artikel vir die gebaar vir BEER (BSL⁵) (Brien 1992: 182-183)

Let daarop dat daar 'n verskil is tussen die ASL- en BSL-gebare vir BEER, sowel as vir die SASL-gebare in figuur 6.

Die noterings het eenvoudig nie posgevat as algemene skryfvoorme van gebaretaal nie. Die rede hiervoor is waarskynlik dat 'n gebruiker eers moet leer hoe om die tekens te ontsyfer en die sisteme vereis dus 'n hoë vlak van 'n spesifieke soort visuele geletterdheid. 'n Mens sou kon argumenteer dat geskrewe en gesproke taal ook 'n hoë mate van spesifieke of gespesialiseerde geletterdheid vereis voordat die gebruiker dit op 'n sinvolle manier kan ontsluit, maar die meeste gebruikers van gebaretaal is eerder geneig om 'n meer intuïtiewe, voor die hand liggende vorm van visuele voorstelling te verkies, waarskynlik weens die beperkte toepassing van hierdie sisteme: 'n persoon wat in 'n geskrewe taal geletterd is, kan elke dag met groot gemak en in enige situasie lees of skryf, maar die noteringsisteme word feitlik uitsluitlik in woordeboeke gebruik, en is gewoonlik nie moontlik om sonder die hulp van gespesialiseerde programmatuur te produseer nie.

Elektroniese gebaretaalwoordeboeke

Die ontwikkeling van elektroniese gebaretaalwoordeboeke bied 'n oplossing vir die probleme wat met voorstelling ondervind word en het akkurate voorstelling moontlik en bekostigbaar gemaak deur middel van vooruitgang in videoprosessering en rekenaar-tegnologie. Sulke woordeboeke kan digitale video's van die gebare verskaf, wat die gebare se aktiwiteite veel beter voorstel, insluitende die dele wat nie op die hande is nie, soos mondbewegings en

gesigsuitdrukking, alhoewel hulle (nog) nie die voorstelling van die gebruik van driedimensionele ruimte toelaat nie (Zwitsers 2010: 451). Met die vin-nige vooruitgang wat ten tyde van hierdie skrywe ten opsigte van driedimen-sionele tegnologie gemaak word — dink maar aan al die 3D-televisies, -teaters, -rekenaarspeletjies, ens. wat reeds beskikbaar is — sal dit egter nie veel langer 'n leemte bly nie.

Die meeste, waarskynlik eerder alle, gebaretaalwoordeboeke bied 'n woordsoekmetode. In gedrukte gebaretaalwoordeboeke is daar gewoonlik (a) 'n alfabetiese lys geskrewe lemmas met 'n verwysing na die toepaslike gebaar, of (b) die gebare is gerangskik as vertaalekwivalente van die lemmas in die geskrewe taal. In elektroniese gebaretaalwoordeboeke is dit ook moontlik om deur die databasis te soek met behulp van 'n geskrewe woord en dan 'n artikel as soekresultaat te vind waarin die geskrewe woord as vertaalekwivalent van die gebaarlemma aangebied word of andersom. In die era van elektroniese lek-sikografie is dit egter ook meer en meer die geval dat elektroniese gebaretaal-woordeboeke 'n gebaarsoekmetode aanbied.

'n Begrip van die fonemiese eienskappe van gebare laat ons sien dat elke gebaar uit sekere elemente of parameters bestaan, nl. handvorm, beweging, palmoriëntering en plek (in die gebareruimte). Sommige gebaretaalwoorde-boeke, soos die *Auslan Dictionary* (Johnston 1989) maak van een of meer van hierdie kenmerke gebruik om 'n vloeiendiagramtipe soekproses vir gebare te skep: die gebruiker word eerstens gevra of die gebaar wat gesoek word, met een of twee hande gemaak word. Indien dit twee hande is, word gevra of die hande simmetries of asimmetries ten opsigte van hul vorm en beweging is. Dan word die gebruiker gevra om die handvorm te identifiseer, sowel as die plek waar die gebaar in die gebareruimte afspeel. Na afloop van hierdie soekproses behoort die gebruiker 'n kort lysie moontlike gebare te hê om van te kies.

Dit is standaardpraktyk onder die meeste gebaretaalwoordeboeke wat vandag aanlyn of digitaal beskikbaar is, om gebruik te maak van 'n soekmetode volgens handvorm (en o.a. dikwels ook plek waar die gebaar geartikuleer word). Enkele voorbeelde van tweetalige internetwoordeboeke — waarin dit dus moontlik is om gebare op te soek asook woorde — van verskillende gebaretale word ter illustrasie genoem. Die soektogte wat hieronder genoem word, is na willekeur gedoen en nie met enige spesifieke gebaar of soekresultaat in gedagte nie, aangesien die skrywer geen kennis van die betrokke gebaretale het nie.

Die internetwoordeboek van Deense Gebaretaal (Center for Tegnsprog 2008) bied 'n totaal van vier soekriteria vir gebare aan (afgesien van 'n woord-soektog): die handvorm van (1) een of (2) twee hande kan gekies word, (3) die plek waar die gebaar geartikuleer word, kan ingesluit word asook (4) die tema waartoe die gebaar behoort. Die gebruiker is nie verplig om van al vier soekcri-teria gebruik te maak nie, alhoewel die soekresultate aansienlik verminder hoe meer kriteria geselekteer word. Deur bv. slegs die handvorm van een hand as die C-handvorm te kies, word 42 gebare as soekresultate aangebied. Deur die gesig/kop as die plek van die gebaar te selekteer, verminder die resultate na 10. Deur "Natuur" as die tema te selekteer, word 3 gebare in die soekresultate

gelewer. Die woordeboek bevat ongeveer 2 000 gebare, wat 'n relatiewe klein aantal lemmas in vergelyking met die meeste woordeboeke van ander gebare-tale is, maar dit dien as 'n goeie aanduiding van die effektiwiteit van 'n traps-gewyse of verfynbare soektog waar soekriteria bygevoeg kan word indien die resultate te veel is. Verfynbare soektogte, veral in die geval van 'n groot aantal soekresultate, is een van Almind (2005: 39) se ontwerp kriteria vir elektroniese woordeboeke. Die voorbeelde van soekprosesse wat hieronder uit ander tweetalige gebaretaalwoordeboeke genoem word, voldoen almal aan hierdie vereiste.

Die *Digitaal Vlaamse Gebaretaal-Nederlands/Nederlands-Vlaamse Gebaretaal woordenboek* (Van Herreweghe et al. 2004) bevat ongeveer 7 458 gebarelemmas en bied ook handvorm en plek/aanraking as soekriteria aan. Deur aanvanklik slegs een handvorm te kies, nl. die F-hand, word 19 soekresultate gelewer. Deur die kop/gesig as plek te selekteer en "aanraak" as die tipe aanraking (in vergelyking met die ander tipes aanraking wat voorsien word, nl. "geen", "gryp", "tussen/in", "slaan", "vee" en "vryf") word die resultate na 'n totaal van 1 gebaar verminder.

Die internetwoordeboek van Finse Gebaretaal (Kuurojen Liittory 2003) bied vier soekriteria vir gebare aan, nl. handvorm, aantal hande (een of twee), plek en beweging. Deur slegs die C-handvorm as soekriterium te selekteer, word 17 gebare in die soekresultate gelewer. Indien die soekprosedure aangepas word om na eenhandige gebare met die C-handvorm te soek, word 9 gebare verkry. Deur een stap verder te gaan en die neutrale gebareruimte as plek te selekteer, word die soeklys na 2 resultate verkort. Dit is nie duidelik hoeveel gebare in die woordeboek opgeneem is nie, en alhoewel die soektog redelik intuïtief volgens die prentjies op die bladsy kon geskied, word al die teks op die webblad slegs in Fins aangebied.

Die internetwoordeboek van Italiaanse Gebaretaal (e-LIS) (Institute for Specialised Communication and Multilingualism 2005) bevat op die oomblik ongeveer 200 gebare en bied ook 'n soekprosedure volgens die parameters handvorm, oriëntasie, plek en beweging aan. 'n Aanvanklike soektog na gebare met die C-handvorm lewer 21 resultate ('n maksimum van agt resultate word per bladsy vertoon). Ten tyde van 'n besoek aan die webblad op 1 September 2011 het die verfyning van die soektog skynbaar nie korrek gefunksioneer nie, aangesien 'n soektog na die C-handvorm met 'n palmoriëntasie na links steeds 21 resultate gelewer het, asook 'n soektog met die bogenoemde parameters plus die boonste gedeelte van die borskas as plek. Dit is natuurlik heeltemal moontlik dat die spesifieke gebruiker se gebrekkige kennis van die Italiaanse instruksies die soektog laat misluk het eerder as 'n fout op die webblad self.

Die internetwoordeboek van Nieu-Seelandse Gebaretaal (McKee et al.) bevat ongeveer 4 000 NZSL-gebare. Die gebruiker kan 'n soektog doen met behulp van 'n Engelse woord, 'n gebaar (volgens handvorm en plek) en ook 'n gevorderde soekfunksie, wat behalwe die selektering van handvorme en plek ook 'n seleksie uit 'n lys tematiese onderwerpe en gebruiksetikette (argaïes/verouderd, neologisme, obsean, informeel of ongewoon) bied. Deur die velde handvorm, plek en gebruiksetiket leeg te laat terwyl slegs 'n tematiese kate-

gorie gekies word, is dit ook moontlik om die totale aantal artikels in elk van die tematiese kategorieë (met 'n maksimum van 9 resultate per bladsy en in alfabetiese volgorde) te sien. Die beperking van die aantal soekresultate wat op 'n bladsy vertoon word, voldoen aan Almind se vyfde ontwerpvereiste vir elektroniese woordeboeke (2005: 39).

Die aanlyn woordeboek van Oostenrykse Gebaretaal (LedaSila) (Zentrum für Gebärdensprache und Hörbehindertenkommunikation) bied benewens die gebruiklike soekriteria plek, handvorm en tema ook die "tipe gebaar" (eenhandig of tweehandig, simmetries of asimmetries) en die landstreek waarin die gebaar voorkom as soekriteria aan. Dit word nie aangedui presies hoeveel gebare in totaal in die woordeboek opgeneem is nie, maar 'n soektog na alle eenhandige gebare lewer 7 066 resultate. 'n Volgende soektog waarby die D-handvorm (geslote vuis met uitgestrekte wysvinger) ingesluit word, lewer 3 resultate.

Soos uit die bostaande bespreking gesien kan word, is dit algemene praktyk in gebaretaalleksikografie om 'n gebaar volgens handvorm op te soek en maak die elektroniese formaat van gebaretaalwoordeboeke dit aansienlik makliker om 'n tweetalige tweerigtingwoordeboek van gebaretaal en 'n geskrewe taal saam te stel. Die elektroniese soekproses volgens handvorm is ook baie meer gebruikersvriendelik as die soekproses aan die hand van notering wat bv. in Stokoe et al. se woordeboek (1976) gebruik is.

Die soektog na gebare op grond van gebareparameterwaardes is dus veel makliker in 'n elektroniese as 'n gedrukte woordeboek. Die woordeboeksamplers moet steeds parameterwaardes aan die gebarevideo's koppel, maar daar is nie die tipiese probleem van rangskikking nie. Die gebruikers van so 'n woordeboek moet weg van die gebaar as 'n geheel abstraher om die gebareparameters in 'n soektog te kan gebruik, maar die aanbieding van sulke waardes kan op 'n duidelike, gebruikersvriendelike manier gedoen word. Die aanbieding van gebare is baie duideliker as in enige gedrukte woordeboek, aangesien 'n video, sowel as 'n tweedimensionele prentjie gesien kan word (Zwitserlood 2010: 453, 454).

Dit is harde werk om sulke omvattende woordeboeke saam te stel, en om daardie rede is daar op die oomblik net 'n handjievul, alhoewel die getal stadig-aan groei. Baie aanlyn gebaretaalwoordeboeke is steeds eenrigting, wat beteken hulle gee slegs gebarevertaling vir woorde van 'n gesproke taal in 'n een-tot-een-verhouding, soos bv. die *American Sign Language Browser* (Michigan State University 1997) (Zwitserlood 2010: 454), en sonder verdere inskrywings vir die lemma in die gebaretaal, soos die *Gallaudet Dictionary of American Sign Language* (Valli 2005). Laasgenoemde woordeboek verskaf in die boekvorm slegs 'n alfabetiese lys Engelse soekwoorde met 'n bygaande skets, sonder enige verdere data. Op die meegaande CD-ROM lei die geskrewe soekwoord uit die alfabetiese lys na 'n video van die gebaar, en in sommige gevalle word sinonieme (in die gesproke taal) verskaf, maar geen verdere data word in ASL gegee nie. Dit is ook nie moontlik om 'n gebaar op te soek nie.




Elektroniese woordeboeke het steeds 'n redelike klein databasis, met ongeveer 2 000 tot 5 000 lemmas (Zwitserlood 2019: 454).

'n Model vir 'n nuwe elektroniese gebaretaalwoordeboek

Dowes verteenwoordig 'n groep teikengebruikers wat nog nie oor 'n gevestigde woordeboekkultuur beskik nie. Die tipiese Dowe persoon — daar is natuurlik altyd uitsonderings — behaal slegs lae vlakke van geletterdheid in die geskrewe/gesproke taal, wat hulle tweede taal is. Die ideaal is dus om 'n woordeboekkultuur te skep by 'n groep jong gebruikers, nl. leerders in die grondslagfase, deur middel van 'n tweetalige tweerigtinggebaretaalwoordeboek wat spesifiek op hulle belangstellings en behoeftes gerig is. Fourie (2013) se model is ontwerp vir die De la Bat Skool vir Dowes in Worcester, spesifiek vir leerders in die grondslagfase. Daar word beplan dat die voorgestelde grondslagfase-woordeboek deur verdere woordeboeke vir die intermediêre en senior fases opgevolg en aangevul sal word, sodat leerders deur hulle hele skoolloopbaan toegang tot 'n woordeboek sal hê wat geskik is vir hulle behoeftes. Die skool beskik tans oor geen skoolwoordeboek nie en gebruik geskrewe klaslyste, wat slegs die Afrikaanse teikenwoorde in die onderskeie leerareas of temas lys. Die geskrewe lyse maak dit egter nie moontlik om 'n visuele skakel te vorm met die betrokke konsep wat deur 'n woord uitgedruk word nie, en ook nie om aan te dui watter gebaar of gebare by die konsep en woord pas nie. Daar bestaan dus 'n behoefte aan 'n elektroniese produk wat 'n visuele soekmetode vir 'n konsep bied, sowel as 'n soekmetode vir 'n gebaar en 'n woord.

Daar is heelwat geskrewe woordeboeke op die mark wat op grondslagfaseleerders gemik is en wat ook tematiese soekareas of kategorieë insluit, aangesien dit makliker vir jong gebruikers is om konsepte op te soek binne die leerareas wat in die klaskamer behandel word. Voorbeelde van sulke woordeboeke sluit in die *Longman Grondslagfasewoordeboek Afrikaans/Engels* (2010) (voorts LGFW) en die *Oxford First Bilingual Dictionary Afrikaans + English* (2010). Woordeboeke soos dié verskaf gewoonlik 'n indeks in albei tale agterin die woordeboek, en in die geval van die Longman-woordeboek is daar ook 'n Afrikaanse alfabetiese afdeling waarin 'n definisie vir elke woord tesame met 'n voorbeeldsin in Afrikaans, asook 'n Engelse vertaling van die voorbeeldsin wat die gebruiker help om sy of haar addisionele taal te verstaan en te gebruik. Prentjies word ook in hierdie afdeling gebruik om woorde te kontekstualiseer. Woorde wat tot spesifieke tematiese kategorieë behoort of daartoe sou kon behoort (maar nie noodwendig in die betrokke tematiese kategorieë ingesluit is nie), word daarheen terugverwys sodat die alfabetiese woordelys steeds gekontekstualiseer bly.

Fourie (2013) se model maak voorsiening vir 'n elektroniese woordeboek wat op drie maniere soekbaar is: volgens konsep (deur 'n versameling tematiese kategorieë wat deur prentjies voorgestel word), volgens gebaar (deur 'n proses van parameterkeuses) en volgens woord. Die soekresultate sal altyd identies wees, ongeag die soekmetode. Op hierdie manier ontstaan 'n trippellemma wat bestaan uit 'n ostensiewe lemma wat deur die **prentjie** verteenwoordig word, 'n tweede lemma wat deur die **gebaar** (as skets en/of video) verteenwoordig word en 'n derde lemma wat uit die geskrewe **woord** bestaan soos voorgestel in figuur 12.

			lemoen
---	---	---	---------------

Figuur 12: Hipotetiese voorstelling van die voorkoms van die drievoudige lemma (Fourie 2013: 349)

Die model stel ook 'n parallelle mikro-argitektuur voor waarin nie net die lem-mata langs mekaar aangebied word nie, maar ook die definisies; sien figuur 13.

			lemoen
		(Video van definisie in gebaretaal).	'n Lemoen is 'n groot, sappige vrug met 'n dik skil.

Figuur 13: Hipotetiese aanbieding van 'n definisie in albei tale (Fourie 2013: 358)

In die bostaande voorbeeld is slegs die gebruik van parallelle definisies in albei tale gedemonstreer, nie die volledige mikrostruktuur nie. Die geskrewe definisie is uit die LGFW (2010: 55). Die gebaretaaldefinisie en die geskrewe definisie sal met mekaar moet ooreenstem om die inhoud van die mikrostruktuur tweetalig te maak. Voorbeeldsinne kan op 'n soortgelyke manier hanteer word.

Verdere besonderhede van die model sal in komende publikasies bekend gemaak word. Op hierdie stadium is dit belangrik om kennis te neem dat die model voorsiening maak vir 'n situasie waarin 'n Dowe gebruiker nie in enige geskrewe taal geletterd hoef te wees om 'n woordeboek in sy of haar eie taal te gebruik nie. Terselfdertyd sal 'n horende gebruiker in staat wees om 'n gebaar op te soek waarvan die betekenis onbekend is, soos wanneer 'n gebaar in 'n gesprek of onbekende konteks opgemerk word. Gebruikers sal egter ook die opsie hê om van konsep na gebaar of woord te beweeg deur middel van 'n visuele tematiese soekmetode. Dus hoef gebruikers eintlik in geen taal geletterd te wees om die woordeboek effektief te gebruik nie.

Slot

Die ontwikkeling van gebaretaalleksikografie was in die verlede grootliks onder-

hewig aan verskillende vooroordele en beperkinge. Dit was dikwels nie *nodig* geag om 'n tweetalige woordeboek saam te stel waarin dit ook moontlik is om 'n gebaar op te soek nie aangesien tweetalige gebaretaalwoordeboeke grotendeels gemik was op horende persone wat die taal wou of moes leer. Daarby was dit baie moeilik vanweë die aard van gedrukte woordeboeke en die gebrek aan 'n ortografiese vorm van gebaretaal, om gebare op 'n manier voor te stel én te rangskik sodat hulle opgesoek kan word sonder om van 'n geskrewe soekterm gebruik te maak. Daar is wel notasiesisteme ontwikkel in 'n poging om hierdie probleem te omseil, maar hierdie sisteme het nooit by gewone gebruikers wyd byval gevind nie. Die gebruik van elektroniese media en platforms om gebaretaalwoordeboeke te ontwikkel en aan te bied het egter gesorg vir 'n revolusie in gebaretaalleksikografie. Dit is nie meer ongewoon om elektroniese gebaretaalwoordeboeke te vind waarin nie net 'n woord nie, maar ook 'n gebaar soekbaar is, gewoonlik volgens een of meer van die gebaar se parameters, soos handvorm. Sulke woordeboeke maak dit moontlik vir 'n Dowe gebruiker om direkte toegang tot sy of haar eerste taal (gebaretaal) te bekom sonder enige kennis van die tweede taal (die geskrewe taal). Aangesien daar nog nie 'n gevestigde woordeboekkultuur onder Dowe gebruikers bestaan nie, sal dit die ideale geleentheid wees om 'n jong generasie teikengebruikers van 'n woordeboek te voorsien wat op hulle spesifieke behoeftes en belangstellings gerig is. Fourie (2013) voorsien 'n model vir 'n elektroniese tweetalige grondslagfasewoordeboek van Suid-Afrikaanse Gebaretaal en Afrikaans waarin dit nie net moontlik is om van 'n woord na 'n gebaar of van 'n gebaar na 'n woord te beweeg nie, maar ook van 'n konsep na 'n gebaar of woord. Behalwe woord- en gebaarsoekmetodes word daar ook in die model voorsiening gemaak vir 'n visuele, tematiese soekmetode. Dit maak dit vir die gebruiker moontlik om toegang tot die woordeboek te verkry sonder om gebruik te maak van enige van die twee tale wat aangebied word. Die soekresultate vertoon identies ongeag die soekmetode wat gebruik word, wat verseker dat die drie soekmetodes volledig met mekaar sluit en dat elke soekmetode onafhanklik van die ander gebruik kan word om toepaslike soekresultate te lewer.

Eindnote

1. Die gebruik van die kleinletter in doof dui op oudiologiese of patologiese doofheid, terwyl die hoofletter in Doof dui op kulturele Doofheid.
2. SASL — South African Sign Language (Suid-Afrikaanse Gebaretaal). Engelse afkortings is in die teks behou omdat dit die bekendste en gebruiklikste in die literatuur is.
3. Die NME's dra geweldige grammatikale waarde — sonder gesigsuitdrukking is die meeste gebare ongrammatikaal en vir 'n Dowe onverstaanbaar. Dit is ook gewoonlik een van die sterk punte van kritiek teen die meeste gebarewoordeboeke dat NME's nie genoegsaam uitgebeeld word nie.
4. ASL — American Sign Language (Amerikaanse Gebaretaal).
5. BSL — British Sign Language (Britse Gebaretaal).

Bibliografie

- Almind, R.** 2005. Designing Internet Dictionaries. *Hermes, Journal of Linguistics* 34: 37-54.
- Brien, D. (Red.).** 1992. *Dictionary of British Sign Language/English*. Londen: Faber&Faber.
- Center for Tegnsprog.** 2008. *Ordbog over Dansk Tegnsprog*. (The Danish Sign Language Dictionary.) <http://www.viittomat.net/http://www.tegnsprog.dk/>.
- Costello, E.** 1998. *Random House Webster's American Sign Language Dictionary*. New York: Gallaudet University Press.
- Fourie, H.** 2013. 'n *Leksikografiese model vir 'n elektroniese tweetalige grondslagfasewoordeboek van Suid-Afrikaanse Gebaretaal en Afrikaans*. Ongepubliseerde D.Litt.-proefskrif. Stellenbosch: Universiteit Stellenbosch.
- Gouws, R., L. Potgieter en S. Burgess (Reds.).** 2010. *Longman Grondslagfasewoordeboek Afrikaans/Engels*. Kaapstad: Maskew Miller Longman.
- 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.
- Institute for Specialised Communication and Multilingualism.** 2005. *Electronic Bilingual Dictionary Italian Sign Language-Italian*. http://elis.eurac.edu/index_en.
- Johnston, T.** 1987. *Preliminary Signing Dictionary of Australian Sign Language*. Adelaide: TAFE National Centre for Research and Development.
- Johnston, T.** 1989. *Auslan Dictionary: A Dictionary of Australian Sign Language (Auslan)*. Adelaide: TAFE National Centre for Research and Development.
- Kuurojen Liittory.** 2003. *Suvi Suomalaisen viittomakielen verkkosanakirja*. (Finnish Sign Language Dictionary.) <http://suvi.viittomat.net>.
- Landau, S.I.** 1984. *Dictionaries. The Art and Craft of Lexicography*. New York: The Scribner Press.
- Malm, A. (Red.).** 1998. *Suomalaisen viittomakielen perussanakirja*. Helsinki: Libris Oy.
- 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>.
- Mesthrie, R., J. Swann, A. Deumert en W. L. Leap (Reds.).** 2000. *Introducing Sociolinguistics*. Edinburgh: Edinburgh University Press.
- Michigan State University.** 1997. *American Sign Language Browser*. <http://aslbrowser.commtechlab.msu.edu/index.html>.
- Nieder-Heitmann, N.** 1980. *Praat met die dowe: 'n Visuele handboek met gestandaardiseerde gebare vir Dowe in Suid-Afrika*. Pretoria: Staatsdrukker.
- Penn, C., D. Ogilvy-Foreman, D. Simmons and M. Anderson-Forbes.** 1992-1994. *Dictionary of Southern African Signs for Communicating with the Deaf*. 5 dele. Pretoria: Raad vir Geesteswetenskaplike Navorsing.
- Prillwitz, S., R. Leven, H. Zienert, T. Hanke en J. Henning.** 1989. *HamNoSys. Version 2.0. Hamburg Notation System for Sign Language: An Introductory Guide*. Hamburg: Signum Verlag.
- Radutzky, E. (Red.).** 1992. *Dizionario bilingue elementare della Lingua Italiana dei Segni*. Rome: Edizioni Kappa.
- Schermer, T.** 2004. *Sign Language Lexicography*. Stichting Nederlands Gebarencentrum. <http://www.gebarencentrum.nl/publicaties.asp>.

SignWriting for Sign Languages. 2011. www.signwriting.org [5 Junie 2011].

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.

Van Hereweeghe, M., S. Slembrouck en M. Vermeerbergen. 2004. *Digitaal Vlaamse Gebarentaal–Nederlands/Nederlands–Vlaamse Gebarentaal woordenboek*. <http://gebaren.ugent.be/>.

Zentrum für Gebärdensprache und Hörbehindertenkommunikation. s.j. *LedaSila — Lexical Database for Sign Languages*. <http://ledasila.uni-klu.ac.at>.

Zwitserlood, I. 2010. Sign Language Lexicography in the Early 21st Century and a Recently Published Dictionary of Sign Language of The Netherlands. *International Journal of Lexicography* 23(4): 443–476.

Die leksikografiese aanbieding en behandeling van vaste uitdrukkings*

Rufus H. Gouws, *Departement Afrikaans en Nederlands,
Universiteit Stellenbosch, Stellenbosch, Suid-Afrika
(rhg@sun.ac.za)*

Opsomming: Vir leksikograwe is dit belangrik om in hulle besluite oor die insluiting van vaste uitdrukkings in hulle woordeboeke vertrou te wees met die status van vaste uitdrukkings as volwaardige leksikale items. Hierdie status behoort 'n invloed te hê op die manier waarop vaste uitdrukkings aangebied en behandel word. Na aanleiding van die behandeling in bepaalde woordeboeke word daar in hierdie artikel aangevoer dat vaste uitdrukkings lemmastatus in woordeboeke moet kry en wel deur as sublemmata opgeneem te word in 'n geneste teksblok wat geheg word aan die artikel van 'n lemma wat 'n sleutelwoord uit die vaste uitdrukking verteenwoordig. Die behandeling moet van 'n omvattender aard wees as wat tans die geval is in algemene woordeboeke. 'n Fokus op leksikografiese funksies vra om die insluiting van sekere aanduidertipes in die behandeling. Dit sluit onder meer etimologiese data in, asook eksplisiete leiding ten opsigte van kultuurgebonde vaste uitdrukkings.

Sleutelwoorde: AANDUIDER, DEELARTIKELTRAJEK, GIDSELEMENT, IDIOME, KOGNITIEWE FUNKSIE, KONTEKSLEIDING, KOTEKSLEIDING, KULTUURGEBONDE ITEMS, LEKSIKOGRAFIESE BEHANDELING, LEKSIKOGRAFIESE FUNKSIES, LEMMATEKEN, LEMMATIESE ADRESSERING, NES, NIELEMMATIESE ADRESSERING, NIS, SUBLEMMA, VASTE UITDRUKKINGS

Abstract: The Lexicographic Presentation and Treatment of Fixed Expressions. In their decisions regarding the inclusion of fixed expressions in their dictionaries it is important for lexicographers to be familiar with the status of fixed expressions as fully-fledged lexical items. This status should have an influence on the way in which fixed expressions are presented and treated. With reference to the treatment allocated to fixed expressions in a few dictionaries it is argued in this paper that they should be awarded lemma status in dictionaries. It is proposed that fixed expressions should be included as sublemmata, presented in a nested text block that is attached to the article of a lemma representing a keyword from the fixed expression. Their treatment should be more comprehensive than what is currently found in prevailing general dictionaries. A focus on lexicographic functions demands the inclusion of certain item types in the treatment of fixed expressions. This includes etymological data as well as much more explicit guidance with regard to culture-bound lexical items.

* Hierdie artikel is 'n verwerking van 'n referaat gelewer by die internasionale kongres "EURO-PHRAS 2012 Maribor: Phraseology and Culture", aangebied deur die European Society of Phraseology, by die Universiteit Maribor, Maribor, Slowenië, 27-31 Augustus 2012.

Keywords: COGNITIVE FUNCTION, CONTEXTUAL GUIDANCE, COTEXTUAL GUIDANCE, CULTURE-BOUND ITEMS, FIXED EXPRESSION, GUIDING ELEMENT, IDIOMS, ITEM, LEMMA SIGN, LEMMATIC ADDRESSING, LEXICOGRAPHIC FUNCTIONS, LEXICOGRAPHIC TREATMENT, NEST, NICHE, NONLEMMATIC ADDRESSING, PARTIAL ARTICLE STRETCH, SUBLEMMA

1. Inleiding

Vaste uitdrukkings word in 'n verskeidenheid woordeboektypes opgeneem, aangebied en behandel, onder meer in woordeboeke met vaste uitdrukkings as enigste teiken van leksikografiese behandeling asook algemene eentalige en tweetalige woordeboeke gerig op 'n weergawe en behandeling van items uit die algemene leksikon van die betrokke taal. In hierdie artikel word daar slegs aandag gegee aan vaste uitdrukkings in algemene woordeboeke. Die term *vaste uitdrukking* word met 'n breë toepassing gebruik, naamlik om te verwys na idioome, spreekwoorde, geveleude woorde, ensovoorts, vergelyk in hierdie verband Gouws (1989). Die term *algemene woordeboek* moet breedweg geïnterpreteer word om na al daardie woordeboeke te verwys wat op 'n verteenwoordigende versameling items uit die algemene leksikon van 'n gegewe taal fokus.

Die bespreking in hierdie artikel is beperk tot gedrukte woordeboeke want hierdie woordeboeke word steeds gebruik, beplan en saamgestel en moet steeds 'n stewige metaleksikografiese basis hê. Talle aspekte wat ter sprake kom, is egter ook ter sake vir elektroniese woordeboeke. Vir meer inligting oor vaste uitdrukkings in elektroniese woordeboeke kan Bergenholtz, Bothma en Gouws (2011) geraadpleeg word.

In die ontwikkeling van die teoretiese leksikografie kan drie belangrike fases onderskei word wat elk 'n eie fokusgebied het, te wete 'n fokus op die taalkundige inhoud van woordeboeke (die sogenaamde Zgusta-fase), 'n fokus op die strukture van woordeboeke (die sogenaamde Wiegand-fase) en 'n fokus op die leksikografiese funksies van woordeboeke (die sogenaamde Bergenholtz en Tarp-fase), vergelyk in hierdie verband ook Gouws (2005) en Gouws en Prinsloo (2005). Ondanks die fokus op verskillende aspekte en ondanks die soms opponerende standpunte tydens hierdie fases is dit noodsaaklik om die wesenlike belang en tersaaklikheid van al drie hierdie fases vir die hedendaagse leksikografieteorie te erken. In die beplanning en samestelling van enige woordeboek moet die leksikograaf deeglik aandag gee aan besluite oor die inhoud, strukture en leksikografiese funksies van die beplande woordeboek.

Die gebruikersperspektief dwing elke leksikograaf wat beoog om 'n nuwe woordeboek saam te stel om die volgende belangrike vraag te beantwoord, naamlik "Wat wil ek hê, moet my gebruiker met die beplande woordeboek kan doen?" 'n Antwoord op hierdie vraag behoort te lei tot 'n vasstelling van die leksikografiese funksie(s) van die woordeboek. In ooreenstemming met die funksie(s) moet die data wat ingesluit moet word, gekies word en die vasstel-

ling van die dataverspreidingstruktuur lei dan tot die vasstelling van 'n verskeidenheid strukture met behulp waarvan die gekose data op die bes moontlike wyse aangebied, behandel en bereik kan word sodat die gebruiker die nodige inligting aan die data-aanduider kan onttrek.

Die gekose benadering tot inhoud, struktuur en funksies moet op 'n konsekwente en uitputtende manier toegepas word. Dit moet ook bepalend wees vir die keuse, aanbieding en behandeling van vaste uitdrukings in algemene een- en tweetalige woordeboeke. Van hierdie aspekte met betrekking tot vaste uitdrukings kom in hierdie artikel aan bod.

2. Die keuse van vaste uitdrukings

Die keuse van vaste uitdrukings wat in algemene woordeboeke opgeneem moet word, mag nie op 'n arbitrêre manier geskied nie. Dit moet met inagneming van die woordeboektype, die funksies en die teikengebruikers van die beplande woordeboek gedoen word. In baie woordeboeke word die keuse van vaste uitdrukings regstreeks beïnvloed deur die status wat leksikograwe aan hulle gee, naamlik as óf makro- óf mikrostrukturele inskrywings van die betrokke woordeboek. As makrostrukturele aanduiders word vaste uitdrukings op grond daarvan gekies dat hulle volwaardige leksikale items van die betrokke taal is en dus kwalifiseer om as lemmata opgeneem te word, dit wil sê as volwaardige behandelingseenhede en dus as niegeadresseerde aanduiders. As mikrostrukturele items is hulle aanduiders wat gesien word as deel van die behandeling van 'n ander lemma. Hulle is as aanduiders geadresseer aan die makrostruktuuraanduider en as behandelingseenhede is hulle hoogstens die teiken van prosedures van nielemmatiese adressering.

Woordeboeke vertoon dikwels 'n inkonsekwentheid deur nie dieselfde teoreties gebaseerde kriteria vir die keuse van makro- en mikrostruktuur-inskrywings te gebruik nie. Te dikwels gebruik leksikograwe bepaalde kriteria vir die keuse van makrostruktuurelemente maar laat die toepassing van daardie kriteria na in die besluite oor die keuse van mikrostruktuurelemente. Ter illustrasie hiervan: gebruiksfrekwensie speel 'n belangrike rol in die keuse van lemmakandidate, maar hierdie criterium speel 'n ondergeskikte rol in die keuse van byvoorbeeld die betekenisparafrases of die kollokasies wat in die behandeling van 'n bepaalde lemma aangebied word. Waar vaste uitdrukings as mikrostrukturele aanduiders aangebied word, is hulle deel van 'n artikel met 'n lemma as gidselement wat ooreenstem met 'n sleutelwoord uit die vaste uitdrukking. 'n Vaste uitdrukking soos *Waar die hart van vol is, loop die mond van oor* sal tipieserwys ingesluit word in die artikel van die lemmateken *hart*. Waar die artikelstruktuur voorsiening maak vir 'n soeksone vir vaste uitdrukings probeer leksikograwe vaste uitdrukings kry wat die woord wat deur die lemmateken voorgestel word as sleutelwoord bevat. Pogings om die soeksone van vaste uitdrukings te beset, lei dikwels daartoe dat die leksikograaf nie die beginsel van gebruiksfrekwensie toepas in die keuse van mikrostruktuurin-

skrywings nie. Waar een of meer vaste uitdrukkings met die verlangde sleutelwoord in die betrokke taal gevind word, word hulle in die gleuf vir vaste uitdrukkings opgeneem om die latente soeksone te bevolk en daardeur daartoe by te dra dat 'n uitgebreide mikrostruktuur verwesenlik word. Volgens die voorwoord van die *Tweetalige Skoolwoordeboek*, 'n tweetalige skoolwoordeboek met Afrikaans en Engels as taalpaar, poog die woordeboek om daardie woordeskat-items op te neem wat dikwels deur die teikengebruikers van die woordeboek teëgekomp word. In die artikel van die lemma *hart* is die vaste uitdrukking *iemand 'n hart onder die riem steek* opgeneem. Dit is 'n verouderde Afrikaanse uitdrukking wat nie 'n kandidaat vir opname in hierdie woordeboek moes gewees het indien gebruiksfrekwensie werklik 'n kriterium was nie. As gebruiksfrekwensie 'n bepalende rol speel in die keuse van leksikale items as lemmakandidate behoort dit ook 'n rol te speel in die keuse van daardie leksikale items, soos vaste uitdrukkings, wat nie lemmastatus het nie maar eerder as mikrostrukturele aanduiders opgeneem word. Nogmaals behoort leksikografiese konsekwentheid 'n deurslaggewende rol te speel.

Die keuse van vaste uitdrukkings vir opname in woordeboeke moet volgens dieselfde dataverkrygingskriteria geskied as wat gebruik word vir die opname van ander leksikale items as behandelingseenhede in die betrokke woordeboek. In hierdie verband moet vaste uitdrukkings dus beoordeel word na aanleiding van hulle optrede as volwaardige leksikale items en nie as meerwoordige konstruksies wat die kombinatoriese vermoë illustreer van die leksikale item wat deur die lemmateken van die artikel waarin die vaste uitdrukking optree, verteenwoordig word nie.

3. Die leksikografiese aanbieding van vaste uitdrukkings

3.1 Makro- of mikrostrukturele elemente

Die insluiting van vaste uitdrukkings as mikrostrukturele aanduiders in 'n artikel waarvan die woord wat as gidselement optree, ooreenstem met 'n woord in die vaste uitdrukking is sowel linguisties as leksikografies onaanvaarbaar, vergelyk Gouws (1989). Dit misken die status van vaste uitdrukkings as volwaardige leksikale items. Hierdie soort aanbieding bring mee dat die behandeling van die vaste uitdrukking met behulp van prosedures van nielemmatiese adresering gedoen word en dit lei daartoe dat die vaste uitdrukking 'n minder prominente behandelingseenheid word.

Om hulle status as volwaardige leksikale items te weerspieël en om te verseker dat hulle 'n behandeling kry wat gelykwaardig is aan dié van ander tipes leksikale items wat as lemmata opgeneem word, moet die plasing van vaste uitdrukkings as behandelingseenhede verhoog word en die vaste uitdrukkings wat gekies word om in 'n woordeboek opgeneem te word, moet as lemmakandidate oorweeg word.

Die lemmatisering van meerwoordige leksikale items is nie onproblema-

ties nie. Dit geld ook vaste uitdrukings. Om praktiese redes, byvoorbeeld die feit dat vaste uitdrukings nie altyd 'n vaste eerste komponent het nie, is 'n prosedure vasgelê om vaste uitdrukings op so 'n manier in woordeboeke in te sluit dat hulle in of geheg aan die artikel van 'n sleutelwoord uit die vaste uitdrukking verskyn. Hierdie metode kan redelik suksesvol uitgevoer word en hierdie tradisie sou ten dele behou kon word, maar dit verg 'n deeglike ondersoek na die posisie, aanbieding en status van 'n vaste uitdrukking tydens die uitvoering van hierdie metode. Dit is belangrik dat die vaste uitdrukking op so 'n manier aangebied moet word dat daar geen twyfel bestaan oor sy status as hetsy makro- hetsy mikrostrukturele element nie. In die besluitneming hieroor kan daar nogmaals aandag gegee word aan die insluiting van vaste uitdrukings vanuit die perspektief van leksikografiese strukture.

3.2 In 'n deelartikeltrajek

Gouws (2010) pleit vir die insluiting van vaste uitdrukings in 'n deelartikeltrajek wat geheg is aan maar nie ingesluit is nie in die artikel van 'n lemma wat 'n sleutelwoord uit die vaste uitdrukking verteenwoordig en waarvolgens die alfabetiese posisionering van die vaste uitdrukking kan geskied. Deur hulle insluiting in 'n afsonderlike deelartikeltrajek, selfs al bevat hierdie deelartikeltrajek slegs een artikel met 'n vaste uitdrukking as sekondêre maar onmiddellike gidselement, word die status van vaste uitdrukings as makrostruktuurelemente bevestig en benadruk. Sommige voorbeelde uit Gouws (2010) sal ook hier bespreek word.

Een van die onbevredigendste maniere waarvolgens vaste uitdrukings in woordeboeke behandel kan word, is deur middel van 'n gegroepeerde mikrostrukturele kluster in 'n artikelgleuf wat vaste uitdrukings en voorbeeldmateriaal in dieselfde kluster bevat sonder om hierdie verskillende tipes aanduiders hoegenaamd van mekaar te onderskei. Die hieropvolgende artikel uit *Groot Woordeboek/Major Dictionary* vertoon so 'n primitiewe mikrostruktuur sonder 'n mikro-argitektuur:

baan, (s) (**bane**), course, path, way, race (of moon), orbit (stars); trajectory (shell); court (*tennis*); rink (*bowls, skating*); floor (dancing); track (*athl.*); breadth, width, panel (cloth); permanent way, roadbed; lane (traffic); circuit (*electr.*); guide-way; gore (skirt); ~ **BREEK**, pave the way; pioneer; *op die ~ BRING*, bring on the tapis; *op 'n GLADDE ~*, on a slippery path; *iem. v.d. ~ KNIKKER*, oust someone; *op die LANGE ~*, postponed indefinitely; shelved; 'n ~ **OPSKOP**, kick up a row; *die ~ OPEN*, open the dance; *die ~ RUIM*, clear the way; clear the floor; *VAN die ~ wees*, be shelved; *VRYE ~ hê*, have been given the right of way; *v.d. ~ WEES*, have been put off (shelved); ...

Soms word vaste uitdrukings so geplaas dat hulle as aanduiders elk in 'n eie semantiese subkommentaar van die artikel van 'n woordlemma optree. Vergelyk in hierdie verband die volgende uittreksels uit die artikel van die lemma-

teken *ear*¹ in die *Collins Dictionary of the English Language*:

ear¹ (...) *n.* **1.** the organ of hearing and balance ... **2.** the outermost cartilaginous part of the ear ... **6.** an object resembling the external ear ... **12.** a flea in one's ear. *Informal.* a sharp rebuke. ... **26.** wet behind the ears. *Informal.* youthful in manner or attitudes ...

Al is dit so dat die vaste uitdrukkings aangebied word as behandelingseenhede, impliseer hulle plasing hier steeds 'n mikrostrukturele en semantiese verhouding tot die lemma wat as gidselement van die spesifieke artikel optree. Toegang tot so 'n vaste uitdrukking is selfs vir die kundige woordeboekgebruiker nie onproblematies nie.

3.3 Artikel-intern versus artikel-ekstern

Vaste uitdrukkings word in verskillende posisies geplaas met betrekking tot die artikel van die lemma wat ooreenstem met die sleutelwoord uit die vaste uitdrukking. Sommige woordeboeke het 'n artikel-interne soeksonne wat gereserveer is vir vaste uitdrukkings en duidelik gemerk is met behulp van 'n struktuurmerker. Vergelyk in hierdie verband byvoorbeeld die artikel van die lem-mateken *pineapple* in *The New Oxford Dictionary of English*:

pineapple ► **noun 1** a large juicy tropical fruit consisting of aromatic edible yellow flesh surrounded by a tough segmented skin and topped with a tuft of stiff leaves.
2 ...
3 *Informal* a hand grenade.
PHRASES **the rough end of the pineapple** *Austral./NZ informal* a situation in which someone receives unfair or harsh treatment.
ORIGIN late Middle English ...

Hierdie soeksonne het 'n artikel-interne posisie en die vaste uitdrukkings word as mikrostrukturele inskrywings aangebied. Die gebruik van so 'n afsonderlike artikelgleuf verhoog toegang tot die vaste uitdrukkings en die struktuurmerker "PHRASES" skep 'n kitstoegangstruktuur. 'n Vergelykbare benadering kan gevind word in HAT, die *Verklarende Handwoordeboek van die Afrikaanse Taal*:

gat¹ (-e; gaatjie)

1 Opening in of holte deur of in 'n voorwerp: 'n *Gat in hout boor*. 'n *Gat in 'n sokkie stop*. 'n *Gat in jou tand hê*. 'n *Gat grawe om 'n boom te plant*. *Die Groot Gat by Kimberley*. 'n *Loer-, lug-, man-, sleutelgat*. **2** Uitgeholde ruimte onder die grond: *Molsgat*. *Erdvarke bly bedags in gate*. ∇ *Jakkalse het gate en voëls het neste, maar die Seun van die mens het nie eers 'n rusplek vir sy kop nie* (Matt. 8:20 NV). **3** Holte of duik in 'n landskap: *Die dorp lê in 'n gat*. **4** ... **8** Gebrek, leemte: *Daar is groot gate in sy kennis*. **UITDR.**: 'n *Gat in jou beursie hê*, spandabel wees. 'n *Gat in die dag slaap*, tot laat in die dag. *Iets*

gate uit geniet (geselst.), dit baie geniet. *Vir 'n ander 'n gat grawe en self daarin val*, self in die moeilikheid beland wat jy vir 'n ander berei het. *In die gate hou* (veroud.), dophou, in die oog hou. *Iemand 'n gat in die kop praat*, hom ompraat, omhaal – gewoonlik negatief gebruik. *Gate ('n gat) uit speel*, baie goed, lekker speel. **ga'terig**

Hierdie soeksone, aangedui deur die struktuurmerker "UITDR." ('n afkorting vir "uitdrukking"), bevat 'n kluster van vaste uitdrukings. Op hierdie soeksone volg 'n artikelnis met slegs een artikel, naamlik 'n artikel met die sublemma *gaterig* as gidselement. Dit is nie sonder meer duidelik of hierdie kluster vaste uitdrukings 'n artikel-interne of 'n artikel-eksterne posisie het nie en daarom is dit nie sonder meer duidelik of die vaste uitdrukings mikro- al dan makrostrukturele aanduiders is nie. Hulle word in 'n gleuf aangebied wat geskei is van die semantiese subkommentare van die betrokke artikel en hulle word as behandelingseenhede aangebied, al is dit slegs 'n beperkte behandeling wat aan elke vaste uitdrukking gegee word. Alhoewel die plasing van die kluster die indruk skep dat dit geïntegreerd is in die artikel van die lemmateken *gat*¹ kan dit ook beskou word as 'n gegroepeerde nes wat die artikelnis wat vir onverklaarde enkelwoord komplekse lemmata voorsien is, voorafgaan. Hierdie nestingsprosedure verteenwoordig 'n afwyking van die strak alfabetiese ordening wat die betrokke woordeboek kenmerk. Toegang tot die nes geskied via die vertikaal-geordende hooflemma van die voorafgaande artikel wat hier optree as artikel-eksterne primêre gidselement van elke vaste uitdrukking in die betrokke kluster. Toegang tot die individuele vaste uitdrukings geskied deur middel van 'n sekondêre alfabetiese prosedure, want die ordening van die vaste uitdrukings in die kluster word gedoen volgens die alfabetiese waarde van 'n verdere woord uit die uitdrukking; nie die sleutelwoord wat as primêre gidselement optree nie. Deur hierdie vaste uitdrukings as makrostruktuur-aanduiders te beskou, kan die toegang tot elke vaste uitdrukking verhoog word deur die toepassing van 'n niegegroepeerde ordening, vergelyk in hierdie verband ook Bergenholtz, Tarp en Wiegand (1999) wat die deelartikeltrajek skei van die voorafgaande hoofartikel en wat ook 'n beter toepassing van die mikro-argitektuur daarstel. Vergelyk ook Gouws (2006):

gat¹ (-e; gaatjie)

1 Opening in of holte deur of in 'n voorwerp: 'n *Gat in hout boor*. 'n *Gat in 'n sokkie stop*. 'n *Gat in jou tand hê*. 'n *Gat grawe om 'n boom te plant*. *Die Groot Gat by Kimberley*. 'n *Loer-, lug-, man-, sleutelgat*. **2** Uitgeholde ruimte onder die grond: *Molsgat*. *Erdvarke bly bedags in gate*. ∇ *Jakkalse het gate en voëls het neste, maar die Seun van die mens het nie eers 'n rusplek vir sy kop nie* (Matt. 8:20 NV). **3** Holte of duik in 'n landskap: *Die dorp lê in 'n gat*. **4 ... 8** Gebrek, leemte: *Daar is groot gate in sy kennis*.

UITDR.: 'n **Gat in jou beursie hê**, spandabel wees.

'n **Gat in die dag slaap**, tot laat in die dag.

Iets gate uit geniet (geselst.), dit baie geniet.

Vir 'n ander 'n gat *grawe* en self daarin val, self in die moeilikheid beland wat jy vir 'n ander berei het.

In die gate *hou* (veroud.), dophou, in die oog hou.

Iemand 'n gat in die *kop* praat, hom ompraat, omhaal – gewoonlik negatief gebruik.

Gate ('n gat) uit *speel*, baie goed, lekker speel.

Al is hierdie vaste uitdrukkings geheg aan die artikel van die lemmateken *gat*¹ is hulle nie aan daardie lemma geadresseer nie. Trouens, hulle het geen adres nie, en dit is 'n tipiese kenmerk van makrostruktuuraanduiders. Daarbenewens tree die vaste uitdrukkings ook nie op as mikrostruktuuraanduiders wat aangebied word as deel van die behandeling van die gegewe lemmateken nie. Hulle insluiting in die nabyheid van die spesifieke lemmateken is slegs om praktiese en toegangsredes met die lemmateken *gat*¹ waarvan die waarde met betrekking tot die kluster vaste uitdrukkings verminder word tot 'n verwyderde artikel-eksterne gidselement wat nodig is vir 'n deel van die eksterne toegangsproses tot die vaste uitdrukkings. Hierdie niegegroepeerde nes vaste uitdrukkings vorm 'n deelartikeltrajek, om meer presies te wees, 'n deelartikeltrajek van geneste artikels met 'n verwyderde deelartikeltrajek-eksterne primêre gidselement, en die nes word ingelei deur 'n lemma-eksterne aanduidertipe-identifiserende struktuurmerker, naamlik die inskrywing "UITDR." Elke vaste uitdrukking in die kluster tree as gidselement van sy eie artikel op. Aangesien die lemma van die voorafgaande hoofartikel die eerste of primêre gidselement is op grond daarvan dat toegang tot elke vaste uitdrukking via daardie gidselement moet geskied, verteenwoordig die vaste uitdrukkings as lemmata sekondêre gidselemente. Daar kan dus geargumenteer word dat die vaste uitdrukkings as makrostruktuuraanduiders aangebied word maar wel as sublemmata omdat toegang tot hierdie lemmata afhanklik is van die voorafgaande hooflemma as primêre gidselement. Dit is belangrik om daarop te let dat die aanbieding as 'n sublemma nie 'n mindere makrostrukturele status of 'n beperkter leksikografiese behandeling as ten opsigte van 'n hooflemma verstrekk, impliseer nie.

Die manier waarop die vaste uitdrukkings in die hersiene uittreksel uit HAT aangebied is, gee erkenning aan hulle status as volwaardige leksikale items en aan hulle optrede as makrostrukturele behandelingseenhede. Hierdie status kwalifiseer vaste uitdrukkings vir 'n uitvoerige mikrostrukturele behandeling.

4. Die behandeling van vaste uitdrukkings

4.1 Verskillende leksikografiese funksies

Die behandeling van enige lemma in 'n woordeboek moet 'n poging wees om die leksikografiese funksies van daardie woordeboek te help bevredig. Daarom moet die datatipes en die aard en omvang van hulle aanbieding in enige woor-

deboek bepaal word deur die leksikografiese funksies van daardie spesifieke woordeboek. In 'n mono- of polifunksionele woordeboek moet daar 'n parallel bestaan tussen die behandeling van vaste uitdrukkings en ander lemmata. Die verstekdataverspreidingsprogram mag anders wees as gevolg daarvan dat vaste uitdrukkings nie noodwendig al dieselfde aanduidertipes in hulle behandeling nodig het as wat in die behandeling van enkelwoordlemmata die geval is nie. 'n Aanduider wat die uitspraak aandui, is byvoorbeeld dikwels nie nodig in die behandeling van 'n vaste uitdrukking nie alhoewel dit as deel van die verpligte mikrostruktuur van die artikels van enkelwoordlemmata beskou mag word. Die behandeling van vaste uitdrukkings moet so volledig as moontlik wees met betrekking tot die datakategorieë wat nodig is vir 'n optimale ontsluiting van inligting oor die betrokke vaste uitdrukking en vir 'n bevrediging van die tersaaklike leksikografiese funksies.

In algemene woordeboeke is die behandeling van vaste uitdrukkings van veel beperkter en minderwaardiger gehalte as dié van enkelwoordlemmata. In eentalige verklarende woordeboeke kry 'n mens dikwels 'n kort betekenisparafraze en in tweetalige woordeboeke óf 'n vaste uitdrukking in die doeltaal wat as vertaalekwivalent aangebied word óf 'n betekenisparafraze in gevalle waar daar geen ekwivalente vaste uitdrukking in die doeltaal bestaan nie. Dit verteenwoordig dikwels die volle omvang van die behandeling.

Ander aanduidertipes is ook nodig om in die behoeftes van die tipiese woordeboekgebruiker te voorsien. Waar 'n verklarende woordeboek 'n kognitiewe funksie het, bevat die artikels dikwels ook 'n aanduiding wat die etimologie gee van die woord wat deur die lemmateken verteenwoordig word. Dit is 'n datatipe wat ongelukkig nie dikwels genoeg gevind word in die behandeling van vaste uitdrukkings in algemene verklarende woordeboeke nie. Gebruikers raadpleeg dikwels 'n woordeboek vanweë 'n kognitiewe behoefte met betrekking tot vaste uitdrukkings en hulle is dikwels geïnteresseerd in die herkoms van die betrokke vaste uitdrukking. Nogmaals is leksikografiese konsekwentheid nodig. Wat vir die enkelwoordlemma geld, behoort ook waar nodig die vaste uitdrukking te geld. Gegee die behoefte aan etimologiese leiding by vaste uitdrukkings, soos blyk uit die navrae van woordeboekgebruikers aan uitgewers en leksikograwe, sou so 'n aanduiding van etimologie 'n vaste inskrywing in die verstekbehandeling van vaste uitdrukkings kon word; al is dit nie deel van die verstekdata-oordrag in ander lemmatipes se artikels nie.

4.2 Kommunikatiewe funksie

Die kommunikatiewe funksie sluit drie subfunksies in, te wete teksresepsie, teksproduksie en vertaling, vergelyk Tarp (2008). Waar teksproduksie die geldende funksie is, het gebruikers redelik uitvoerige leiding nodig, byvoorbeeld met behulp van koteks- en konteksinskrywings. Die konteksinskrywings behoort 'n wye verskeidenheid stilistiese etikette in te sluit en die kotekstuele inskrywings moet aanduiders bied wat tipiese gebruiksvoorbeelde van die aan-

duider wat deur die lemma voorgestel word, weergee. Sowel een- as tweetalige woordeboeke laat te dikwels na om konteks- en koteksaanduiders te gee in die behandeling van vaste uitdrukkings. Dit is moontlik omdat vaste uitdrukkings veral as mikrostrukturele aanduiders optree en spesifiek as sou hulle kotekstuele leiding bied met betrekking tot die woord wat deur die eenwoordlemma-teken verteenwoordig word.

As meerwoordige leksikale items bevat vaste uitdrukkings sekere verpligte elemente maar soms kan hulle ook opsionele elemente bevat, vergelyk Combrink (1989). In die geval van teksproduksie as funksie is dit belangrik dat woordeboekgebruikers daarvan bewus gemaak moet word of 'n gegewe vaste uitdrukking of sy vertaalekwivalent ook opsionele elemente bevat. Dit raak byvoorbeeld ook die optrede of afwesigheid van 'n ontkenningvorm in 'n vaste uitdrukking. Dit is naamlik belangrik om te weet of 'n bepaalde vaste uitdrukking 'n negatiewe al dan 'n positiewe sinswaarde het, vergelyk Ponelis (1979). Vergelyk die volgende Afrikaanse vaste uitdrukkings:

- 1 *die/n bobbejaan agter die bult gaan haal*
- 2 *iets nie oor die hart kry nie*
- 3 *nie al jou kalwers in die hok hê nie*
- 4 *wie nie hoor nie moet voel.*

Die eerste voorbeeld het 'n positiewe sinswaarde maar die vaste uitdrukking kan ook genegatiewer word en steeds as vaste uitdrukking gebruik word, byvoorbeeld *nie die bobbejaan agter die bult gaan haal nie*. Die tweede voorbeeld het 'n negatiewe sinswaarde wat as sodanig gebruik word. Dit het normaalweg nie 'n variant met 'n positiewe sinswaarde nie, behalwe wanneer die vaste uitdrukking in 'n sin deur sekere bywoorde voorafgegaan word. Vergelyk in hierdie verband die volgende voorbeelde:

- *? *Ek kon dit oor my hart kry om die hondjie weg te gee.*
Ek kon dit beswaarlik oor my hart kry om die hondjie weg te gee.

Voorbeeld 3 het 'n negatiewe sinswaarde en word nooit positief gebruik nie. In normale Afrikaans kry 'n mens dus nie 'n sin soos die volgende nie:

- **Jy kan staatmaak op wat hy sê want hy het al sy kalwers in die hok.*

Voorbeeld 4 bevat sowel 'n negatiewe as 'n positiewe komponent maar ook hier is geen verandering van sinswaarde moontlik nie.

Hierdie tipe inligting is nuttig wanneer dit om teksproduksiehulp gaan. Gebruiksvoorbeelde waarin die vaste uitdrukkings optree, is waardevolle leksikografiese leiding. Die leksikograaf kan van die standpunt uitgaan dat dit die verstekwaarde is dat 'n positiewe sinswaarde ook genegatiewer kan word en andersom. Waar vaste uitdrukkings aan hierdie verstekwaarde voldoen, hoef dit nie eksplisiet gestel te word nie, maar daar sou 'n voorbeeld met sowel die positiewe as die negatiewe sinswaarde gegee kon word. Waar daar wel beper-

kings ten opsigte van die sinswaarde bestaan en daarom met betrekking tot die aard van die sintaktiese patroon van 'n vaste uitdrukking, behoort dusdanige leiding in 'n woordeboek se behandeling aangebied te word. Voorbeeldsinne kan help, maar die leksikograaf sou ook 'n gebruiksnote, as artikel-interne binneteks, kon verstrek om die nodige leiding aan gebruikers te verskaf.

4.3 Betekenisverhoudinge

Gebruikers raadpleeg dikwels 'n woordeboek om inligting oor die betekenis van 'n gegewe lemma te onttrek. In 'n eentalige verklarende woordeboek vind die oordrag van betekenisleiding primêr deur middel van die aanduiders van die betekenisparafraze plaas. Bykomende aanduiders wat ook 'n rol in hierdie verband kan speel, is onder meer aanduiders van sinonieme en antonieme ter verstrekking van bepaalde betekenisverhoudings waarin die leksikale item wat deur die lemma verteenwoordig word, kan optree. In die leksikale semantiek is dit reeds duidelik gestel dat absolute sinonieme heel skaars is. In die meeste gevalle waar daar sprake is van sinonimie gaan dit om verhoudings van gedeeltelike sinonimie. Gedeeltelike sinonieme kan mekaar in sommige gebruiksomgewings vervang maar verskil van mekaar in terme van hulle onderskeie polisemiese paradigmas, hulle styl, register en ook gebruiksoptrede. Sinonimie is nie 'n verhouding wat slegs tussen woorde geld nie. Dit geld tussen leksikale items en daarom kan vaste uitdrukkings ook deel hê aan sinonimiese verhoudings. Sinonimie geld ook oor die grense van verskillende tipes leksikale items heen. Vaste uitdrukkings kan daarom sowel woorde as ander meerwoordige leksikale items as sinonieme hê.

In die artikels van enkelwoordlemmata word sinonieme dikwels aangedui en wel met die doel om as teksproduksiehulp die gebruiker die moontlikheid tot 'n meer gevarieerde taalgebruik te bied. Waar vaste uitdrukkings in 'n woordeboek met 'n teksproduksiefunksie ingesluit word, behoort die gebruiker soortgelyke hulp van die woordeboek te kan verwag as wat in die artikels van enkelwoordlemmata gevind word.

Gedeeltelike en absolute sinonieme, sowel woorde as ander vaste uitdrukkings, behoort as deel van die behandeling van die vaste uitdrukking gegee te word, maar verskille in styl, register en gebruik asook beperkings in die betekenisoorvleueling moet duidelik aangetoon word.

Afrikaans het 'n verskeidenheid vaste uitdrukkings met die betekenis "om dood te gaan". Hierdie uitdrukkings strek vanaf die formele *die tydelike met die ewige verwissel* tot die hoogs informele *om die lepel in die dak te steek*. In die behandeling van hierdie vaste uitdrukkings moet daar onder meer 'n betekenisparafraze verstrek word maar dan ook een of meer sinonimiese vaste uitdrukkings. Elke vaste uitdrukking met hierdie betekenis hoef nie telkens in die behandeling van elk gelys te word nie. Een vaste uitdrukking kan op grond van gebruiksfrekwensie gekies word om 'n volledige lysing van die sinoniemiese vaste uitdrukkings te bevat. In die artikels van die ander vaste uitdruk-

kings kan daar dan telkens 'n kruisverwysing na hierdie lemma gegee word. Meer data as die blote kruisverwysingsmerker is egter nodig om gebruikers bewus te maak van die tersaaklike styl- en registerbeperkings. Hier kan die benutting van 'n stelsel van leksikografiese etikettering 'n belangrike rol speel.

Waar 'n woordeboek teksproduksie as funksie het, kan dit ook belangrik wees dat gebruikers ingelig moet word oor woorde en vaste uitdrukkings, ook gedeeltelike sinonieme, waarmee 'n bepaalde vorm maklik verwar kan word. Die behandeling wat in 'n artikel gebied word, behoort in sulke gevalle nie tot 'n geïsoleerde aanduiding van sinonieme beperk te wees nie. Afrikaans het die vaste uitdrukkings *nie 'n vinger verroer nie* en *nie 'n voet versit nie*. Hierdie twee vaste uitdrukkings is gedeeltelike sinonieme met die betekenis "om niks te doen nie". Albei vaste uitdrukkings behoort op hulle onderskeie alfabetiese plekke in die woordeboek opgeneem te word. 'n Aanduider van die betekenisparafraze van hierdie vaste uitdrukkings soos "om niks te doen nie" wat heel dikwels in Afrikaanse woordeboeke gevind word, is nie voldoende om suksesvolle prosedures van teksproduksie te verseker nie. Voorbeeldmateriaal wat as koteksaanduiders verstrekkend word, is byvoorbeeld nodig maar sulke aanduiders sal nie noodwendig voldoende wees om gebruikers bewus te maak van die geldende gebruikbeperkings nie. Bykomende leksikografiese data is nodig. Dit kan byvoorbeeld gedoen word deur 'n aanduider, geadresseer aan die gedeeltelike sinoniem, wat die gebruikbeperking van die gegewe sinoniem met betrekking tot die vaste uitdrukking waaraan die gedeeltelike sinoniem geadresseer is, verstrekkend. Alhoewel albei die genoemde vaste uitdrukkings die betekenis "om niks te doen nie" het, is die aard van die "niks" nie dieselfde nie. In *nie 'n vinger verroer nie* verwys die "niks" tipieserwys na 'n vorm van hulp of bystand, byvoorbeeld:

Ek moes die werk alleen doen want hy het nie 'n vinger verroer om my te help nie.

In *nie 'n voet versit nie* verwys die "niks" eerder na fisiese beweging, byvoorbeeld:

Sy was die hele tyd in haar kamer en het nooit 'n voet versit nie.

Die insluiting van sinonieme as aanduiders in die behandeling van vaste uitdrukkings verhoog die gehalte van die mikrostrukturele dekking en dit lei tot 'n beter balans tussen die leksikografiese behandeling van vaste uitdrukkings en enkelwoordlemmata en dra ook by tot 'n omvattender uitvoering van die leksikografiese funksies wat vir 'n bepaalde woordeboek vasgestel is. Aanduiders van sinonieme behoort ook in die behandeling van vaste uitdrukkings in tweetalige woordeboeke verstrekkend te word.

4.4 Kultuurverskille

Waar die taalgemeenskappe van die taalpaar van 'n tweetalige woordeboek tot

verskillende kulture behoort, kan dit dikwels tot probleme lei in die behandeling van vaste uitdrukkings. Vir 'n gegewe betekenis kan 'n spesifieke taal 'n vaste uitdrukking hê wat 'n blote vertaling is van 'n vaste uitdrukking wat ook in verskeie ander tale, insluitende die tweede taal van die betrokke woordeboek, gebruik word. Die spesifieke taal kan egter ook 'n kultuurspesifieke vaste uitdrukking hê met dieselfde betekenis as die vaste uitdrukking wat 'n vertaling uit ander tale is. In 'n veeltalige land soos Suid-Afrika kan dit dikwels die geval wees. Tweetalige woordeboeke behoort sowel die algemene as die kultuurspesifieke uitdrukking te bevat met 'n duidelike aanduiding dat die een 'n kultuurspesifieke beperking het en dus nie in die tweede taal vertaal moet word nie. Vir die bekende Engelse vaste uitdrukking *the tip of the iceberg* is daar 'n Afrikaanse ekwivalent, naamlik die vertaling *die puntjie van die ysberg*. Hierdie ekwivalent behoort in 'n tweetalige woordeboek met Engels en Afrikaans as taalpaar verstrek te word in die behandeling van die gegewe Engelse vaste uitdrukking. Ysberge is nie so bekend in Suid-Afrika nie; seekoeie wel. Die betekenis van *die puntjie van die ysberg* kan in Afrikaans ook geleksikaliseer word as die vaste uitdrukking *die oortjies van die seekoei*. In die behandeling van *the tip of the iceberg* moet hierdie vaste uitdrukking ook verstrek word, maar dit moet geëtiketteer word as 'n kultuurspesifieke leksikale item.

Nie net die kommunikatiewe funksie nie maar ook die kognitiewe funksie van tweetalige woordeboeke vereis 'n uitvoeriger behandeling van vaste uitdrukkings. In die behandeling van kultuurspesifieke leksikale items, in hulle aanbieding as hetsy enkelwoord- hetsy meerwoordige lemmata, is bykomende inskrywings van groot waarde om die kulturele inhoud van sulke items aan te toon. Die gebruik van artikel-interne ingevoegde binnetekste of van tekskassies kan help met die vereenduidiging van kultuurgebonde vaste uitdrukkings in die brontaal. Xhosa het die vaste uitdrukking *yaqhawuk'imbeleko* wat letterlik vertaal kan word as "die riem het gebreek", met die betekenis dat daar 'n versteuring in of verbreking van 'n verhouding was of dat daar 'n skeiding was tussen mense of dinge wat met mekaar vereenselwig word. Afrikaans het geen vaste uitdrukking as ekwivalent vir hierdie Xhosa uitdrukking nie en die behandeling hiervan in 'n tweetalige woordeboek sal deur middel van 'n betekenisparafrase wat as surrogaatekwivalentaanduider aangebied word, moet wees. Vir teksresepsiedoeleindes behoort dit voldoende te wees. Waar die woordeboek egter (ook) 'n kognitiewe funksie het, sal 'n blote surrogaatekwivalent nie verseker dat die werklike doel van die woordeboek bereik word nie. In die tradisionele Xhosa-gemeenskap dra vrouens dikwels hulle babas op die rug en die baba word aan die vrou of aan haar rok vasgebind. Wanneer hierdie riem waarmee die baba aan haar vasgebind is per ongeluk breek, word 'n baie intieme verhouding versteur of gebreek. Om die omvang van die emosionele waarde te begryp wat deur die betrokke Xhosa- vaste uitdrukking geleksikaliseer word, is meer leiding nodig as dit wat deur 'n surrogaatekwivalent wat maar 'n beperkte betekenisparafrase gee, gebied word. 'n Tekskassie of 'n artikel-interne ingevoegde binneteks kan gebruik word om data oor die kulturele

waarde van hierdie vaste uitdrukking oor te dra. Indien die betrokke woordeboek 'n raamstruktuur vertoon, sou 'n alternatiewe moontlikheid wees om in die artikel van die vaste uitdrukking, naas die surrogaatekwivalent, 'n kruisverwysing te gee wat die gebruiker lei na 'n buiteteks in die agtertekste-afdeling van die woordeboek waar 'n uitvoeriger bespreking van alle kultuurgebonde leksikale items van albei die behandelde tale verstrekkend word. So 'n behandeling in 'n buiteteks vestig 'n transtekstuele benadering tot leksikografiese funksies, vergelyk Gouws en Steyn (2005), en bevestig 'n uitputtende toepassing van die benadering tot leksikografiese funksies wat vir die betrokke woordeboek vasgestel is. Deur sowel vaste uitdrukkinge as ander lemmata van die sentrale teks in 'n enkele tematies-gemotiveerde buiteteks te betrek, bewys nogmaals die status wat in die woordeboek aan vaste uitdrukkinge verleen word.

'n Tradisionele aanbieding van vaste uitdrukkinge as mikrostrukturele aanduiders lei normaalweg nie tot 'n uitvoerige behandeling nie. Deur vaste uitdrukkinge as makrostrukturele aanduiders te erken en hulle artikels as kandidate vir 'n gevarieerde artikel- en mikrostruktuur te beskou, word die moontlikheid geskep vir 'n verteenwoordigende en bevredigende leksikografiese behandeling en dit bring mee dat die woordeboek 'n geldige weerspieëling van 'n groter deel van die leksikon bied wat in sy teikengebied val.

5. Ter afsluiting

As makrostruktuuraanduiders en volwaardige behandelingseenhede verg vaste uitdrukkinge 'n nuwe oordeel oor hulle artikelstruktuur asook oor die mikrostrukturele program en dataverspreidingstruktuur van die betrokke woordeboek. Die inhoud, struktuur en funksies wat vir artikels met enkelwoordlemmata beplan word en as die verstekaanbieding van die woordeboek dien, moet sistematies en volledig toegepas word en ook op subwoordelike en meerwoordige lemmata, insluitend vaste uitdrukkinge, gerig word. Leksikografe moet besef dat vaste uitdrukkinge volwaardige behandelingseenhede is en nie as deel van die behandeling van 'n woordlemma aangebied moet word nie.

Die keuse, aanbieding en behandeling van vaste uitdrukkinge stel leksikografe voor 'n wye reeks uitdagings. Die kwessies wat in hierdie artikel bespreek is, is slegs die oortjies van die idiomatiese seekoei.

Acknowledgement

This work is based on the research supported in part by the National Research Foundation of South Africa (Grant specific unique reference number (VID) 85434).

The Grantholder acknowledges that opinions, findings and conclusions or recommendations expressed in any publication generated by the NRF sup-

ported research are those of the author, and that the NRF accepts no liability whatsoever in this regard.

Bibliografie

- Bergenholtz, H., T.J.D. Bothma en R.H. Gouws.** 2011. A Model for Integrated Dictionaries of Fixed Expressions. Kozen, I. en K. Kosen (Reds.). 2011. *Electronic Lexicography in the 21st Century: New Applications for New Users*: 34-42. Ljubljana: Trojina, Institute for Applied Slovene Studies.
- Bergenholtz, H., S. Tarp en H.E. Wiegand.** 1999. Datendistributionsstrukturen, Makro- und Mikrostrukturen in neueren Fachwörterbüchern. Hoffman, 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, Bd./ Vol. 2*: 1762-1832. Berlyn: Walter de Gruyter.
- Combrink, J.G.H.** 1989. Wesenlike elemente, ekstras en vulsels in die aangawe van idiome in tweetalige woordeboeke. Botha, T.J.R. (Red.). 1989. *Leksikografie en leksikologie*: 55-77. Menlopark: Serva Uitgewers.
- Eksteen, L.C. et al. (Reds.).** 1997¹⁴. *Groot Woordeboek Afrikaans-Engels/Engels-Afrikaans / Major Dictionary Afrikaans-English/English-Afrikaans*. Kaapstad: Pharos.
- Gouws, L.** 2006. *Die behandeling van vaste uitdrukkings in tweetalige woordeboeke: 'n Hulp vir vertalers*. Ongepubliseerde M.Phil.-skripsie. Stellenbosch: Universiteit van Stellenbosch.
- Gouws, R.H.** 1989. *Leksikografie*. Pretoria/Kaapstad: Academica.
- Gouws, R.H.** 2005. Meilensteine auf dem historischen Weg der Metalexikographie. *Lexicographica* 21: 158-178.
- Gouws, R.H.** 2010. Fixed Word Combinations as Second Level Treatment Units in Dictionaries. Ďurčo, P. (Red.). 2010. *Feste Wortverbindungen und Lexikographie*: 51-63. Berlyn: Walter de Gruyter.
- Gouws, R.H. en D.J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.
- Gouws, R.H. en M. Steyn.** 2005. Integrated Outer Texts: A Transtextual Approach to Lexicographic Functions. Barz, I., H. Bergenholtz en J. Korhonen (Reds.). 2005. *Schreiben, Verstehen, Übersetzen und Lernen: Zu ein- und zweisprachigen Wörterbüchern mit Deutsch*: 127-136. Frankfurt a.M./Bern/New York/Parys: Peter Lang.
- Hanks, P. (Red.).** 1985². *Collins Dictionary of the English Language*. London/Glasgow: Collins.
- Odendal, F.F. en R.H. Gouws (Reds.).** 2000⁴. *Verklarende Handwoordeboek van die Afrikaanse Taal*. Midrand: Perskor.
- Pearsall, J. (Ed.).** 1998. *The New Oxford Dictionary of English*. Oxford: Clarendon Press.
- Ponelis, F.A.** 1979. *Afrikaanse sintaksis*. Pretoria: Van Schaik.
- Tarp, Sven.** 2008. *Lexicography in the Borderland between Knowledge and Non-knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Tübingen: Max Niemeyer.
- Tweetalige Skoolwoordeboek/Bilingual School Dictionary*. 2012. Kaapstad: Pharos Woordeboeke.

Das lexikographische Beispiel in der Parömiographie. Formen und Funktionen

Vida Jesenšek, *Abteilung für Germanistik, Philosophische
Fakultät, Universität Maribor, Maribor, Slowenien (vida.jesensek@siol.net)*

Zusammenfassung: Der vorliegende Beitrag bespricht das lexikographische Beispiel im Bereich der Parömiographie (Sprichwortlexikographie). Im Fokus stehen Kriterien zur Identifizierung, Festlegung und Selektion jener kontextueller Sprachdaten, die in der Funktion des lexikographischen Beispiels den lexikographischen Prozess in der modernen Parömiographie systematisch und zweckgemäß unterstützen können. Es wird die Idee verfolgt, dass das lexikographische Beispiel in der Zeit der korpusempirischen Lexikographie, die meist auch elektronische lexikographische Produkte hervorbringt, seinen Status und damit seine Funktionalität geändert hat. Von der traditionellen primär dokumentierend-illustrativen Funktion und Unterstützung der individuellen Sprachkompetenz des Lexikographen wird es zum Ausgangs- und Mittelpunkt des lexikographischen Arbeitsprozesses. Es wird zur empirischen Basis für die lexikographische Beschreibung sprachlicher Phänomene und kann erst daraufhin in seiner tradierten illustrativ-dokumentierenden Funktion mikrostrukturell integriert werden. Interdisziplinär und unter Einbeziehung der phraseologie- und parömiologietheoretischen Kenntnisse zur Semantik, Pragmatik und Grammatik der Sprichwörter werden sodann Aussagen zu Qualitätsmerkmalen von Textstellen ausgearbeitet, anhand deren potentielle lexikographische Beispiele im Bereich der Parömiographie identifiziert, systematisch ausgewertet und selektiert werden können. Abschließend werden die Akzeptabilität und Operationalisierbarkeit der festgelegten Qualitätsmerkmale diskutiert sowie einige weiterführende Forschungsfragen angesprochen. Die Überlegungen gründen auf den Erfahrungen bei der Erarbeitung eines mehrsprachigen parömiographischen Produktes, welches als Dokumentation des aktuellen Sprichwortgebrauchs und zugleich als Lern- und Lehrmaterial in fremdsprachlichen Lernzusammenhängen konzipiert und erarbeitet wurde (SprichWort 2010). Der Artikel will somit zu einer seit 1977 geforderten (vgl. Wiegand 1977) und bisher im Bereich der Parömiographie noch nicht realisierten Ausarbeitung einer Theorie des lexikographischen Beispiels beitragen.

Stichwörter: LEXIKOGRAPHISCHES BEISPIEL, BELEG, BELEGBEISPIEL, LEXIKOGRAPHIE, PHRASEOGRAPHIE, PARÖMIOGRAPHIE, SPRICHWORTLEXIKOGRAPHIE, KORPUSLEXIKOGRAPHIE, METHODOLOGIE, METHODIK, FUNKTIONEN, SPRICHWORT, IDENTIFIZIERUNG, SELEKTION

Abstract: The Lexicographic Example in Paremiography. Forms and Functions. This paper discusses the lexicographic example within the scope of paremiography (proverb lexicography). The focus is on criteria for the identification, determination and selection of those contextual language data that can, in the functioning of the lexicographic example, systematically and appropriately support the lexicographic process in modern paremiography. The idea is pur-

sued that during the time of corpus empirical lexicography, which mostly also yielded electronic products, its status and thus its functionality has been altered. From the traditional primarily documentary illustrative function and support of the individual competence of the lexicographer, it became the starting- and central point of the lexicographic working process. It became the empiric base for the description of linguistic phenomena and can only then be microstructurally integrated into its traditional illustrative documentary function. Interdisciplinarily and through inclusion of the phraseological and paremiographical theoretical knowledge of semantics, pragmatics and grammar of the proverbs, assertions are then developed on the quality characteristics of text passages with the help of which potential lexicographic examples within the scope of paremiography can be identified, systematically evaluated and selected. Finally the acceptability and operability of the determined quality characteristics are discussed as well as some further research questions addressed. The considerations are based on the experiences from the development of a multilinguistic paremiographical product, that was conceptualized and developed as documentation of the actual proverbial use and also as learning and teaching material in foreign language learning contexts (SprichWort 2010). The article will therefore contribute to the development of a theory of the lexicographic example, since 1977 called for (cf. Wiegand 1977) and until now not yet realized within the scope of paremiography.

Keywords: LEXICOGRAPHIC EXAMPLE, CITATION, CITATION EXAMPLE, LEXICOGRAPHY, PHRASEOGRAPHY, PAREMIOGRAPHY, PROVERB LEXICOGRAPHY, CORPUS LEXICOGRAPHY, METHODOLOGY, METHODICS, FUNCTIONS, PROVERB, IDENTIFICATION, SELECTION

1. Einleitung

Verfolgt man die Geschichte der praktischen Lexikographie, so sieht man schnell, dass lexikographische Beispiele (LB) in Prozessen der Wörterbucherstellung seit immer eine bedeutende Rolle gespielt haben. Im Allgemeinen spricht man von Belegen, die entweder als Belegbeispiele (vgl. u.a. Haß 1991a) oder als Textbelege (vgl. u.a. Schläefer 1990) verstanden werden. Aus historischer Sicht steht die Anwendung von Belegen in der Funktion des LB in unterschiedlichen Traditionen, anhand deren sich ihre jeweils dominante Funktion abspiegelt.¹ Folgerichtig besteht die Funktionalität des LB im Wesentlichen darin, die Existenz eines Lemma-Zeichens zu begründen und seine semantischen, grammatischen, pragmatischen und sonstigen Charakteristika zu illustrieren und dokumentieren. Die Feststellung gilt für allgemeine wie auch für spezielle Wörterbücher und dementsprechend kann auch die historische Parömiographie (Sprichwortlexikographie) in vielen Fällen eine reiche Belegung einzelner Sprichwörter nachweisen (vgl. Mieder 1989: 1034).

Umso überraschender ist die Beobachtung, dass Studien zur Begrifflichkeit und Funktionalität des LB im lexikographietheoretischen Forschungsfeld bis vor kurzem vergleichsweise selten vorkamen (vgl. u.a. Zöfgen 1994).² Eine intensivere Erforschung dieser »lexikographischen Selbstverständlichkeit« ist erst in der letzten Zeit zugegen; das verstärkte Forschungsinteresse kann wohl

als eine unausweichliche Folge der gegenwärtigen korpusempirischen Lexikographie verstanden werden. Dabei rückt in den Vordergrund die sog. „Urbild“-Tradition des LB (Beleg als Ausgangspunkt der Bedeutungsbeschreibung laut Haß 1991a) und das LB wird zum zentralen Ausgangspunkt der lexikographischen Arbeitsprozesse erklärt (vgl. u.a. Āurčo 2006, Hein 2012, Jesenšek 2009 und 2010, Steyer 2010 und 2012, Dräger 2011, Klosa, Töpel und Koplenig 2012). Diese zentrale Position gilt umso mehr für die lexikographische Bearbeitung von festen Wortverbindungen unterschiedlicher Ausprägung, darunter auch für Sprichwörter (SW). Werden SW als spezifische, formal und semantisch komplexe Phänomene, als eigenständige Wortschatz- und/oder Texteinheiten angesehen, so sind sie ohne eine systematisch-analytische Beobachtung und Auswertung ihres textuell und situativ eingebetteten Gebrauchs nur ansatzweise (wenn überhaupt) lexikographisch beschreibbar. Die korpusempirische Parömiographie zielt aber nicht nur darauf, SW anhand der empirischen Datenbasis ausführlich beschreiben zu können; sie hat primär in Aussicht, durch die Auswertung empirischer Sprachdaten „verdeckte Strukturen /zu identifizieren/, die sich dem individuellen Wahrnehmungshorizont bis dato verschlossen haben“ (Steyer 2012: 287). Anhand einer repräsentativen empirischen Datenbasis kann einerseits bekanntes Wissen über die SW nachgeprüft werden, andererseits — und das wird hier als entscheidender Mehrwert betrachtet — wird dadurch ermöglicht, das gesamte Variations- und Transformationspotenzial von SW zu ermitteln und zudem neue SW zu identifizieren (Steyer 2012: 288f.).³ Auswirkungen auf die Auffassung des LB-Status im lexikographischen Arbeitsprozess liegen somit auf der Hand: korpusermittelte Textausschnitte, die SW enthalten, werden zur Grundlage und zum Ausgangspunkt deren lexikographischen Erfassung und fungieren sodann eventuell auch als LB, die die jeweils vorgesehenen Mikrostrukturen systematisch unterstützen vermögen. Somit rücken zwei wesentliche Fragen in den Vordergrund: (1) welche Qualitäten sollen Textausschnitte nachweisen, damit sie in beiden Funktionen in parömiographischen Arbeitsprozessen adäquat auftreten können und (2) welche ist die Methodik, wonach entsprechende lexikographisch relevante Textausschnitte in der Fülle der verfügbaren textuellen Korpusdaten identifiziert, ausgewertet und ausgewählt werden sollen. Nachdem in Kap. 2 die Begrifflichkeit und Funktionalität des LB in Kürze dargelegt werden, wird in Kap. 3 auf beide Fragestellungen näher eingegangen. Die Überlegungen gehen auf die empirischen Sprachdaten aus dem deutschen Teil der mehrsprachigen parömiologischen Datenbank namens SprichWort (2010) zurück und wollen exemplarisch verstanden werden.⁴

2. Zum Begriff des lexikographischen Beispiels

In diesem Kapitel wird das LB in seiner Begrifflichkeit kurz besprochen. Während in Abschnitt 2.1 definitorische und funktionale Aspekte im Allgemeinen erörtert werden, stehen in Abschnitt 2.2 Formen und Funktionen des

LB in der Parömiographie (und Phraseographie) im Vordergrund der Betrachtung.

2.1 Definition und Funktionen

Generalisierende und weit gefasste Definitionen des LB findet man in allen relevanten Referenzwerken zur Lexikographie; stellvertretend dazu sei die Definition aus Engelberg und Lemnitzer (2008: 230) genannt:

Das l. B. /lexikographische Beispiel/, auch Gebrauchsbeispiel oder Verwendungsbeispiel genannt, ist ein authentischer Beleg oder ein auf die Sprachkompetenz des Lexikographen gestütztes Syntagma, welches ein sprachliches Phänomen oder eine Verwendungsweise eines sprachlichen Zeichens illustriert.

Hervorgehoben werden zwei Aspekte des LB: seine demonstrierend-illustrierende Funktionalität und gleichermaßen seine vorhandene oder fehlende Authentizität. Zöfgen (1994) definiert das LB vor dem Hintergrund der Lernlexikographie und betont seine semantikbezogenen Dimensionen; vom LB will er nur dann reden, »wenn es sich um objektsprachliche, eindeutig auf der *parole*-Ebene angesiedelte Äußerungen handelt, bei denen das monosemierte Lemma-Zeichen in einem semantisch relevanten Kontext gezeigt wird« (Zöfgen 1994: 184).⁵

Weitere Definitionspräzisierungen bringen lexikographietheoretische Arbeiten zu lexikographischen Mikrostrukturen hervor. Nach Wiegand (1989) lassen sich LB an den Elementen der Mikrostruktur mit dem Angabenstatus⁶ erschließen. Bezogen auf die mikrostrukturelle Beschaffenheit traditioneller Printwörterbuchformate werden mehrere Klassen von solchen Angaben festgelegt und definiert. Deutlich ist eine begriffliche Unterscheidung zwischen Beispiel und Beleg: mit den *Beispielangaben* werden »Kotexte (meistens für das Lemmazeichen und seine Formen) genannt«, während *Belegangaben* zu den wörterbuchbasisbezogenen Angaben gehören (Wiegand 2006: 98). Fungieren Belegangaben als solche, anhand deren LB erschließbar sind, so werden sie *Belegbeispielangaben* (auch *Belegtextangaben*) genannt. Der Aspekt der vorhandenen oder fehlenden Authentizität spiegelt sich in der Unterscheidung von *Belegbeispielangaben* und *Kompetenzbeispielangaben* ab; hierzu kommen auch *Corpusbeispielangaben*, die für weitere Ausführungen noch besonders vom Interesse sein werden.

Wie bereits angedeutet, findet man bis vom kurzem relativ wenige theoretisch ausgelegte Arbeiten zum LB. Auf die Forderung Wiegand's aus dem Jahr 1977, man müsste sich mit dem LB theoretisch und systematisch auseinandersetzen,⁷ wurde das Augenmerk hauptsächlich auf seine (lerndidaktische) Funktionalität und anschließend auf die Zusicherung seiner Qualität ausgerichtet.

Eine Theorie des LB weist hauptsächlich zwei Komponenten aus, eine deskriptive und eine normative Komponente (vgl. Wiegand 1977). Die deskriptive Theoriekomponente meint die Korrelation des LB zu anderen Elementen einer konkreten Wörterbuchartikel- oder Datenbankeintragsstruktur, wogegen

mit der normativen Theoriekomponente „bestimmt werden /soll/, was ein gutes lexikographisches Beispiel ist“ (Harras 1989: 607); man befragt also die wünschenswerten Eigenschaften von authentischen und/oder konstruierten Sprachdaten („Wünschbarkeiten“, um mit Hermanns (1988) und Harras (1989) zu sprechen), anhand deren sie an der lexikographischen Relevanz gewinnen.

Über die potentielle Funktionalität von LB wurde bisher verschiedentlich diskutiert. So bezieht sich Martin (1989) auf die LB-Funktionen in einsprachigen Wörterbüchern,⁸ während Zöfgen (1994) eine auf die pädagogischen Ziele ausgerichtete Funktionstypologie „des (Satz)beispiels“ mit zehn „idealtypischen“ Fällen erarbeitet.⁹ Funktional ausgerichtet sind auch Arbeiten von Nikula (1986) und Cowie (1989). Beide wollen die vorhandene oder fehlende Authentizität der LB vor dem Hintergrund der Wörterbuchfunktionen beurteilen. Während authentische Belege vor allem zur illustrierenden Demonstration des sprachlich Varianten bzw. Besonderen geeignet seien, dienen konstruierte LB eher zur Illustration des Prototypischen. Cowie (1989) will verschiedene LB-Funktionen darüber hinaus in Abhängigkeit von der aktiven bzw. passiven Ausgerichtetheit des jeweiligen Wörterbuchs verstanden wissen.

Somit wären wir bei der benutzerorientierten Beurteilung der LB-Funktionen, die auch in Hermanns (1988) stark zum Ausdruck kommt. Auf die Frage, wozu ein LB dient, wird aus der Benutzerperspektive geantwortet: ein LB sei „ein Modell, nach dem ich mich beim Sprechen und Verstehen richten kann“ (Hermanns 1988: 176), ein LB hat die Funktion, dass es „die mögliche Verwendung eines sprachlichen Zeichens zeigt“ (Hermanns 1988: 175). Die tradierte demonstrierend-illustrierende Funktion des LB — wonach dieses anderen Angaben, vor allem der Bedeutungsangabe hierarchisch untergeordnet ist — wird jedoch auch aus der umgekehrten Perspektive betrachtet, wenn er schreibt: „die Regel /Bedeutungsangabe/ dient dazu, das Beispiel verständlich zu machen“ (Hermanns 1988: 176).

Dadurch zeigt sich eine Reorientierung in der theoretischen Betrachtung des LB. Die gegenwärtig zum Standard gewordene Korpusbasiertheit der Lexikographie (vgl. u.a. Prinsloo 2009, Müller-Spitzer und Engelberg 2011) und der parallele Medienwechsel zur elektronischen und Internetlexikographie haben zur Folge, dass methodisch-methodologische Fragestellungen zur korpusempirischen Suche nach potentiellen LB zentral geworden sind. Im Fokus stehen Aspekte der korpusanalytischen Identifikation, Auswertung und Auswahl von kontextuellen Sprachdaten, die in lexikographischen Prozessen so einsetzbar sind, dass auf ihrer Basis ein Lemmazeichen (hier ein SW) ausführlich und zuverlässig bearbeitet werden kann und dass sie folgerichtig in der demonstrierend-illustrierenden Funktion als LB auftreten können (vgl. u.a. Klosa 2005, Āurčo 2006, Steyer 2004, 2010 und 2012, Dräger 2011). Vgl. dazu Dräger (2011: 5): »Der Beleg erhält damit einen neuen Stellenwert, indem er nicht mehr nur belegend aufgeführt wird, sondern zum Informationsgewinn beiträgt«, er gehört zum »Kernbereich eines Wörterbuchs«.¹⁰

Eine vorläufige Schlussfolgerung zur Begrifflichkeit und Funktionalität

des LB lässt sich nun folgendermaßen formulieren: LB sind Elemente lexikographischer Mikrostrukturen (bezogen auf Wörterbuchartikel traditioneller Wörterbuchformate und/oder auf Strukturen der Datenbankeinträge). Deren Funktionalität besteht traditionellerweise darin, system- und gebrauchszugehörige Eigenschaften eines Lemmazeichens zu *präsentieren* und zu *illustrieren* und dadurch seine in der Regel vielfältige sprachliche Existenz zu *begründen* und zu *dokumentieren*. Ermittelt werden die LB anhand der jeweiligen Wörterbuchbasis; neuerdings anhand der quantitativ-qualitativen analytischen Auswertung umfangreicher Textkorpora.¹¹ In der gegenwärtigen korpusempirischen Lexikographie gebührt dem LB allerdings eine zentrale Position, da es zum prinzipiellen Ausgangspunkt der lexikographischen Beschreibung eines sprachlichen Ausdrucks erklärt wird. Folgerichtig sind vorrangig Aspekte der korpusbasierten Sprachdatenermittlung zu klären, um eine operationalisierbare Methodik der Identifikation und Auswahl von lexikographisch relevanten Textausschnitten — potentiellen LB sicherzustellen.¹²

2.2 Das lexikographische Beispiel in der Parömiographie

Auch für die Parömiographie (und Phraseographie) gilt die Beobachtung, dass systematische Auseinandersetzungen mit dem LB bisher hauptsächlich vor dem methodisch-theoretischen Hintergrund der korpusempirischen Lexikographie entstanden sind. Das heißt jedoch nicht, dass Phraseographen die Relevanz des LB früher nicht erkannt und stellenweise auch nicht explizit betont hätten. Schemann (1989: 1025f.) hat sich dazu sehr kritisch geäußert:

Die Erkenntnis, daß zu einem idiomatischen Wörterbuch Beispiele gehören, hat sich /.../ gerade in den letzten Jahren zunehmend durchgesetzt. Allerdings sind sie oft /.../ das Gegenteil von dem, was ein Beispiel sein soll: eine Situierung des Idioms und seine gesamte Kontextkonstellation, in der seine semantische, seine pragmatische und seine stilistische Funktion als Einheit anschaulich greifbar wird.

In den wenigen (germanistischen) Arbeiten, die sich ausdrücklich mit der lexikographischen Behandlung von SW auseinandersetzen (vgl. Mieder 1984 und 1999, Grzybek 1992, Kispál 2000 und 2007, Kühn 2003, Ďurčo 2005) wird die Problematik des LB hauptsächlich aus der wörterbuchkritischen Perspektive behandelt (vgl. Mieder 1999, Kispál 2000),¹³ wogegen seine ausdrückliche Thematisierung im Rahmen der Sprichwortlexikographie m.W. noch aussteht. Eine informative Übersicht über die Parömiographie und ausgewählte sprichwortspezifische Beschreibungsaspekte gibt Kispál (2007). Erwähnt wird die Möglichkeit, dass durch Verwendungsbeispiele und Belege SW-Charakteristika dem Wörterbuchbenutzer nahegelegt werden können, festgehalten wird aber auch, dass dies in den Sprichwörterbüchern eher selten praktiziert wird.

Mit Recht sehen Ďurčo (2006) und Kispál (2007) in der empirischen Erforschung des (kon)textuellen Vorkommens der SW eine erforderliche Basis

für die Weiterentwicklung der Sprichwortlexikographie.¹⁴ Die Möglichkeit, von den korpusempirischen Sprachdaten auszugehen und dazu noch das Medium zu wechseln, bringt für die moderne Parömiographie entscheidende prozess- und produktbezogene Konsequenzen mit sich. Einerseits wird durch die Aufhebung der Platzbeschränkung eine komplexe, mehrdimensionale und von einer prinzipiell offenen Anzahl von LB unterstützte lexikographische SW-Präsentation ermöglicht; andererseits basiert die lexikographische Bearbeitung von SW auf kontextuellen korpusermittelten Sprachdaten, die nach einer systematischen quantitativ-qualitativen Auswertung als LB in lexikographische Mikrostrukturen integriert werden können. Der ausschlaggebende Aussagewert solcher Sprachdaten besteht darin, realitätsnahe Einblicke in den aktuellen Sprichwortgebrauch zu geben und somit eine adäquate und komplexe lexikographische Bearbeitung zu ermöglichen.¹⁵ Hein (2012) konnte an konkreten Beispielen einleuchtend nachweisen, welche Wechselverhältnisse sich im lexikographischen Prozess zwischen einem Textbeleg als Basis für die Ausformulierung einer Angabe und einem Textbeleg in der LB-Funktion herstellen, wenn der Lexikograph konsequent mit korpusermittelten Sprachdaten arbeitet.¹⁶ Obgleich auch mit Problemen konfrontiert, ermöglicht eine konsequente Arbeit mit Korpusdaten die Identifikation von inhaltlichen und formalen Charakteristika eines SW; der sog. „semantische Kern“ eines SW, seine konnotativ-pragmatischen Bedeutungsanteile samt illokutiv-kommunikativem Potenzial und ebenso strukturell-grammatische Eigenartigkeiten können weitgehend ermittelt werden. Es ist zu erwarten, dass dadurch die problematischsten aber auch die zentralen Aspekte der Phraseographie samt Parömiographie (i.e. semantische und pragmatische Angaben samt Beispielen und Belegen; vgl. dazu Kühn 2003 und Burger 2009) zufriedenstellend gedeckt werden können.

Im Zentrum weiterer Überlegungen steht somit die Fragestellung, welche Eigenschaften kontextuelle Korpusdaten in Form von Redeteilen bzw. Textteilen (im Folgenden werden sie *Textstellen* benannt) nachweisen müssen, um als LB in der praktischen Parömiographie adäquat und zweckmäßig fungieren zu können. Aus der Sicht der deskriptiven Theoriekomponente des LB lässt sich diese vorerst ziemlich allgemeine Fragestellung folgendermaßen genauer formulieren:

Welche Eigenschaften müssen Textstellen haben, damit sie in der Erarbeitung parömiographischer Produkte so anwendbar sind, dass sie in demonstrativ-illustrierender Funktion mit jeweils geplanten und verschiedenartigen mikrostrukturellen Elementen (Angaben) konsistent, sinnvoll und zweckgemäß korrelieren und folgerichtig als LB integriert werden können?

Für die korpusempirische Lexikographie sind jedoch in erster Linie solche Qualitätsmerkmale von Textstellen relevant, die es ermöglichen, diese zur Basis und zum Ausgangspunkt der parömiographischen Bearbeitung eines SW zu erklären. Somit lässt sich die Fragestellung weiterhin spezifizieren:

Welche Eigenschaften müssen Textstellen haben, damit sie im parömiogra-

phischen Prozess so anwendbar sind, dass sie als empirische Ausgangsbasis dienen, anhand deren Parömiographen lexikographische Aussagen zu einzelnen SW erarbeiten können?

Anhand der Sprachdaten aus dem deutschen Teil der SprichWort-Datenbank (vgl. SprichWort 2010) wird nachfolgend versucht, die gefragten Eigenschaften zu systematisieren. Die lexikographische Perspektive wird dabei sinnvoll und zweckmäßig durch die Perspektive der Phraseologie- und Parömiologieforschung ergänzt, da beide Disziplinen bisher bedeutende, allerdings in der praktischen Lexikographie noch immer wenig berücksichtigte Kenntnisse zur Semantik, Pragmatik und Grammatik von SW hervorbringen konnten.¹⁷

3. Das lexikographische Beispiel in einem mehrdimensionalen Beschreibungsmodell für SW

Wie in Abschnitt 2.1 ausgeführt, sind mit der deskriptiven Theoriekomponente des LB seine systematischen Zusammenhänge mit einzelnen Angaben innerhalb lexikographischer Mikrostrukturen gemeint. Folgerichtig sollten LB wörterbuchartikelintern mit einzelnen Angaben sinnvoll korrelieren. Des Weiteren wurde gesagt, dass korpusempirisch ermittelte und systematisch ausgewertete Textstellen als Grundlage für die Angabenformulierung zweckmäßig fungieren sollten. Es handelt sich somit um eine reziproke Perspektiven-Relation, die im Folgenden anhand eines parömiographischen mehrdimensionalen Beschreibungsmodells für SW näher betrachtet wird.

Die gegenwärtige lexikographische Bearbeitung von SW ist bekannterweise mangelhaft; es wird auf verschiedene Defizite hingewiesen, dies vorrangig hinsichtlich der Erfassung semantisch-pragmatischer und grammatischer Daten zum SW-Gebrauch.¹⁸ Dies mag bestimmt auch an den tradierten Wörterbuchformaten liegen, die eine umfassende und mehrdimensionale lexikographische Darstellung wegen Platzbegrenztheit in der Regel nicht oder nur teilweise ermöglichen und funktional gesehen auch nicht immer anstreben. Es hat sich inzwischen allerdings die Einsicht durchgesetzt, dass eine fachgerechte lexikographische Beschreibung von SW in speziellen parömiologischen Produkten nur mehrdimensional und hochgradig strukturiert erfolgen kann (vgl. u.a. Ďurčo 2005 und 2006, Hrisztova-Gotthardt 2010, Jesenšek 2010 und 2012). Es wurden bisher mehrere komplexe Beschreibungsmodelle vorgeschlagen, darunter das Modell von Ďurčo (1992 und 2005), welches auch in der hier aktuellen SprichWort-Datenbank angewandt wurde. Modellgerecht werden semantische, pragmatische und formalgrammatische Charakteristika eines SW anhand mehrerer Beschreibungsparameter lexikographisch erfasst. Der Zweck der LB, die im Modell vorausgesetzt werden, besteht darin, lexikographische Angaben zu einzelnen Beschreibungsparametern (im Sinne der Regelformulierungen) so zu unterstützen, dass dadurch die LB-eigene demonstrativ-illustrierende Funktion realisiert werden kann. Werden sie methodisch transparent korpusempirisch ermittelt, so fungieren sie zugleich als Grundlage für

die Angabenformulierung.

Das Wesen einer derartigen mehrdimensionalen SW-Beschreibungsstruktur machen drei in sich komplexe Angabenbereiche aus: der Angabenbereich zur Semantik, der Angabenbereich zur Pragmatik und der Angabenbereich zur Grammatik. Vor dem Hintergrund der inzwischen recht umfangreichen Kenntnisse über die Komplexität der SW-Semantik, über die pragmatischen Gebrauchspräferenzen und/oder -restriktionen sowie über die syntagmatisch-syntaktischen und formal-strukturellen Charakteristika einzelner SW, lassen sich notwendige Eigenschaften der korpusermittelten Textstellen annehmen, um zum Ausgangspunkt für die Ausformulierung der lexikographischen Angaben zu werden und zugleich als LB dieselben Angaben gezielt unterstützen zu können. Anhand der konkreten Daten aus der SprichWort-Datenbank (und anhand der deutschen SW) werden sie im Folgenden in der Form von exemplarischen Aussagen dargelegt.¹⁹

3.1 Der semantische Angabenbereich

Der Angabenbereich zur Semantik betrifft vorrangig die denotativen Bedeutungsanteile eines SW. Daraus ergibt sich die Aussage (I):

(I) **Eine Textstelle ist dann als LB geeignet, wenn es Ausdrücke/Redeteile enthält, die proto- und stereotypische Eigenschaften dessen zeigen, was das SW benennt:**

- (1) *„Wer A sagt, muss auch B sagen“, sagt uns beispielsweise, dass etwas Begonnenes auch weitergeführt werden muss — und fällt der zweite Schritt auch noch so schwer.²⁰*

Die Textstelle (1) in der Funktion eines LB korreliert mit der Bedeutungsangabe (1a) zum Lemma *Wer A sagt, muss auch B sagen*, anhand deren Informationen zur Bedeutung des SW nachvollziehbar sind:

- (1a) *Sagt man dafür, dass begonnene Dinge auch zu Ende geführt oder die Konsequenzen eigener Handlungen auch getragen werden sollten.*

Die Aussage lässt sich auch eingehender ausformulieren, z.B. dann, wenn einzelne SW-Bedeutungsanteile zu einer besseren Bedeutungerschließung beitragen können, vgl.:

(I a) **Eine Textstelle ist dann als LB geeignet, wenn es Ausdrücke/Redeteile enthält, die enzyklopädische, kulturelle oder historische Aspekte des SW-Inhalts zeigen:**

- (2) *Immer wieder zitiert wird der Spruch „Auge um Auge, Zahn um Zahn“ im 2. Buch Mose (21, 24) — und auch wenn, was Theologen*

betonen, diese Aussage nur als Gebot zur Wiedergutmachung, nicht als Aufforderung zur Rache gemeint sein mag, so ist dieser Vers doch nicht ganz zufällig sehr missverständlich.

Die Textstelle (2) in der Funktion eines LB korreliert mit der Angabe (2a) zum Lemma *Auge um Auge, Zahn um Zahn*, anhand deren Informationen zum kulturhistorischen Hintergrund des SW nachvollziehbar sind:

- (2a) *In den Korpusbelegen wird häufig auf die biblische Herkunft des Sprichworts verwiesen.*

3.2 Der pragmatische Angabenbereich

Der Angabenbereich zur Pragmatik deckt die jeweilige Gebrauchsspezifika eines SW und kann sich in vielerlei Hinsicht zeigen.²¹ Sie betrifft möglicherweise Sprachhandlungen, die anhand eines SW vollzogen werden (können), weiter präferierte und/oder restringierte Verwendung in bestimmten Situationen und Kontexten, kommunikativen Domänen, Textsorten, eventuelle themengebundene Verwendung u.a.m.²² Daraus ergibt sich zunächst eine generelle Aussage zur Geeignetheit einer Textstelle als LB zur SW-Pragmatik:

- (II) Eine Textstelle ist dann als LB geeignet, wenn es Ausdrücke/Redeteile enthält, die pragmatische Aspekte eines SW zeigen.**

Sie lässt sich weiterhin so spezifizieren, dass jeweils einzelne pragmatische SW-Charakteristika genannt werden:

- (II 1) Eine Textstelle ist dann als LB geeignet, wenn es Ausdrücke/Redeteile enthält, die das sprachhandlungsbezogene Potential eines SW zeigen:**

- (3) *Nun können wir gemeinsam tanzen und können mit Sicherheit behaupten, dass das Tanzen unser größtes Hobby ist. Also: Wenn es am Anfang nicht klappt — nur nicht aufgeben — Übung macht den Meister.*

Das LB (3) korreliert mit der pragmatischen Angabe (3a) zum Lemma *Übung macht den Meister*, anhand deren die durch das SW realisierte Sprachhandlung ERMUTIGEN erschließbar ist:

- (3a) *In bestimmten Korpusbelegen wird mit dem Sprichwort eine Ermutigung ausgedrückt, geduldig zu sein und nicht zu schnell aufzugeben.*

- (II 2) Eine Textstelle ist dann als LB geeignet, wenn es Ausdrücke/Redeteile enthält, die das Vorkommen eines SW in einer präferierten kommunikativen Domäne zeigen:**

- (4) *Vorbeugen ist besser als heilen. Ausgewogenes Essen, genügend*

Schlaf, viel trinken, Bewegung im Freien, Freude und Zufriedenheit — alle diese Eigenschaften helfen, den Körper gegen Krankheiten zu stärken.

Das LB (4) korreliert mit der Angabe (4a) zum Lemma *Vorbeugen ist besser als heilen*, anhand deren Informationen zu der typischen Verwendungsdomäne erschließbar sind:

- (4a) *In den Korpusbelegen kommt das Sprichwort häufig in der Domäne Gesundheit vor.*

(II 3) Eine Textstelle ist dann als LB geeignet, wenn es Ausdrücke/Redeteile enthält, die das Vorkommen eines SW in einer präferierten Textsorte zeigen:

- (5) *Alles was Sie anpacken gelingt. Waage: **Wer wagt, der gewinnt**. Venus schenkt Ihnen viele zauberhafte Stunden mit Ihrem Herzblatt.*

Das LB (5) korreliert mit der Angabe (5a) zum Lemma *Wer wagt, gewinnt*, anhand deren Informationen zum textsortenpräferierten Gebrauch nachvollziehbar sind:

- (5a) *In bestimmten Korpusbelegen wird das Sprichwort in Horoskopen verwendet.*

(II 4) Eine Textstelle ist dann als LB geeignet, wenn es Ausdrücke/Redeteile enthält, die das präferierte themenbezogene Vorkommen eines SW zeigen:

- (6) *Wer auch immer für den Spruch „**Wer schön sein will, muss leiden**“ verantwortlich ist, hatte vermutlich nicht die ästhetisch-plastische Chirurgie vor Augen. Dennoch trifft der Satz. Jede Schönheitschirurgie ist ein Eingriff in den Körper und kann mitunter fatale Folgen haben.*

Das LB (6) korreliert mit der Angabe (6a) zum Lemma *Wer schön sein will, muss leiden*, anhand deren Informationen zum themaspezifischen Gebrauch nachvollziehbar sind:

- (6a) *In den Korpusbelegen wird das Sprichwort häufig in Verbindung mit dem Thema Schönheitschirurgie verwendet.*

(II 5) Eine Textstelle ist dann als LB geeignet, wenn es Ausdrücke/Redeteile enthält, die das Vorkommen eines SW in einer präferierten Situation zeigen:

- (7) *[...] dem Klub droht der Zerfall. Die Spieler haben sich aufgegeben,*

obwohl noch neun Spiele zu absolvieren sind. Bonhof will sich damit nicht abfinden und forderte sie auf, Zeichen zu setzen. [...] Goalie Robert Enke und Stürmer Markus Feldhoff haben schon ihren Abschied verkündet, die Fans dankten es ihnen mit dem Spruchband: "Die Ratten verlassen das sinkende Schiff".

Das LB (7) korreliert mit der Angabe (7a) zum Lemma *Die Ratten verlassen das sinkende Schiff*, anhand deren die Einschränkung des SW-Gebrauchs auf eine typische Gebrauchssituation erschließbar ist:

(7a) *Mit diesem Sprichwort wird immer eine Kritik am Verhalten der Verantwortlichen ausgedrückt.*

(II 6) Eine Textstelle ist dann als LB geeignet, wenn es Ausdrücke/Redeteile enthält, die eine dem SW immanente wertende Sprechereinstellung zeigen:

(8) *Der Schulwart [...] weiß ein Lied von den Zuständen in seiner Schule zu singen. Die Fassade wurde bereits renoviert. Aber **außen hui, innen pfui**: Die Duschen mußten vor etwa neun Monaten gesperrt werden, Fliesen fielen von den Wänden, [...] Wasserabflüsse waren verstopft — 1000 Schüler können sich nach dem Turnunterricht nicht duschen.*

Das LB (8) korreliert mit der Angabe (8a) zum Lemma *Außen hui, innen pfui*, anhand deren eine negativ wertende Sprechereinstellung erschließbar ist:

(8a) *Sagt man leicht abwertend, wenn etwas oder jemand an der Oberfläche gut oder vielversprechend erscheint, bei genauerer Betrachtung aber nicht ist.*

3.2 Der grammatische Angabenbereich

Der Angabenbereich zur Grammatik betrifft einerseits formal-strukturelle und andererseits syntagmatisch-syntaktische SW-Eigenschaften. Vorrangig interessiert hierbei das SW-eigene Variabilitätspotenzial; die Variabilität betrifft möglicherweise sowohl die formale SW-Struktur als auch seine syntaktisch-syntagmatische textuelle Einbindung. Vom lexikographischen Interesse ist somit ihr (als usuell angesehenes) lexikalisches, morphologisches, syntagmatisches und syntaktisches Transformationspotenzial. Daraus ergibt sich zunächst eine generelle Aussage zur Geeignetheit einer Textstelle als LB zur SW-Grammatik:

(III) Eine Textstelle ist dann als LB geeignet, wenn es Ausdrücke/Redeteile enthält, die grammatische Aspekte eines SW zeigen.

Sie lässt sich weiterhin wiederum spezifizieren, und zwar so, dass jeweils ein-

zelse grammatische SW-Charakteristika genannt werden:

(III 1) Eine Textstelle ist dann als LB geeignet, wenn es Ausdrücke/Redeteile enthält, die das lexikalische Variationspotenzial eines SW zeigen (Komponentenvariabilität):

- (9) *„Stille Wasser gründen tief“, urteilt ein Sprichwort. Ohne die Wahrheitstiefe ergründen zu können, sei die Bedeutung erläutert: Wahrt jemand äußerlich den Schein, wirkt sehr ruhig, ist aber im Innern ganz anders, dann wird er oder sie oft als „stilles Wasser“ eingestuft. Verbreitet ist das (Vor)urteil für gewöhnlich gegenüber einsilbigen Frauen.*

Das LB (9) korreliert mit der Angabe (9a) zum Lemma *Stille Wasser sind tief*, anhand deren die usuelle Variabilität des Komponentenbestands eines SW erschließbar ist:

- (9a) *Stille Wasser gründen tief.*

(III 2) Eine Textstelle ist dann als LB geeignet, wenn es Ausdrücke/Redeteile enthält, die das flexionsmorphologische Variationspotential eines SW zeigen (Personalformen, Tempus, Modus, Numerus):

- (10) *Der Frankfurter Gastronom hat seine eigene Apfeltheorie entwickelt: „Die geraden Jahre sind gute, die ungeraden schlechte Erntejahre.“ Die Ausnahme bestätigt die Regel: „Es kann auch manchmal anders sein.“*

Das LB (10) korreliert mit der Angabe (10a) zum Lemma *Ausnahmen bestätigen die Regel*, anhand deren flexionsgrammatisch variable SW-Formen (hier Numerus) erschließbar sind:

- (10a) *Die Ausnahme bestätigt die Regel.*

- (11) *Die Scherben brachten in diesem Fall kein Glück. Bei einem Pkw, der in der Limburger Straße 29 in Diez abgestellt war, wurde in der Zeit von Dienstag, 8. Mai, 17 Uhr bis Mittwoch, 9. Mai, 8 Uhr der linke Außenspiegel nach unten gedrückt, wodurch der Spiegel mit Halterung zerbrach.*

Das LB (11) korreliert mit der Angabe (11a) zum Lemma *Scherben bringen Glück*, anhand deren flexionsgrammatisch variable Textrealisierung (hier Tempus) erschließbar ist:

- (11a) *In den Korpusbelegen wird das Sprichwort häufig in der Vergangenheitsform in andere Sätze eingebettet.*

(III 3) Eine Textstelle ist dann als LB geeignet, wenn es Ausdrücke/Redeteile

enthält, die das syntaktisch-syntagmatische Transformationspotential eines SW zeigen (Passiv-, Imperativ-, Nominalisierungstransformation, Verwendung mit Modalverb, Verwendung als Relativsatz u.a.):

- (12) *Nach 22 Jahren Vorstandsarbeit, davon 16 Jahre im Amt des Vorsitzenden, hat Wolfgang Müller nicht mehr für den Posten des Vorsitzenden der Bibliser Feuerwehr kandidiert. "Es hat sich in den letzten Jahren gezeigt, dass **man nicht gleichzeitig auf zwei Hochzeiten tanzen kann**" [...] Die Belastung seines Amtes als Kreisbrandinspektor stelle erhöhte Anforderungen an ihn, sodass er sich außerstande sehe, so für den Verein zu arbeiten, "wie ich es eigentlich möchte".*

Das LB (12) korreliert mit der Angabe (12a) zum Lemma *Man kann nicht auf zwei Hochzeiten tanzen*, anhand deren eine als usuell angesehene syntaktisch variable Textrealisierung (hier eingeleiteter Teilsatz) erschließbar ist:

- (12a) *In den Korpusbelegen wird das Sprichwort häufig als Teilsatz verwendet und mit ‚dass‘ eingeleitet.*
- (13) *Die Niederlage in Salzburg? Abgehakt! Aber nicht vergessen. Das Hagebau Team Tirol träumte vor dem heutigen dritten Kampf gegen Salzburg von **süßer Rache**.*

Das LB (13) korreliert mit der Angabe (13a) zum Lemma *Rache ist süß*, anhand deren eine als usuell angesehene syntaktisch variable Textrealisierung (hier Konversion) erschließbar ist:

- (13a) *In den Korpusbelegen wird das Sprichwort häufig nicht als Satz, sondern als Satzglied verwendet.*
- (14) ***Erst die Arbeit, dann das Vergnügen**: So hielten es jüngst die Mitglieder des [...] Ortsvereins Remagen. Sie schlossen an ihre Jahreshauptversammlung [...] einen gemütlichen Nachmittag mit musikalischer Begleitung an.*

Das LB (14) korreliert mit der Angabe (14a) zum Lemma *Erst die Arbeit, dann das Vergnügen*, anhand deren eine durch die SW-Position gegebene themeneinleitende Funktion erschließbar ist:

- (14a) *In den Korpusbelegen übernimmt das Sprichwort häufig eine thematisch einleitende Funktion und wird entsprechend mit einem Doppelpunkt markiert.*

4. Fazit und Ausblick

Anhand des LB im Bereich der Parömiographie wurde im Beitrag die Idee verfolgt, dass dieses in der Zeit der korpusempirischen Lexikographie seinen Sta-

tus geändert hat: Von der primären und tradierten dokumentierend-illustrativen Funktion und somit der Unterstützung der individuellen Sprachkompetenz des Lexikographen wird es zum Ausgangs- und Mittelpunkt der lexikographischen Arbeitsprozesse. Es wird zur empirischen Datenbasis für die lexikographische Erfassung sprachlicher Phänomene (hier der Sprichwörter) erklärt und es kann zugleich in Korrelation mit dem jeweilig aktuellen Mikrostrukturenprogramm seine tradierte dokumentierend-illustrative Funktion systematisch realisieren. Somit rücken in den Vordergrund des Forschungsinteresses methodisch-methodologische Fragen zu Identifikation, qualitativ-quantitativer Auswertung und Selektion von kontextuellen Korpusdaten (hier *Textstellen* genannt), die in der Funktion eines LB auftreten können. Dadurch schließe ich mich Schnörch und Storjohann (2012) an, wonach eine kritisch-analytische, gegenstandsbezogene und kompetenzgeleitete Auswertung und Überprüfung von korpusempirischen Daten notwendig ist, wenn in der praktischen Lexikographie Qualität und Zuverlässigkeit angestrebt werden.

Die Ausführungen in Kap. 3 gingen von der phraseologisch-parömiologischen Perspektive aus; es wurde versucht, die phraseologietheoretischen Kenntnisse zum (deutschen) SW und das korpusermittelte Wissen davon, wie SW im (gegenwärtigen) textuellen Gebrauch vorkommen, auf eine systematische Art und Weise mit der Begrifflichkeit des LB in Verbindung zu setzen. Obgleich exemplarisch hat es sich deutlich gezeigt, dass im Fall der modernen hochgradig strukturierten digitalen parömiographischen Produkte die einst angestrebte Multifunktionalität von LB (vgl. Hermanns 1988: 177ff.) nicht mehr vorrangig vom Interesse ist. Vielmehr von Bedeutung sind feinkörnige und auf einzelne Dimensionen der Semantik, Pragmatik und Grammatik ausgerichtete Aspekte der LB, wodurch eine ausdifferenzierte und strukturierte parömiographische Beschreibung und jeweils spezifische illustrierende Belegung ermöglicht wird. Es ist weiterhin zu erwarten, dass die empirischen Korpusdaten neue Beobachtungen und Erkenntnisse zum textuellen SW-Gebrauch werden evidentieren und somit weiterführende Untersuchungen anregen können.

Weiterführende Untersuchungen sind aber auch in parömiographietheoretischer Hinsicht notwendig. Die obige Diskussion der deskriptiven Theoriekomponente des LB im Bereich der Parömiographie eröffnet manche Fragen, auf die man im Detail eingehen sollte. Folgende drei seien hier hervorgehoben:

- Es wäre zu prüfen, inwiefern die obigen Aussagen zur prinzipiellen Geeignetheit von Textstellen als LB fungieren zu können in der praktischen Lexikographie akzeptabel und in konkreten lexikographischen Prozessen operationalisierbar sind. Man kann hierbei mit gutem Grund annehmen, dass semantisch-pragmatische Aspekte des LB eher sprachübergreifend sind, während grammatikbezogene Aspekte einzelsprachlich stärker differieren und somit einzelsprachspezifisch festzulegen sind.
- Dringend zu prüfen wäre die normative Theoriekomponente des LB, die

hier nicht aufgegriffen wurde. Die Fragestellungen zur Verständlichkeit der korpusermittelten empirischen Sprachdaten, zur potenziell notwendigen lexikographischen Intervention, zu den angestrebten Qualitätsmerkmalen von LB (Lebensnähe, alltägliche bzw. erkennbare Kommunikationskonstellationen, Angemessenheit und Kontrolle der Lexik u.a.) sind jedoch nur mit Bezug auf konkrete lexikographische Produkte und unter Einbeziehung der Wörterbuchbenutzungsforschung zu beantworten (vgl. hierzu neuerdings Potgieter 2012).

- Es ist notwendig, das Thema mit Bezug auf verschiedene Wörterbuchtypen, -funktionen und -formate zu erforschen, damit LB spezifische Funktionen in Bezug auf das Informationsprogramm des jeweilig zu erstellenden Wörterbuchs wahrnehmen und minimalistische bis maximalistische Mikrostrukturenprogramme unterstützen können.

Anmerkungen

1. Haß (1991a) nennt hierzu vier Traditionen: die „Urbild“-Tradition (Beleg als Ausgangspunkt der Bedeutungsbeschreibung), die Nachweis-Tradition (Beleg als Wahrhaftigkeitsbeweis), die Illustratio-Tradition (Beleg als Veranschaulichung einer Regel) und die Vorbild-Tradition (Beleg als Vorbild zur Nachahmung). Vgl. dazu Zöfgen (1994: 155f.) und seine Diskussion zur Wichtigkeit der sog. belegenden Lexikographie in Europa seit dem 18. Jh.
2. Darüber hinaus stellt man fest, dass auch die Wörterbuchkritik dieses Thema eher spärlich angeht und — bezogen auf einzelne Wörterbücher — ziemlich unsystematisch bewertet (vgl. u.a. Jehle 1990).
3. Vgl. Ďurčo (2006: 4): »große Korpora helfen bei der Suche nach Standardformen und -varianten sowie nach sprichwörtlichen Transformationen«.
4. Die Datenbank ist im Rahmen des gleichnamigen EU-geförderten Forschungsprojekts entstanden (2008-2010). Das Projekt wurde unter der Leitung der Autorin dieses Beitrags (Universität Maribor, Slowenien) und unter Mitwirkung der Universität des Hl. Kyrill und Methodius Trnava (Slowakei), der Universität Zlin (Tschechien), der Universität Szeged (Ungarn), der Technischen Universität Graz (Österreich) und des IDS Mannheim (Deutschland) durchgeführt. Es ging primär um die Konzipierung und Erarbeitung einer online-Lernplattform zur Vermittlung von Sprichwörtern. Die SprichWort-Plattform besteht aus drei Komponenten: einer mehrdimensionalen und mehrsprachigen SprichWort-Datenbank, den Didaktischen Inhalten und der SprichWort-Community (www.sprichwort-plattform.org). Mit je unterschiedlicher Gewichtung sind Projektprodukte auf Fremdsprachenlerner, Fremdsprachenvermittler, Entwickler von e-Lernmaterialien, Parömiologen, Parömiographen und Sprichwort-Interessierte gerichtet. Die einbezogenen Sprachen sind Deutsch, Slowenisch, Slowakisch, Tschechisch und Ungarisch, die zueinander jeweils Nachbarsprachen sind. Unter den innovativen Projektdimensionen u.a. das Modell für eine komplexe und mehrdimensionale lexikographische Bearbeitung von SW betont werden. Für eine ausführliche Projektbeschreibung vgl. Jesenšek (2010).
5. Man sieht, dass das Thema vorrangig im Rahmen der pädagogischen Lexikographie (vgl. u.a. Drysdale 1987, Jehle 1990, Zöfgen 1994, Potgieter 2012) und größerer lexikographischer bzw.

lexikographisch-didaktischer Vorhaben aufgegriffen wird (vgl. u.a. Schlaefer 1990, Haß 1991 und 1991a, Abel 2000, Klosa 2005, Jesenšek 2009, Dräger 2011, Klosa, Töpel und Koplenig 2012).

6. Angaben sind „funktionale lexikographische Textsegmente, die entweder zum Lemmazeichen oder zu artikelinternen Angaben oder zu Angaben in anderen Artikeln oder zu Textteilen in den Außentexten des Wörterbuches in einer Angabebeziehung stehen, und deren genuiner Zweck darin besteht, daß der potentielle Benutzer aus ihnen entweder lexikographische Informationen über den Wörterbuchgegenstand oder solche über diesen und die Wörterbuchform erschließen kann“ (Wiegand 1989: 427).
7. Vgl. Wiegand (1977: 102): »Die Lexikographie benötigt eine Theorie des lexikographischen Beispiels. Diese muß zeigen, wie die bedeutungserläuternde /.../ Funktion lexikalischer Paraphrasen systematisch durch Beispieltypen unterstützt werden kann«.
8. Unterschieden werden *syntagmatische, paradigmatische, definitorische, rhetorisch-pragmatische, beweisende, metasprachliche, enzyklopädische, ideologische* und *Vorbildfunktionen* des LB (vgl. Martin 1989: 601ff.).
9. Diese reichen von den *Pseudo-, enzyklopädischen* und *definierenden Beispielen* (die Letzteren will er eigentlich nicht als LB verstanden wissen) über *bedeutungsdeterminierende* und *bedeutungsdifferenzierende Beispiele* bis zu den *kollokationsäquivalenten, grammatikillustrierenden* und *bedeutungsindifferenten Beispielen* (vgl. Zöfgen 1994: 186ff.).
10. Diese Betrachtungsperspektive ist allerdings nicht neu. In der Geschichte der Lexikographie spiegelt sie sich in der Zusammenstellung umfangreicher Belegarchive und genauso in redaktioneller Betreuung der lexikographischen Prozesse ab (vgl. dazu die »Urbild«-Funktion bei Haß 1991a), nur scheint es, dass die oben skizzierte funktional und lerndidaktisch ausgerichtete metalexikographische Diskussion vom Ende des 20. Jh. andere Schwerpunkte gesetzt hat. Vgl. z.B. die Beobachtung zur englischen Lexikographie des 18. Jh. in Atkins und Rundell (2008: 543): »Johnson's dictionary thus embodies the principle that languages should be described on the basis of objective evidence of their use — and this, in a sense, is the primary function of examples of usage: as a source of data from which lexicographers construct their entries. Attaching examples to definitions is a separate process«.
11. Zwischen dem sog. *Konsultationsparadigma* (korpusbasiert) und *Analyseparadigma* (korpusdriven) (vgl. Steyer 2004) wird hier nicht explizit unterschieden, zumal in der Praxis (und das gilt auch für das hier zugrundeliegende Projekt SprichWort) beide Vorgehen nicht unbedingt voneinander isoliert angewandt sondern im wechselseitigen Verhältnis zueinander stehen können (vgl. Heine 2012: 178).
12. Ein weites LB-Verständnis, wie in Harras (1989: 608) dargelegt und wonach unter LB auch Angaben von Kollokationen, Phrasemen, syntaktischen Strukturformeln des Typs *sich vor jmdm., etw., um jmdn. ängstigen* aufgefasst werden, wird hier allerdings nicht vertreten. Das hiesige Verständnis von LB deckt sich allgemein gesehen mit dem Harras'schen „Verwendungsbeispiel“ bzw. mit dem „Satzbeispiel“ Zöfgens (1994).
13. Ebenso knapp in der Reflexion des LB sind manche phraseologische Wörterbücher, u.a. das repräsentative phraseologische Wörterbuch des Deutschen Duden Redewendungen (2008). In den Benutzungshinweisen (2008: 23) wird man hauptsächlich über die technischen Aspekte von »Beispielen und Belegen« informiert (Platzierung, Quellenangaben), während ihre Funktionalität sehr allgemein dargelegt wird: sie bestehe in der Veranschaulichung des Gebrauchs des dazugehörigen Lemmas, somit in der Ergänzung der Bedeutungsangaben

- und folgerichtig der helfenden Unterstützung im DaF-Bereich.
14. Nur so ließe sich nämlich die alte Tradition bewältigen, die Heine (2012: 173) kurz und treffend zusammenfasst: »Die Parömiographie des 20. Jahrhunderts basiert überwiegend auf älteren Sammlungen«.
 15. Somit wird die Annahme, dass Sprichwörter über die (gegenwärtigen) Textkorpora nicht oder nur schwer zugänglich seien, wie das in Heine (2012: 181) prognostiziert wird, nicht vertreten. Die Erfahrungen im Projekt SprichWort (2010) sprechen dafür, dass korpusermittelte SW-Daten (für alle beteiligten Sprachen, darunter auch für Deutsch) für den lexikographischen Prozess durchaus zugänglich und lexikographisch relevant sind.
 16. Die Arbeit Hein's entstand in Anlehnung an das Projekt SprichWort, im dessen Rahmen die lexikographische SW-Bearbeitung ausschließlich anhand der Auswertung von korpusermittelten Sprachdaten verlief.
 17. Die Software-Tools zur Identifikation von »guten« Kandidaten zum LB, sog. GDEX (Good Dictionary Examples vgl. u.a. Husák et al. 2008 und Kilgariff et al. 2008), die etwa im Rahmen von Sketch Engine (<http://the.sketchengine.co.uk>) entwickelt werden, werden in diesem Beitrag ausgeklammert, vor allem deshalb, weil GDEX bisher m.W. nur für Englisch validiert wurde, der vorliegende Beitrag jedoch von der deutschen Sprache ausgeht.
 18. Vgl. hierzu Āurčo (2005: 144f.): »Ein- und zweisprachige Wörterbücher weisen große Defizite hinsichtlich der semantischen, morphosyntaktischen und kommunikativ-pragmatischen Informationen zur Parömiologie auf /.../ Es gilt zu prüfen, /.../ um welche Aspekte die Beschreibung trotz der notwendigen Wörterbuchökonomie in den Lexika erweitert werden sollte, um adäquate Informationen für den aktiven Gebrauch zu liefern /.../ Es fehlen dabei präziser konstruierte Paraphrasen, die zeigen würden, in welchem Umfang bzw. mit welchen Einschränkungen die Paraphrase ihre Gültigkeit hat und welche zusätzlichen differenzierenden semantischen Merkmale im Umfeld um den semantischen Kern noch eine Rolle spielen.«
 19. Es sei allerdings betont, dass die vorliegende Auflistung keinen Anspruch auf Vollständigkeit erhebt; es werden lediglich die wesentlichen Aspekte der lexikographischen SW-Beschreibung ausgewählt und anhand der deutschen SW exemplifiziert.
 20. Alle exemplarischen Textstellen sind korpusermittelt und entstammen dem deutschen Teil der SprichWort-Datenbank (2010). Unter einfacher Nummerierung (1) werden jeweils datenbankeigene LB angegeben, darauf folgt die entsprechende Bezugsangabe aus dem jeweiligen SW-Eintrag (1a). Auf Quellenangaben, die in der Datenbank konsequent angeführt sind, wird hier verzichtet.
 21. Zu den pragmatischen Potenzialen von SW gibt es inzwischen recht viel (germanistische) Literatur, vgl. u. a. Lüger 1999, Umurova 2005, Lewandowska 2008, Burger 2010, Hoffmann 2012.
 22. Die hier präsentierten Daten betreffen die ausgewählte parömiologische Datenbank (SprichWort). Weitere korpusanalytische Untersuchungen würden eventuelle zusätzliche od. andersartige od. einzelsprachspezifische Besonderheiten des SW-Gebrauchs ans Licht bringen können.

Literatur

- Abel, Andrea. 2000. Das lexikographische Beispiel in der L2-Lexikographie (am Beispiel eines L2-Kontext- und Grundwortschatzwörterbuches). *Deutsch als Fremdsprache* 37(3): 163-169.
- Atkins, B.T. Sue und Michael Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford/

New York: Oxford University Press.

- Burger, Harald.** 2009. Semantische Aspekte der deutschen Phraseographie: die aktuelle Praxis — allgemeine und phraseologische Wörterbücher im Vergleich. Mellado Blanco, Carmen. (Hrsg.). 2009. *Theorie und Praxis der idiomatischen Wörterbücher*: 23-44. Lexicographica. Series Maior 135. Tübingen: Max Niemeyer.
- Burger, Harald.** 2010. *Phraseologie. Eine Einführung am Beispiel des Deutschen*. 4. neu bearbeitete Auflage. Berlin: Erich Schmidt Verlag.
- Cowie, Anthony P.** 1989. The Language of Examples in English Learners' Dictionaries. James, Gregory (Hrsg.). *Lexicographers and their Works*: 55-65. Exeter Linguistic Studies 14. Exeter: University of Exeter Press.
- Dräger, Marcel.** 2011. *Der phraseologische Wandel und seine lexikographische Erfassung. Konzept des „Online-Lexikons zur diachronen Phraseologie (OLdPhras)“*. Dissertation. Freiburg: Albert-Ludwigs-Universität.
- Drysdale, Paul D.** 1987. The Role of Examples in a Learner's Dictionary. Cowie, Anthony P. (Hrsg.). 1987. *The Dictionary and the Language Learner. Papers from the EURALEX Seminar at the University of Leeds, 1–3 April 1985*: 213-223. Lexicographica. Series Maior 17. Tübingen: Max Niemeyer.
- Duden Redewendungen.** 2008. *Wörterbuch der deutschen Idiomatik*. 3., überarbeitete und aktualisierte Aufl. 2008. Hrsg. von der Dudenredaktion. Mannheim/Leipzig/Wien/Zürich: Dudenverlag.
- Đurčo, Peter.** 1992. Ein Matrixmodell für ein- und zweisprachige phraseologische Wörterbücher. Mair, Christian und Manfred Markus (Hrsg.). 1992. *Neue Ansätze in der Kontrastiven Linguistik. Proceedings of the Conference Held at the Leopold-Franzens-University of Innsbruck, Austria, 10–12 May 1991. Vol. I*: 261-269. Innsbruck: Universität Innsbruck.
- Đurčo, Peter.** 2005. *Sprichwörter in der Gegenwartssprache*. Trnava: Univerzita sv. Cyrila a Metoda v Trnave.
- Đurčo, Peter.** 2006. Methoden der Sprichwortanalysen oder Auf dem Weg zum Sprichwörter-Optimum. Buhofer, Häcky Annelies und Harald Burger (Hrsg.). 2006. *Phraseology in Motion I. Methoden und Kritik. Akten der Internationalen Tagung zur Phraseologie (Basel, 2004)*: 3-20. Hohengehren: Schneider Verlag.
- Engelberg, Stefan und Lothar Lemnitzer.** 2008. *Lexikographie und Wörterbuchbenutzung*. 3. Aufl. Tübingen: Stauffenburg.
- Grzybek, Peter.** 1992. Probleme der Sprichwort-Lexikographie (Parömiographie): Definition, Klassifikation, Selektion. Meder, Georg und Andreas Dörner (Hrsg.). 1992. *Worte, Wörter, Wörterbücher. Lexikographische Beiträge zum Essener Linguistischen Kolloquium*: 195-223. Tübingen: Max Niemeyer.
- Harras, Gisela.** 1989. Zu einer Theorie des lexikographischen Beispiels. Hausmann, Franz Josef et al. (Hrsg.). 1989: 607-614.
- Haß, Ulrike.** 1991. Textkorpora und Belege. Methodologie und Methoden. Harras, Gisela, Ulrike Haß und Gerhard Strauß (Hrsg.). 1991. *Wortbedeutungen und ihre Darstellung im Wörterbuch*: 212-286. Berlin/New York: Walter de Gruyter.
- Haß, Ulrike.** 1991a. Zu Bedeutung und Funktion von Beleg- und Kompetenzbeispielen im Deutschen Wörterbuch. Kirkness, Alan, Peter Kühn und Herbert Ernst Wiegand (Hrsg.). 1991. *Studien zum Deutschen Wörterbuch von Jacob Grimm und Wilhelm Grimm. Bd. II*: 535-594.

Tübingen: Max Niemeyer.

- Hausmann, Franz Josef et al. (Hrsg.).** 1989. *Wörterbücher. Ein internationales Handbuch zur Lexikographie/Dictionaries. An International Encyclopedia of Lexicography/Dictionnaires. Encyclopédie internationale de lexicographie.* Handbücher zur Sprach- und Kommunikationswissenschaft 5.1. Berlin/New York: Walter de Gruyter.
- Hein, Katrin.** 2012. Zugang zu Sprichwortbedeutung und -gebrauch mit Hilfe von Korpora. Steyer, Kathrin (Hrsg.). 2012. *Sprichwörter multilingual. Theoretische, empirische und angewandte Aspekte der modernen Parömiologie:* 315-340. Tübingen: Narr.
- Heine, Antje.** 2012. Empirische Methoden der Phraseologieforschung im Wandel der Zeit. Prinz, Michael und Ulrike Richter-Vapaatalo (Hrsg.). 2012. *Idiome, Konstruktionen, „verblümete Rede“.* Beiträge zur Geschichte der germanistischen Phraseologieforschung: 165-185. Stuttgart: S. Hirzel Verlag.
- Hermans, Fritz.** 1988. Das lexikographische Beispiel. Ein Beitrag zu seiner Theorie. Harras, Gisela (Hrsg.). 1988. *Das Wörterbuch. Artikel und Verweisstrukturen:* 161-195. Düsseldorf/Bielefeld: Schwann-Bagel/Cornelsen-Velhagen u. Klasing.
- Hoffmann, Sarah.** 2012. *Argumentative Strukturen in Sprichwörtern.* Bern u.a.: Peter Lang.
- Hrisztova-Gotthardt, Hrisztalina.** 2010. *Vom gedruckten Sprichwörterbuch zur interaktiven Sprichwortdatenbank. Überlegungen zum linguistischen und lexikographischen Konzept mehrsprachiger Sprichwortdatenbanken.* Bern u.a.: Peter Lang.
- Husák, Milos et al.** 2008. GDEX: Automatically Finding Good Dictionary Examples in a Corpus. Bernal, Elisenda und Janet DeCesaris (Hrsg.). 2008. *Proceedings of the Thirteenth EURALEX International Congress, Barcelona, Spain, 15–19 July 2008:* 425-432. Barcelona: Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada.
- Jehle, Günter.** 1990. *Das englische und französische Lernerwörterbuch in der Rezension. Theorie und Praxis der Wörterbuchkritik.* Lexicographica. Series Maior 30. Tübingen: Max Niemeyer.
- Jesenšek, Vida.** 2009. Phraseologische Wörterbücher auf dem Weg zu Phraseologiedatenbanken. Mellado Blanco, Carmen (Hrsg.). 2009. *Theorie und Praxis der idiomatischen Wörterbücher.* Lexicographica. Series Maior 135: 65-81. Tübingen: Max Niemeyer.
- Jesenšek, Vida.** 2010. Sprichwörter im Netz. Eine Internet-Lernplattform für das Sprachenlernen. Mellado Blanco, Carmen et al. (Hrsg.). 2010. *La fraseografía del S. XXI. Nuevas propuestas para el español y el alemán:* 125-148. Berlin: Frank & Timme.
- Jesenšek, Vida.** 2012. Sprichwörter aus (kontrastiv-)linguistischer, lexikografischer und didaktischer Sicht: zum Projekt SprichWort. Steyer, Kathrin (Hrsg.). 2012. *Sprichwörter multilingual. Theoretische, empirische und angewandte Aspekte der modernen Parömiologie:* 275-268. Tübingen: Narr.
- Kilgarriff, Alan et al.** 2008. GDEX: Automatically Finding Good Dictionary Examples in a Corpus. Bernal, Elisenda und Janet DeCesaris (Hrsg.). 2008. *Proceedings of the Thirteenth EURALEX International Congress, Barcelona, Spain, 15–19 July 2008:* 425-432. Barcelona: Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada.
- Kispál, Tamás.** 2000. Einige deutsche Sprichwort-Wörterbücher aus der Sicht des DaF-Lerners. Bernáth, Árpád und Gunther Dietz (Hrsg.). 2000. *Jahrbuch der ungarischen Germanistik 1999:* 163-181. Budapest: Gesellschaft ungarischer Germanisten/Bonn: Deutscher Akademischer Austauschdienst.
- Kispál, Tamás.** 2007. Sprichwörteransammlungen. Burger, Harald et al. (Hrsg.). 2007. *Phraseologie. Ein internationales Handbuch der zeitgenössischen Forschung. 1. Halbband:* 414-423. Berlin/New York:

Walter de Gruyter.

- Klosa, Annette.** 2005. Belege in *elexico*. Haß, Ulrike (Hrsg.). 2005. *Grundfragen der elektronischen Lexikographie. *elexico* — das Online-Informationssystem zum deutschen Wortschatz*: 96-104. Berlin/New York: Walter de Gruyter.
- Klosa, Annette, Antje Töpel und Alexander Kopleinig.** 2012. Zur Funktion und Rezeption von Belegen — Ergebnisse einer Benutzungsstudie zum Onlinewörterbuch *elexico*. *Sprachwissenschaft* 37(1): 93-123.
- Kühn, Peter.** 2003. Phraseme im Lexikographie-Check. Erfassung und Beschreibung von Phrasemen im einsprachigen Lernerwörterbuch. *Lexicographica* 19: 97-118.
- Lewandowska, Anna.** 2008. *Sprichwort-Gebrauch heute: ein interkulturell-kontrastiver Vergleich von Sprichwörtern anhand polnischer und deutscher Printmedien*. Berlin u.a.: Peter Lang.
- Lüger, Heinz-Helmut.** 1999. *Satzwertige Phraseologismen. Eine pragmalinguistische Untersuchung*. Wien: Edition Praesens.
- Martin, Robert.** 1989. L'exemple lexicographique dans le dictionnaire monolingue. Hausmann, Franz Josef et al. (Hrsg.). 1989: 599-607.
- Mieder, Wolfgang.** 1984. Geschichte und Probleme der neuhochdeutschen Sprichwörterlexikographie. Wiegand, Herbert Ernst (Hrsg.). 1984. *Studien zur neuhochdeutschen Lexikographie*: 307-358. Hildesheim: Georg Olms.
- Mieder, Wolfgang.** 1989. Grundzüge einer Geschichte des Sprichwortes und der Redensart. Besch, Werner et al. (Hrsg.). 1989. *Sprachgeschichte. Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung Bd. 3*: 2559-2569. Berlin/New York: De Gruyter.
- Mieder, Wolfgang.** 1999. *Sprichwörter/Redensarten — Parömiologie*. Heidelberg: Julius Gross Verlag.
- Müller-Spitzer, Carolin und Stefan Engelberg.** 2011. Elektronische Lexikografie zwischen Grammatik und Lexikon. Engelberg, Stefan, Anke Holler und Kristel Proost (Hrsg.). 2011. *Sprachliches Wissen zwischen Lexikon und Grammatik*: 559-572. Berlin/New York: De Gruyter.
- Nikula, Henrik.** 1986. Wörterbuch und Kontext. Ein Beitrag zur Theorie des lexikalischen Beispiels. Weiss, Walter, Herbert Ernst Wiegand und Marga Reis (Hrsg.). 1986. *Textlinguistik contra Stilistik? Wortschatz und Wörterbuch. Grammatische und pragmatische Organisation von Rede?*: 187-192. Tübingen: Niemeyer.
- Potgieter, Liezl.** 2012. Example Sentences in Bilingual School Dictionaries. *Lexikos* 22: 261-271.
- Prinsloo, Daniel J.** 2009. The Role of Corpora in Future Dictionaries. Nielsen, Sandro und Sven Tarp (Hrsg.). 2009. *Lexicography in the 21st Century. In Honour of Henning Bergenholtz*: 181-206. Amsterdam/Philadelphia: John Benjamins. .
- Schemann, Hans.** 1989. Das phraseologische Wörterbuch. Hausmann, Franz Josef et al. (Hrsg.). 1989: 1019-1032.
- Schlaefel, Michael.** 1990. Praktische Fragen der Beleglexikographie am Beispiel des Deutschen Wörterbuchs von Jacob und Wilhelm Grimm. Schützeichel, Rudolf und Peter Seidensticker (Hrsg.). 1990. *Wörter und Namen. Aktuelle Lexikographie. Symposium Schloß Rauischholzhausen, 25.-27. September 1987*: 139-154. Marburg: Hitzeroth.
- Schnörch, Ulrich und Petra Storjohann.** 2012. Ein Korpus als Garant zuverlässiger lexikografischer Informationen? Eine vergleichende Stichprobenuntersuchung. Vatvedt Fjeld, Ruth und Julie Matilde Torjusen (Hrsg.). 2012. *Proceedings of the 15th EURALEX International Congress 2012, Oslo, Norway, 7-11 August 2012*: 310-322. Oslo: Universitetet i Oslo, Institutt for lingvistiske og nordiske studier.

- SprichWort.** 2010. *Eine Internetplattform für das Sprachenlernen*. www.sprichwort-plattform.org.
- Steyer, Kathrin.** 2004. Kookurenz, Korpusmethodik, linguistisches Modell, lexikografische Perspektiven. Steyer, Kathrin (Hrsg.). 2004. *Wortverbindungen — mehr oder weniger fest*. Institut für deutsche Sprache Jahrbuch 2003: 87-116. Berlin/New York: De Gruyter.
- Steyer, Kathrin.** 2010. Korpusbasierte Phraseographie. Neue empirische Methoden und Beschreibungsformen. Mellado Blanco, Carmen et al. (Hrsg.). 2010. *La fraseografía del S. XXI. Nuevas propuestas para el español y el alemán*: 249-277. Berlin: Frank & Timme.
- Steyer, Kathrin.** 2012. Sprichwortstatus, Frequenz, Musterbildung. Parömiologische Fragen im Lichte korpusmethodischer Empirie. Steyer, Kathrin (Hrsg.). 2012. *Sprichwörter multilingual. Theoretische, empirische und angewandte Aspekte der modernen Parömiologie*: 287-314. Tübingen: Narr.
- Umurova, Gulnas.** 2005. *Was der Volksmund in einem Sprichwort verpackt ... : moderne Aspekte des Sprichwortgebrauchs anhand von Beispielen aus dem Internet*. Frankfurt am Main u.a.: Peter Lang.
- Wiegand, Herbert Ernst.** 1977. Nachdenken über Wörterbücher: Aktuelle Probleme. Drosdowski, Günther, Helmut Henne und Herbert Ernst Wiegand (Hrsg.). 1977. *Nachdenken über Wörterbücher*: 51-102. Mannheim/Wien/Zürich: Lexikographisches Institut.
- Wiegand, Herbert Ernst.** 1989. Der Begriff der Mikrostruktur: Geschichte, Probleme, Perspektiven. Hausmann, Franz Josef et al. (Hrsg.). 1989: 409-462.
- Wiegand, Herbert Ernst.** 2006. Angaben, funktionale Angabezusätze, Angabetexte, Angabestrukturen, Strukturanzeiger, Kommentare und mehr. Ein Beitrag zur Theorie der Wörterbuchform. *Lexikographica* 21: 202-379.
- Zöfgen, Ekkehard.** 1994. *Lernerwörterbücher in Theorie und Praxis. Ein Beitrag zur Metalexikographie mit besonderer Berücksichtigung des Französischen*. Lexicographica. Series Maior 59. Tübingen: Max Niemeyer.

Introducing a New Lexicographical Model: AlphaConceptual⁺ (and How it Could Be Applied to Dictionaries for Luganda)

Deo Kawalya, *KongoKing Research Group, Department of Languages and Cultures, Ghent University, Ghent, Belgium; and School of Languages, Literature and Communication, Makerere University, Kampala, Uganda (kawalyad@arts.mak.ac.ug)*

and

Gilles-Maurice de Schryver, *KongoKing Research Group, Department of Languages and Cultures, Ghent University, Ghent, Belgium; and Department of African Languages, University of Pretoria, Pretoria, South Africa (gillesmaurice.deschryver@UGent.be)*

Abstract: In this article we explore the possibility of amalgamating the semasiological (i.e. alphabetical), onomasiological (i.e. conceptual) and visual approaches to dictionary compilation, here termed an alphaconceptual⁺ (i.e. alphaconceptual 'plus') dictionary, using Luganda as a brief case study. Such a dictionary would combine the strong points of alphabetical and conceptual lexicography, with all entries also linked to relevant picture plates. In Section 1 we expound on the history of Luganda lexicography, highlighting the different types of dictionaries in the language since the early 1900s. Section 2 is an exposition of semasiological and onomasiological lexicography. In Sections 3 and 4 we study the actual dictionary market and scholarly lexicographic literature, in Africa and the rest of the world respectively. In Section 5 a case for language-independent alphaconceptual⁺ lexicography is argued, and its proposed compilation approach is sketched out in Section 6, followed by the conclusion in Section 7.

Keywords: SEMASIOLOGICAL LEXICOGRAPHY, ONOMASIOLOGICAL LEXICOGRAPHY, THESAURUS, THEMATIC DICTIONARY, TOPIC DICTIONARY, SYNONYM DICTIONARY, PICTURE DICTIONARY, VISUAL DICTIONARY, NEW LEXICOGRAPHICAL MODEL, ALPHACONCEPTUAL⁺, LUGANDA

Ekifunze: Okwanjula Enkola ya Kannankuluze Empya: Ennawalifu-Bitegeero⁺ (n'engeri gy'eyinza okukozesebwamu mu nkuluze z'Oluganda).

Mu lupapula luno tutunuulidde obusobozi bw'okugatta enkola z'okuwandiika enkuluze omuli ey'ekiwalifu, ey'ebitegeero n'ennyinyonnyozabifaananyi, wano kye tutuumye enkuluze ennawalifubitegeero⁺, nga tukozesa Oluganda ng'ekyokulabirako. Enkuluze ey'engeri eno eba eja kugatta emiganyulo gya kannankuluze ow'ekiwalifu n'ow'ebitegeero, ng'ennyinzi zonna era ziyungiddwa ku bifaananyi ebizituukirako. Mu Kitundu 1 tulombojja ebyafaayo bya kannankuluze w'Oluganda, nga tumenya ebika by'enkuluze eby'enjawulo mu Luganda okuviira ddala ku

ntandikwa y'emyaka gya 1900. Ekitundu 2 kinnyonyola kannankuluze ow'ekiwalifu n'ow'ebitegeero. Mu Kitundu 3 ne 4 tutunuulira akatale k'enkuluze kennyini mu Afrika n'ebiwandiiko by'ekiyivu ku nkuluze okwetooloola ensi yonna. Mu Kitundu 5 tuwa ensonga lwaki enkuluze ennawalifubitegeero* yeetaagibwa mu Luganda, ne tuteesa n'engeri gy'eyinza okuwandiikibwamu mu Kitundu 6, ne tuzzaako okukubira mu Kitundu 7.

Ebigambo Ebikulu: KANNANKULUZE OW'EKIWALIFU, KANNANKULUZE OW'E-BITEGEERO, WABITEGEERO, ENKULUZE NNAMIRAMWA, ENKULUZE Y'EMITWE, ENKULUZE Y'ENFAANANYAMAKULU, ENKULUZE Y'EBIFAANANYI, ENKULUZE ENNYINYONNYOZABIFAANANYI, ENKOLA YA KANNANKULUZE EMPYA, ENNAWALIFUBITEGEE-RO*, LUGANDA

1. Luganda lexicography

The oldest known lexicographical work in Luganda, i.e. Le Veux's Luganda–French vocabulary of 1917 is also regarded as one of the pioneering works in Bantu lexicography (Benson 1964: 73). This was followed by Kitching and Blackledge's Luganda–English dictionary of 1925 that was later revised by Mulira and Ndawula in 1952. Although these pioneering works have been referred to as dictionaries and are, no doubt, an invaluable effort in the foundation of lexicography in Luganda, one would still consider them as simple glossaries, for their entries are comprised of only two or at most three pieces of information, i.e. the headword, part of speech and a target-language gloss. A more developed (but still bilingual) Luganda–English dictionary was written by Snoxall in 1967, followed by another Luganda–English dictionary by Murphy in 1972. Even a brief look at these two dictionaries reveals that Murphy drew a lot from Snoxall, so much so that it would take a critical eye to notice the (subtle) differences. For a period of close to four decades after Murphy's dictionary, no Luganda dictionary was compiled, up to the late 2000s when Kiingi et al.'s 1 000-page *Enkuluze y'Oluganda ey'e Makerere* was published in 2007. This was the first-ever monolingual dictionary to be written in Luganda. A number of bilingual dictionaries have recently been compiled as well, including *A Concise Luganda–English Dictionary* by Bagunywa et al. in 2009 and *Fountain Standard Luganda–English Dictionary* by Bukenya and Kamoga, also in 2009.

Compilers of dictionaries in Luganda have thus concentrated on bilingual lexicography, with only one attempt at a monolingual reference work so far. Secondly, and more importantly, all these dictionaries have followed the same old tradition of alphabetical arrangement, without due regard to other advantageous forms of arrangement. There is but one exception to this trend: Kawalya's unpublished BA dissertation (2003), which includes a thematically arranged dictionary of Kiganda culture. The intrinsic contribution of this work notwithstanding, that dictionary cannot be representative of the entire language since it only deals with a section of the language, i.e. culture. An MA project (Kawalya

2006), developing the theoretical notions of Kiingi (2005) on conceptual lexicography for Luganda, does not provide a full solution to the absence of conceptual lexicographical works in the language either: it stops at categorizing the vocabulary of Luganda into conceptual clusters.

A modern dictionary for Luganda that is not straitjacketed by the alphabetic ordering would however be a very welcome addition to the dictionary landscape. As it stands, the well over five million first and second-language speakers of Luganda (Lewis et al. 2013) are ill-served with basic A-to-Z bilingual dictionaries on the one hand (most of them considerably dated at that), and a voluminous single-volume A-to-Z monolingual dictionary on the other. Writing in the 2010s, the dire need for a modern Luganda dictionary could and should of course translate into an electronic product. Not assuming automatic speech recognition (ASR), being able to spell remains a requirement to access an electronic product, but the contents should not necessarily be presented in alphabetical order. In an electronic product there is also room for more and different types of multimedia to support the presentation of the lexical information (for an in-depth overview of the possibilities, cf. De Schryver 2003). It is precisely this state of affairs, and the options offered by the modern digital world, that set us on a course to devise a new lexicographical model with which new types of dictionaries could be compiled — for Luganda, or for any other language(s).

2. Semasiological vs. onomasiological dictionaries

For a long time, the principle of alphabetization has gone hand in hand with the understanding of lexicography. Lest it be forgotten, the first monolingual dictionary in the English language even honoured the alphabet in its title: *A Table Alphabeticall* (Cawdrey 1604), and had to explain the process to its readers:

If thou be desirous (gentle Reader) rightly and readily to vnderstand, and to profit by this Table, and such like, then thou must learne the Alphabet, to wit, the order of the Letters as they stand, perfectly without booke, and where euery Letter standeth: as (b) neere the beginning, (n) about the middest, and (t) toward the end. Nowe if the word, which thou art desirous to finde, begin with (a) then looke in the beginning of this Table, but if with (v) looke towards the end. Againe, if thy word beginne with (ca) looke in the beginning of the letter (c) but if with (cu) then looke toward the end of that letter. And so of all the rest. &c.
(Cawdrey 1604: fol. A4v)

In many published works, the definition of dictionary continues to be found inseparable from the term alphabet. To this effect, Jackson (1988: 208) notes: "Ask anybody for a definition of the word dictionary and the term alphabet will most likely feature in it." He quotes the *Collins Pocket English Dictionary* which defines dictionary as "a book of alphabetically listed words in a language with definitions, pronunciations, etc." Most dictionaries have indeed

been compiled following this principle.

It is true that alphabetical arrangement of entries is advantageous to the user, mainly in one aspect, the convenience of reference or accessibility (for those who know the alphabet, and on the condition that a strict alphabetical ordering is followed, and not, for example, a phonemic classification). It is easy for the users of the dictionary to locate the items they are looking for, because entries are arranged following their alphabetical progression. In this way, a dictionary is basically serving a decoding purpose (Sierra 2000), while, according to Hartmann (1983), a large number of dictionary users need dictionaries for encoding purposes too. To the latter category of dictionary users, traditional (alphabetical) dictionaries may not be the most helpful (Sierra 2000).

The possibility of transforming traditional semasiological dictionaries into onomasiological dictionaries — i.e. "dictionaries that are used for finding a word from an idea" (Sierra 2000: 224) — has been upheld by many, e.g. Kipfer (1986), Sierra (2000), Burkhanov (2005). The need to see it as a 'transformation' is somehow ironic, however, as the 'topical tradition' is of course much older than the alphabetic one (cf. Hüllen 1999: 36).

Alphabetical dictionaries have conquered the market. This fact will not change in the near future. One consequence of this is that they have also conquered the attention of scholars. Yet, it seems high time that non-alphabetical dictionaries should at least regain their history. (Hüllen 1999: 21)

Although the biggest advantage of semasiological dictionaries seems to be centred on accessibility, Kipfer (1986: 55) notes that a lot of information is still hidden behind the structure of alphabetized headwords and that this is seriously disadvantageous to users who are interested in the differences between related words. Urdang (quoted in Kipfer 1986)¹ is disappointed that lexicographers have not been more resourceful in developing methods for organizing and accessing dictionaries. According to his survey, the main arrangements available besides the alphabetical one include: chronological (day-by-day and year-by-year with, in certain cases, a cross-grid of general thematic categories); indexed (words and phrases, topics, and content); thematic (usually supplemented by an index); rhyming; etymological; and structured (e.g., the Roget-type thesaurus with an index) (Kipfer 1986: 58). According to Urdang, of these, only thematic and structured formats use concepts as the basis of their methodology and attempt to go from "meaning" to "word". However, these approaches also have shortcomings. For example, Kawalya (2006: 8) notes that Roget's thesaurus — if not the first,² certainly the most famous among dictionaries compiled in a principle other than alphabetization — relies heavily on an alphabetical index for accessibility. Furthermore, the numerical code assigned to individual words in the index is shared by all words within the same cluster. This implies that the code 503 for *mad* for example (in Roget et al. 1966) will not take the user directly to *mad* but rather to the cluster title "Insanity" where *mad* is found together with over 750 other words, within which the user has to look for *mad*.

On this note, Kipfer (1986) argues a case for electronic onomasiological dictionaries, contending that large print dictionaries "would be too costly in formats other than alphabetical, both to prepare and print traditionally; they would require indexes and be very cumbersome to use" (p. 61). She summarizes the main advantage of the onomasiological approach over the traditional semasiological dictionary as lying in transparency, by providing access through each of the categories rather than just one (p. 64). Kipfer further argues that the most important advantage of an (onomasiological) electronic dictionary lies in the ability to ask for all entries that have a certain word in their definitions. For example, if one searched for all entries with the word *start* in the definition, one would be able to get words such as *begin*, *commence*, *embark*, *initiate*, *instigate* and *institute* (assuming that the word *begin* was used in the definitions of all these words).

3. Beyond the semasiological dictionary: the market for African-language dictionaries

Even a brief review of the available dictionaries for the African languages reveals that most are basically A-to-Z treatments of the general language. For some regions, however, especially in Central and South Africa, thematic approaches are quite popular, an activity typically undertaken by terminographers compiling terminology lists. For Cilubà, for instance, De Schryver (1999: 25) lists 38 dictionaries published over the course of a century (between 1897 and 1998), about a quarter of them with onomasiological features. These include a dictionary of synonyms (Nzongola 1967), and thematic dictionaries that focus on linguistic terminology (Bunduki 1975), economic and social activities (Maalu-Bungi et al. 1983), a classification of the general language (Yukawa 1992), hygiene (Kadima 1995), road signs (Kadima and Katende wa Ndaya 1995), and education (Kadima et al. 1995). Often, thematic dictionaries are built around illustrations (drawings, photos, etc.), an activity at which South African dictionary publishers excel. Examples include Pharos's *Multilingual Illustrated Dictionary* (Bennett and Tsoeu 2006), which covers English, Afrikaans, Zulu, Xhosa, Northern Sotho, Southern Sotho, and Tswana, or OUPSA's *First Bilingual Dictionary* series, which pairs English to each of the other official South African languages, ten dictionaries in all (Paizee et al. 2007–2009). Hybrid dictionary concepts are also being tried out: Maskew Miller Longman released the *Longman Foundation Phase Bilingual Dictionaries* (Various 2010), a series of five dictionaries spearheaded by R.H. Gouws and D.J. Prinsloo, in which English is paired with Afrikaans, Zulu, Xhosa, Northern Sotho and Tswana respectively. Each dictionary consists of four sections: a bilingual picture dictionary built around themes, an A-to-Z dictionary with definitions and examples in the African language and equivalents resp. translations in English, and two indexes, one for each language. Four approaches to what are basically

similar dictionary contents make this hybrid series rather cumbersome to use. The problem, of course, is that one is constantly hitting the limitations of paper: there is only so much one can present on a single page, and there is only so much one can pack into creative search routes for paper dictionaries.

In the Great Lakes region, at least two attempts were made to streamline the process. Kawalya (2006), in his MA dissertation, attempted a conceptual categorization of the Luganda lexicon using a Situational-Role Theory and imitating the biological taxonomy of living things. The underlying idea being that such a categorization would be so natural as to be able to dispense with indexes. In testing the hypothesis, Kawalya used the lexicon of Luganda, as reflected in Kiingi et al.'s (2007) *Enkuluze y'Oluganda ey'e Makerere*, considering the primary senses of the lexemes therein, and then fitting individual lexical items under various formulas, which could help in assigning an item to appropriate clusters. The lexicon of Luganda fitted into 52 conceptual clusters, which were grouped into eight major taxonomic categories that were given labels borrowed from the biological taxonomy of living things, i.e. super kingdom, kingdom, phylum, class, order, family, genus, and species. Brief tests of the outcome revealed that the conceptual clustering, especially the formulaic aspects, was too abstract to be of widespread practical use. A similar attempt in a closely related language, Runyankore-Rukiga, is that of Uganda's President, Y.K. Museveni and colleagues.³ In their *Katondoozi y'Orunyankore-Rukiga* 'Thesaurus of Runyankore-Rukiga' (Museveni et al. 2012), the entire world was categorized into eleven broad clusters. Unlike Kawalya, Museveni et al. used words to label the clusters, i.e.:

- Of the heavens, earth and cosmology;
- Of plants;
- Of animals;
- Of humankind;
- Of man and language;
- Of social behaviour in closely related communities;
- Of agriculture and preparation of foods and beverages;
- Of livestock farming, milk and meat;
- Of work and the world of skills;
- Of dressing, ornaments, perfumery and general hygiene; and
- Of travel on land, water and in the air.

Unfortunately, we feel that the items in this thesaurus cannot be easily accessed, exactly because there is no alphabetical index. The classification may make a lot of sense according to Uganda's President own perception of the world, but that does not automatically make it universal. One thus needs to know, either by intuition or otherwise, to which cluster a given word belongs in order to be able to access the respective list of words under the desired cluster. Also, and similar to Roget's thesaurus, words are simply listed without any definitions.

The search for innovative access routes to lexical information thus remains on, at least for the African languages. Sadly, compared to a decade ago, the lexicographic community with an interest in the African languages seems to devote ever less of their intellectual efforts to onomasiological approaches. This can be seen from Figure 1, in which various onomasiological key concepts are plotted for each volume of *Lexikos*, from Volume 1 (1991) up to Volume 22 (2012).

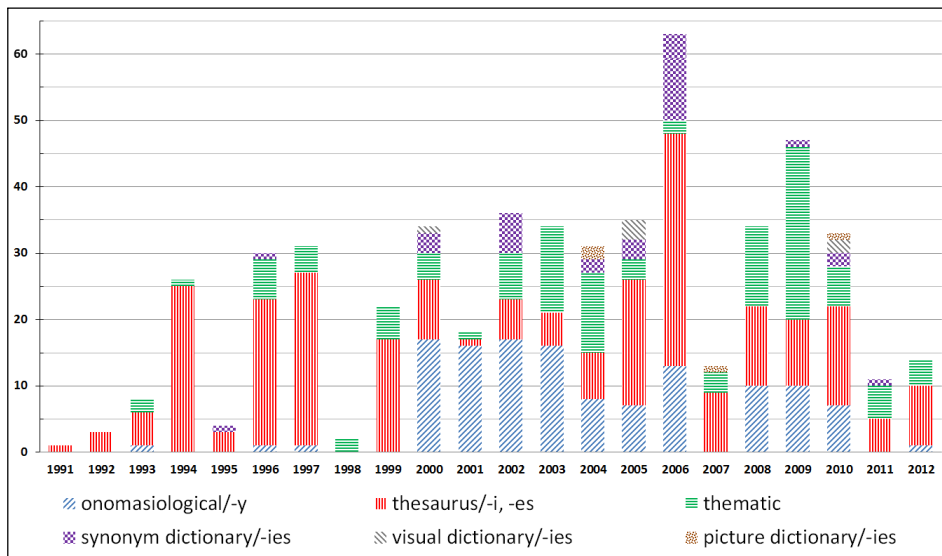


Figure 1: Onomasiological concepts in *Lexikos* (actual occurrences per volume).

One firstly notices that the sum of the various terms averages only about 24 occurrences per volume, with the mentions in Volume 16 (2006) exceptionally high. Also, the technical terms "onomasiological/onomasiology" only start to be used in earnest as of the year 2000. On the whole, references to "thesaurus/thesauri, thesauruses" gradually give way to references to the more generic "thematic". Finally, discussions of the "synonym dictionary/-ies" are rather random, while the "visual dictionary/-ies" and "picture dictionary/-ies" are truly undertreated in *Lexikos*. In order to appreciate the trend, Figure 2 shows the normalised totals for the same onomasiological terms (expressed as hits per 100 000 words in each volume), as well as the five-year moving average to obtain an even better idea of the true trend. The trendline in Figure 2 clearly shows that while interest in the onomasiological dictionary steadily grew up to and including 2006 in *Lexikos*, interest has declined fast ever since. This, in our view, is a sad development, especially in this day and age of digital solutions.

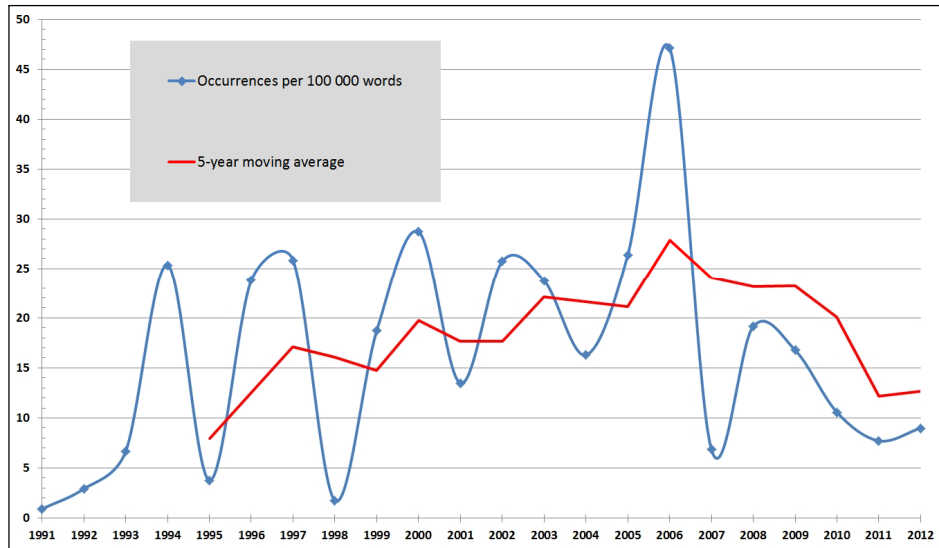


Figure 2: Onomasiological concepts in *Lexikos* (normalised occurrences per volume, and trendline).

4. Beyond the semasiological dictionary: the international dictionary market

In order to move from Africa to the rest of the world, we can start by comparing the trendline arrived at in Figure 2, with the trendline for the same terms as seen in a lexicographic reference corpus of 20 million words that spans the entire world (for more on that lexicographic reference corpus, cf. De Schryver 2012). The result is shown in Figure 3.

From Figure 3 it may be seen that attention to onomasiological aspects is not only consistently higher in the lexicographic reference corpus, but also presently, following a dip during the 1990s, getting ever more popular, unlike what was seen for *Lexikos*. Assuming that more scholarly activity in this field also results in more advanced onomasiological traits in published dictionaries, we conducted the following experiment. On a single day in November 2011, we visited a well-stocked bookshop in the West and checked the variety of dictionaries available, taking pictures of interesting onomasiological features.⁴ We use both our findings and a selection of the photos in the discussion that follows.

We first noted that modern paper dictionaries don't just split along the dichotomy semasiological vs. onomasiological, but that picture dictionaries interact with both of these, resulting in a trichotomy.

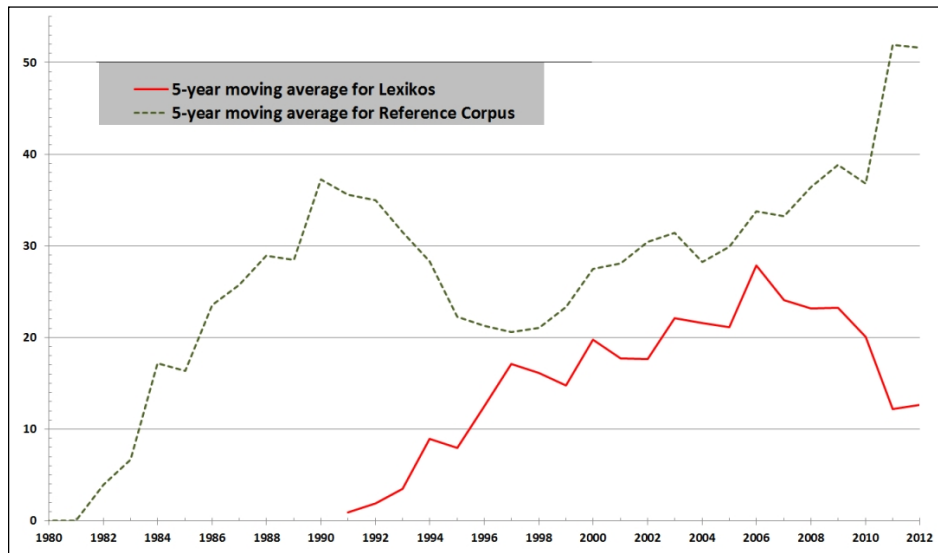


Figure 3: Onomasiological concepts in *Lexikos* vs. the same concepts in a 20-million-word lexicographic reference corpus (trendlines are the five-year moving averages of the normalised occurrences per 100 000 words).

Examples of "pure" semasiological and "pure" onomasiological dictionaries abound, but picture dictionaries tend to have many features of onomasiological dictionaries and often also of semasiological ones. At the same time, it also quickly became clear that semasiological dictionaries include ever more onomasiological features, and that, vice versa, onomasiological dictionaries include ever more semasiological features. We started in the children section, moved to the travel section, then on to the language acquisition section, to end up at the "proper" (advanced/adult) dictionary section.

- The *Beeldwoordenboek Engels* (Van Humbeeck 2001): This is a bilingual picture dictionary English–Dutch, arranged around various themes, which also includes example sentences in English, a mini-grammar and two indexes. Not all collections of drawings bring together similar concepts, however. The pages with "Useful words", for instance, mainly bring together function words and some prepositions, as seen in Addendum 1. These are clearly daunting concepts for a picture dictionary to convey, and pages like these are close to "pure" picture dictionaries: no single concept or even onomasiological field brings the material together; if anything, it is the *function* these words have in the language that brings them together.
- The *Arabic–English Visual Bilingual Dictionary* (Salih et al. 2009): This is basically a picture dictionary showing actual photos grouped around certain concepts and topics, with words in two languages assigned (often by

means of lines) to their respective photos. There are also two indexes. For example, a plate of crockery and cutlery has pictures of (and their words for) mug, coffee cup, plate, bowl, teapot, jug, glassware, fork, spoon, knife, etc. No definitions are given. See Addendum 2. On the cline "pure picture dictionary – pure onomasiological dictionary", a dictionary such as this one sits right in the middle of the continuum.

- The *Thematische woordenschat Engels* (Häublein et al. 2011): This is an example of a thematic dictionary, where the lexicon is arranged around themes and subthemes. Within each subtheme the words are grouped in a logical rather than an alphabetical order (there is also an index), and given a brief lexicographic treatment. Within the subtheme "Free time, hobby, game", for example, one of the groups of words is: *leisure* n., *recreation* n. uncount form, and *relaxing* adj. See Addendum 3. The lexicographic treatment includes word-class assignation, pronunciation, occasionally combinations and/or examples, and translations in a separate column.
- *Le Nouveau Petit Robert: Dictionnaire Alphabétique et Analogique de la Langue Française* (Rey-Debove and Rey 2011): This well-known product of French lexicography hints at both semasiology and onomasiology in the title already (i.e. *alphabétique* 'alphabetical' vs. *analogique* 'analogical'). Although basically an A-to-Z dictionary, and entirely picture-less, it goes to great lengths, throughout the dictionary, to indicate the various lexical relations. In Addendum 4, for instance, one finds, at the end of the article for *agile*, the label "CONTR.", followed by *gauche*, *lent*, *lourd* — the latter three being antonyms of the headword. Under the same article, at sense 1, one finds, following a right-pointing arrow, the words *leste*, *souple*, *vif* — these three being closely related words (synonyms) of the headword in its first sense. Or still, at the article for *agissements*, the label 'FAM.' at the end signifies that the words that follow, *combine* and *magouille*, are synonyms in 'familiar' registers only (i.e. not to be used in 'solemn' situations).
- The *Macmillan Dictionary & Thesaurus* (Rundell et al. 2013): As one moves closer to the traditional A-to-Z type of dictionary, the advanced section in the bookshop holds an unusually large selection of especially English learners' dictionaries. In addition to highly developed A-to-Z sections, all these reference works also contain thematic sections and illustrations. As the latest big-publisher entrant in the series, the Macmillan version could have been featured here, but Macmillan has since decided that "it will no longer be publishing dictionaries in book form" (EURALEX News, 2 November 2012). The move from print to online indeed beautifully allows the full and seamless merging of "the dictionary" and "the thesaurus" (Calzolari 1988), as seen in the online *Macmillan Dictionary & Thesaurus*, and illustrated in Addendum 5. In this online environment, one can not only either consult the dictionary or consult the thesaurus, but also move between the two — and this is crucial — *on sense level*. Starting at the first

sense of *agile*, as shown in Addendum 5a, one can thus literally pull in extra thesauric information for that particular sense (5b), and dig even further into the thesaurus (5c). In an online environment everything is indexed and hyperlinked, allowing for a smooth navigation of all the data, both semasiological and onomasiological, through mere clicks.⁵ There are, however, no illustrations in this dictionary, placing this type right in the middle of the continuum "pure semasiological dictionary – pure onomasiological dictionary".

- The *Collins COBUILD Advanced Dictionary* (Sinclair et al. 2009): This is an alphabetically arranged dictionary with small picture templates, provided on selected pages throughout the dictionary. These are in the form of so-called "Picture Dictionary" boxes and "World Web" boxes. This dictionary also has "Thesaurus" boxes incorporated within some entries throughout the dictionary, to provide synonyms of particular senses of the headword. See Addendum 6. Clearly, not all entries are illustrated with pictures, and even in cases where picture plates are provided, they are far from exhaustive; for example a picture plate for 'bird' has only a handful of birds.⁶
- The *Oxford Learner's Thesaurus: A Dictionary of Synonyms* (Lea et al. 2008): As the name indicates, this work deals with synonyms only. Full dictionary treatments of each major entry and of its synonyms are all provided, again in logical rather than alphabetical order. See Addendum 7a. There are almost no pictures except for a few scattered templates. These few picture templates are interesting, though, because we see mini-articles complete with part-of-speech and sense-number indication as well as phrases being linked to a picture. See Addendum 7b. Accessibility is reinforced by topic and alphabetical indexes at the end of the thesaurus.

What this brief overview shows is that there are increased attempts by A-to-Z lexicographers to include ever-more onomasiological features and illustrations, that thesauri are turned inside out and mimic fully-fledged A-to-Z dictionaries, and that there are at the same time attempts by picture-template developers to include features more typical of both semasiological and onomasiological dictionaries. Semasiological, onomasiological and picture dictionaries are but the extreme ends on continua between them, of course, so what dictionary makers are (no doubt mostly unconsciously) doing is to try finding the optimal point for a particular target user group in the three-dimensional space created by those continua.

5. Alphaconceptual+ lexicography

If we now reduce today's lexicographic reality to its most basic representation, then we can visualise the situation at the base as consisting of a triangle with word-based, concept-based and picture-based as the vertices (corners) of that

triangle. Viewed like this, one immediately recognizes the famous Semantic or Semiotic Triangle of Ogden and Richards (1923: 11), reproduced in Figure 4, an idea which can be traced back to (who else) Aristotle.

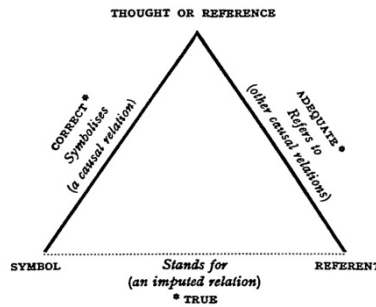
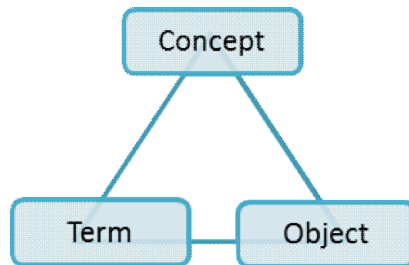


Figure 4: Semantic or Semiotic Triangle of Ogden and Richards (1923: 11).

One way to read this is to say that a graphical form (SYMBOL), say the spelling "c o m p u t e r", symbolises the idea (THOUGHT or REFERENCE), of a computer, which refers to the non-linguistic entity being implied (REFERENT), that is the actual computer in the real world.

Expressed in ISO terms, this could become as shown in Figure 5.



- **Object:** "Any part of the perceivable or conceivable world." [ISO 1087: 2000]
Objects can be material (e.g. a certain house, Tower Bridge) or immaterial (speed, pain, freedom, process). [cf. DIN 2330, p. 3]
- **Concept:** "A unit of thought constituted through abstraction on the basis of properties common to a set of **objects**. The semantic content of a concept can be re-expressed by a combination of other and different concepts, which may vary from one language or culture to another." [ISO 5963: 1985]
- **Term:** "Designation of a defined **concept** in a special language by a linguistic expression." [ISO 1087: 2000]

Figure 5: Semantic or Semiotic Triangle in ISO terms (cf. eCoLoTrain 2007).

Or thus in lexicographic terms as shown in Figure 6, with the word and concept levels corresponding to semasiological and onomasiological dictionaries respectively, and picture dictionaries — or in a digital environment, any multimedia dictionaries — the closest thing to the actual referent in the real world.

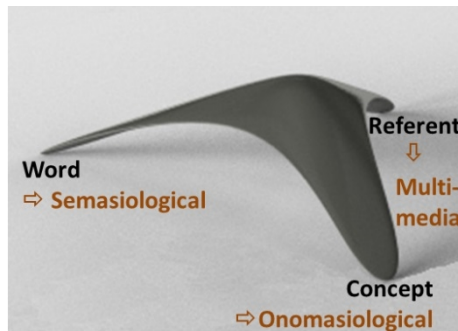


Figure 6: Semasiological, onomasiological and multimedia vertices in lexicography.

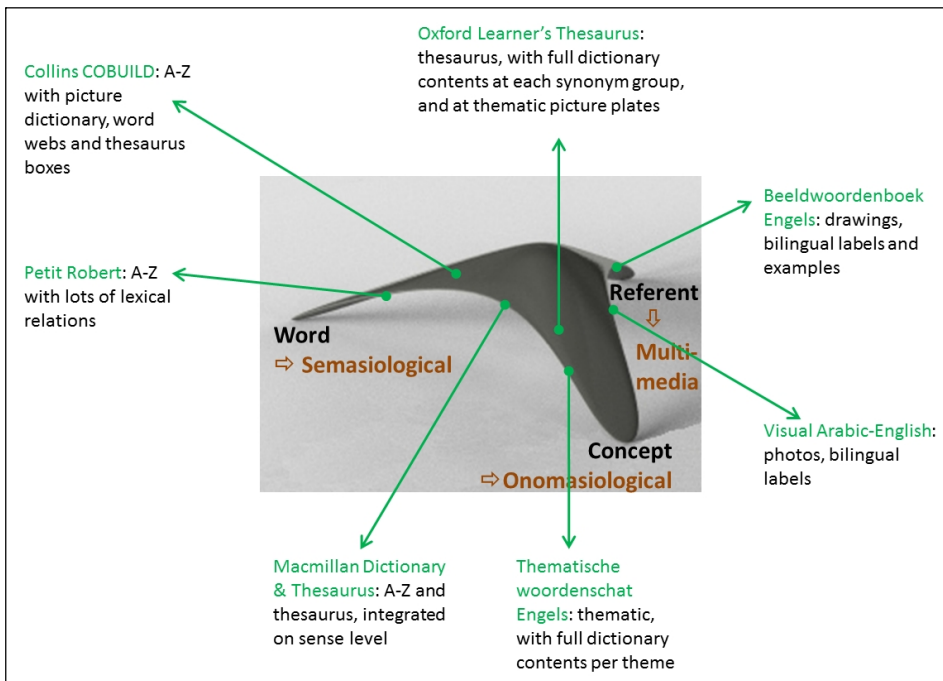


Figure 7: Using the lexicographical draped tripod to catalogue dictionaries.

Rather than a plain triangle, the three continua between the three vertices are allowed to blend into one another, creating the three dimensional surface shown. The visualisation is three-dimensional rather than two-dimensional, as the idea is to visualize the positive multiplication effect of combining features: the higher on the draped tripod — all other things being equal — the better the lexicographic product. This is best illustrated by placing the seven dictionaries and thesauri discussed in Section 4 on this shape, as done in Figure 7.

Five of the seven dictionaries combine features of just two vertices, and are thus positioned on the edge of the three dimensional shape. The distance from a vertex and the height from the base are meaningful to interpret the type of dictionary one is dealing with. Two dictionaries, the *Collins COBUILD Advanced Dictionary* (which stands for any of the Big Five learners' dictionaries) and the *Oxford Learner's Thesaurus: A Dictionary of Synonyms* move away from the edge, onto the surface, indicating that they combine features from the three vertices. The position again reveals the exact mix.

At this point we are in a position to introduce our new lexicographical model, which is the logical furtherance of the combination of the three lexicographical approaches in a digital environment — semasiological, onomasiological and multimedia — with which we end up near the apex of the lexicographical draped tripod. We have baptized this model **alphaconceptual+** (with "+" pronounced "plus"), as it amalgamates the alphabetical (i.e. the approach that departs from the word) and the conceptual (i.e. the approach that departs from the idea), to which a multimedia level is added. It needs to be stressed that what is envisaged is an *integrated* amalgamation, not the mere juxtaposition (or even automatic indexation and hyperlinking) of the various levels. Graphically, an alphaconceptual+ dictionary may be visualised as seen in Figure 8.

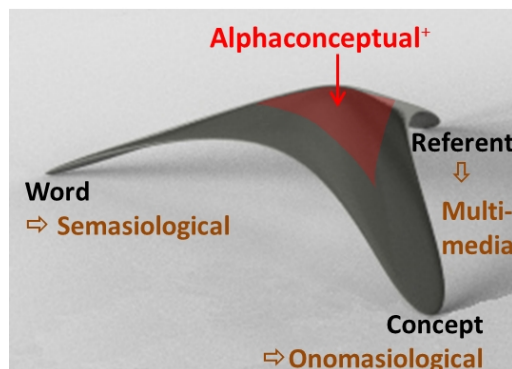


Figure 8: Visual representation of an alphaconceptual+ dictionary.

6. Compiling an alphaconceptual+ dictionary for Luganda

In an alphaconceptual+ dictionary, the search for lexical information starts from either the orthographic word, from the concept, or from a picture plate. Rather than a proof of concept, we will briefly outline how the compilation of such a dictionary could be undertaken for Luganda. The crux of the method, however, is that the entire lexicon would be built around picture plates such as the one seen in Addendum 7b, the electronic version of which is shown in Addendum 7c. As a starting point the classification proposed in Kawalya (2006) would be checked against a recently built Luganda corpus, and modified based on this corpus evidence. Rather than obscure, formulaic labels, a *hierarchy* (ontology) of concrete labels would be used. These labels — at least one, typically several — would be assigned to each sense as appropriate. The point is thus that one may very well compile an A-to-Z dictionary, as long as the onomasiological information is added as one proceeds, and that these two types of information thus end up to be fully integrated. For each group of headwords at the lowest level in the label hierarchy, a sensible number of picture plates would then be composed, to which the headwords would be linked. This aspect, too, would need to be prepared in the same dictionary compilation environment, to assure a seamless integration. In professional lexicography software such as TLex (De Schryver 2011a), it is already possible to change the view from a traditional A-to-Z list, to one based on labels, or indeed, to one where the lexicon is grouped around picture plates. An actual digital dictionary product would thus only need to replicate such a compilation environment, at which point the user will also truly be dealing with an alphaconceptual+ dictionary. In order to test the various proposals in this section, work has begun on the corpus-driven compilation of an alphaconceptual+ dictionary for Luganda, cf. Addendum 8.⁷

7. Conclusion

An alphaconceptual+ dictionary is advantageous because it combines the virtues of the traditional semasiological dictionaries, those of onomasiological dictionaries, and those of picture dictionaries, while at the same time solving the shortcomings of each of those types in isolation. A user can approach the dictionary from the alphabetical angle and proceed to the typical lexicographic information. From there, users can also move on to (or concurrently inspect) the thesaural aspects. Because all entries are likewise linked to appropriate picture plates, users also have the bi-directional option of either moving from word or concept to picture or from picture to word or concept, where either choice enhances the understanding of the word or concept in question. In an electronic environment there is also no need for alphabetical indexes where the thematic information is listed. This considerably reduces the time and stress involved in moving back and forth connecting the words in the index to the words in the main body of the dictionary. The starting point and indeed all

preliminary dictionary compilation work was the result of our wish to produce a new, modern type of dictionary for Luganda, for which there is a dire need. The model itself is of course language-independent.

Endnotes

1. Although Kipfer (1986) labels Urdang's 1986 'Grossetto Workshop' paper as "forthcoming", when selected papers of that landmark workshop were eventually published, as Walker et al. (1995), Urdang's paper was not included.
2. Wilkins (1668) invented the principle on which the Roget's thesaurus is based two centuries before Roget (cf. Hüllen 2004: 285; De Schryver 2008: 422).
3. That Uganda's President, Y.K. Museveni, is an adept and accomplished lexicographer became clear when he launched the first monolingual Lusoga dictionary (Nabirye 2009) in Kampala on 8 October 2010 (cf. e.g. Jaramogi 2010, De Schryver 2011). On that occasion his laudatio included an in-depth analysis of the power of African languages and the need for dictionaries for these languages — a talk that could easily have passed as a keynote at Afri-lex.
4. The shop visited was the FNAC, in the city of Ghent, Belgium, on 25 November 2011.
5. Unfortunately, the indexing and hyperlinking is typically an automated process, which results in quite a large number of nonsensical links. For example, clicking on "Bailey" in the example sentence *Bailey became increasingly aggressive in his questioning of the witness*, at sense 1 of *aggressive*, leads to the definition for *bailey* as "an area of land between the outer and inner walls of a castle". Likewise, clicking on "becoming" in the first definition for *quell*, "to get rid of unpleasant thoughts or feelings, or to prevent them from becoming stronger", leads to the following definitions for *becoming*: "1. making you look attractive; 2. suitable or correct for you or the situation that you are in". Clearly, not what it should be. (On the cluttered page, one does find a box with "Related dictionary definitions", however, where *become* VERB and *become of* PHRASAL VERB are listed, but that means far more clicks where a single one should have sufficed. Plus, at *bailey*, the only other option is a link to 'the old Bailey', i.e. the Central Criminal Court in London, which in combination with the example sentence under *aggressive* will only confuse even more.) In making the index, the dictionary data should have been lemmatised, so nouns at least lead to entries for nouns, verbs to entries for verbs, etc. This will avoid that a click on "drink" in *food and drink* (under *help* VERB) will lead to the verb *drink*, or that a click on "helping" in *He'd been helping himself ...* (in the same article) will lead to the noun *helping*.
6. Even though the picture plate is labelled "bird", this may not be a fair assessment. The front matter explains: "'Word Webs' present topic-related vocabulary through encyclopedia-like readings combined with stunning art, creating opportunities for deeper understanding of the language and concepts." (p. viii). This particular picture plate is actually about the *Sinornithosaurus*, an avian dinosaur, from which modern birds may have evolved.
7. At this stage, the label ontology has not yet been (re)developed; the aim was to see if the TLex software could handle the various aspects of the model.

References

- Bagunywa, A.M., S.S. Kyakulumbye, S.W. Muwonge and W. Ssentooogo.** 2009. *A Concise Luganda–English Dictionary*. Kampala: Fountain Publishers.
- Bennett, J. and N. Tsoeu.** 2006. *Multilingual Illustrated Dictionary*. Cape Town: Pharos.
- Benson, T.G.** 1964. A Century of Bantu Lexicography. *African Language Studies* 5: 64-91.
- Bukenya, A. and L. Kamoga.** 2009. *Fountain Standard Luganda–English Dictionary*. Kampala: Fountain Publishers.
- Bunduki, K.-N.** 1975. *Essai de lexique linguistique français–ciluba*. Lubumbashi: CELTA.
- Burkhanov, I.** 2005. Requirements for a Learner-Oriented Onomasiological Dictionary of the General Vocabulary. *Lexicographica* 21: 122-130.
- Calzolari, N.** 1988. The Dictionary and the Thesaurus Can Be Combined. Evens, M.W. (Ed.). 1988. *Relational Models of the Lexicon. Representing Knowledge in Semantic Networks*. Studies in Natural Language Processing: 75-96. Cambridge: Cambridge University Press.
- Cawdrey, R.** 1604. *A Table Alphabeticall, conteyning and teaching the true writing, and vnderstanding of hard vsuall English wordes, borrowed from the Hebrew, Greeke, Latine, or French. &c. With the interpretation thereof by plaine English words, gathered for the benefit & helpe of Ladies, Gentlewomen, or any other vnskilfull persons. Whereby they may the more easilie and better vnderstand many hard English wordes, which they shall heare or read in Scriptures, Sermons, or elsewhere, and also be made able to vse the same aptly themselues*. London: Edmund Weaver.
- De Schryver, G.-M.** 1999. *Bantu Lexicography and the Concept of Simultaneous Feedback, Some Preliminary Observations on the Introduction of a New Methodology for the Compilation of Dictionaries with Special Reference to a Bilingual Learner's Dictionary Cilubà–Dutch*. Unpublished M.A. Dissertation. Ghent: Ghent University.
- De Schryver, G.-M.** 2003. Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography* 16(2): 143-199.
- De Schryver, G.-M.** 2008. Book Review: P. Hanks, ed. 2008. *Lexicology: Critical Concepts* (Routledge, six volumes, 2 793 pages). *International Journal of Lexicography* 21(4): 419-438.
- De Schryver, G.-M.** 2011. *AFRILEX President's Annual Report: 2010–2011*. Presented at the AGM of the AFRILEX 2011 Conference, 6 July 2011. Available online at: http://afrilex.africanlanguages.com/presreport1011_text.pdf (text), http://afrilex.africanlanguages.com/presreport1011_slides.pdf (photos) [last accessed on 10 September 2013].
- De Schryver, G.-M.** 2011a. Why Opting for a Dedicated, Professional, Off-the-shelf Dictionary Writing System Matters. Akasu, K. and S. Uchida (Eds.). 2011. *ASIALEX 2011 Proceedings. LEXICOGRAPHY: Theoretical and Practical Perspectives. Papers Submitted to the Seventh ASIALEX Biennial International Conference, Kyoto Terrsa, Kyoto, Japan, August 22–24, 2011*: 647-656. Kyoto: The Asian Association for Lexicography.
- De Schryver, G.-M.** 2012. Trends in Twenty-five Years of Academic Lexicography. *International Journal of Lexicography* 25(4): 464-506 + 42 pages of supplementary material online.
- DIN = Deutsches Institut für Normung e.V.** Available online at: <http://www.din.de/> [last accessed on 10 September 2013].
- eCoLoTrain.** 2007. *TERMI — Basic Principles*. Available online at: <http://ecolotrain.uni-saarland.de/index.php?id=1814&L=1> [last accessed on 10 September 2013].
- EURALEX News.** 2012. *End of Print Dictionaries at Macmillan*, 2 November 2012. Available online at: <http://www.euralex.org/news/end-of-print-dictionaries-at-macmillan/> [last accessed on 10 September 2013].

- Gavira, A. (Senior Editor) et al.** 2005. *Spanish–English Visual Bilingual Dictionary*. London: Dorling Kindersley Limited.
- Hartmann, R.R.K.** 1983. The Bilingual Learner's Dictionary and its Uses. *Multilingua — Journal of Cross-Cultural and Interlanguage Communication* 2(4): 195-201.
- Häublein, G., R. Jenkins and Y. Pullens.** 2011. *Thematische woordenschat Engels*. Amsterdam: Inter-taal.
- Hüllen, W.** 1999. *English Dictionaries 800–1700: The Topical Tradition*. Oxford: Clarendon.
- Hüllen, W.** 2004. *A History of Roget's Thesaurus: Origins, Development, and Design*. Oxford: Oxford University Press.
- ISO** = *International Organization for Standardization*. Available online at: <http://www.iso.org/> [last accessed on 10 September 2013].
- Jackson, H.** 1988. *Words and Their Meaning*. London: Longman.
- Jaramogi, P.** 2010. *President Museveni launches Lusoga Dictionary*. Available online at: <http://www.mediacentre.go.ug/details.php?catId=4&item=1036> [last accessed on 10 September 2013].
- Kadima, B.A.** 1995. *L'information sanitaire, lexique français–ciluba, ciluba–français*. Kinshasa: Éditions Universitaires Africaines.
- Kadima, B.A. and Katende wa Ndaya.** 1995. *Signalisation routière, lexique français–ciluba, ciluba–français*. Kinshasa: Éditions Universitaires Africaines.
- Kadima, B.A., T.T. Mukendi and B. Mpyoi.** 1995. *Terminologie grammaticale et pédagogie [sic], lexique français–ciluba, ciluba–français*. Kinshasa: Éditions Universitaires Africaines.
- Kawalya, D.** 2003. *Enkuluze y'Obuwangwa bw'Abaganda Eyeesigamiziddwa ku Bitegeero*. Unpublished B.A. Dissertation. Kampala: Makerere University.
- Kawalya, D.** 2006. *Conceptual Lexicography: A Case of Luganda*. Unpublished M.A. Dissertation. Kampala: Makerere University.
- Kiingi, K.B.** 2005. *Lexical Taxonomy*. Available online at: <http://luganda.canadianwebs.com/LEXICAL%20TAXONOMY.pdf> [last accessed 10 September 2013].
- Kiingi, K.B. (Chief Compiler) et al.** 2007. *Enkuluze y'Oluganda ey'e Makerere*. Kampala: Fountain Publishers.
- Kipfer, B.N.** 1986. Investigating an Onomasiological Approach to Dictionary Material. *Dictionaries* 8: 55-64.
- Kitching, A.L. and G.R. Blackledge.** 1925. *A Luganda–English and English–Luganda Dictionary*. London: Society for Promoting Christian Knowledge.
- Lea, D. (Chief Editor) et al.** 2008. *Oxford Learner's Thesaurus: A Dictionary of Synonyms*. Oxford: Oxford University Press.
- Le Veux, H.** 1917. *Premier essai de vocabulaire luganda–français d'après l'ordre étymologique*. Maison-Carrée (Algiers): Imprimerie des Missionnaires d'Afrique.
- Lewis, M.P., Simons, G.F. and C.D. Fennig (Eds.).** 2013. *Ethnologue: Languages of the World, Seventeenth edition*. Dallas, TX: SIL International. Available online at: <http://www.ethnologue.com/> [last accessed 10 September 2013].
- Maalu-Bungi, L.L., N. Kazadi, K. Badibanga and B. Mpyoi.** 1983. *Lexiques thématiques de l'Afrique centrale (LETAC), Zaïre, Cilubà, 1. Activités économiques et sociales*. Paris : ACCT-CERDOTOLA.
- Mulira, E.M.K. and E.G.M. Ndawula.** 1952. *A Luganda–English, English–Luganda Dictionary*. Second edition, first published in 1925. London: Society for Promoting Christian Knowledge.
- Murphy, J.D.** 1972. *Luganda–English Dictionary*. Washington, DC: The Catholic University of America Press.

- Museveni, Y.K., M.J.K. Muranga, G. Gumoshabe and A.N. Muhoozi.** 2012. *Katondoozi y'Orunyankore-Rukiga*. Kampala: Fountain Publishers.
- Nabirye, M.** 2009. *Eiwanika ly'Olusoga. Eiwanika ly'aboogezi b'Olusoga n'abo abenda okwegwa Olusoga*. Kampala: Menha Publishers.
- Nzongola, P.R.K.** 1967. *Dictionnaire des synonymes tshiluba*. Luebo: J. Leighton Wilson Press.
- Ogden, C.K. and I.A. Richards.** 1923. *The Meaning of Meaning. A Study of The Influence of Language upon Thought and of The Science of Symbolism*. London: Kegan Paul, Trench, Trübner & Co.
- Paizee, D. et al.** 2007–2009. *First Bilingual Dictionary* [series of ten dictionaries]. Cape Town: Oxford University Press Southern Africa.
- Rey-Debove, J. and A. Rey.** 2011. *Le Nouveau Petit Robert: Dictionnaire Alphabétique et Analogique de la Langue Française*. Paris: Dictionnaires Le Robert.
- Roget, P.M.** 1852. *Roget's Thesaurus*. London: Longman.
- Roget, P.M., J.L. Roget and S.R. Roget.** 1966. *Roget's Thesaurus of English Words and Phrases: Arranged and Classified so as to Facilitate the Expression of Ideas and to Assist Literary Composition*. Brooklyn, NY: P. Shalom.
- Rundell, M. et al.** 2013. *Macmillan Dictionary & Thesaurus*. Available online at: <http://www.macmillandictionary.com/> [last accessed 10 September 2013].
- Salih, S. (Translator) et al.** 2009. *Arabic–English Visual Bilingual Dictionary*. London: Dorling Kindersley Limited.
- Sierra, G.** 2000. *The Onomasiological Dictionary: A Gap in Lexicography*. 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*: 223–235. Stuttgart: Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.
- Sinclair, J.M. (Founding Editor-in-Chief) et al.** 2009. *Collins COBUILD Advanced Dictionary*. Glasgow: Harper Collins Publishers.
- Snoxall, R.A.** 1967. *Luganda–English Dictionary*. London: Oxford University Press.
- Urdang, L.** 1986. Lexicographic Practice as a Source of Information for Designing Dictionaries Intended for Electronic Access. Unpublished paper presented at the *Workshop on 'Automating the Lexicon: Research and Practice in a Multilingual Environment'*, Marina di Grosseto, Italy, 19–23 May 1986.
- Van Humbeeck, S.** 2001. *Beeldwoordenboek Engels, Snel en vlot 1000 Engelse woorden leren* [Original title: *Mon Premier Larousse d'Anglais*]. Aartselaar: Zuidnederlandse Uitgeverij.
- Various.** 2010. *Longman Foundation Phase Bilingual Dictionary* [series of five dictionaries]. Cape Town: Maskew Miller Longman.
- Walker, D.E., A. Zampolli and N. Calzolari (Eds).** 1995. *Automating the Lexicon: Research and Practice in a Multilingual Environment*. Oxford: Oxford University Press.
- Wilkins, J.** 1668. *An Essay Towards a Real Character and a Philosophical Language*. London: S. Gellibrand.
- Yukawa, Y.** 1992. *A Classified Vocabulary of the Luba Language*. Tokyo : ILCAA.

Addendum 1: Beeldwoordenboek Engels (Van Humbeeck 2001: 9)

<p>only slechts 'Only one, please!'</p>	<p>perhaps misschien Perhaps it will rain, perhaps not.</p>	<p>with met, samen met Cedric with Fred.</p>
<p>or of 'Chocolate or vanilla?'</p>	<p>very heel The teacher is very angry.</p>	<p>without zonder Cedric without Fred.</p>
<p>yes ja</p>	<p>no neen</p>	<p>from uit, van Pedro comes from Mexico.</p>
		<p>to naar The baby is going to the park.</p>


Addendum 2: *Arabic-English Visual Bilingual Dictionary* (Salih et al. 2009: 65)



Addendum 3: Thematische woordenschat Engels (Häublein et al. 2011: 191)

Vrije tijd, ontspanning, sport, spel 10

10 Vrije tijd, ontspanning, sport, spel

10.1 Vrije tijd, hobby, spel  track 31

relaxation <i>n</i> [ˌrɪːləkˈseɪʃən]	ontspanning
relax <i>v</i> [rɪˈlæks]	(zich) ontspannen
time <i>n uncount</i> [taɪm]	tijd
What do you like to do most in your leisure time ? <i>BE</i>	Wat doe jij het liefst in je vrije tijd?
In his kind of job, Geoffrey doesn't get much free time .	Met een dergelijke baan heeft Geoffrey maar weinig vrije tijd.
In our spare time we like to help out at our church.	In onze vrije tijd helpen we graag in de kerk.
leisure <i>n</i> [ˈleɪʒə]	vrije tijd
recreation <i>n uncount form</i> [ˌrekreɪˈeɪʃən]	recreatie; tijdverdrijf
Steve's main form of recreation is playing golf.	Steve brengt zijn vrije tijd vooral door met golfen.
relaxing <i>adj</i> [rɪˈlæksɪŋ]	ontspannend
activity <i>n</i> [ækˈtɪvətɪ]	activiteit, bezigheid
event <i>n</i> [ɪˈvent]	gebeurtenis, evenement
Our town fair is a popular summer event .	De kermis in onze stad is een populair zomerevenement.
a cultural/social event	een cultureel/sociaal evenement
cultural <i>adj</i> [ˈkʌltʃərəl]	cultureel
culture <i>n uncount</i> [ˈkʌltʃə]	cultuur
leisure centre <i>BE n</i> [ˈleɪʒəˌsentə]	(sport- en) recreatiecentrum
recreation center <i>AE n</i> [ˌrekreɪˈeɪʃənˌsentə]	(sport- en) recreatiecentrum
recreational <i>adj</i> [ˌrekreɪˈeɪʃənəl]	recreatie-, vrije tijds-
facilities <i>n pl</i> [fəˈsɪlɪtɪz]	faciliteiten; voorzieningen
The town has built new recreational facilities which are open to residents and visitors.	De stad heeft nieuwe recreatievoorzieningen die open zijn voor zowel inwoners als toeristen.
fun <i>n uncount</i> [fʌn]	plezier
What do you do when you want to have fun ?	Wat doe jij als je plezier wilt maken?
entertain <i>v</i> [ˌentəˈteɪn]	gastvrij ontvangen; vermaken
We love to entertain friends at home.	We nodigen thuis graag vrienden uit.
entertainment <i>n uncount</i> [ˌentəˈteɪnmənt]	vermaak

Addendum 4: *Le Nouveau Petit Robert: Dictionnaire Alphabétique et Analogique de la Langue Française* (Rey-Debove and Rey 2011: 49)

AGHA [aga] n. m. var. **AGA** – 1535 ◊ mot turc « chef » ■ Officier de la cour du sultan, dans l'ancienne Turquie. ◊ En Algérie, Chef au-dessus du caïd. *Des aghas.* ◊ *Agha Khan*, titre du chef spirituel des musulmans ismaéliens.

AGILE [aʒil] adj. – XIV^e ◊ latin *agilis*, de *agere* ■ **1** Qui a de la facilité et de la rapidité dans l'exécution de ses mouvements. > **leste, souple, vif.** « *La bohémienne dansait [...] agile, légère.* » **HUGO.** « *Le clown agile.* » **VERLAINE.** *Les doigts agiles du pianiste.* ■ **2** FIG. Prompt dans les opérations intellectuelles. *Un esprit agile.* ■ **CONTR.** Gauche, lent, lourd.

AGILEMENT [aʒilmɑ̃] adv. – XIV^e ◊ de *agile* ■ Avec agilité.

AGILITÉ [aʒilite] n. f. – XIV^e ◊ latin *agilitas* ■ Qualité de ce qui est agile. > **aisance, légèreté, rapidité, souplesse, vivacité.** *Ils « se hissaient le long des murs avec une agilité et une malice toutes simiesques »* **BARRÈS.** FIG. *L'agilité de sa pensée.* ■ **CONTR.** Gaucherie, lenteur, lourdeur.

AGIO [aʒjo] n. m. – 1679 ◊ italien *aggio* ■ **1** vx Différence entre la valeur nominale et la valeur réelle d'échange d'une monnaie, d'un effet; le profit que cette différence permettait de réaliser à l'agioteur. ■ **2** MOD. Rémunération (intérêt, commission) perçue par une banque, un intermédiaire à l'occasion de certaines opérations. *Frais d'agio pour découvert bancaire.*

A GIORNO - GIORNO (À)

AGIOTAGE [aʒjɔtaʒ] n. m. – début XVII^e ◊ de *agioter*, de *agio* ■ HIST. « L'étude et l'emploi de manœuvres les moins délicates pour produire des variations inattendues dans le prix des effets publics et tourner à son profit les dépouilles de ceux qu'on a trompés » [Mirabeau]. > **spéculation.** *Se livrer à l'agiotage* (**AGIOTER** v. intr. (1)).

AGIOTEUR, EUSE [aʒjɔtœr, øz] n. – début XVII^e ◊ de *agioter*, de *agio* ■ HIST. Personne qui se livrait à l'agiotage. > **spéculateur.** ◊ Aujourd'hui, Spéculateur utilisant des informations obtenues plus ou moins malhonnêtement pour influencer le cours des valeurs à son profit. > **initié.**

ÂGISME [aʒism] n. m. – 1985 ◊ de *âge*, par anal. avec *racisme* ■ Discrimination envers toute personne âgée. ■ **CONTR.** Jeunisme.

AGISSANT, ANTE [aʒisɑ̃, ɑ̃t] adj. – 1584 ◊ de *agir* ■ Qui agit effectivement, se manifeste par des effets tangibles. > **actif, effectif, efficace.** *Force agissante.* « *Une foi vive et agissante.* » **BOURDALOUE. *Une minorité agissante.* ■ **CONTR.** Inactif, inefficace.**

AGISSEMENTS [aʒismɑ̃] n. m. pl. – 1794 ◊ de *agir* ■ Suite de procédés et de manœuvres condamnables. > **intrigue, machination, manigance, menées; FAM. combine, magouille.**

AGITATEUR, TRICE [aʒitɑtœr, tris] n. – 1792; « cocher » 1520, sens du latin *agitor*; « représentant de l'armée parlementaire anglaise » 1687 ◊ de *agiter* ■ **1** Personne qui crée ou entretient l'agitation politique ou sociale. > **factieux, meneur, trublion.** *Les agitateurs royalistes de la Vendée.* « *De dangereux agitateurs, tel Oulianov, le futur Lénine.* » **BAINVILLE.** ■ **2** n. m. (1838) Instrument de laboratoire, dispositif servant à agiter des liquides, brasser des mélanges. — *Agitateur jetable pour boisson.* > **FAM. touillette.**

AGITATION [aʒitasjɔ̃] n. f. – 1355 ◊ latin *agitatio* ■ **1** État de ce qui est agité, parcouru de mouvements irréguliers en divers sens. > **trouble, turbulence.** « *l'agitation de l'eau.* » **ROUSSEAU.** « *Les habitants avaient l'agitation d'une ruche inquiétée.* » **HUGO.** « *cette agitation [...] qui secoue les boulevards à la sortie des théâtres.* » **MAUPASSANT.** > **animation, branle-bas, grouillement, remue-ménage.** — **PHYS.** *Agitation thermique*: mouvement spontané et continu des molécules les unes par rapport aux autres, qui augmente avec la température. ■ **2** État d'une personne en proie à des émotions et à des impulsions diverses, et qui ne peut rester en repos. > **fièvre, nervosité.** « *À mes jours d'agitation succédaient des jours de torpeur.* » **FRANCE.** *Son agitation augmentait avec l'attente.* ◊ **PSYCHIATR.** Manifestation extérieure d'un état d'excitation psychique et motrice. ■ **3** Mécontentement d'ordre politique ou social se traduisant par des manifestations, des revendications, des troubles. > **effervescence, fermentation, FAM. grogne, remous.** *L'agitation étudiante.* ■ **CONTR.** 1 Calme, paix, repos.

AGITATO [aʒitato] adv. – 1791 ◊ mot italien « agité » ■ **MUS.** Indication de mouvement musical, de caractère passionné, tourmenté.

Addendum 5: *Macmillan Dictionary & Thesaurus* (Rundell et al. 2013)

(5a) The dictionary entry *agile* in this online dictionary:

agile - definition

ADJECTIVE  Pronunciation /'ædʒaɪl/

- able to move quickly and easily
The ferret is an agile hunter.
 Thesaurus entry for this meaning of agile
- able to think quickly, solve problems, and have new ideas
an agile mind
 Thesaurus entry for this meaning of agile

(5b) Clicking on "Thesaurus" in (5a) shows basic thesaurus information:

agile - thesaurus entry

ADJECTIVE

1 able to move quickly and easily

Synonyms or related words for this sense of agile

Words used to describe fast movement: *fast, quick, swift, speedy, nimble, brisk, high-speed, agile, deft, at full speed/tilt/pelt/throttle...* [more](#)

2 able to think quickly, solve problems, and have new ideas

Synonyms or related words for this sense of agile

Words used to describe intelligent or wise people: *genius, intelligent, wise, brilliant, clever, bright, brainy, discerning, clear-sighted, no/nobody's fool...* [more](#)

(5c) Clicking on "more" in (5b) presents the full thesaurus information for a particular sense (here sense 1):

Words used to describe fast movement

fast ADJECTIVE

able to move quickly

quick ADJECTIVE

able to move fast or do something fast

swift ADJECTIVE

moving quickly

speedy ADJECTIVE

happening very quickly

nimble ADJECTIVE

able to move quickly and easily

brisk ADJECTIVE

moving or acting quickly

high-speed ADJECTIVE

moving or operating very quickly

agile ADJECTIVE

able to move quickly and easily

deft ADJECTIVE

deft movements are made quickly and with skill

at full speed/tilt/pelt/throttle

as fast as possible

Addendum 6: Collins COBUILD Advanced Dictionary (Sinclair et al. 2009: 491)

eco-
491
economist

Word Web eclipse

When the **earth** passes between the **sun** and the **moon**, we see a **lunar eclipse**. When the moon passes between the sun and the earth, we see a **solar eclipse**. A total eclipse of the sun happens when the moon covers it completely. In the past, people were frightened of eclipses. Leaders of some civilizations understood eclipses. They pretended to control the sun in order to gain the respect of their people. The next total eclipse of the sun visible in the United Kingdom will be 23 September, 2090.



is no longer noticed because the second thing gets all the attention. □ [be v-ed + by] *The gramophone had been eclipsed by new technology such as the compact disc.*
→ see Word Web: **eclipse**

eco- /iːkəʊ-/ PREFIX **Eco-** combines with nouns and adjectives to form other nouns and adjectives which describe something as being related to ecology. □ ...*the eco-horror of the North Sea oil spill.*

eco-friendly ADJ **Eco-friendly** products or services are less harmful to the environment than other similar products or services. □ ...*eco-friendly washing powder.*

ecological /iːkɒlədʒɪkəl/ [N] ADJ [ADJ N] **Ecological** means involved with or concerning ecology. □ *Large dams have harmed Siberia's delicate ecological balance.* •**ecologically** /iːkɒlədʒɪkli/ ADV [ADV adj]/-ed, ADV after v □ It is **ecological** to run and **ecologically sound**. [ADJ] [ADJ N] **Ecological** groups, movements, and people are concerned with preserving the environment and natural resources, so that they can be used in a sensible way, rather than being wasted. □ *Ecological groups say that nothing is being done to tackle the problem.*

ecologist /iːkɒlədʒɪst/ (ecologists) [N-COUNT] An **ecologist** is a person who studies ecology. □ *Ecologists argue that the benefits of treating sewage with disinfectants are doubtful.*
[N-COUNT] An **ecologist** is a person who believes that the environment and natural resources should be preserved and used in a sensible way, rather than being wasted. □ *In the opinion polls the ecologists reached 20 per cent.*

ecology /iːkɒlədʒi/ (ecologies) [N-UNCOUNT] **Ecology** is the study of the relationships between plants, animals, people, and their environment, and the balances between these relationships. □ ...*a senior lecturer in ecology.* [N-VAR] When you talk about the **ecology** of a place, you are referring to the pattern and balance of relationships between plants, animals, people, and the environment in that place. □ [+ of] ...*the ecology of the rocky Negev desert in Israel.*
→ see air

Word Link e = electronic : e-book, e-commerce, e-mail

e-commerce N-UNCOUNT **E-commerce** is the same as **e-business**. [BUSINESS] □ ...*the anticipated explosion of e-commerce.*

economic /iːkənɒmɪk, ɡk-/ [ADJ] [usu ADJ N] **Economic** means concerned with the organization of the money, industry, and trade of a country, region, or society. □ ...*Poland's radical economic reforms.* □ *The pace of economic*

growth is picking up. •**economically** /iːkənɒmikli, ɡk-/ ADV [ADV adj]/-ed, ADV after v □ ...*an economically depressed area.*
□ *Economically and politically, this affair couldn't come at a worse time.* [ADJ] If something is **economic**, it produces a profit.
□ *The new system may be more economic but will lead to a decline in programme quality.*

economical /iːkənɒmɪkəl, ɡk-/ [ADJ] [oft ADJ to-inf] Something that is **economical** does not require a lot of money to operate. For example a car that only uses a small amount of petrol is **economical**. □ ...*plans to trade in their car for something smaller and more economical.* □ It is more **economical** to wash a full load. •**economically** ADV [ADV after v] □ *Services could be operated more efficiently and economically.* [ADJ] Someone who is **economical** spends money sensibly and does not want to waste it on things that are unnecessary. A way of life that is **economical** does not need a lot of money. □ ...*ideas for economical housekeeping.* [ADJ] [usu v-link ADJ] **Economical** means using the minimum amount of time, effort, or language that is necessary. □ *His gestures were economical, his words generally mild.* •**economically** ADV □ *Burn's novel, vividly and economically written, is a sombre reflection on fame and its cost.*

Thesaurus economical Also look up:
ADJ. cost-effective, inexpensive [careful, frugal, practical, thrifty]

Word Link ics = system, knowledge : economics, electronics, ethics

economics /iːkənɒmɪks, ɡk-/ [N-UNCOUNT] **Economics** is the study of the way in which money, industry, and trade are organized in a society. □ *He gained a first class Honours degree in economics.* [N-VAR] → see also **home economics** [N-UNCOUNT] The **economics** of a society or industry is the system of organizing money and trade in it. □ [+ of] ...*the economics of the third world.* □ *He is regarded as a committed supporter of a radical free-market economics policy.*
→ see Word Web: **economics**

economies of scale N-PLURAL **Economies of scale** are the financial advantages that a company gains when it produces large quantities of products. [BUSINESS] □ *Car firms are desperate to achieve economies of scale.*

economist /iːkənɒmɪst/ (economists) [N-COUNT] An **economist** is a person who studies, teaches, or writes about economics.

Word Web economics

The study of **economics** explores how a society distributes its **wealth**. This subject is divided into two main areas: **macroeconomics** and **microeconomics**. **Macroeconomics** looks at how a society as a whole handles money, **capital**, and **commodities**. **Microeconomics** focuses on individuals and businesses. A key microeconomic principle is the law of **supply and demand**. This theory says that prices of **goods** and **services** are based on a balance between two factors. The first is how much of something is available (supply). The second is how much people are willing to pay for it (demand).



Addendum 7a: Oxford Learner's Thesaurus: A Dictionary of Synonyms (Lea et al. 2008: 466)

enchanted /ɪn'tʃɑ:ntrɪd; AmE -'tʃæntrɪd/ (written) placed under a spell (= magic words that have special powers): *The children made their way through the enchanted forest, their hearts beating fast.* ❶ **Enchanted** is often used in children's stories to talk about places where fairies and other magical creatures live.

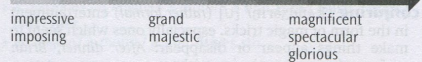
occult /ə'kʌlt; 'bʌkʌlt; AmE 'ɑ:k-/ [only before noun] (written) supernatural, especially when it is connected with evil and secrecy: *They assured us that their organization did not follow occult practices.*

transcendental /træn'sendntɪl/ [usually before noun] (written) going beyond the limits of human knowledge, experience or reason, especially in a religious or spiritual way: *He said that it was transcendental meditation that brought him inner peace.* ❶ **Transcendental** is most often used to talk about an experience, meditation, reflection or a state.

other-worldly (written) concerned with spiritual thoughts and ideas rather than with ordinary life: *There was an other-worldly quality to her performance that night.* ❶ **Other-worldly** is usually used to talk about strange experiences, noises, states or qualities, which give the impression of being in a dream-world.

magnificent *adj.* See also the entries for BEAUTIFUL 2 and WONDERFUL

magnificent • impressive • spectacular • grand • glorious • majestic • imposing
These words all describe things that are extremely attractive and make you feel admiration.



PATTERNS AND COLLOCATIONS

- ▶ a magnificent / an impressive / a spectacular / a glorious / a majestic **sight**
- ▶ a magnificent / an impressive / a spectacular / a glorious **display / array**
- ▶ magnificent / impressive / spectacular / grand / glorious / majestic **scenery / views**
- ▶ magnificent / spectacular / glorious **countryside**
- ▶ a magnificent / an impressive / a spectacular **waterfall**
- ▶ a magnificent / spectacular / glorious **coastline / sunset**
- ▶ magnificent / spectacular / majestic **mountains**
- ▶ a magnificent / an impressive / a grand / a majestic / an imposing **building**
- ▶ a magnificent / a grand / a majestic / an imposing **castle**
- ▶ a magnificent / a grand / an imposing **palace / staircase**
- ▶ glorious / majestic **beauty**
- ▶ **quite** magnificent / impressive / spectacular / grand
- ▶ **truly / rather** magnificent / impressive / spectacular / grand / glorious
- ▶ **absolutely** magnificent / spectacular / glorious
- ▶ **very** impressive / spectacular / grand / majestic / imposing

magnificent /mæg'nɪfɪsnt/ extremely beautiful in a way that makes you feel wonder and admiration: *The Taj Mahal is a magnificent building.* ◊ *She looked magnificent in her wedding dress.* ◊ *It was an absolutely magnificent performance.*
▶ **magnificence** *noun* [U]: *the magnificence of the scenery*
▶ **magnificently** *adv.*: *The city boasts a wealth of magnificently preserved temples and palaces.*

impressive making you feel admiration, because it is very large, attractive or well or expensively made or built: *A large portico provides a suitably impressive entrance to the chapel.* ◊ *This is one of the most impressive novels of recent years.* ❶ **Impressive** is often used when you feel great admiration for a place or thing but do not necessarily find it beautiful. See also **impressive** → IMPRESSIVE

spectacular /spek'tækjələ(r)/ (especially of scenery or a performance) extremely beautiful and impressive: *The coastal road has spectacular scenery.* ◊ *In the evening, there will be a spectacular display of fireworks.* See also **spectacular** → IMPRESSIVE, **brehtaking** → AMAZING

grand impressive and large or important: *It's not a very grand house.* ◊ *The wedding was a very grand occasion.* ❶ When **grand** is used to describe a thing, it is usually sth relating to a building or part of a building such as a house, villa, hotel, castle, palace, staircase or entrance. When **grand** is used to describe an occasion, it is usually one that involves rich people or the spending of a large amount of money. See also **grandeur** → LUXURY

glorious /glɔ:'riəs/ extremely beautiful and impressive: *We sat on the beach and gazed at the glorious sunset.* ◊ *Both her daughters had glorious red hair.* ❶ **Glorious** is used especially to describe things that are bright and connected with summer or the colours of the sun – yellow, orange or red. See also **glory** → LUXURY

majestic /mə'dʒestɪk/ very impressive because of its size or beauty: *The college is close to Edinburgh's majestic castle.* ◊ *The Rockies are majestic in size.* ❶ **Majestic** is usually used to describe large and impressive buildings such as castles, mansions or hotels, or high natural features such as mountains or cliffs.

▶ **majestically** *adv.*: *The cliffs rise majestically from the ocean.*

imposing /ɪm'pəʊzɪŋ; AmE -'pouz-/ impressive to look at; making a strong impression: *The Tower House is not a particularly grand or imposing building.* ◊ *She was a tall imposing woman.* ❶ **Imposing** people or things are likely to make you feel respectful, but also rather small or unimportant.

main *adj.* See also the entry for TOP

main • major • key • central • principal • chief • prime • primary • number one • predominant
These words all describe sb/sth that is the largest or most important of its kind.

PATTERNS AND COLLOCATIONS

- ▶ to be key / central **to** sth
- ▶ a / the main / major / key / central / principal / chief / prime / primary / number one / predominant **concern**
- ▶ a / the main / major / key / central / principal / chief / prime / primary / predominant **purpose / source / factor**
- ▶ a / the main / major / key / central / principal / chief / prime / primary **aim / focus / function / objective / task / reason / consideration**
- ▶ a / the main / major / principal / chief / prime / primary **object**
- ▶ a / the main / major / key / principal / chief / prime / primary / number one **cause**
- ▶ a / the main / major / principal / chief / prime / predominant **effect**
- ▶ a / the main / major / key / central / principal / chief / prime **attraction**
- ▶ a / the main / major / key / central / principal / predominant **theme**
- ▶ a / the main / major / key / central / principal / prime / primary / predominant **role**
- ▶ a / the main / major / principal / prime / number one **contender**
- ▶ the main / chief / prime / number one **suspect**
- ▶ a / the main / major / principal **road / town / city**
- ▶ the main / key **thing** is to...
- ▶ to be of major / key / central / prime / primary **importance**

main [only before noun] largest or most important: *Be careful crossing the main road.* ◊ *We have our main meal at lunchtime.* ◊ *The main course was roast lamb.* ◊ *Reception is in the main building.* ◊ *Please use the main entrance.* ◊ *Poor housing and unemployment are the main problems.* ◊ (spoken) *The main thing is to remain calm.*
▶ **mainly** *adv.*: *They eat mainly fruit and nuts.* ◊ *The population almost doubles in August, mainly because of the film festival.*

Addendum 7b: Oxford Learner's Thesaurus: A Dictionary of Synonyms (Lea et al. 2008: 896)

896

Conflict

disagree



argue *verb*
argue about sth
argument *noun* 1
a fierce argument
debate *noun*
a growing debate
conflict *verb*
conflicting opinions
controversial *adj.*
a controversial issue
disagree *verb*
disagree with sb/sth


oppose



be against sb/sth *adj.*
I'm against the idea
challenge *verb*
challenge the leader
oppose *verb*
oppose a plan
opposition *noun*
opposition to the plans

enemy *noun*
know your enemy
tension *noun*
racial tensions
remove *verb* 2
remove sb from power
revolution *noun* 1
a popular revolution


fight



aggressive *adj.* 1
aggressive behaviour
attack *verb* 1
violently attacked
fight *noun*
get into a fight
fight *verb* 2
fight an attacker

hit *verb* 2
hit sb on the head
trouble *noun* 1
crowd trouble
violent *adj.*
a violent attack
thug *noun*
an armed thug

war



army *noun*
the enemy army
attack *noun* 1
a bomb attack
attack *verb* 2
attack a city
break out *phrasal verb*
war breaks out
guerrilla *noun*
a guerrilla war

invade *verb*
invade a country
soldier *noun*
a good soldier
war *noun*
go to war
fight *verb* 1
fight a battle

resolving conflict



negotiate *verb*
negotiate on sb's behalf
negotiator *noun*
an independent negotiator

resolve *verb*
resolve the crisis
peace *noun*
restore peace
contract *noun*
sign a contract

Addendum 7c: *Oxford Learner's Thesaurus: A Dictionary of Synonyms* (Lea et al. 2008; CD-ROM version, here a picture plate ("Topic map") can only be accessed through the text of the thesaurus)



Addendum 8: Work in progress on a (bilingual) alphaconceptual+ dictionary for Luganda

The screenshot shows a software interface for creating a dictionary entry. The main window displays the entry for 'abakulu' with its grammatical information and a detailed definition in English. The interface includes a toolbar at the top with icons for file operations, editing, and search. Below the toolbar, there are several panels: a 'References' panel on the left, a 'Bilingual References' panel at the bottom left, and a 'Senses' panel on the right. The 'Senses' panel contains a table with columns for 'Sense (L)', 'Subject Label', 'Position', 'Format (F3)', and 'Sense (L): Class'. The table lists various senses for the word, such as 'Economics', 'Family', 'Fight', 'Food', 'Geography', 'Government', 'Group', 'Health', 'History', 'Information', 'Language', 'Life Orientation', 'Matis', 'Mood', 'Movement', 'Music', 'Owning', 'Part of', and 'People'. The main text area contains the following entry:

abakulu * pl. noun 1/2 Singular = omukulu¹ [GOVERNMENT] ► officials
 ♦ Wabula abaami be abakulu teyabaagala era teyabassamu kitubwa. • But he never liked his officials and he did not respect them. [Empisa] <Freq: 714> <# Texts: 86> abakulu <Rank: 228>

Article has cross-references to >>

omukulu * noun 1/2 [1st abakulu] [GOVERNMENT] ► official ♦ Kabaka anatekanga omukulu mu buki kigo oba lusisira mu Buganda, awafugibwa Omumubi omukulu. • The King will always put an official in every parish or hut in Buganda, where a Mumubi official leads from. [Basekabaka] <Freq: 751> <# Texts: 78> omukulu <Rank: 216>

abakulu adjective cl. 2 1 [AGE] ► old; mature ♦ Akawuka kano kasigala nga kabongootedde mu busimu bwo mu mubiri naye katambula ne kaleeta kisiipi naddala mu bantu abakulu abasussizza mu myaka 60. • This virus remains paralysed in the nervous system of the body but it moves and causes herpes zoster especially in old people over 60 years. [ED110513-Omusawo] 2 [PEOPLE] ► leading ♦ Bwe baawulira ebigambo ebyo omukulu wa yeeakaalu ne bakabona abakulu ne basoberwa mu bwataliikirivu bwabwe ekigambo kino nga bwe kijja okubuna. • When they heard those words the leader of the church and the leading priests wondered about how this word could be spread. [Ebikolwa]

abalenzi pl. noun 1/2 Singular = omulenzi [PEOPLE] ► boys ♦ Edda abalenzi baawasanga okuva ku myaka 16. • Long ago, boys used to marry from 16 years on. [Embaga] <Freq: 175> <# Texts: 38> abalenzi <Rank: 879>

abamu * exclusive quantitative pronoun cl. 2 QUANTITY ► some <Freq: 672> <# Texts: 147> abamu <Rank: 250>

abantu *** pl. noun 1/2 Singular = omuntu [PEOPLE] ► people ♦ Omumtama ogulwanirira ensi abantu baagulina. • People have the heart of fighting for the country. [Ebibuuzo ku Kimala] <Freq: 7165> <# Texts: 316> abantu <Rank: 16>

Expectation Levels in Dictionary Consultation and Compilation*

Inge Kosch, *Department of African Languages,
University of South Africa, Pretoria, South Africa*
(koschim@unisa.ac.za)

Abstract: Dictionary consultation and compilation is a two-way engagement between two parties, namely a dictionary user and a lexicographer. How well users cope with looking up words in a Bantu language dictionary and to what extent their expectations are met, depends on their consultation skills, their knowledge of the structure of a Bantu language and the dictionary design. There are also demands that the lexicographer places on the user. These demands are of varying degrees, ranging from low-level expectations, where basic look-up skills according to the letters of the alphabet are assumed to have been mastered already (as needed when using a word-based dictionary), to medium-level expectations (such as being able to look up words in a stem-based dictionary), to high-level demands (where intuitive dictionary skills no longer suffice). A fine balance must be struck between the user's guidelines and the presumed dictionary consultation skill of the target user. This article endeavours to look into the matter of various levels of expectation of Bantu language dictionaries, both from the perspective of the user and the compiler.

Keywords: DICTIONARY CONSULTATION, DICTIONARY COMPILATION, EXPECTATION LEVELS, CONSULTATION SKILLS, WORD-BASED DICTIONARY, STEM-BASED DICTIONARY, DICTIONARY DESIGN, TARGET USER, LEXICOGRAPHER, LEMMATISATION

Opsomming: Vlakke van verwagting in woordeboekgebruik en -samestelling. Die raadpleging en samestelling van 'n woordeboek behels 'n wisselwerking tussen twee partye, naamlik 'n woordeboekgebruiker en 'n leksikograaf. Hoe goed gebruikers vaar wanneer hulle woorde in 'n Bantoetaalwoordeboek opsoek en die mate waarin daar aan hulle verwagtinge voldoen word, hang af van hulle eie naslaanvaardighede, hulle kennis van die Bantoetaalstruktuur en woordeboekstruktuur. Daar is ook sekere vereistes wat die leksikograaf aan die gebruiker stel. Hierdie vereistes is op verskillende vlakke, wat strek van laevlakverwagtinge waar aangeneem word dat basiese naslaanvaardighede volgens die letters van die alfabet reeds bemeester is (soos nodig wanneer 'n woord in 'n woord-gebaseerde woordeboek opgesoek moet word), tot mediumvlakverwagtinge (waar woorde volgens 'n stam-gebaseerde woordeboek nageslaan moet word), tot hoëvlakverwagtinge (waar intuïtiewe woordeboekvaardighede nie meer toereikend is nie). 'n Fyn balans moet gevind word tussen die gebruikersleiding en die veronderstelde naslaanvaardighede van die teikengebruiker. Hierdie artikel poog om die kwessie van verskillende vlakke van verwagting van Bantoetaalwoordeboeke te beskou, beide uit die perspektief van die gebruiker en die samesteller.

* This article was presented as a paper at the Seventeenth Annual International Conference of the African Association for Lexicography (AFRILEX), which was hosted by the Department of African Languages, University of Pretoria, Pretoria, South Africa, 2–5 July 2012.

Sleutelwoorde: WOORDEBOEKRAADPLEGING, WOORDEBOEKSAMESTELLING, VLAKKE VAN VERWAGTING, NASLAANVAARDIGHEDE, WOORD-GEBASEERDE WOORDEBOEK, STAM-GEBASEERDE WOORDEBOEK, WOORDEBOEKONTWERP, TEIKENGEBRUIKER, LEKSIKOGRAAF, LEMMATISERING

1. Introduction

Dictionary consultation and compilation is a two-way engagement between two parties, namely a dictionary user at the consumer end and a lexicographer at the production end. Both the consumer and the producer have mutual expectations when consulting or compiling a dictionary respectively. The nature of these expectations and the level at which they are pitched will be the focus of this article.

Most casual users of dictionaries may not have any other expectation besides establishing the meaning of a word which, according to Svensén (1993: 14), is the highest ranked type of information for which a dictionary is consulted. This is followed by the need to establish correct spelling, pronunciation, grammar, application in context, synonyms and etymology. After just a few consultation attempts, users will be able to tell whether a dictionary is user-friendly or not by the degree of satisfaction or frustration experienced during their look-up activity and the time it took to retrieve required information. How well they cope with looking up words in a Bantu language dictionary and to what extent their expectations are met, are largely dependent on the following three factors, namely

- their consultation skills, that is, their previous exposure to dictionary pedagogy,
- their knowledge of the structure of a Bantu language, and
- the dictionary design itself.

Not all users have been exposed to dictionaries to the same extent and hence they will experience different degrees of difficulty when consulting a dictionary. Abecassis (2007: 248) rightly states that "there is not a prototypical user, but a whole variety of users, with different cultural backgrounds which may affect their abilities".

Users of Bantu language dictionaries, especially inexperienced second language learners, are invariably guided by previous encounters with non-Bantu language dictionaries and will expect to find the required information in the first place they look, namely in the appropriate alphabetical slot according to the first letter of the word. However, they may be confronted with an unknown dictionary design owing to the agglutinative structure of the Bantu languages, which calls for a specialised approach to lemmatisation, different from the European languages, for example.

Lexicographers consciously and purposefully design their dictionaries with certain target users in mind. The demands they place on users' consultation abilities to successfully retrieve information may range from low-level, to medium-level, to high-level expectations.

2. Levels of expectation

2.1 Low-level expectations

At a low level of expectation the lexicographer can assume that the user, at the very least, has mastered basic look-up skills according to the letters of the alphabet. This requirement is so basic that it is taken for granted among literate people. In the context of the Bantu languages, this means that a user would be able to easily find items that have been lemmatised according to their first letter in a predictable alphabetical order — the so-called word-based approach. The advantage of this approach is that the consultation success is immediate and almost guaranteed without the user first having to resort to the user's guidelines. The word-based arrangement of items makes such dictionaries very user-friendly, particularly for beginners and pupils. At the same time, however, it obscures the connection between inflectionally or derivationally related lemmas that are removed from each other alphabetically as a result of phonological changes as in the case of *-leka* 'try' under the letter 'L' versus its deverbative *tekô* 'test' under the letter 'T' in Northern Sotho.

A word-based dictionary, though user-friendly and popular among beginners, has its limitations as a printed work. Owing to space and cost considerations it can impossibly include all the lemmas in a language. However, such challenges are no obstacle to modern technology, which has revolutionised dictionary compilation and consultation by electronic means as stated by Prinsloo (2005: 11): "The great capacity and speed characteristic of electronic products combined with enhanced query and data retrieval technology, indeed pave the way to a new generation of dictionaries unimagined in the paper-dictionary era."

2.2 Medium-level expectations

In agglutinating languages it is customary to compile dictionaries in terms of stems. The fact that expansions containing prefixes and or infixes must be searched for under the appropriate stem, according to Haas (1967: 49) "requires considerably more grammatical sophistication on the part of the user than is ever required, for example, on the part of the user of the ordinary English or French dictionary. Words found in actual texts would have to be analyzed before they could be looked up".

A stem approach assumes that the user is familiar with the morphological structure of words to the extent of being able to identify and then dismiss the

prefixes when looking up lemmas under the first letter of the stem, for example, to access the meaning of *seatla* 'hand' in Northern Sotho, the user is expected to truncate the prefix *se-* and look up the meaning of the word under the first letter of the stem, namely 'A'.

Most dictionaries using the stem approach actually follow a mixed approach of stem and word lemmatisation, as done in the *Comprehensive Northern Sotho Dictionary* of Ziervogel and Mokgokong (1975). Word lemmatisation is applied to nouns with irregular or non-overt class prefixes. These nouns are entered according to the first letter of the word, for example, *mmuši*, 'ruler', a noun in class 1, will be found under the letter 'M', while *pono* 'vision', a noun in class 9, will be lemmatised under the letter 'P'. The look-up success is not immediate, however, as the meaning is not given under *mmuši* or *pono*. Instead a cross-reference is given to the stems *-buša* 'rule' and *-bona* 'see' respectively from which these words have been derived. Though user-friendly, this method is time-consuming as it requires two look-up actions instead of one before the required information is obtained. The mixed design, whereby a word approach is used within a dictionary which is overridingly stem-based, constitutes a compromise to facilitate access by the not so advanced user. Lemmatisation according to stems is less user-friendly than lemmatisation according to words, but, in the opinion of Ziervogel and Mokgokong (1975: Preface, p. 87) it is the only scientific method. This scientific method has come at a price, though, namely the loss of accessibility, and the rise in the degree of difficulty as well as the time in finding desired information.

2.3 High-level expectations

High-level expectations can be regarded as those where users are required to apply skills that are language and dictionary specific and where intuitive dictionary skills no longer suffice. Referring once more to the example *mmuši*, it is interesting to note that Endemann (1911) took the scientific approach of stem lemmatisation to the extreme, as he does not list this word at all under its first letter, but instead only under the first letter of the stem. Without knowledge of the phonological changes in the Sotho languages or without a prior in-depth familiarisation with the front matter, the user is at a loss as to where to look up the meaning of this word.

The necessity to complete paradigms may also be experienced by a user as a medium- to high-level challenge. In a paper dictionary completion of paradigms is a practical requirement for space- and cost-saving reasons. Entries are often omitted from the central list, if the lexicographer is of the opinion that the user can complete the relevant paradigms, based on directives given in the user's guidelines or in the treatment of articles themselves. As a case in point, most dictionaries supply limited entries of reflexive forms, and only include those of which the meaning is not simply an accumulation of the meaning of the verb plus 'self', as in the case of *-ithuta* ('study' < teach oneself) and *-itshwara* ('behave' < carry oneself). To avoid duplication of verb stems Snyman, Shole

and Le Roux (1990), for example, have generally omitted verb stems with the reflexive prefix *i-* in the central list. However, for users to be able to find the meaning of reflexive verb stems, the compilers have provided the necessary support in the form of a table of sound changes in the front matter. Kriel (1983), on the other hand, lists a number of reflexive forms under the letter 'I', but provides no instructions on how to look up the meaning of non-listed reflexive forms. In the *English-Zulu Dictionary* by Doke, Malcolm and Sikakana (1971) reflexive forms of verbs are not listed under *zi-*. Instead, the function of this morpheme is briefly mentioned under the entry *zi-* with one example, *ukuzibona* 'to see oneself'. For the rest, the user is expected to find the meaning of reflexive forms by truncating the prefix *zi-*. This is incidentally not such a daunting task in Zulu as in the Sotho languages, where phonological changes accompany the prefixation of the reflexive morpheme.

As far as deverbatives are concerned, Endemann (1911: 36) explains that users can create nouns on their own according to a set pattern. In his front matter he gives one example of a regular formation and one where sound strengthening is involved in the initial sound of a verb stem as in the case of prefix-less classes. Expecting users to form their own deverbatives based on these limited instructions, can probably count as a high-level activity as no further support is provided at that point and no cross-reference is given to page 15 where sound strengthening is explained. Ziervogel and Mokgokong (1975: 117-123), on the other hand, offer a more comprehensive overview of the sound changes in their front matter. Items that users would expect to find in the central list should only be phased out of the central list, if their formation is sufficiently covered in the front matter.

A fine balance must be struck between the instructions given and the presumed dictionary consultation skills of the target user. The effort of the user to access the required information, including consultation of the outer text, should not be perceived as outweighing the informative value gained from the research result. A user-friendly dictionary would enable a user to derive 'missing' forms in a paradigm with the maximum amount of support and the minimum amount of effort. This once again underscores the fact that the front matter is pivotal for effective dictionary consultation, but it is unfortunate that lexicographers can generally not rely on the average dictionary users to consult the guidelines as they "usually move directly to the central list where they try to solve the problem that motivated their search. Too often the search does not lead them to the required destination because the user is not familiar with the system of the dictionary" (Gouws and Prinsloo 2005: 85).

One could also regard as medium- to high-level expectations those where users are required to cope with practices which they do not anticipate, such as the unconventional access alphabet in Ziervogel and Mokgokong (1975) where combinations of letters serve as article stretches. Without an understanding of this design, the user will try to look for a word starting with KG (for example *kgomo* 'head of cattle') before words starting with KO (for example *kolobe* 'pig') under the letter 'K'.

Compilers should make sure that scientific needs are not served to the detriment of the practical needs of users. The level of expectation should be pitched at the level of the target user. In this vein Gouws and Prinsloo (2005: 39) endorse that "[m]odern dictionaries are judged by the success with which the *user* is able to retrieve the sought information and not in the first instance by 'linguistic achievement'". Thus the ultimate test of a good dictionary is how well it can be used. This can be established from extensive empirical user research with or without preliminary drafts of (sections of) the dictionary. Often the compiler can only gauge whether the demands placed on the user were realistic once the dictionary is in active use. Lessons learnt can then be implemented in the revision of the dictionary.

3. Pre-empting users' queries and needs

Dictionaries may seem to provide "too little" information in some respects and "too much" in others. The compiler should address such apparent inconsistencies in the front matter or within article treatments, thereby pre-empting queries most likely to come up in the minds of users. Lexical gaps and semantic elaboration are among the issues that would need to be brought to the user's attention.

3.1 Lexical gaps

Any perceived gaps in the central list should be accounted for in a lucid manner, such as the practice of paradigm completion as a space-saving device. The formation of reflexive forms of verb stems is a case in point as discussed in section 2.3.

Lexicographers can choose not to include certain lemmas for various reasons. In the *South African Multi-language Dictionary and Phrase Book* (Reader's Digest 1991: 6) the compilers, anticipating queries about the absence of some entries, offer the following as an explanation:

When using the word lists, it is important to remember that no language translates literally into another — and where no simple translation exists in a particular African language, our contributors have chosen not to translate the word in that language.

If a word does not appear in a dictionary, the reason may not necessarily be that it was overlooked. The selection of words may seem arbitrary, but it is governed by incidence of use and need in the fields of thought of the particular society who speaks the language.

3.2 Semantic elaboration

Dictionary compilation is steeped in culture and therefore the lexicographer

has to take cognisance of all possible meanings of a word if he/she wants to be culturally correct. To this end the user is in some cases provided with unexpected (but important) semantic elaboration.

Although Martin (1967: 156) states that "[i]n the interest of conciseness, we should aim at a single translational equivalent whenever possible", semantic elaboration is unavoidable in some cases. The foreign language learner might, for example, quickly want to look up a translation equivalent for the word 'brother', only to find that a simple equivalent does not exist. Instead, he/she is confronted with a list of possible words, wherein the meaning of 'brother' is shown to be determined by this person's relation to other people in the family, e.g. younger brother, elder brother, mother's brother, father's elder brother, father's younger brother, etc. (e.g. Snyman, Shole and Le Roux 1990: 226-227). The user is thus confronted with more than the expected information and this may slow down successful retrieval of the meaning of the desired word. However, as stated by Nkabinde (2003: 180): "Culture is indispensable in the description of a language in general and in the compilation of a dictionary in particular".

Obviously, in a monolingual dictionary, a lengthy treatment of a lemma like 'brother' would be the norm, as users of the dictionary consult it for this very purpose. Abecassis (2007: 251) points out that high proficiency learners (unlike low proficiency learners) often give preference to monolingual dictionaries "when bilingual dictionaries do not suffice for finding extra information on a word meaning". The relative article length assigned to words like 'brother' is a reflection of its importance in the society.

4. Conclusion

This article considered various levels of expectation, both from the perspective of the user and the compiler. At a low level of expectation basic look-up skills according to the letters of the alphabet are assumed to have been mastered already (as needed when using a word-based dictionary). At a medium level of expectation users are expected to be able to look up words in a stem-based dictionary. High-level expectations are those where the user has to be able to look up words which have undergone phonological changes or to complete paradigms where lemmatisation is limited. The latter is acceptable as long as there are appropriate guidelines given in the front matter. The fine balance that was referred to between the assumed dictionary consultation skill of the user and the expectations of the compiler is aptly summarised as follows by Svensén (1993: 16): "It is not merely the information needs of different categories of user that are of importance. An equally important factor is their ability to find and make use of the information given. Here the lexicographer is forced into a difficult balancing act. On the one hand, the dictionary must contain as much information as possible without becoming unnecessarily bulky; on the other hand, this information must not be so compressed that the intended user will not be able to utilize it."

References

- Abecassis, M.** 2007. Is Lexicography Making Progress? On Dictionary Use and Language Learners' Needs. *Lexikos* 17: 247-258.
- Doke, C.M., D.McK. Malcolm and J.M.A. Sikakana.** 1971. *English-Zulu Dictionary*. Johannesburg: Witwatersrand University Press.
- Endemann, K.** 1911. *Wörterbuch der Sothosprache*. Abhandlungen des Hamburgischen Kolonialinstituts. Band VII. (Reihe B. Völkerkunde, Kulturgeschichte und Sprachen. Band 4.) Hamburg: L. Friederichsen & Co.
- Gouws, R.H and D.J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.
- Haas, M.R.** 1967. What Belongs in a Bilingual Dictionary? Householder, F.W. and S. Saporta (Eds.). 1967. *Problems in Lexicography*: 45-50. Bloomington: Indiana University.
- Kriel, T.J.** 1983. *Pukuntšu Woordeboek: Noord-Sotho-Afrikaans, Afrikaans-Noord-Sotho*. Third edition. Goodwood: Nasionale Boekdrukkery.
- Martin, S.E.** 1967. Selection and Presentation of Ready Equivalents in a Translation Dictionary. Householder, F.W. and S. Saporta (Eds.). 1967. *Problems in Lexicography*: 153-159. Bloomington: Indiana University.
- Nkabinde, A.C.** 2003. The Implications of Culture for Dictionaries of the African Languages. *Lexikos* 13: 168-182.
- Prinsloo, D.J.** 2005. Electronic Dictionaries Viewed from South Africa. *Hermes, Journal of Linguistics* 34: 11-35.
- Reader's Digest.** 1991. *South African Multi-Language Dictionary and Phrase Book*. Cape Town: The Reader's Digest Association South Africa.
- Snyman, J.W., J.S. Shole and J.C. le Roux.** 1990. *Dikišinare ya Setswana-English-Afrikaans Dictionary/Woordeboek*. Pretoria: Via Afrika.
- Svensén, B.** 1993. *Practical Lexicography. Principles and Methods of Dictionary-Making*. Oxford/New York: Oxford University Press.
- 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.

New Proposals for the Design of Integrated Online Wine Industry Dictionaries

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

Abstract: The specialised lexicographic treatment of oenology and viticulture usually consists in the compiling of articles describing the language of wine in general language dictionaries, or presenting professional knowledge of wine in specialised multi-field dictionaries and encyclopedias. This treatment also encompasses the terminological compilation of single field dictionaries describing the language and/or knowledge of wine. Lexicographically speaking, all this is but a fraction of the complete picture. Indeed, the specialised lexicography of oenology and viticulture is multifaceted and goes far beyond the above-mentioned types of articles and dictionaries. It includes a broad range of both online and printed lexicographically structured information tools, such as wine guides, atlases, companions, oenological websites, and mobile applications. In line with this growing interest for lexicographically structured information tools on oenology, wine making, and wine tasting, this article argues for an expansion of lexicography dealing with wine: It explains how a new lexicographic information tool, in this case OENOLEX Burgundy, a French monolingual online wine dictionary commissioned by the Burgundian wine industry, offers various functions and usage modes to its users, including access to multimodal data, and how it differs from a comparable South African online wine dictionary commissioned by the South African wine industry. In line with the findings of this comparative study, lexicographic proposals are subsequently formulated. The first proposal is that such online wine industry dictionaries should be developed in accordance with specific user situations and needs, and make use of a monofunctional design and an adaptive user interface. The second proposal is that such online wine industry dictionaries should not stand alone, as they could benefit from a full integration with the websites of the wine industry, and consequently transform into genuine lexicographic information tools.

Keywords: SPECIALISED LEXICOGRAPHY, SPECIALISED DICTIONARIES, ONLINE SPECIALISED DICTIONARIES, SUBJECT-FIELD-BASED DICTIONARIES, USER-SITUATION-BASED DICTIONARIES, FUNCTION THEORY OF LEXICOGRAPHY, WINE INDUSTRY DICTIONARIES, CONSULTATION, NAVIGATION, ACCESS MODES, MULTIMODALITY, USER NEEDS PARADIGM, WEBSITE INTEGRATION

Opsomming: Nuwe voorstelle vir die ontwerp van geïntegreerde aanlyn wynbedryfwoordeboeke. Die gespesialiseerde leksikografiese behandeling van die wynkunde en wynbou bestaan gewoonlik in die samestelling van artikels wat die taal van wyn in algemene taalwoordeboeke beskryf of die vakkennis van wyn in gespesialiseerde multiveldwoordeboeke en -ensiklopediëe aanbied. Hierdie behandeling omvat ook die terminologiese samestelling van enkelveldwoordeboeke wat die taal en/of kennis van wyn omskryf. Leksikografies gesproke is

dit alles maar 'n klein deeltjie van die geheelbeeld. Die gespesialiseerde leksikografie van die wynkunde en wynbou is inderdaad ryk geskakeerd en gaan ver verby die bogenoemde soort artikels en woordeboeke. Dit sluit 'n wye reeks van sowel aanlyn as gedrukte leksikografies gestruktureerde inligtingsgereedskap in, soos wyngidse, -atlasse, -maatskappye, wynekundige webtuistes en mobiele toepassings. Ooreenkomstig hierdie groeiende belangstelling vir leksikografies gestruktureerde inligtingsgereedskap oor wynkunde, wynmaak en wynproe, bepleit hierdie artikel 'n uitbreiding van die leksikografie wat handel oor wyn: Dit verduidelik hoe 'n nuwe leksikografiese inligtingswerktuig, in hierdie geval OENOLEX Burgundy, 'n Franse eentalige aanlyn wynwoordeboek wat 'n opdrag van die Boergondiese wynbedryf is, verskillende funksies en gebruiksmodusse aan sy gebruikers bied, insluitende toegang tot multimodale data, en hoe dit verskil van 'n vergelykbare Suid-Afrikaanse aanlyn wynwoordeboek wat 'n opdrag is van die Suid-Afrikaanse wynbedryf. Ooreenkomstig die bevindinge van hierdie vergelykende studie, word leksikografiese voorstelle vervolgens geformuleer. Die eerste voorstel is dat sulke aanlyn wynbedryfwoordeboeke ontwerp behoort te word in ooreenstemming met spesifieke gebruikersituasies en -behoefte en gebruik te maak van 'n monofunksionele ontwerp en 'n aanpasbare gebruikerskoppelvlak. Die tweede voorstel is dat sulke aanlyn wynbedryfwoordeboeke nie alleen behoort te staan nie, aangesien hulle kan baat by 'n volle integrasie met die webtuistes van die wynbedryf, en gevolglik verander in werklike leksikografiese inligtingsgereedskap.

Slutelwoorde: GESPECIALISEERDE LEKSIKOGRAFIE, GESPECIALISEERDE WOORDEBOEKE, AANLYN GESPECIALISEERDE WOORDEBOEKE, ONDERWERPSVELDGEBASEERDE WOORDEBOEKE, GEBRUIKERSITUASIEGEBASEERDE WOORDEBOEKE, FUNKSIETEORIE VAN DIE LEKSIKOGRAFIE, WYNBEDRYFWOORDEBOEKE, RAADPLEGING, NAVIGASIE, TOEGANGSMODUSSE, MULTIMODALITEIT, GEBRUIKERSBEHOEFTEPARADIGMA, WEBTUISTE-INTEGRASIE

1. Introduction: The specialised lexicography of oenology and viticulture within the user needs paradigm

As a subject field, wine represents the world's universal cultural heritage. In terms of economy, it is an influential industry representing a large number of stakeholders such as wine growers, wine makers, and wine merchants. As an object of consumption, wine is probably becoming one of the most popular lifestyle-determined subjects of interest and curiosity in the minds of global hyperconsumers of today (Lipovetsky 2006). Paper-based wine dictionaries (DLV 2007), wine encyclopedias, wine atlases, wine companions and wine guides (GHV 2010) have become increasingly popular and are now being rapidly replaced by online resources, such as professional online wine dictionaries (Winetech and Sawis 2012), huge encyclopedic articles and portals (Wikipedia 2013), and online wine guides (GHV 2013), to name just a few. The transition from the paper form and the main principles of online lexicography have been thoroughly described in Granger and Paquot (2012), Fuertes-Olivera and Bergeholtz (2011a and 2011b), Tarp (2012) and Fuertes-Olivera and Tarp (2014), all of whom advocate user-oriented solutions based on user needs. Also, the

Internet as a prolific source of information on wine provides access to a vast number of word lists, glossaries, and online wine dictionaries of varying quality. The development of mobile information technologies for smartphones has led to a vast expansion of the volume and types of data which users look for to satisfy their information needs. It has also brought about a multiplication of access modes (Leroyer 2011; Leroyer and Gouws 2009). A smartphone App such as Drync Wine Free and Pro (2013) gives fast and easy access to information about 1,2 million wines, 100.000 score cards written by wine tasting oenologists, and a plethora of oenological information, including advice on how to buy wine and virtually manage one's own wine cellar.

Common to all these information resources, perhaps with the exception of the Drync wine App, is the fact that they seem to privilege a subject field perspective rather than the user needs perspective. In other words, it seems that the presentation of data supporting communication about wine or knowledge of wine *per se* tends to become the aim of the resource. The process of data selection is normally guided by a description of wine as a subject field, or a range of major subject-fields and a large number of associated subfields. From a functional perspective, and in line with Fuertes Olivera and Tarp (2014), one should think that compiling a wine dictionary should start with an analysis of the social situations in which information needs about wine can lead to the consultation of specific kinds of wine dictionaries or other wine-related lexicographic information tools. This calls for an approach based on user needs and user situations, one in which lexicographic functions have a major role to play. Should the tool help its intended users to communicate about wine, to acquire knowledge of wine, to interpret non-verbal signs and codes such as the complex information provided by wine labels, or to carry out wine-related mental and physical operations, such as wine tasting? The implementation of the functions should start with the lexicographic identification of the unique social situations which are characteristic of the information society, then analyze the complex of specific information problems and information needs related to these situations, and determine the extent to which potential information needs arising from the situations can be met and gratified by a lexicographic information tool. In many cases, the lexicographic situation will only have been partly, if at all, recognised as such before.

This article consists of three parts. First, it presents the Function Theory of specialised lexicography and confirms the pre-eminence of the social situation in the extra-lexicographic world. Second, a South African online wine industry dictionary is reviewed as a case of a predominantly subject-field-based dictionary, and proposals for improvement are made. Third, a new user-situation- and user-need-based dictionary of wine, currently under construction, is presented, and new proposals for the development of online wine dictionaries in general are finally formulated. Although different in scope, the two online dictionaries share the same functions, which is to help users in training situations.

The basic lexicographic methodology underlying the present research was

previously introduced in Leroyer (2008 and 2011), and in Gouws and Leroyer (2009) (regarding the transformation of guides into lexicographic information tools), and in Leroyer and Gouws (2009) (regarding the compilation of a new kind of website integrated online wine guide and dictionary).

In conclusion, it is claimed that user-based specialised lexicography taking the specific wine-related social situations as a point of departure for dictionary design and dictionary compiling could have a substantial impact on the development of integrated online wine industry dictionaries of tomorrow.

2. The lexicographic functions of specialised dictionaries

There are four basic elements that constitute the core of the Function Theory of lexicography (Fuertes-Olivera and Tarp 2014; Tarp 2008a and 2008b). Dictionaries and other lexicographic tools are built to satisfy punctual information needs. As tools, they have functions which are essential to their genesis and to their use. According to Bergenholtz and Tarp 2010, a lexicographic function is defined as:

- (a) the satisfaction (through the selection and presentation of easily accessible data)
- (b) of the specific types of lexicographically relevant needs that may arise
- (c) in a specific type of potential user
- (d) in a specific type of extra-lexicographic situation.

Information needs are lexicographically relevant in so far as they can be satisfied by consultation. Bergenholtz and Tarp (2010: 30) state: *Not all situations are relevant for lexicography; only situations in which needs may arise that can be satisfied by consulting dictionaries.* This statement should actually be slightly modified by adding: *and other lexicographic information tools.* Accordingly, the lexicographer's task is to pinpoint the specific user needs in the specific use-situations, and to establish the user profiles. This analysis should lead to the design of a unique dictionary concept, in which selection and presentation of easily accessible data will lead to the satisfaction of the needs.

User-based specialised lexicography begins by acknowledging the primacy of the situation. A definition of lexicography could thus be that it is a discipline dedicated to the construction of functional lexicographic information tools for the satisfaction of information needs in specific types of social situations in the extra-lexicographic world, combining types of social situations, types of user needs, and types of users. The primacy of the social situation can then lead to the invention of new lexicographic concepts and the development of corresponding functional configurations, the implementation of innovative data selection and presentation principles with integrated solutions, and the implementation of innovative access options on multiple platforms and termi-

nals, including mobile solutions.

From a functional perspective, subject-based specialised lexicography normally includes three fundamental, major steps:

- (a) In the **recognition phase**, the lexicographer takes the recognition of a pre-existing subject-field and associated subfields as a point of departure (cf. Bergenholtz and Kaufmann 1997).
- (b) In the **transformation phase**, the lexicographer critically transforms existing subject-field-based dictionary concepts into new ones by applying adequate lexicographic functions.
- (c) Finally, in the **implementation phase**, the lexicographer compiles a lexicographic information tool according to the functions.

Situation-based specialised lexicography is also characterised by three fundamental steps, but the point of departure is the social situation itself, i.e. the user needs, not the subject-field:

- (a) In the **identification phase**, the specialised lexicographer takes the identification of a unique social situation as a point of departure.
- (b) In the **invention phase**, the specialised lexicographer can invent new configurations of the lexicographic functions.
- (c) Finally, in the **innovation phase**, the lexicographer can compile an innovative lexicographic tool taking advantage of the new functional configurations.

The above stated two sets of lexicographic principles — subject based and situation based — will now be applied in the comparative study of two online wine industry dictionaries in the following sections.

3. **Subject-field-based wine industry dictionary: The SAWIS South African Trilingual Dictionary of Wine**

The South African Trilingual Wine Industry Dictionary — SATWID (Winetech and SAWIS 2013) is an online wine industry dictionary of oenology and viticulture. It contains the central wine terminology used by the South African wine industry and covers numerous subject fields and subfields related to wine. All terms are lemmatised in English and translated into Afrikaans and Xhosa. The terminology of wine tasting, for instance, includes 228 entries. All terms are provided with subject field labels indicating the specific field or subfield to which they belong. For example, the subject field "viticulture" includes a number of subfields, such as "organic growth and production", "soil science", "plant biotechnology", "vine viruses" and "plant improvement". The subject field "oenology" encompasses the terminology of production technology, bottling,

packaging and distribution, as well as microbiology. This subject field approach is confirmed by a subject-field-based access to data through the list of subject fields: By clicking on a specific subject field label/link from the list in the right panel, users can get access to the complete list of terms included in the field. Linguistic information, including information on part of speech is also provided, but only in English so far. It is foreseen that the information will also be provided in Afrikaans and Xhosa at a later stage of the project.

The metalexigraphic description of SATWID, i.e. the introduction and instructions, seems to contain a number of inconsistencies as far as user profiles, target users, and functions are concerned. According to the compilers of the dictionary, one of its purposes is to assist its intended users in learning situations in order to fulfil their educational needs:

Emerging producers and people who do not have training or a background in the industry can be assisted by this dictionary. It fulfils a training and introductory function and can fulfil the educational needs of the user with explanations of terms that are used in the industry.

In line with Fata (2010), one should expect then that the SATWID should make provision for assistance in connection with cognitive situations, as knowledge acquisition is the central element of any training program (see also Fuertes-Oliviera 2010). However, the SATWID is defined as communication-oriented, and no mention at all is made of cognitive functions:

The dictionary is multi-functional, i.e. it can be used for text production, text reception and text translation.

Whereas the intended users, in accordance with the genuine purpose of the dictionary, are clearly identified as semi-experts in the background information on the dictionary:

Emerging producers and people who do not have training or a background in the industry,

the actual users are said to consist of both experts (the first four categories listed below) and semi-experts (the following five ones); moreover, wine lovers probably belong in many cases to the category of lay people:

Viticulturists and oenologists, producers and wine makers, viticulture and oenology students and lecturers, public relations officers and information officers, wine marketers, wine writers, wine lovers.

These users have very different profiles and are engaged in a large number of various social situations related to wine (wine growing, wine making, learning about wine, promoting wine, selling wine, writing about wine — or simply being interested in wine in general and having a desire for information in the case of wine lovers), in which lexicographic information needs can arise. It might be feared then that the fulfilment of the lexicographic aims cannot be

achieved, because the different user profiles have basically very different lexicographic information needs owing to their diverging background, general culture and level of knowledge, and because they are engaged in radically different user situations. The functional shortage of the SATWID is probably due to a systematic, subject-based approach, which must have been commissioned by the industry for which the dictionary is a tool of interlingual communication and standardisation. The dictionary is not really user-based; it is a list of expert words for experts of different fields. As for language planning purposes, including terminologisation patterns and solutions (for Xhosa), text production and translation, the data prove clearly insufficient. Grammatical information is limited to parts of speech, and the dictionary does not offer any information on collocations. Only terms of the wine industry are lemmatised in English and translated into Afrikaans and Xhosa (English **white wine**, Afrikaans *witwyn*, Xhosa *iwayini emhlophe*), although a small number of culture-dependent items are also included, such as *wine route* (see below).

The article presented below illustrates one of the basic problems of the subject-based approach. The meaning of the lemma **wine route** is categorized and specified by the label 'commerce', but strangely enough not by the label 'wine tourism' or 'oenotourism', which would be much more appropriate. Besides, no references are made to South African wine routes, so the information value of the article is limited for those users who participate in a wine training program and definitely need such information to acquire knowledge of this crucial aspect of the South African wine industry and wine culture.

wine route

noun

(Subject: Commerce)

tourist destination in a specific area linking a group of wine farms.

Afrikaans **wynroete**

Xhosa **indlela yewayini**

One of the most striking features of the SATWID is the fact that the lexicographic data are largely generic (according to its compilers: 'physiology and fruit of the vine, winemaking processes') in so far as they are subject-field-based. The data are not really specific to South Africa, apart from the fact that the dictionary consists of English lemmas and articles translated into Afrikaans and Xhosa. As such, the SATWID can be effectively used for language planning purposes, included standardisation, but there seems to be paradoxically very little specific South African information in it: For instance, it does not provide any encyclopedic information on South African varieties of grapes (Bukettraube, Cape Riesling, Hanepoot, Roobernet, or Pinotage, the local cross between Pinot Noir and Cinsaut, which is unique to South Africa) nor on any other crucial cultural aspects of viticulture (ethical trade, history, soils, climates, wine of origin legislation). The SATWID offers a link in the right panel, *List of varieties >>*, but this link unfortunately does not give direct access to the

list. The access is unnecessarily intricate. Users have to click on a series of various links and pick up the right heading in the navigation panel before they finally can open a PDF document included in a wine-of-origin scheme with all varieties. The varieties should have been lemmatised in the first place.

On the basis of this short review of the SATWID, it is hereby proposed to:

- (a) expand, in line with recommendations formulated by Carstensen (1989) and Leroyer and Bergenholtz (2014), the lemma list with culture bound words and expressions which are truly constitutive of the South African wine industry and its unique wine culture; this would include, e.g., the lemmatisation of grape varieties, wine routes, estates, etc.
- (b) expand and adapt definitions to the intended target user groups (semi-experts) for text reception functions and for learning, by including relevant encyclopedic explanations and illustrations to the short definitions of the terms,
- (c) include collocations and translation of collocations in all three languages to reinforce the text translation functions,
- (d) include extended grammatical information to reinforce the text production functions,
- (e) include text examples to reinforce the text production functions,
- (f) provide *internal links* to synonyms to secure easier and faster access in connection with the text production functions,
- (g) include multimodal data types (illustrations, videos) to improve the cognitive functions, particularly in learning situations,
- (h) make the incorporation of the SATWID into the SAWIS homepage (SAWIS 2013) even more visible than it is now in order to secure faster and easier access to the dictionary itself,
- (i) include direct links to the different pages and websites of the SAWIS homepage to support cognitive functions in learning situations, or even to the pages of commercial resources such as Wine.co.za 2013, and
- (j) provide, whenever relevant in text reception situations, direct links from terms used in the texts of the SAWIS homepage to the definitions provided in the SATWID (opening in new windows).

The above recommendations, which fall into the categories of data selection (more specific data for the foreseen user-situations), access to the data (faster and easier access to the data), and multimodal data (integration of the homepage), would contribute to transform the SATWID, which at present is a predominantly subject-field based dictionary, into a truly user-situation based dictionary.

4. User-situation-based wine dictionary: OENOLEX Burgundy

OENOLEX Burgundy (OB) is an international lexicographic co-operation project between the University of Burgundy in Dijon (France) and Aarhus University (Denmark). The project was commissioned by the BIVB (VDB 2013), the branch organisation of the Burgundy wine industry. The goal of the BIVB was to develop an information tool aimed at the promotion of communication and knowledge about Burgundy wine. More precisely, the decision was made to develop a lexicographic information tool aimed at the information needs of its intended users, in this case sommelier (wine waiter) students and other students following wine tasting courses at the Burgundy wine school. OB was originally commissioned by the BIVB as a so-called *nuancier lexicographique*, meaning a kind of lexicographic 'colour card' of terms, expressions, and other codes used by the Burgundy wine industry in connection with wine tasting. No matter how strange it may sound for the designation of a lexicographic information tool, what was needed was a lexicographic information tool, no matter what it was called, in line with the statements of Fuertes-Olivera and Tarp (2014):

Specialised lexicographical works may be published under a whole range of different names such as *dictionary*, *encyclopaedia*, *encyclopaedic dictionary*, *lexicon*, *vocabulary*, *glossary*, *terminological database*, *knowledge bank*, *resource*, *tool*, etc. At the end of the day, it is not these names but the functions of the respective products that determine whether or not they can be considered specialised lexicographical works.

OB includes four types of user situations, all related to wine tasting situations, and their related information needs:

- (a) Cognitive situations in which the intended users need to acquire knowledge on the act of wine tasting itself, on the language used to verbalise this act, including knowledge on relations between categories of wine tasting concepts, or on the knowledge of the wine being tasted in order to contextualise the tasting experience.
- (b) Communicative situations in which the intended users need assistance in connection with the reception or production of oral texts about wine tasting. The lexicographic challenge here is to provide alternative access routes to the data, so that the user can verbalise their subjective experience of wine tasting properties, one of the main objectives of the teaching being to acquire the competences necessary to make a professional evaluation of the wine.
- (c) Operative situations in which the intended user needs assistance in order to perform the operation of wine tasting.
- (d) Interpretive situations in which the intended users need assistance in

order to interpret the types of legal and commercial information printed on the labels of the wine bottles being tasted; the Burgundy wine industry makes use of a very intricate system of "appellations" and "climats" and "lieux dits"[†].

Wine tasting evaluation is the result of three phases, which are based on three distinctive sensory experiences: visual, olfactory, and gustatory evaluations. In the following section, the four types of user situations identified for OB are briefly presented, and the solutions chosen by the compilers of OB to satisfy user needs. Together, these solutions contribute to the realisation of the evaluative function, which is the genuine purpose of OB.

4.1 Communicative situations

In OB, the relevant communicative situations are wine tasting situations in which potential tasters of Burgundy wine are confronted with a reception or a production problem in an oral communication situation (to understand speech during a wine tasting lesson or presentation, or to produce an oral text and verbalise sensations in a professional way).

All lexical units directly connected with the evaluative act of tasting Burgundy wines are lemmatised and provided with the following data fields, shown in Table 1 below:

lemma
lemma subject class
lemma grammar
definition
collocations
synonyms
antonyms
examples
references to specific wines on the BIVB website
references to relevant web pages on the BIVB website

Table 1: Communicative situations

However, the lexicographic problem in connection with wine tasting is more complex. It is necessary to create access routes to resources that are not directly accessible through lemmas, as users (learners) do not (yet) know how to verbalise their evaluation. OB will provide a systematic access to lists of relevant

wine tasting items, making it possible, for instance, to redirect the search from basic colours to multiple colour shades, from basic fragrances to detailed fragrances, including the evaluation of qualities and faults, or from specific wines to their characteristics. Access is made possible through the systematic organisation of wine tasting concepts, each concept (here **floral** in table 2 below) being linked to classes of concepts, and to specific wines. The user can also click on the wine itself in order to be re-routed to the website itself, which contains full descriptions of the specific wines (a, b, c, d, e, ...)

impressions
1. olfactory impressions 1.1. aromas 1.1.1. floral 1.1.1.1. jasmine → wine a , b , c , d , e , ... violet → wine a , b , c , d , e , ... rose → wine a , b , c , d , e , ...
Web pages

Table 2: Classification of wine tasting terms and linking to specific wines on the website

Wine qualifiers are also lemmatised, and they contain the same data fields as shown above in Table 1. Again, lemmas are linked to specific wines on the website:

generic qualities
delicate → wine a , b , c , d , e , ... distinguished → wine a , b , c , d , e , ... elegant → wine a , b , c , d , e , ... fresh → wine a , b , c , d , e , ... graceful → wine a , b , c , d , e , ... masculine → wine a , b , c , d , e , ... powerful → wine a , b , c , d , e , ... rich → wine a , b , c , d , e , ... robust → wine a , b , c , d , e , ...

Table 3: Lemmatisation of qualifiers and linking to specific wines on the website

4.2 Cognitive situations

It appears from a close study of course descriptions at the *École des vins de Bourgogne* (2013) that knowledge involved in wine tasting is complex and can be divided into five types: declarative knowledge, schematic knowledge, procedural knowledge, experiential knowledge, and ultimately, evaluative knowledge.

Declarative knowledge encompasses knowledge of the identity of the wine (status of the 'appellation contrôlée', position of the vineyard, varieties of grapes used, kind of vinification, ranking of the vintage and evolution in time, etc.). OB collects these data and addresses them to toponymic lemmas. All types of 'appellations contrôlées' (Regional, Villages, Premiers Crus, Grands Crus) as well as individual producers are lemmatised. All tasting characteristics are registered, and linking to wine tasting tutorials is provided:

Chablis grand cru
<p>Visual White: pure green-gold in colour, evolving to a light yellow with age. A wine with fine keeping properties (10–15 years, sometimes more).</p>
<p>Nose Intense mineral scents (flint) but with room for lime-flowers, dried fruits, almond and a discreet touch of honey. Field mushroom gives a strongly characteristic touch to this wine.</p>
<p>Palate On the palate, acidity and fat are in perfect balance, as are liveliness and dryness. It has all the charm of an inimitable and authentic wine.</p>
<p>Global evaluation This is the jewel in the crown of the chablis range, richly nuanced by variations in "climat".</p>
<p>Link to wine tasting instruction video (tutorials) on the website of BIVB</p>

Table 4: Data addressed to specific wines

Schematic knowledge encompasses knowledge of relations between wine tasting terms and regions, appellations, and specific wines. In OB, assistance is provided by wine tasting data being addressed to specific wines (cf. table 3 and 4 above).

Procedural knowledge is the knowledge of the act of wine tasting itself: choice of the right type of glass, quantity of wine to be poured in it, serving temperature, etc. OB lemmatises all keywords connected to this kind of knowledge. Access will be supported by organising the data allowing systematic access through searches via keywords and table of contents (cf. table 5).

Since experiential knowledge is gained through personal experience, this type of knowledge cannot be represented in OB. OB cannot explain what it feels like to experience wine tasting, because tasting is such a complex and individual sensory experience. It might be argued, however, that the use of OB supports the gathering of experience, and hereby the acquisition of experiential knowledge.

Evaluative knowledge is the ultimate competence needed for wine tasting, and also the educational goal. It involves a range of cognitive processes in which the students become gradually capable of evaluating the wines: appraising, assessing, comparing and contrasting, criticising, grading, interpreting, and ranking them.

4.3 Interpretive functions

In OB, the relevant interpretive situations are situations in which the intended user needs help in a wine tasting situation in order to interpret information and information conventions printed on the label. This information, regulated by law, also contains a commercial dimension (promotion of wine) and is extremely complex in Burgundy. It is planned that OB would provide direct access to the interpretation of wine labels using visual recognition software and make it available on smartphones. This will be achieved by the indexing of wine label data, both as generic data (for the recognition and interpretation of the main types of appellations) and as specific data (for the recognition of specific wines and wine growers).

4.4 Operative situations

Operative situations are situations in which the intended user needs immediate assistance to carry on the act of wine tasting. The data includes online instructions, recommendations, etc., and is presented in the form of an online wine tasting multimodal manual (texts and videos). Access is obtained through table of contents, questions and answers, and an indexation of related keywords, as shown in Table 5 below:

questions	index of keywords
What kind of glass?	glass shape volume quality

How much to pour into the glass?	pour
Which temperature?	temperature
How to hold on the glass?	hold
How to swirl the glass?	swirl
Where to look in the glass?	look rims sides

Table 5: Questions and index of keywords

5. Dual access mode in OB

Online wine dictionaries like the wine glossary of VDB 2013 demonstrate two kinds of access to the data, namely consultation and navigation. Consultation, i.e. making queries to search and retrieve the data, is the default mode of dictionary use. Consultation is aimed at the appropriation of data, while navigation is aimed at the exploration of data. Navigation, in which the user follows hyperlinks to navigate between articles and pages, is normally the privilege of online dictionaries, but it should be noticed that it is also possible to navigate in paper dictionaries (Nau 2007), mainly by following internal links to other articles or to outer texts, or by simply reading articles for pleasure as a kind of lexicotainment.

5.1 Consultation

The consultation mode is characterised by the "I search" user behaviour. Consultation is motivated by a problem-based information need and aimed at the search for a solution to the encountered problem. The lexicographic relevance lies in the satisfaction of the information need. The point of departure of the search lies in a pre-formulated query. For instance, when a reader of a text on wine making encounters the term "astringency" the meaning of which is unclear them, they can make a search in the glossary on the VDB website, either analogically (through the alphabetic macrostructure of the glossary) or digitally (by typing the term in the query window and launching the search engine, which will return the information from the database). Consultation is goal-oriented because it is always aimed at getting a specific answer. The number of access routes (marked out by means of data structuring devices) is also limited. The access perspective is ruled by the lexicographer in so far as the answer to the query is given beforehand. Consultation is a *modus operandi* in which information is retrieved from relevant data presentation. Consultation belongs to the functionalist paradigm of information science, in which success

or failure of information searches is measured after the use of the dictionary. Hereby the capacity of the dictionary — or lack of it — to satisfy the needs and fulfil the function is also measured. OB is designed to allow for this type of access in the foreseen user situations. Table 6 below shows the main characteristics of consultation:

CONSULTATION	
motivation	problem-driven information need
relevance	satisfaction of need
behaviour	search for data
point of departure of query	preformulated
form of access	goal-oriented
data structuring	mainly relational
number of access routes	limited
dominating perspective	mainly lexicographer
information paradigm	functionalism

Table 6: Consultation mode in OB

5.2 Navigation

Navigation is the opposite of consultation. It is characterised by the "I do not search, I find" user behaviour and motivated by a pleasure-driven desire for information. In OB, it is achieved, as explained above, through the multiple interlinking of the dictionary itself (the database) and the website of the BIVB:

NAVIGATION	
motivation	pleasure-driven desire for information
relevance	satisfaction of curiosity
behaviour	discovery of data
point of departure of query	none
form of access	exploratory
data structuring	mainly associative

number of access routes	potentially unlimited
dominating perspective	mainly user
information paradigm	constructivism

Table 7: Navigation mode

Relevance lies in the satisfaction of the curiosity which drives the user to explore the data. There is no pre-formulated query or point of departure. The data are structured into associative networks, with a potentially unlimited number of access routes, and the user exposes themselves to data retrieval through discoveries being made along these routes. Access is of the exploratory kind, as neither a goal nor a logical path is needed. Access is guided by links and headings on the website, and routes are subject to fortuitous encounters generating pleasure. Navigation is ruled by the user perspective, as the access process itself is central, including obstacles, diversions and dead ends that can tease the user's curiosity and redirect them. Navigation belongs to the constructivist paradigm of information. In OB, the functions involved in navigation are cognitive-oriented, as the purpose here is to acquire new knowledge, or to check on already existing knowledge in connection with wine tasting training programmes at the *École des vins de Bourgogne*.

6. Conclusion

In this article, it has been shown how an online subject-field-based wine dictionary (Van der Merwe 2008) could benefit from an expansion of its data and from integration with the resources of the websites of the South African wine industry. It could subsequently be transformed into an integrated information tool in order to satisfy even more efficiently the specific information needs of the intended user groups, particularly students and wine lovers. It has also been shown how a user-situation-based specialised lexicographic information tool, in this case OB, gathers multimodal data and provides monofunctional access to the data in order to satisfy the information needs of its intended users. Through integration with the website of the wine industry, OB offers a dual access mode — consultation and navigation — in order to help its users to get access to, or discover, the data that will help them learn how to verbalise the subtle nuances of wine tasting experience and to relate their knowledge of wine to the tasting itself in order to contextualise and enrich the experience. In this respect, OB is completely in line with what Béjoint (2010: 386) has in mind when predicting the transformation potential of online lexicography:

Probably the dictionary as we know it is on its way out, and we will see the emergence of new kinds of tools, reference tools encompassing more than the dictionary, containing other kinds of information and providing a better treatment of the more traditional presentations.

OB is also in line with Prinsloo (2005: 11), who depicts this transformation even more accurately:

The great capacity and speed characteristics of electronic products, combined with enhanced query and data retrieval technology, indeed pave the way to a new generation of dictionaries unimagined in the paper-dictionary era.

In co-operation with the professionals of the wine industry, a proposal is made here to design online wine industry dictionaries that are truly integrated into the websites of the wine industry, wine boards and authorities, or with the websites of providers of wine tourism experiences (wine routes, visit to estates etc.). Websites of wine industry and wine tourism are indeed characterised by user-situation-based content management systems aiming at the satisfaction of the heterogeneous information needs of wine students, customers, lovers, visitors, etc. On such websites, data selection — and presentation — is to a large extent ruled by a user-situation-based (read: customer- and market-oriented) strategy: who is interested, in what, for what reasons, when, and how — being the crucial questions that have been answered in order to cater for the needs of the intended user groups. Such websites are not only sources of inspiration for lexicographic innovation, but also valuable information resources that can directly contribute to developing the functional quality of online wine industry dictionaries. Finally, it is worth noticing that the lexicographic proposals formulated in this article are not specific to wine industry. They also apply, without limitations, to the development of lexicographically structured information tools in other kinds of industries, sectors of the economy, corporations, and businesses: banking, insurance, real estate, medicine, technology and engineering, tourism, etc. Naturally, data differ, and so do priorities given to the specific user-situations involved, but the overall lexicographic principles for the online integration of user-situation identification, data selection, data access and data presentation remain identical.

Note

- † As explained in the English version of VDB 2013: "The vineyard of Burgundy is made up of small plots known as climats and lieux-dits, pieced together like a mosaic. These give the Burgundy wine region its special character and rich diversity."

Literature

Dictionaries and other lexicographically structured information tools

DLV = Coutier, M. 2007. *Le dictionnaire de la langue du vin*. Paris: CNRS.

Drync Wine Free and Pro. 2013. *Find and Manage Wine Online or on Your Mobile*. <http://www.drync.com/> (19/11/2010).

- École des vins de Bourgogne.** 2013. <http://www.ecoledesvins-bourgogne.com> [accessed 30 August 2013]
- GHV** 2010. *Guide Hachette des Vins*. Paris: Hachette pratique.
- GHV** 2013. *Guide Hachette des Vins*. Paris: Hachette pratique. <http://www.hachette-vins.com/le-guide-hachette-des-vins/> [accessed 30 August 2013]
- SAWIS.** 2013. *S.A. Wine Industry Information and Systems NPC*. <http://www.sawis.co.za/> [last seen 1 June 2013]
- VDB.** 2013. *Vins de Bourgogne*. Le site officiel des vins de Bourgogne. Bureau Interprofessionnel des Vins de Bourgogne. Beaune Cedex. <http://www.vins-bourgogne.fr/> [accessed 1 June 2013].
- Winetech and SAWIS.** 2012. *South African Trilingual Wine Industry Dictionary* <http://www.sawis.co.za/dictionary/index.php> [accessed 1 June 2013].
- Wikipedia.** 2013. *Portail de la vigne et du vin*. http://fr.wikipedia.org/wiki/Portail:Vigne_et_vin [accessed 1 June 2013].
- Wine.co.za.** 2013. *Virtual Home of South African Wine*. <http://www.wine.co.za/> [last seen: 1 June 2013]

Theoretical contributions

- Béjoint, H.** 2010. *The Lexicography of English*. Oxford: Oxford University Press.
- Bergenholtz, H. and U. Kaufmann.** 1997. Terminography and Lexicography. A Critical Survey of Dictionaries from a Single Specialised Field. Tarp, S. (Ed.). 1997. Specialised Lexicography and Terminology. *Hermes* 18: 91-125.
- Bergenholtz, H. and S. Tarp.** 2010. LSL Lexicography or Terminography? The Lexicographer's Point of View. Fuertes-Olivera, P.A. (Ed.). 2010: 27-36.
- Carstensen, S.** 1989. Zur Darstellung kulturgebundener Wörter in der zweisprachigen Lexikographie. *Kopenhagener Beiträge zur Germanistischen Linguistik* 25: 5-21.
- Fata, I.** 2010. The Bilingual Specialised Translation Dictionary for Learners. Fuertes-Olivera, P. and H. Bergenholtz. 2010. *Specialised Dictionaries for Learners*: 83-103. *Lexicographica*. Series Maior 136. Berlin/New York: De Gruyter.
- Fuertes-Olivera, P.A. (Ed.).** 2010. *Specialised Dictionaries for Learners*. *Lexicographica*. Series Maior 136. Berlin/New York: De Gruyter.
- Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.).** 2011a. *e-Lxicography: The Internet, Digital Initiatives and Lexicography*. London/New York: Continuum.
- Fuertes-Olivera, P.A. and H. Bergenholtz.** 2011b. Introduction: The Construction of Internet Dictionaries. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011: 1-16.
- Fuertes-Olivera, P.A. and S. Tarp.** 2014. *Theory and Practice of Specialised Online Dictionaries: Lexicography versus Terminography*. Berlin/New York: De Gruyter. [to appear]
- Gouws, R. and P. Leroyer.** 2009. Verhoogde leksikografiese toeganklikheid in die oorgang van 'n toeristewoordeboek na 'n toeristegids as naslaanbron. *Tydskrif vir Geesteswetenskappe* 49(1): 145-159.
- Granger, S. and M. Paquot (Eds.).** 2012. *Electronic Lexicography*. Oxford: Oxford University Press.
- Leroyer, P.** 2008. Du support d'information à l'outil lexicographique: la lexicographisation du guide touristique. Bernal, E. and J. DeCesaris (Eds.). 2008. *Proceedings of the XIII EURALEX Interna-*

- tional Congress, Barcelona, 15–19 July 2008*: 879-886. Sèrie Activitats 20. Barcelona: Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada.
- Leroyer, P.** 2011. Change of Paradigm: From Linguistics to Information Science and from Dictionaries to Lexicographical Information Tools. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011: 121-140.
- Leroyer, P. and H. Bergenholtz.** 2014. Métalexicographie culturelle, fonctions lexicographiques et finalité pragmatique. *Études de linguistique appliquée*. [To appear.]
- Leroyer, P. and R. Gouws.** 2009. En termes de vin — Modularisation lexicographique du guide œnotouristique en ligne. *Revue Française de Linguistique Appliquée* 14(2): 99-116.
- Lipovetsky, G.** 2006. *Le bonheur paradoxal. Essai sur la société d'hyperconsommation*. Paris: Gallimard.
- Nau, J.-Y.** 2007. Revue de Coutier, Martine. Dictionnaire de la langue du vin. *Le Monde des livres*, 21 September 2007.
- Prinsloo, D.J.** 2005. Electronic Dictionaries Viewed from South Africa. *Hermes* 34: 11-35.
- Tarp, S.** 2008a. *Lexicography in the Borderland between Knowledge and Non-knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Lexicographica. Series Maior 134. Tübingen: Max Niemeyer.
- Tarp, S.** 2008b. The Third Leg of Two-legged Lexicography. *Hermes. Journal of Language and Communication Studies* 40: 117-131.
- Tarp, S.** 2012. Online Dictionaries: Today and Tomorrow. *Lexicographica* 28(1): 253-268.
- Van der Merwe, M.F.** 2008. Wine and Words: A Trilingual Wine Dictionary for South Africa. *Lexikos* 18: 337-348.

How Dictionary Users Choose Senses in Bilingual Dictionary Entries: An Eye-Tracking Study

Robert Lew, *Department of Lexicography and Lexicology, Faculty of English,
Adam Mickiewicz University in Poznań, Poland*
(rlew@amu.edu.pl)

Marcin Grzelak (grzelak.m.r@gmail.com)
and

Mateusz Leszkowicz, *Department of Pedutology, Faculty of Educational
Studies, Adam Mickiewicz University in Poznań, Poland*
(mateuszl@amu.edu.pl)

Abstract: We use modern eye-tracking technology to scrutinize the process of sense and equivalent selection in polysemous bilingual entries. Our study subjects, intermediate and advanced Polish learners of English, consulted 26 Polish-to-English dictionary pages prompted with a sentence translation task. Throughout the task, an eye-tracking device unobtrusively recorded their gaze patterns, which are analyzed and discussed. Both successful and unsuccessful searches are examined. Also, we assess the potential of eye-tracking technology in the study of dictionary use.

Keywords: EYE TRACKING, BILINGUAL DICTIONARY, ENTRY NAVIGATION, INNER ACCESS, MICROSTRUCTURE, POLYSEMOUS ENTRIES, SENSE INDICATION, ENGLISH LANGUAGE, POLISH LEARNERS

Opsomming: Hoe woordeboekgebruikers betekenis in tweetaligewoordeboekinskrywings kies: 'n oogvolgstudie. Ons gebruik moderne oogvolgtegnologie om die proses van betekenis- en ekwivalensiekeuse in poliseme tweetalige inskrywings te ondersoek. Ons studiepersone, intermedieëre en gevorderde Poolse aanleerders van Engels, het 26 Poolse-na-Engelse woordeboekbladsye geraadpleeg vir die doel van 'n sinsvertalingstaak. Gedurende die hele taak het 'n oogvolgtoestel onopsigtelik hul kykpatrone geregistreer wat ontleed en bespreek word. Sowel suksesvolle as onsuksesvolle soektogte word beskou. Ook beoordeel ons die potensiaal van oogvolgtegnologie by die studie van woordeboekgebruik.

Sleutelwoorde: OOGVOLGING, TWEETALIGE WOORDEBOEK, INSKRYWINGNAVIGASIE, BINNETOEGANG, MIKROSTRUKTUUR, POLISEME INSKRYWINGS, BETEKENISAAN-
DUIDING, ENGELSE TAAL, POOLSE AANLEERDERS

1. Introduction

Researchers working in the area of dictionary use have always wanted to be able to observe which parts of dictionary entries dictionary users are paying attention to when engaging in dictionary consultation, and in what sequence. In search of clues as to which particular portions of an entry are used and when, investigators have resorted to introspective techniques based on self reports. These might variously involve underlining, think-aloud protocols, or written self-recording sheets.

In studies which employ underlining (e.g. Bogaards 1998; Lew and Dziemianko 2006; Lew 2010), users are usually asked to physically mark in pencil the particular fragments of the dictionary entry which they happen to be consulting while engaged in a task calling for lexicographic support. There are several quite serious downsides to this technique. One is due to attentional factors: asking someone to underline text in a dictionary ties up a portion of their attention, which is otherwise busy with both the dictionary consultation task and the primary task for which dictionary assistance is sought. Another problematic issue is the degree to which the effort of monitoring which parts of a dictionary are being used affects the very way in which dictionary users make use of a dictionary. It is quite likely that the distortion is substantial, as they need to be constantly aware of the monitoring aspect, as they attend to registering details of their dictionary consultations. Finally, there is a real danger of some of the consultation activity being left unrecorded: participants fail to underline as expected, as they become focused on the main task.

Similar problems beset the use of think-aloud protocols for recording dictionary consultation (Al-Besbasi 1991; Mackintosh 1995; Wingate 2002). Here again we are likely to obtain a self-conscious, incomplete, and distorted picture of consultation behaviour. The degree of success with this technique depends substantially on the skills of the participants in following the protocol.

Written protocol sheets are another option (e.g. Harvey and Yuill 1997). But if completed during the consultation, they tend to be even more intrusive and distracting than underlining. As their complexity naturally reflects the structural involvedness of dictionaries themselves, the quality of the data returned is questionable. In contrast, retrospective protocols completed after dictionary consultation will fail to record the interesting detail due to memory limitations.

All of the above options suffer from problems which severely diminish the validity of recording dictionary consultation in any of these ways. Until very recently, there had been no way to collect reliable information on which sections of the dictionary entry the user was consulting and in what sequence. But such an option became a possibility with the advent of modern eye-tracking technology. Today's equipment allows non-intrusive monitoring of participants' gaze, yielding insights into the patterns of dictionary consultation.

This study uses eye tracking to look into the process of inner access, that is

entry-internal navigation. Our participants were instructed to look for the sense which held an English translation equivalent appropriate for the context of the Polish sentence cue presented with the entry. As eye tracking has been used very little in dictionary user studies so far, another goal of the study is to examine the applicability of this technique to the study of dictionary entry navigation.

The design of the study will be explained in section 4, and its results will be presented in section 5. Before we get to the study itself, however, some background on eye movement research will be given below (section 2), focusing on its application within dictionary user studies, followed in section 3 by an overview of previous studies dealing with sense selection. As we want to leave as much space as possible to presenting the results of the study, we try to keep the overview sections brief. Readers wishing to learn more about eye tracking are invited to consult Rayner (1998), whereas Nesi and Tan (2011) offer a comprehensive overview of findings on sense selection.

2. Eye tracking in dictionary user research

Eye tracking, also known as gaze tracking or eye movement recording (EMR), is by no means a new technique: it has been in sporadic use for over 100 years now. But it was only recently that advances in technology made it an attractive and affordable option for researchers in many domains, most importantly in the examination of various aspects of reading and visual processing, human-computer interaction and web design.

Eye tracking covers a cluster of related techniques for monitoring and recording fine movements of the eye as an indication of where the subject is looking at a given time, as well as the sequencing of gaze movements across some visual-perceptual space.

Human visual perception normally involves a series of intermittent *fixations*, during which the location of the gaze is relatively stable, and *saccades*, which are quick movements in between the consecutive fixations, when no significant visual processing takes place. Gaze behaviour is usually interpreted as reflecting perception, on the strength of the *eye-mind assumption* (Just and Carpenter 1980). A fixation then is assumed to represent perceptual and cognitive processing of stimuli. In reading a text, longer fixations imply longer processing, possibly due to increased attention, such as when facing some difficulty. Gaze *regressions* in reading are movements against the normal text orientation: in English, regressions are movements to the left, as in any left-to-right writing language. More extensive regressions may take the gaze back to a previous line. These are often indicative of the reader backtracking during reading in order to resolve a processing problem. Parameters of eye movement such as gaze duration, saccade length or search time are believed to correlate well with processing complexity (Duchowski 2007). Eye fixations reflect the encoding of

information based on the stimuli being viewed. Research on reading, mostly English texts by fluent native speakers, typically finds mean fixation duration for single fixations to be on the order of 225 milliseconds for silent reading and 275 milliseconds for reading aloud, although specific values vary both individually and with text difficulty (Rayner 1998: 373, 2009). There are no reliable data as yet on eye movement parameters during the reading of a dictionary entry, which is a formally structured special text, qualitatively different from normal reading matter.

Only a handful of studies of dictionary use have so far employed eye-tracking technology. Three of these studies are due to Henrik K hler Simonsen. Simonsen (2009a) investigated gaze patterns and gaze duration of users consulting an online Danish accounting dictionary, looking for evidence of differences in reference behaviour associated with different lexicographic functions, or modes of using the dictionary (L1 knowledge acquisition, L1 production, L1 reception, and L1-to-L2 translation).

In the same year, Simonsen (2009b) compared gaze parameters of professional translators working with vertical and horizontal data presentations in an internet dictionary. He was also concerned with general methodological issues of viability of eye tracking for studying internet dictionary consultation. In Simonsen (2011), the author further explored the applicability of the analysis options typically present in eye movement data analysis software for dictionary user reference behaviour.

A study by Kaneta (2011) looked at the frequency and duration of reference to illustrative examples in two forms of digital entry presentations: unfolded (flat) and folded (layered), in both monolingual and bilingual entries. Rather predictably, when illustrative examples were hidden from the initial view, they were consulted less often than in a complete presentation.

Tono (2011) used eye-tracking technology to examine a number of variables related to look-up behaviour. This work is especially relevant to the present study, as his investigation focused on entry navigation devices (menus and signposts). It will be summarized in the following section.

3. Previous studies of sense navigation

Tono (1984) was a pioneering work addressing the issue of how users select senses in entries. In this study, dictionary users exhibited a tendency to pick the first sense of a bilingual entry and ignore the remainder of the entry, unless the first sense did not fit in an obvious way. A number of subsequent studies focused on entry navigation devices in the form of (entry-initial) menus and (sense-initial) signposts, typically in monolingual entries (Tono 1992, 1997; Bogaards 1998; Tono 2001; Lew and Pajkowska 2007; Lew 2010; Nesi and Tan 2011; Tono 2011), though not exclusively (Lew and Tokarek 2010). Most studies have confirmed the value of signposts, both in terms of helping users find the

right sense, and in terms of speed. Signposts were found to be more effective than menus by Lew (2010) and Nesi and Tan (2011), but Tono (2011) reports the opposite.

An important challenge to previous findings on the advantage of entry-initial senses came from Nesi and Tan (2011), who found entry-*final* senses to be at least as salient as entry-initial ones. These disparate results are not necessarily contradictory if we allow for the fact that entry navigation strategies may be contingent on several factors, including the user's reference skills, proficiency in language, type and form of the dictionary, properties of the lexical item being looked up, and the task which prompted dictionary consultation. It is quite possible that the final-sense advantage arose in this case as part of a specific consultation strategy of fairly experienced dictionary users consulting English monolingual learners' dictionaries in comprehension tasks. Such users may have discovered through continued dictionary work that the most relevant senses are usually found towards the bottom of polysemous entries, as the most frequent senses at the top are usually familiar.

There are very few studies investigating sense identification in bilingual dictionaries (Lew and Tokarek 2010). Sense guidance in L2→L1 bilingual dictionaries is largely achieved by virtue of the fact that entries feature equivalents in the users' native language. Thanks to their salience in the respective senses, such entries can usually be scanned quite efficiently, and the need for additional navigation aids is diminished. In contrast, L1→L2 dictionaries feature equivalents in a language of which the user has only partial knowledge. Reading, and especially scanning, foreign language text is obviously less efficient than in one's native language. Also, many of the L2 equivalents given in an entry will not be well known to the dictionary user, and thus provide few clues to meaning. Multi-word expressions and phrases in the source language (L1) may offer useful visual pivots in those entries that cover them, but in order to distinguish between decontextualized equivalents, the better bilingual dictionaries supply sublemmatic guiding elements in the form of sense indicators and equivalent discriminators. Sense indicators in bilingual dictionaries and signposts in monolingual entries are in fact quite close, both structurally and functionally.

A study by Tono (2011) deserves special attention in this context as it is similar to the present study in both its goals and use of an eye-tracking system. Tono attempted to test several variables at a time, and the results are somewhat complex and difficult to interpret unambiguously. Perhaps the most important finding to take out of Tono (2011) is that consultation behaviour is rarely systematic, but tends to be erratic. This suggests that a neat, simple model of dictionary consultation which implicitly underlies many lexicographers' efforts may be too much of an idealization. Tono concludes by calling for further study of the look-up process with the use of eye-tracking systems. The present paper responds to this call.

4. The study

4.1 Aim

The overall aim of this study is to examine how users navigate polysemous bilingual dictionary (L1 to L2) entries in a lexical search scenario induced by a sentence translation task from L1 (Polish) to L2 (English). Further goals are:

- to uncover patterns of dictionary users scanning to locate senses in a bilingual dictionary, as revealed by eye-tracking data;
- to compare successful and unsuccessful searches and explore their possible correlates; and
- to explore the applicability of the eye-movement paradigm to the study of dictionary entry navigation.

4.2 Participants

Participants in the study were ten Polish university students. Half of them were English majors and thus advanced learners of English (CEFR level B2 to C1). These participants made up the high-proficiency (HP) group, and reported using dictionaries on a daily basis. The other five participants were majoring in the following areas: preschool education; modern Greek; architecture and town planning; environmental engineering; and corporate management. These five low-proficiency (LP) students (CEFR level A2 to B1) admitted to using dictionaries several times a week. All recruited subjects had full visual acuity (20/20), some with correction. The eye-tracking system we used is tolerant of correction glasses and (untinted) contact lenses, so this was not an issue.

4.3 Materials

Two sets of polysemous Polish-to-English dictionary entries were adapted from two modern Polish–English bilingual dictionaries: thirteen items from *Nowy Słownik Fundacji Kościuszkowskiej. The New Kosciuszko Foundation Dictionary* (NKFD 2003) and another thirteen items from *Wielki Słownik Angielsko-Polski, Polsko-Angielski PWN-Oxford* (PWNO 2002). We chose these two dictionaries as they are currently the only comprehensive general bilingual dictionaries between Polish and English compiled and published in this century. They are comparable in size and coverage. Both dictionaries use similar techniques and devices for sense guidance, with Polish as the metalanguage in the respective Polish–English volumes. Senses are most typically indicated by means of near-synonyms, hyponyms, collocates or domain labels, which is standard fare in general bilingual dictionaries of high quality. Dictionary page mock-ups were constructed for the twenty-six items, replicating the original typography

closely, but increasing the font size and line spacing so as to make the text comfortable to read on screen from a viewing distance of 60 cm by subjects with normal vision. Each word was presented in a context sentence constructed for the experiment, and placed in a rubric above the dictionary extract which included the entry for the target word. Experimental stimuli were prepared as single-page PNG graphic files at a native Tobii T60 screen resolution (1280 x 1024 pixels) so as to avoid interpolation and aliasing distortion. The top 9% (90 pixels) of the screen included a frame with the sentence cue against a light-grey background. The remainder of the screen held the mock-up dictionary page with the test entry. Shorter entries were accompanied by alphabetically neighbouring entries, as they would in a paper dictionary. A typical stimulus item with a dictionary page mock-up and a sentence cue appeared as in **Figure 1**. There were twenty-six screen pages, each containing a sentence cue with an underlined key word, including thirteen entries from each dictionary.

Wskazówki na tarczy jego zegarka wskazywały godzinę dwunastą.

<p>tarcza <i>f</i></p> <p>1. (rycerska) shield; osłaniać się tarczą przed ciosami przeciwnika to protect oneself from the enemy's blows with a shield</p> <p>2. (połcyjna) (riot) shield</p> <p>3. (z cyframi, liczbami) dial; tarcza zegarka (na rękę) the dial of a. on the watch; (stojącego) the face of the clock; tarcza kompasu a compass dial; tarcza telefoniczna a. telefonu a telephone dial</p> <p>4. Techn. (w maszynie) disc GB, disk US; tarcza tnąca a cutting disc; tarcza szlifierska a grinding wheel; tarcze hamulcowe brake discs</p> <p>5. (ceł) target; strzelać do tarczy to aim at the target; trafić w tarczę to hit the target; trafić w środek tarczy to hit the bullseye; tarcza strzelnicza a shooting target</p> <p>6. (szkolna) school badge</p> <p>7. (herbowa) shield, escutcheon</p> <p>8. (ciał niebieskich) disc GB, disk US; tarcza słoneczna the sun's disc; tarcza Księżycza the face of the moon</p> <p>■ być zwycięsą a. dla kogoś tarczą książk. to shield sb;</p> <p>■ wrócić na tarczy książk. to return defeated; wrócić z tarczą książk. to return victorious; wrócić z tarczą lub na tarczy I will return victorious or die honourably nabrało dla niej nowych treści meeting him gave (a) new meaning to her life</p> <p>tarczowły</p> <p>I adi. pila tarczowa circular saw; hamulce tarczowe disc brakes</p> <p>II m środ., Spot person who checks the scores in target shooting</p>	<p>tarczycę <i>f</i> ^{Anat.} thyroid (gland); nadczynność/niedoczynność tarczycy an overactive/underactive thyroid</p> <p>tarczycowjy <i>adi.</i> ^{Anat.} thyroid <i>attr.</i>; gruczoł tarczycowy the thyroid (gland)</p> <p>Tarenit <i>m</i> (G Tarentu) Geog. Taranto</p> <p>targ</p> <p>I m (G targu)</p> <p>1. (rynek) market; targ koński/zbożowy a horse/grain market; kupiła kury na targu she bought hens at the market; targ staroci an antiques market; pchli targ a flea market; w piątki jest targ w miasteczku Friday is market day in the town</p> <p>2. <i>pot.</i> (o cenie) haggling, bargaining; zapłaciła za ziemniaki bez targu she didn't haggle over the price of potatoes; dobić targu to strike a bargain a. deal</p> <p>3. <i>zw. pl pot.</i> (spory) bargaining <i>U</i>; targi o podwyżkę wage bargaining; po długich targach ustapili they gave up after a lot of haggling</p> <p>II targi <i>pl</i> (wystawa) fair; targi książki a book fair; targi branzowe a trade fair; targi motoryzacyjne a car show; wystawiać swoje towary na targach to display one's goods at a fair; zwiedzać targi to visit an exhibition a. an exposition a. a fair</p> <p>■ targ w targ <i>pot.</i> after much hard bargaining</p> <p>targać ¹ <i>impf</i> → targnąć</p> <p>targać ² <i>impf vt</i></p> <p>1. (wichrzyć) [wiatr] to tousle [włosy] ⇒ potargać</p> <p>2. <i>przest.</i> (rozrywać) to tear; targał na drobne kawałki stare</p>
--	---

Figure 1: A sample stimulus item with a dictionary page.

The twenty-six Polish key words were: kosz, siatka, poślizg, blok, ekspozycja, emisja, język, forma, przedmiot, siła, rakietą, rezerwa, promień, paczka, treść, płyta, album, dyscyplina, figura, legenda, korek, prąd, wpaść, serce, tarcza,

podzielać. The criteria guiding the selection of words and target senses were: familiarity, difficulty, and sense position in the dictionary. We chose familiar Polish words, but in less frequent senses which, though clear to the participants at both levels, would present a challenge in selecting their English equivalents. We wanted participants to focus on picking the correct dictionary sense with its English equivalent rather than puzzle over the meaning of the Polish word. For example, the familiar Polish word *poślizg* was used, not in its default sense 'skid', but in the metaphorically derived sense 'delay'. The context of each sentence cue made it clear which sense was meant, which was verified during the piloting stage.

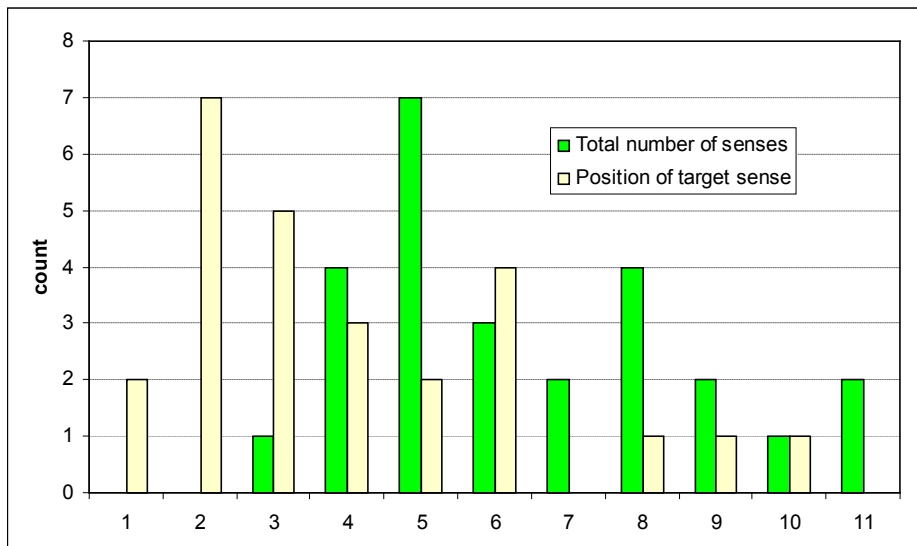


Figure 2: Distribution of the (1) total number of senses in the entry and (2) position of the target sense within the entry.

We have included entries of between three and eleven senses in length, whereas the position of the target sense in the entry ranged between first and tenth. A detailed distribution of the total number of senses in the entry and the position of the target sense is given in **Figure 2**. For example, at $x=5$, the two differently-shaded bars indicate, respectively, that the complete set of 26 items included seven entries of five senses, and there were two entries in which sense five was the target sense. Mean length of entry was 6.5 senses (median = 6). Mean position of target sense was 4.0 (median = 3). Although it was not our intention to compare the two dictionaries, we nevertheless made an effort to balance entry length and target sense position across the dictionaries. Thus, for PWNO and NKFD, respectively, mean length of entry was 6.4 (median = 7) and

6.5 (median = 5), and mean position of target sense was 3.8 (median = 3) and 4.2 (median = 3).

4.4 Apparatus

A Tobii T60 eye tracker was used in the study. The unit has a sampling rate of 60 Hz, a nominal spatial accuracy of at least 0.5 degrees and drift below 0.3 degrees of visual angle. It is equipped with a 17-inch screen with a native resolution of 1280 x 1024 pixels. An important advantage of this model is that it requires no head restraining, allowing the participant a fair amount of head movement. The tracking cameras are integrated in the main unit and are not obvious to the user. The unit looks like an ordinary flat-screen computer monitor. Thanks to these features, the Tobii T60 has high ecological validity, offering participants the look and feel of a regular computer screen, thus a highly naturalistic setting for students accustomed to working with a computer. The software used in the design of the experiment and data collection during the recording sessions was Tobii Studio, version 2.0.8.

4.5 Procedure

The experiment took place in a spacious, daylight university office. A single experimenter (the second author) worked individually with one participant at a time, in several sittings during January 2012. Participants would be seated in front of the Tobii T60 unit placed on top of a work desk, at a viewing distance of about 55 centimetres from the monitor, with the eyes at a level just below the centre of the screen. The unit was connected to two computers: one running the Tobii Studio software to control the tracking unit and collect the data, the other for the researcher to monitor progress. During the experiment, the experimenter was able to monitor the participants' posture on his screen, and correct it if needed.

Before the recording session, each participant was given specific instructions (in Polish) about the procedure. Participants were asked to keep looking at the screen and try not to move their heads too much. Then the eye tracker was calibrated, once for each participant. The participants followed a red dot on the screen with their eyes as it moved around for a few seconds. Once the calibration was successful, the recording could begin.

Each participant was presented with the same twenty-eight screens in turn. The first screen included basic instructions reminding the participants what they should focus on. The last screen indicated the end of the recording. The other twenty-six screens included dictionary pages with sentence cues as described above. The order in which the twenty-six experimental items appeared was randomized to minimize any order effects. For each of the twenty-six items, participants were asked to translate the underlined word by locating the appropriate sense within a polysemous entry presented below the

context sentence. They were asked to speak the translation out loud after they had decided on the correct equivalent. A complete audio transcript of the sessions was made using a digital recorder. We had rejected the option of asking the participants to write down the answers themselves, as this would have made them look away from the monitor and might have disrupted the gaze recording. We did not want to ask them to give the sense number itself, as this might have made them too aware of the sense selection aspect. With the solution adopted, the equivalent itself did not in every case unambiguously indicate which particular sense was chosen, as sometimes different senses shared an equivalent. However, coupled with a subsequent review of eye-scan paths, the sense selected could always be determined with high certainty. The complete procedure was piloted on two students to ensure that all elements worked as expected. The main experiment proceeded smoothly with no problems, yielding good quality eye-tracking data, which were subsequently analyzed. The only slight complication was that the unit we used exhibited spatial accuracy issues for the extreme upper area of the screen. In recording such extreme top values, it tended to offshoot towards the margin of the screen in the vertical dimension. The problem was caught during initial testing (even before the piloting), and so the top strip of the screen was used for the sentence cue. Since twenty-three of the twenty-six sentence cues fit on a single line of text, this did not cause any ambiguity in interpreting the data, and in any case our main interest was not in how participants read the sentence cue, but how they worked with the dictionary excerpt below. Importantly, in the entire area of the screen used for the dictionary mock-up there were no spatial accuracy issues, so analysis by Areas of Interest could proceed without distortion.

4.6 Data analysis

The experiment generated 260 complex searches (data from ten participants, each looking up twenty-six items). Each search was classified as successful or unsuccessful, depending on whether the participant located the contextually correct sense in the correct entry. For a successful search, two conditions had to obtain at the same time: (1) the participant had to provide the correct English equivalent for the sentence cue in the verbal feedback; and (2) an examination of the gaze paths showed fixations on the target sense which coincided with the correct English equivalent.

Complete eye movement data from all 260 searches were collected. Measures used in the analysis included fixation counts and fixation duration. To detect fixations, the Tobii sliding-average algorithm was used as described in Olsson (2007), with the default threshold radius of 35 pixels. This filter setting turned out to be very effective in detecting fixations in both the inner access searches (within the entry) and outer access searches (headword scan). We also tested the ClearView filter at the settings recommended by Gerganov (2007) for translation-related data: 80 milliseconds minimum fixation duration, and 40

pixels fixation radius. This setting worked reasonably well for inner access searches, but missed some of the quicker headword scans which were apparent in the raw data. In order to capture those as well, the ClearView filter had to be set at 40 ms for minimum fixation duration and at least 60 pixels fixation radius. Visual scan paths were generated for all searches to aid in the qualitative assessment of look-up behaviour.

Areas of Interest (AOIs) were defined and plotted manually for all twenty-six dictionary excerpts. AOIs were entered separately for each entry so that in each case they covered the following entry components (refer to **Figure 3**):

- the headword (coded as *hw*);
- each sense (coded as *sN*, where *N* was the relevant sense number); and
- sense-guiding element(s) for each sense and any embedded subsenses or phrases within the sense (coded as *gN*, where *N* was the relevant sense number).

This was done in order to allow the computation of gaze data specific to the structural components of the entries: individual senses and their guiding elements.

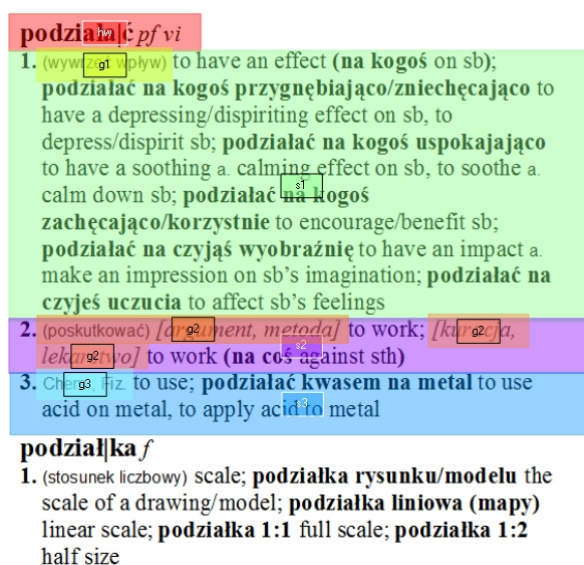


Figure 3: Areas of Interest (AOIs) marking and coding (*hw* = headword; *sN* = sense number *N*; *gN* = guiding element for sense *N*)

Fixation counts and fixation duration were computed for the above entry components, by participants and by items separately. In addition, separate calculations were made for the structural elements of the target sense: the target sense

itself and any of its guiding elements.

The eye tracker also logged the time stamp for every event, so accurate timing data became available as well, and these were used in calculating temporal parameters of the consultations.

The software used in the analysis included: Tobii Studio 2.0.8, OpenOffice Calc 3.4.1, LibreOffice Calc 4.0.0.3, Microsoft Excel 2003, Statistica 8.0, and statistical test applets at <http://www.vassarstats.net> (Lowry 2001–2013).

5. Results and discussion

5.1 Task completion time

The mean time for a single participant to complete the whole task (all 26 items) was 557 seconds. Considering the two groups of participants separately, an average high-proficiency student needed 509 seconds to finish the task, while in the low-proficiency group the mean time spent on the task was 19 per cent longer at 605 seconds. This difference, however, is not statistically significant (one-way ANOVA, $F_{(1, 8)} = 0.94$, $p = 0.36$, partial $\eta^2 = 0.11$), likely due to a small number of subjects coupled with large individual variation. A detailed breakdown of the time taken to complete the whole task is given in **Table 1**.

Table 1: Task completion times for individual participants in the two groups

low proficiency		high proficiency	
participant	total time [sec]	participant	total time [sec]
LP-03	652	HP-01	713
LP-06	615	HP-02	500
LP-07	844	HP-04	346
LP-08	514	HP-05	591
LP-09	401	HP-10	394
LP mean	605	HP mean	509
LP SD	165	HP SD	149

Next, task completion times were computed for each item (see **Table 2**). An average item took 20.1 seconds to complete ($SD = 6.6$ seconds). Mean values computed for NKFD and PWNO items separately were 19.1 and 21.1 seconds, respectively. This difference is not statistically significant (one-way ANOVA, $F_{(1, 24)} = 0.61$, $p = 0.44$) and its effect size is very weak (partial $\eta^2 = 0.02$). In addition, it might be noted that one of the PWNO items (*plyta*) took more than twice the average time to finish, and this item is responsible for the high standard deviation value in the PWNO set. If this one outlier is ignored, the mean for PWNO goes down to 19.2, and becomes virtually identical to the corresponding NKFD figure.

Table 2: Task completion times for individual items and the two dictionaries

NKFD		PWNO	
item	mean time [sec]	item	mean time [sec]
blok	16.8	album	13.1
ekspozycja	20.9	dyscyplina	15.7
emisja	16.2	figura	14.4
forma	23.8	korek	17.0
język	18.9	legenda	13.4
kosz	21.6	paczka	18.5
poślizg	17.9	plyta	43.6
promień	17.0	podzialać	28.5
przedmiot	18.5	prąd	14.6
rakieta	12.8	serce	24.8
rezerwa	14.9	tarcza	21.4
siatka	24.6	treść	20.5
siła	23.9	wpaść	28.8
NKFD mean	19.1	PWNO mean	21.1
NKFD SD	3.67	PWNO SD	8.67

5.2 Sense selection success rates

There were 260 searches overall (26 items x 10 participants). Of those searches, 157 covered all senses in the entry, i.e. there was evidence in the eye-movement data of participants fixating on every sense within an entry. The remaining 103 searches failed to examine one or more senses within an entry. Nineteen percent of all searches (50 out of 260) resulted in the wrong sense being chosen. One might hypothesize that errors in sense selection might be more likely in incomplete searches, and less likely in searches covering all senses, as incomplete searches carry a greater risk of missing the target sense. To assess whether this was actually the case, we tabulated error counts separately for complete and incomplete searches in **Table 3**. Altogether, there were 50 searches resulting in errors of sense selection. Of these, 26 errors occurred in complete searches (an error rate of 17%, or a success rate of 83%), and 24 errors in incomplete searches (an error rate of 23%, success rate 77%). The difference between the proportions of errors in the two types of searches is not statistically significant (Z-test for independent proportions, z-score = -1.349, $p = 0.18$). The error rates were thus quite similar across the two types of searches, suggesting that viewing all senses was no guarantee of getting the sense right. Conversely, incomplete searches could well be successful: these searches typically stopped once the target sense was positively identified.

Table 3: Proportion of searches resulting in erroneous sense selection, when all senses have been examined, and when some senses have been omitted

	all searches	wrong sense selected	correct sense selected	sense selection error rate	sense selection success rate
all senses viewed	157	26	131	17%	83%
not all senses viewed	103	24	79	23%	77%
all searches	260	50	190	19%	81%

5.3 Sense selection success, entry length, and sense position

To investigate the relationship between sense selection success and entry length, we computed separate Pearson correlation coefficients between sense selection success rates and three measures: (1) total number of senses; (2) absolute target sense position; and (3) relative target sense position within the entry. None of these correlations turned out to be statistically significant, and detailed results are given in **Table 4**.

As shown in the top row of **Table 4**, there is a weak positive correlation ($r = 0.28$) between the total number of senses in the entry and sense selection success rate. Therefore, there is no evidence here that it was easier to find the correct sense in shorter entries than in entries with a larger number of senses. The tendency is actually the reverse, although it is not significant ($p = 0.16$).

Table 4: Pearson correlations (r) between the sense selection success rate and (1) total number of senses; (2) absolute target sense position; and (3) relative target sense position. Also given are: coefficient of determination (r^2), and the t -score and p -level for $H_1: r \neq 0$ against $H_0: r = 0$

	r	r^2	t	p
total number of senses	0.28	0.08	1.4	0.16
absolute target sense position	-0.03	0.001	-0.13	0.90
relative target sense position	-0.29	0.08	-1.5	0.16

The second row of **Table 4** reveals a near-zero correlation, or an almost perfect independence, between the success rate and sense number of the target sense. This suggests that the ease of locating the target sense was independent of how

far from the beginning of the entry the relevant sense was located.

In the last row, correlation is given of the success rate with a relative measure of target sense position. This was computed by dividing the target sense number by the total number of senses in the entry. For example, sense number five in an entry of ten senses would have a relative position of $5/10 = 0.5$ (or 50%). This measure exhibits a small degree of negative correlation with the success rate of sense selection. There is thus a slight tendency for target senses that are relatively early in the longer entry to be somewhat easier to find, but this tendency is not statistically significant ($p = 0.16$). This best estimate of the correlation coefficient yields a low coefficient of determination of $r^2 = 0.08$; this means that only about 8% of the success in locating the target sense could be accounted for by the relative placement of the target sense within the entry.

Remarkably, there was a perfect match in all searches between target sense selection success and correct provision of the contextually appropriate English equivalent. This means that once participants were able to locate the correct sense, they had no difficulty extracting the right equivalent. In most cases, high-proficiency participants were able to locate the relevant sense, pick the correct English equivalent and use it in the translation of the sentence cue. Of the five high-proficiency participants, two did not make any sense selection errors at all, one made a single error, another one committed two errors, and the one least successful HP participant made five errors in the 26 items. This translates into an overall error rate of six percent only in the HP group. Selection errors were far more frequent in the low-proficiency group, with 42 erroneous sense selections out of a total of 130 individual lookups, that is 32% of all cases, or over five times the error rate in the high-proficiency group. The errors were more or less evenly distributed across the individual low-proficiency students.

Interestingly, nearly all cases (with only two exceptions) of incorrect selection involved choosing a sense earlier than the target sense. Specifically, in sixteen cases of erroneous sense selections, the sense selected was placed one sense above the target item, and in nineteen cases (the most typical scenario) a sense preceding the target by two senses was chosen. More tellingly, though, of the 50 erroneous selections, 35 (or 70%) involved the participants' choosing the first sense of an entry. This corroborates the results of some previous studies (Tono 1984; Lew 2004), where users were found to pick the first sense if they did not know any better. In contrast, we found no evidence in our data of an advantage of entry-*final* senses which was reported by Nesi and Tan (2011). In fact, not a single erroneous choice in our study involved a participant wrongly selecting an entry-*final* sense. Further, detailed eye-tracking data revealed that users typically proceeded from the top of the entry downwards rather than from the bottom up. This possible difference in reference behaviour may find explanation in the fact that our study examined bilingual entries, unlike in Nesi and Tan (2011), where entries from English monolingual learners' dictionaries

were used. These entries tended to be arranged by sense frequency, with the rarest senses appearing towards the end, and an awareness of this organizing principle might have prompted Nesi and Tan's subjects to work from the final senses upwards. But bilingual dictionaries tend to be organized differently, with the range of applicability of the translation equivalent playing an important role, and in this case the traditional top-down scanning of entries as predominantly used by our participants may represent the better default strategy.

Of the total fifty misidentified senses, eleven errors were made in WSPO, and thirty-nine in NKFD. This might invite the conclusion that WSPO is more user-friendly in its sense guidance. However, we should keep in mind that the experiment was not designed to compare the two dictionaries. In contrast to Tono (2011), we preferred not to present the same headword more than once to the same participant, mindful of the risk of carry-over effects. In our experiment, the two sets of entries from the two dictionaries (thirteen each) had all different headwords. Although we made an effort to match the two sets in terms of sense position and entry length (see section 4.3 above), there are often subtle and unpredictable effects in language data which make some items more difficult to the experimental subjects than others. It may be that a greater number of such difficult items found their way into the NKFD set, negatively impacting consultation success in these cases. Another possible confounding factor is the familiarity of participants with the dictionaries. WSPO is probably the more popular of the two modern comprehensive Polish–English dictionaries, and thus more likely to be known. Having said this, it must be stressed that the two dictionaries are fairly similar in terms of sense discrimination strategies, layout and typography, which makes item difficulty the more probable reason for the observed difference.

5.4 Gaze fixation statistics

Overall mean duration of fixation was 298 milliseconds ($SD = 190$ ms): that is only about 30 per cent longer than the typical value for silent text reading by native speakers (Rayner 1998). This would suggest that, in terms of eye-movement parameters, consulting dictionary entries is not dramatically different from normal text reading, and experience gained in extensive eye-movement research in reading can, with some caution, be drawn on in dictionary user studies. Fixation duration was fairly uniform across our ten participants. Mean per-subject values ranged from 262 ms (participant HP-05) to a high of 319 ms (participant HP-02). Means calculated for each of the groups separately yield 307 ms for low-proficiency and 290 ms for high-proficiency participants, which is a modest difference and not statistically significant (one-way ANOVA, $F_{(1, 8)} = 2.05$, $p = 0.19$, partial $\eta^2 = 0.20$).

Across all items, participants made 9267 fixations within any of the senses. Of these, 22% (2019) were focused on sense guiding elements. This proportion was slightly higher for high-proficiency participants (23% on average) than for

low-proficiency participants (21%). This small difference turned out not to be statistically significant (one-way ANOVA, $F_{(1,8)} = 0.43$, $p = 0.53$, partial $\eta^2 = 0.05$).

In terms of total *dwell time* (i.e. cumulative duration), fixations on guide-words accounted for 23% of the time spent looking anywhere within any of the senses. Here again, a slightly higher proportion of time spent on guiding elements is evidenced in searches by high-proficiency participants (25% of the time on average) than those by low-proficiency participants (21% of the time). Similarly as in the case of fixation counts, this difference was not statistically significant (one-way ANOVA, $F_{(1,8)} = 1.12$, $p = 0.32$, partial $\eta^2 = 0.12$).

The above figures indicate that there was a tendency, albeit not statistically significant, for lower-level participants to fixate for longer periods. With regard to the proportion of attention directed towards sense guiding elements, the two proficiency levels used guiding devices to a similar extent, unlike in Tono's (2011) study. This suggests that sense guidance was universally useful.

5.5 Patterns of look-up behaviour

The main aim of this study has been to examine how dictionary users look up senses in polysemous bilingual entries in translation-induced production, and how the position of the sense affects the process. This major section deals with this issue with the help of two types of visualizations of eye-tracking data: scan paths and heat maps.

A systematic qualitative analysis of the scan paths of all the lookups revealed that by far the dominant strategy was to engage in a systematic scan of the senses, starting at the top of the entry and proceeding in a downwards direction until the last sense was reached. Participants usually scanned rather rapidly through the senses, mostly focusing on their sense indicators, until they reached what they believed was the right sense. At that point, they would normally proceed to read the entire sense rather more carefully. A typical example of such a pattern of consultation behaviour is mapped out in **Figure 4**. In this scan path representation, fixations are shown by dots (the larger the dot, the longer the duration). The numbers in the dots represent their temporal sequencing.

Participants would normally start by reading the sentence cue in Polish (fixations 3-6). The first brief fixation or two would sometimes be elsewhere (just as the new stimulus first appeared), not uncommonly within the central area of the screen (this is normal). Once in the sentence cue, they would soon focus on the underlined word, whose equivalent was sought (here the longer fixation number 7). Next, they would scan the headwords in search of the lemma sign. In this particular case, as is usual in a highly-inflected language such as Polish, they would need to reduce the inflected noun form (*formie*, locative) to its citation form (*forma*, 'form'). Here the scanning did not start at the top left of the dictionary page, where the lemma sign *forma* was actually to be found, but rather in a vertical motion down to the headword *formacja* (fixa-

tion 9). This was not the right lemma, but the subsequent fixations (10, 11) focus on the correct headword *forma*. Next comes a fairly systematic skim over the sense indicators of the consecutive senses (fixations 12-19). Fixations 16 and 17 likely cover two one-line senses each, as the sense indicators lie within the area of foveal vision for these fixation points (at the viewing distance used in the experiment, foveal (= sharpest) vision covers a circle of approximately 2.5 centimetres in diameter, which in the vertical dimension corresponds to about three lines of text in our material).

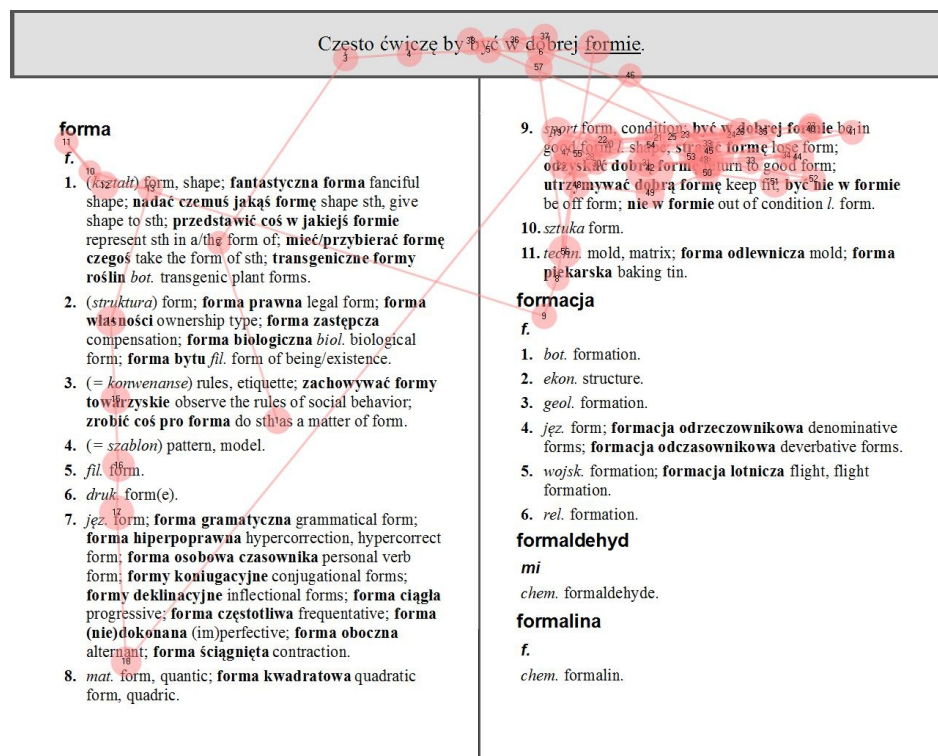


Figure 4: A typical scan path of a successful search.

Having reached sense 9 (our target sense), this high-proficiency participant spends a long time studying the sense, including the multi-word expressions. Much attention is directed (fixations 24-35) at the expression *być w dobrej formie* ('keep fit'), which is actually the expression used in the sentence cue. At this point, the participant makes a detour back to the second half of the sentence cue (fixations 36-38), apparently to check the match of the context against the expression located in sense 9. He then reconfirms the match (39-41), but goes on to read the remaining phrases nevertheless (fixations 42-55). A look at the two remaining senses (fixation 56) and return of the gaze to the sentence cue

conclude this successful search. We should stress again that about 80% of the searches recorded in the experiment were of this nature (though not necessarily going beyond the target sense).

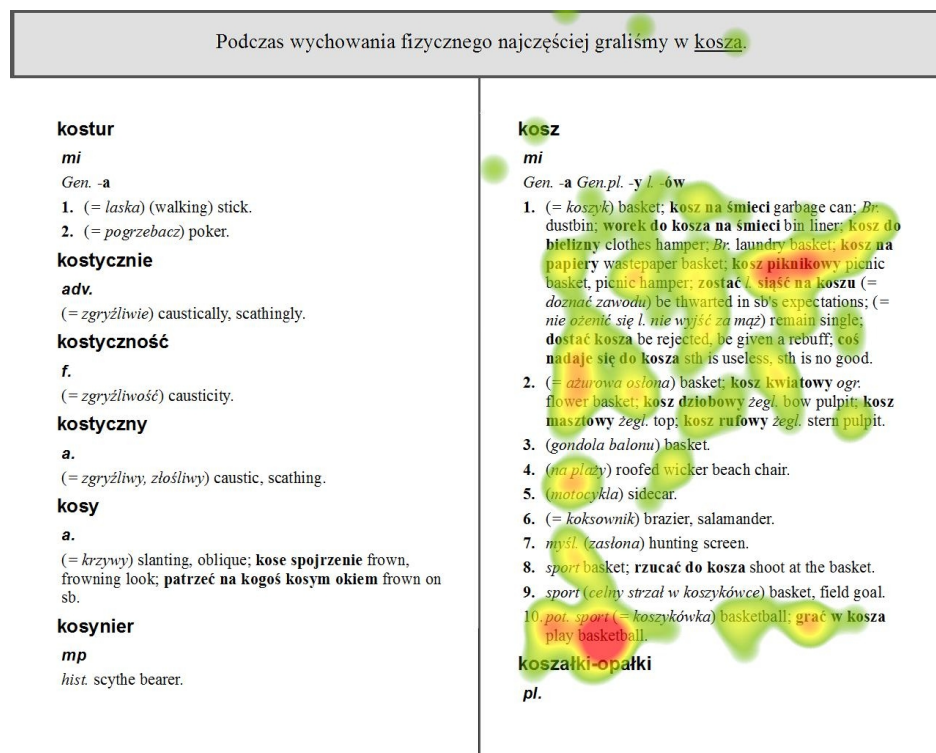


Figure 5: Fixation count heat map of a successful search for an entry-final sense by a low-proficiency participant. An animated version of this figure is available at <http://vimeo.com/59560367>.

Another example of a successful search, this time from a low-proficiency student, is given in **Figure 5**. The figure presents a so-called fixation count *heat map* (an animated version is available at <http://vimeo.com/59560367>). In this type of visualization, the more an area has been looked at, the hotter (redder) its colour (in non-colour print this appears as a darker shade). In this entry (*kosz*, 'basket'), our target was the very final sense (10). It represents an informal use referring to basketball as a game (*koszykówka* in general Polish). The fixation pattern indicates that the participant (LP07) reviewed the respective sense indicators and, having spent some time scanning (twice) sense one and examining some bold-type phrases in the earlier senses, correctly homed in on the final sense, and then identified a nested phrase which corresponds to the expression in the sentence cue, *grać w kosza* 'play basketball'. The challenge of this particu-

lar item lies in the fact that the entry includes, not just one, but three basketball-related senses, all of them marked with the domain label *sport*. Sense 8 refers to the circular net used in basketball, and sense 9 to a goal scored in the game.

However, participants did not necessarily examine all the senses. Just as suggested by Tono (1984), searches stopped at the first sense if the translation seemed to fit. This strategy was quite common among the low-proficiency participants, but only very occasional with high-proficiency students. **Figure 6** illustrates this difference in approach, using the entry *rakieta* ('missile') as an example. In this item, the sentence cue unambiguously referenced a Polish-made anti-aircraft missile.

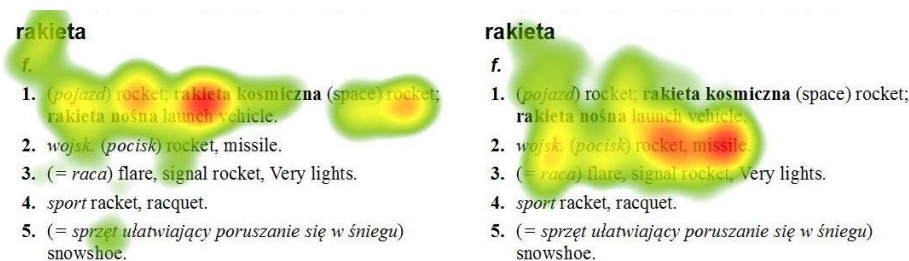


Figure 6: Two cumulative fixation heat maps for the item *rakieta* ('missile'), with the military sense number 2 being the target. On the left, unsuccessful consultations by low-proficiency students; on the right, successful consultations by high-proficiency participants.

The heat map on the left represents cumulative data from the five low-proficiency participants, four of whom opted for the incorrect first sense. On the right, data from high-proficiency participants are visualized: here, four out of five participants selected sense number 2, which best reflects the use of the word in the original sentence cue. Sense 1 is the most general sense and is indicated by the hyponym *pojazd* ('vehicle'). Sense 2 is the specific military sense and has two sense indicators: the abbreviated domain label *wojsk.* for *wojskowość* ('military'), and the near-synonym *pocisk* ('projectile'). The advanced learners were for the most part able to locate this specific military meaning, and the heat map reveals that they had studied the sense indicators. The lower-proficiency participants selected sense 1, which, admittedly, is not a completely wrong-headed choice in this case. Another interesting finding showing up in the heat map is that the first word of the phrase **rakieta kosmiczna** attracted a lot of attention from low-proficiency participants. The likely reason for this is the typography: it was the first bold-type element within this entry. This observation underscores the important role that typography plays in dictionary entries.

The above analysis illustrates an important tendency: while lower proficiency participants were often happy to skip any further senses once they were

reasonably satisfied they had identified the correct choice, participants in the high-proficiency group often preferred to continue looking at the remaining senses as well. This systematic strategy of high-proficiency participants resulted in success in the above case, but in many longer entries it led them to examine the remaining senses even after the correct earlier sense had already been positively identified, possibly wasting some time.

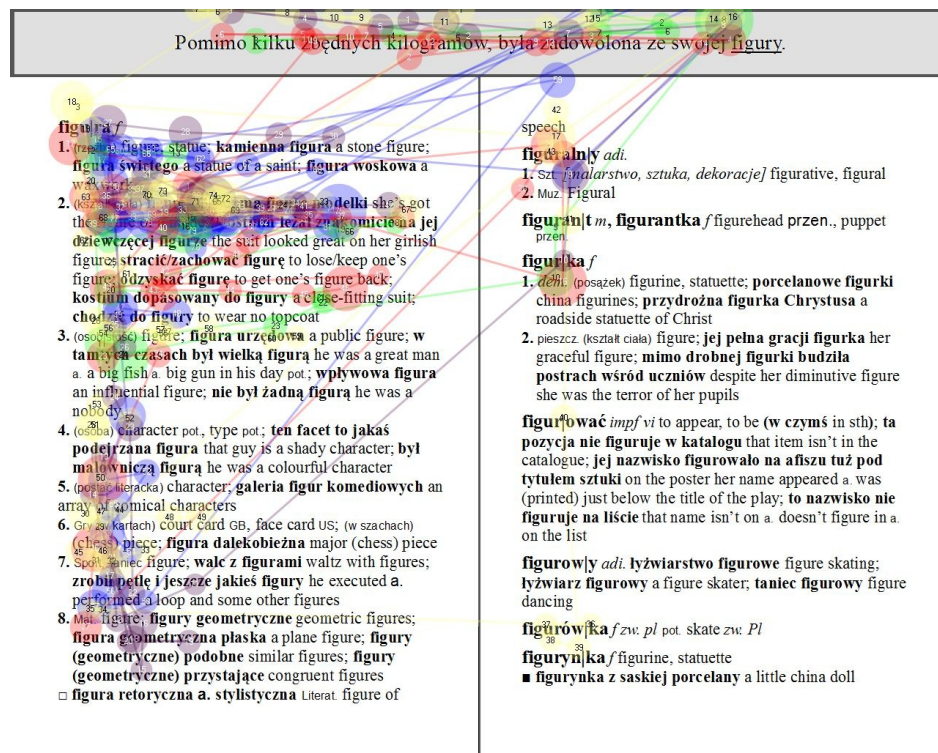


Figure 7: Cumulative gaze plot of the five HP participants consulting the entry for *figura*. Most looked at all senses, with sense 2 being the appropriate one.

A typical scenario illustrating this strategy of high-proficiency students is illustrated in Figure 7, a cumulative gaze plot of the five high-proficiency participants working with the entry for *figura* ('figure, shape'). The plot indicates that four out of five HP subjects examined all eight senses, and one participant viewed sense 3. Thus, all high-proficiency participants looked beyond the target sense 2. Most subsequently went back to the appropriate sense, but having checked the remaining senses first.

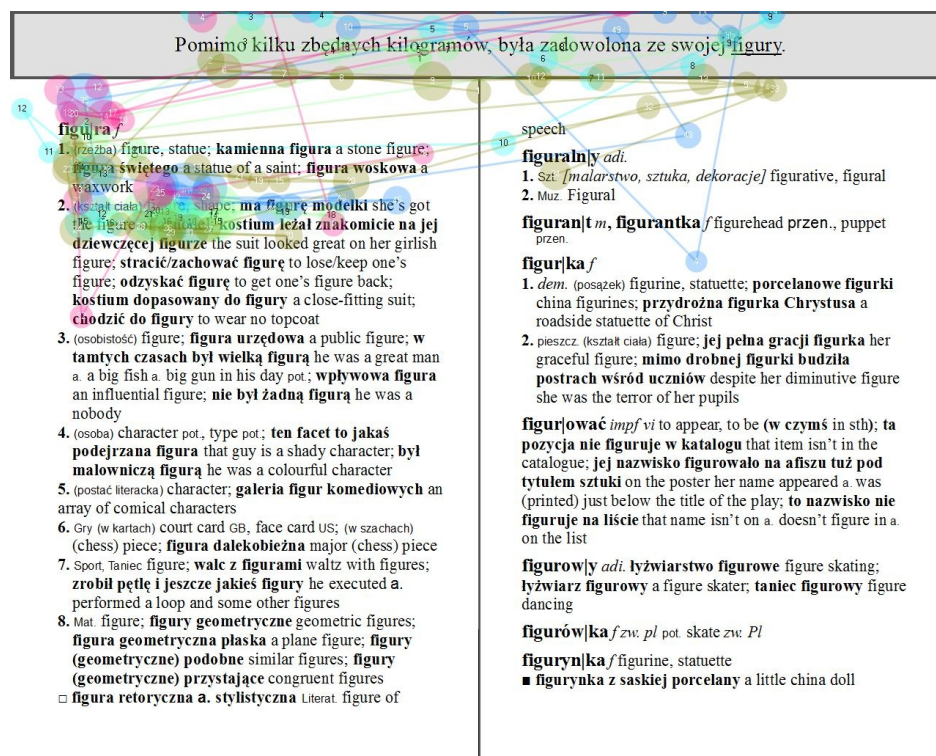


Figure 8: Cumulative gaze plot of the five LP participants consulting the entry for *figura*. All students stopped at the appropriate sense 2.

In such cases, the dominant strategy of our low-proficiency participants may be described as more economical than the one evident in the HP group: they dispensed with examining the remaining senses beyond the sense found relevant to the context of the sentence cue, in this case sense 2. In other words, once they were satisfied with sense 2 being the appropriate one, they moved on. This difference of strategy is evident in many items. On the other hand, low-proficiency students did not work any faster than the high-proficiency group (see section 5.1 above). However, it is possible that they worked more slowly in general and, had they decided to also examine further senses, they would have needed more time than the high-proficiency participants.

An interesting case of a relatively short entry where a substantial proportion of users have missed a specific equivalent is that of *promień* ('radius', as of a circle). Here, the wrong equivalent *ray* was supplied in four instances of unsuccessful searches (Figure 9). The required equivalent was *radius*, as used in geometry. This sense was clearly marked in the entry under sense 3, which held the domain label *geom.*, transparent to a Polish speaker. However, the four unsuc-

cessful users did not get to this sense. Instead, they registered the many instances of *ray* as an equivalent, and presumably concluded from the repeated tokens of *ray* that this equivalent is universal enough and it will do in this case as well.

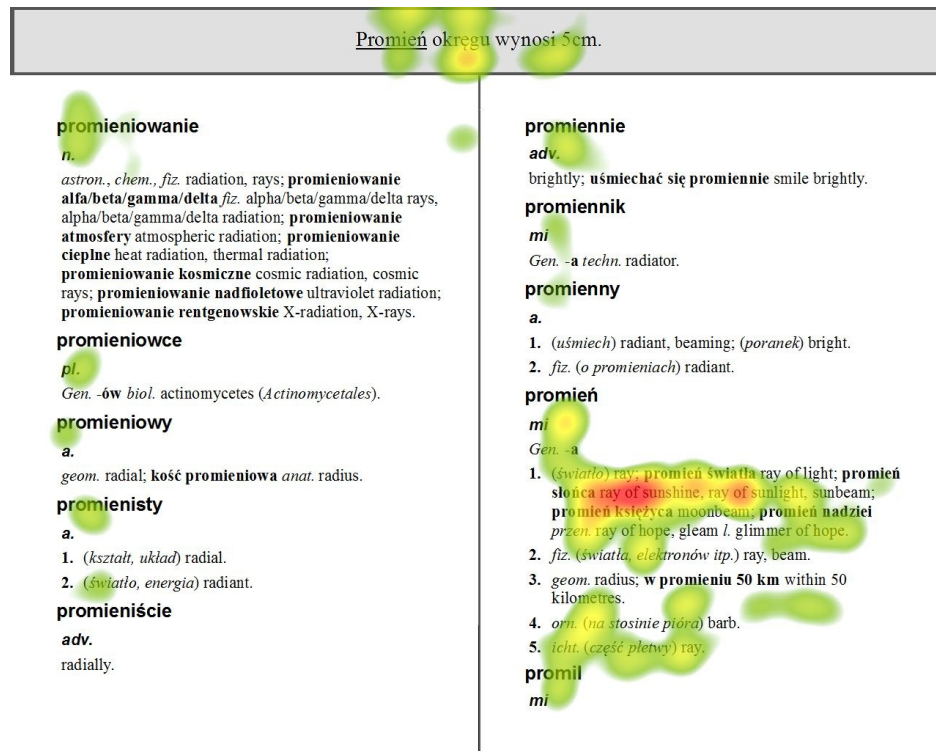


Figure 9: Cumulative fixation count heat map for the four students who gave the wrong equivalent of *promień* ('radius').

A very similar scenario was noted in the case of the item *poślizg* ('delay'). This would suggest that visual prominence of an equivalent which occurs several times within an entry can sometimes override the significance of sense indicators, effectively preventing dictionary users from selecting an isolated equivalent when exposure to several tokens of a more popular equivalent gives them false confidence that the latter will fit in just about any context. It is hard to come up with an acceptable lexicographic solution to this kind of problem. Giving extra salience to such isolated equivalents through typography might be one option, but it is controversial as it might in turn result in an overuse of this particular equivalent. Another avenue is to try to avoid repeating a frequent equivalent too many times, but rather to group together senses served by the same equivalent. Considered from this point of view, an equivalent-struct-

ture of the bilingual entry might be better than a source-language structure (Jarošová 2000; Adamska-Salaciak 2006; Lew 2013).

5.6 Headword scan patterns

Our study focused on the inner access search process, i.e. the search within the entry once it has been identified, with a view to locating the relevant sense. However, since we used page-sized dictionary mock-ups, our data also allows some conclusions with regard to how dictionary users scan the dictionary page in search of the relevant headword. These findings, based mostly on raw (unfiltered) data, will be presented briefly, as we had not planned to investigate this aspect. However, we do want to include them, as they have important methodological implications for the design of similar studies in the future.

Gaze data indicate that our dictionary users typically scanned the headwords in the top-down direction. However, they did not necessarily start their search in the left-hand column (our mock-ups used the usual two-column layout). Instead, they often went from the end of the sentence cue to the top of the right-hand column, presumably because it was closer to the end of the sentence. Moreover, once they had reached the bottom headword in the column without finding the lemma sign sought, they would often cross over to the other column on the same level and then proceed in the upwards direction. These findings suggest that dictionary users, when scanning the headwords on a page, optimize their gaze path so it is as short as possible, even if this goes against the alphabetic sequence.

Another interesting finding was that, whereas scanning for information within the entry resembled patterns found in silent text reading, headword scans patterned differently. They tended to be more rapid and the landing point would often miss the headword, with an ensuing corrective movement, producing loose clusters of gaze points around some lemma signs. Some such clusters went undetected when a default setting of fixation detection filters was used. With the standard ClearView filter, a larger fixation radius (over 60 pixels) had to be used, combined with a low threshold for minimum fixation duration (40 milliseconds). This finding might explain why in Tono's (2011) study a few searches appeared to jump straight to the relevant section, without apparent evidence of the user scanning the page. It may be that the scan was rapid enough to have been filtered out, as a minimum fixation duration of 100 milliseconds was used. When a page is scanned in search of a relevant section, as in a dictionary headword scan, a shorter fixation duration setting should be used.

6. Summary and conclusion

The present study of sense look-up patterns with the use of an eye-tracking system has produced some interesting results. Overall, participants in both

groups performed fairly satisfactorily, having extracted the correct sense from a bilingual entry about four times out of five. As one would expect, many more erroneous choices were made in the lower-proficiency group. Eye-tracking data indicate that in many cases participants chose to examine every single sense in the relevant entry. High-proficiency participants tended to adopt this strategy even after they had identified the target sense; in contrast, low-proficiency participants usually terminated their search having found the relevant sense. Somewhat surprisingly, no evidence was found of shorter entries resulting in better success in selecting the relevant sense; in fact, a reverse tendency was noted. However, there was a weak (but not significant) correlation between the relative position of the target sense within the entry and success, with a tendency for earlier senses to be easier to identify. In addition, whenever the wrong sense was chosen, it was almost invariably one located higher up than the target sense, most commonly the first sense of the entry. This finding points to the special salience of the first sense of an entry, but no similar effect was found in our data for entry-*final* senses. To maximize success, lexicographers should try to place a translation equivalent with the broadest possible range of application in the first sense of a bilingual dictionary entry.

Our data also indicate that elements in bold attract significant attention of the users, and that they tend to interpret repeated occurrence of an equivalent as evidence of its universal application. This, in principle, is a valid inference, but it does make it more likely for users to ignore or miss the less common equivalents. To avoid this effect, the same equivalent should not be repeated too many times, if only there is a way to group lexicographic data so as to avoid such repetition.

An important finding of this study was that sense guiding elements occupy a significant proportion of the users' attention (between one-fifth and a quarter in terms of both fixation counts and relative dwell time). This proportion was quite stable across participants of both proficiency levels, suggesting that sense indicators in bilingual entries do fulfil the purpose for which they were designed.

This study has also tested the application of eye-movement tracking technology to the investigation of dictionary look-up processes, being one of the first to apply this instrumental approach within dictionary user research. The results demonstrate that eye tracking is a highly appropriate technique, as it provides detailed, first-hand information on users' visual scan patterns, both at the outer and inner search stages. From these patterns we infer information on which elements users consulted, how long they dwelled on them, and in what particular sequence, as well as whether they revisited particular elements. This type of data has not been available by any other technique. Unfortunately, the accuracy of currently available physical-object eye-tracking technology is not good enough to capture fine detail of dictionary structure. One solution used here is to use screen-based page mock-ups, but this inevitably detracts from the naturalness of the dictionary use situation.

In order for the use of eye tracking to be maximally useful, certain stan-

dards are needed to guide the design of similar studies in the future. One finding from the present study is the mean duration of fixation during the process of entry consultation. This has turned out to be about 300 milliseconds, that is some 30% longer than for normal text reading by native speakers. As variation is substantial, the minimal fixation threshold of around 80 milliseconds seems a valid cut-off value for entry reading. However, to capture headword scanning on a dictionary page, a threshold as low as 40 milliseconds may be needed, and a higher setting (at least 60 pixels in our case, or about two degrees of visual angle) for the fixation radius.

Overall, eye-tracking technology proves to be a highly fitting and fruitful approach for examining what happens in dictionary consultation, and should be used more widely.

References

Dictionaries

- NKFD:** Fisiak, J., A. Adamska-Salaciak and P. Gąsiorowski (Eds.). 2003. *Nowy Słownik Fundacji Kościuszkowskiej. The New Kosciuszko Foundation Dictionary*. New York: The Kosciuszko Foundation.
- PWNO:** Usiekniewicz, J. (Ed.). 2002. *Wielki Słownik Angielsko-Polski, Polsko-Angielski PWN-Oxford*. Warszawa: PWN.

Other literature

- Adamska-Salaciak, A.** 2006. *Meaning and the Bilingual Dictionary: The Case of English and Polish*. Frankfurt am Main: Peter Lang.
- Al-Besbasi, I.** 1991. *Translation Processes and Dictionary Use*. Ph.D. Thesis. Exeter: University of Exeter.
- Bogaards, P.** 1998. Scanning Long Entries in Learner's Dictionaries. Fontenelle, T., P. Hiligsmann, A. Michiels, A. Moulin and S. Theissen (Eds.). 1998. *EURALEX '98 Actes/Proceedings*: 555-563. Liège: English and Dutch Departments, University of Liège.
- Duchowski, A.** 2007. *Eye Tracking Methodology. Theory and Practice*. London: Springer.
- Gerganov, A.** 2007. Appendix A: Eye Tracking Studies with Tobii 1750 — Recommended Settings and Tests.
- Harvey, K. and D. Yuill.** 1997. A Study of the Use of a Monolingual Pedagogical Dictionary by Learners of English Engaged in Writing. *Applied Linguistics* 18(3): 253-278.
- Jarošová, A.** 2000. Problems of Semantic Subdivisions in Bilingual Dictionary Entries. *International Journal of Lexicography* 13(1): 12-28.
- Just, M.A. and P.A. Carpenter.** 1980. A Theory of Reading: From Eye Fixations to Comprehension. *Psychological Review* 87(4): 329-354.
- Kaneta, T.** 2011. Folded or Unfolded: Eye-Tracking Analysis of L2 Learners' Reference Behavior with Different Types of Dictionary. Akasu, K. and S. Uchida (Eds.). 2011. *Asialex2011 Proceedings Lexicography: Theoretical and Practical Perspectives*: 219-224. Kyoto: Asian Association for Lexicography.

- Lew, R.** 2004. *Which Dictionary for Whom? Receptive Use of Bilingual, Monolingual and Semi-Bilingual Dictionaries by Polish Learners of English*. Poznań: Motivex.
- Lew, R.** 2010. Users Take Shortcuts: Navigating Dictionary Entries. Dykstra, A. and T. Schoonheim (Eds.). 2010. *Proceedings of the XIV Euralex International Congress*: 1121-1132. Ljouwert: Afûk.
- Lew, R.** 2013. Identifying, Ordering and Defining Senses. Jackson, H. (Ed.). 2013. *The Bloomsbury Companion to Lexicography*: 284-302. London: Bloomsbury Publishing.
- Lew, R. and A. Dziemianko.** 2006. Non-Standard Dictionary Definitions: What They Cannot Tell Native Speakers of Polish. *Cadernos de Tradução* 18: 275-294.
- Lew, R. and J. Pajkowska.** 2007. The Effect of Signposts on Access Speed and Lookup Task Success in Long and Short Entries. *Horizontes de Lingüística Aplicada* 6(2): 235-252.
- Lew, R. and P. Tokarek.** 2010. Entry Menus in Bilingual Electronic Dictionaries. Granger, S. and M. Paquot (Eds.). 2010. *eLexicography in the 21st Century: New Challenges, New Applications*: 193-202. Louvain-la-Neuve: Cahiers du CENTAL.
- Lowry, R.** 2001–2013. *Vassarstats. Website for Statistical Computing*. Accessed on <http://www.vassarstats.net/>.
- Mackintosh, K.** 1995. *An Empirical Study of Dictionary Use in Version*. M.A. Thesis. Ottawa: University of Ottawa.
- Nesi, H. and K.H. Tan.** 2011. The Effect of Menus and Signposting on the Speed and Accuracy of Sense Selection. *International Journal of Lexicography* 24(1): 79-96.
- Olsson, P.** 2007. *Real-Time and Offline Filters for Eye Tracking*. M.A. Thesis. Stockholm: Royal Institute of Technology.
- Rayner, K.** 1998. Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychological Bulletin* 124(3): 372-422.
- Rayner, K.** 2009. Eye Movements and Attention in Reading, Scene Perception, and Visual Search. *The Quarterly Journal of Experimental Psychology* 62(8): 1457-1506.
- Simonsen, H.K.** 2009a. Se — og Du Skal Finde: En Eyetracking-Undersøgelse Med Særlig Fokus på De Leksikografiske Funktioner. *Nordiske Studier i Leksikografi* 11. *Rapport Fra Konference Om Leksikografi i Norden. Finland 3.-5. Juni 2009*: 274-288. Tampere: Nordisk forening for leksikografi.
- Simonsen, H.K.** 2009b. Vertical or Horizontal? That Is the Question: An Eye-Track Study of Data Presentation in Internet Dictionaries. Paper presented at the Eye-to-IT Conference on Translation Processes, Sentence Processing and the Bilingual Mental Lexicon, Business School, Copenhagen, Denmark, April 25–29, 2009.
- Simonsen, H.K.** 2011. User Consultation Behaviour in Internet Dictionaries: An Eye-Tracking Study. *Hermes* 46: 75-101.
- Tono, Y.** 1984. *On the Dictionary User's Reference Skills*. B.Ed. Thesis. Tokyo: Tokyo Gakugei University.
- Tono, Y.** 1992. The Effect of Menus on EFL Learners' Look-up Processes. *Lexikos* 2: 230-253.
- Tono, Y.** 1997. Guide Word or Signpost? An Experimental Study on the Effect of Meaning Access Indexes in EFL Learners' Dictionaries. *English Studies* 28: 55-77.
- Tono, Y.** 2001. *Research on Dictionary Use in the Context of Foreign Language Learning: Focus on Reading Comprehension*. Lexicographica Series Maior 106. Tübingen: Niemeyer.
- Tono, Y.** 2011. Application of Eye-Tracking in EFL Learners' Dictionary Look-up Process Research. *International Journal of Lexicography* 24(1): 124-153.
- Wingate, U.** 2002. *The Effectiveness of Different Learner Dictionaries. An Investigation into the Use of Dictionaries for Reading Comprehension by Intermediate Learners of German*. Lexicographica Series Maior 112. Tübingen: Niemeyer.

Gabonese French Dictionaries: Survey and Perspectives*

Paul Achille Mavoungou, *Département des Sciences du Langage,
Université Omar Bongo, Libreville, Gabon*
(*moudika2@yahoo.fr*)

Abstract: This article is a survey study of lexicographic activities with regards to the French language as it is spoken in Gabon, a French-speaking country in central Africa. The French language as it is spoken in Gabon does not only refer to Gabonisms (mostly items derived from native Gabonese languages), but also to standard French, official French (acrolectal level), common French (mesolectal level) and popular French (basilectal level). All four levels are present in French as it is used in Gabon. In other words, the Gabonese have named the realities of their own environment with lexical items coined on their own soil or borrowed from other languages. This has strengthened and enriched French as it is spoken in Gabon. The modern era of Gabonese lexicography has witnessed the production of various dictionaries focusing on this African variety of French. Before 1999, in Gabon and due to the status of French as the sole official language of the country, French dictionaries used in Gabon were mostly French products. Monolingual French dictionaries produced in Gabon by Gabonese scholars include a very limited number of authors, namely: Dodo Bounguendza (2008), Ditougou (2009) and Moussounda Ibouanga (2011).

Despite the fact that these dictionaries were compiled for use by Gabonese, they are differential as they tend to include references of the Parisian standard French. Indeed, a differential dictionary is a dictionary focusing on differences between two language varieties or two similar languages, and current Gabonese French dictionaries are by all means differential. It is this situation, and the ways it can be corrected that this article will focus on. The idea is to produce a complete dictionary of variant B, namely Gabonese French. Such a dictionary will reflect certain emancipation from the French language spoken in France (i.e. Parisian standard French), herein referred to as variant A.

Keywords: GABONESE FRENCH, DIFFERENTIAL DICTIONARIES, EMANCIPATION, STANDARD FRENCH, GABON

Résumé: Les dictionnaires du français du Gabon: État des lieux et perspectives. Cet article fait le point sur les activités lexicographiques en ce qui concerne la langue française telle qu'elle est parlée au Gabon, un pays francophone de l'Afrique centrale. La langue française telle qu'elle est parlée au Gabon ne fait pas seulement référence aux gabonismes (majoritairement des termes dérivés des langues gabonaises locales), mais également au français standard, au français officiel (niveau acrolectal), au français commun (niveau mésolectal) et au français populaire (niveau basilectal). Tous les quatre niveaux sont présents dans le français qu'il est utilisé au

* This article was presented as a keynote address at the Seventeenth Annual International Conference of the African Association for Lexicography (AFRILEX), which was hosted by the Department of African Languages, University of Pretoria, Pretoria, South Africa, 2-5 July 2012.

Gabon. En d'autres termes, les Gabonais ont nommé des réalités de leur propre environnement avec des termes forgés sur leur propre sol ou empruntés à d'autres langues. Ceci a renforcé et enrichi le français tel qu'il est en usage au Gabon. L'ère moderne de la lexicographie gabonaise atteste de la production de différents dictionnaires se focalisant sur cette variété africaine du français. Avant 1999, au Gabon et en raison de la situation du français comme unique langue officielle du pays, les dictionnaires français utilisés au Gabon étaient pour la plupart des produits français. Les dictionnaires monolingues français produits au Gabon concernent un nombre très limité d'auteurs, à savoir: Dodo Bounguendza (2008), Ditougou (2009) et Moussounda Ibouanga (2011).

Malgré le fait que ces dictionnaires ont été compilés pour être utilisés par les Gabonais, ils sont différentiels car ils ont tendance à inclure des références à la norme parisienne du français. En effet, un dictionnaire est dit différentiel s'il met l'accent sur les différences entre les deux variétés d'une même langue ou de deux langues similaires et actuellement les dictionnaires du français du Gabon sont tous des ouvrages différentiels. C'est cette situation et les moyens de la corriger qui constituent l'objet du présent article. L'idée est de produire un dictionnaire complet de la variante B, à savoir: le français du Gabon. Un tel dictionnaire sera le reflet d'une certaine émancipation du français du Gabon par rapport à la langue française parlée en France, autrement dit le français de Paris, langue standard ou norme appelée variante A.

Mots-clés: FRANÇAIS DU GABON, DICTIONNAIRES DIFFÉRENTIELS, ÉMANCIPATION, FRANÇAIS STANDARD, GABON

Introduction

Nowadays French is geographically spread across a significant number of continents and countries. The geographical dispersal of French has brought the language into daily contact with new usages in Europe (Switzerland, Belgium, Luxemburg), in North America (Canada, Louisiana), in Africa (Senegal, Ivory Coast, Republic of the Congo, etc.), in Asia (Vietnam, Cambodia, Laos) and in the Middle East (mainly Lebanon). This has strengthened and enriched the French language as it is spoken in France. It is very important that dictionaries should give an account of the usage of French spoken outside the Hexagon, especially as it influences the form used in France. In any dictionary, the focus should be on one variety of a given language. But when including lexical items restricted to a variety exclusively used by the target users of the dictionary, the lexicographer should indicate this restricted usage by employing a system of geographical labels. A number of geographical labels are used in dictionaries, namely: Britishisms, Americanisms, Canadianisms, Quebecisms, Belgicisms, Africanisms and so on. The distinct nature of English varieties outside Britain has become an acceptable phenomenon. English dictionaries produced in America (American English), Canada (Canadian English), New Zealand (New Zealand English), South Africa (South African English), and so on are part of the world's dictionary landscape. Quite differently, French dictionaries produced outside France are hardly given a warm reception. Let us just mention the case of the lexicographic revolution in Quebec. The lukewarm, not to say openly hostile reception of the *Dictionnaire québécois d'aujourd'hui* (DQA),

edited by Jean-Claude Boulanger (1992) and supervised by Alain Rey of the Robert publishing house, is an example in this regard.

Since its introduction, French has been used in Gabon for more than 150 years and is no longer just a foreign language but also a native language, a second language and an international language. The severe colonial French language policy prohibited the use of any native language in the public domain (Kwenzi-Mikala 1990: 122) and ultimately led to the decay of native languages. Gabon is linguistically diverse; the amalgamation of the many ethnic groups (approximately 40) of the country was made possible by a common language: French. Originally it was the language of the conqueror, but it also became the true *lingua franca* of the country as well as the *language of national unity* after the country's independence in 1960. Since colonial times, the French language in Gabon has developed in various fascinating ways. A vivid national identity had been created. Subtle nuances and colourful images have appeared in Gabonese French.

This article seeks to reflect on the survey of Gabonese French dictionaries and lexica as well as concrete projects that are currently undertaken in order to improve the development of the variety of the French language spoken in Gabon. The whole is concluded by some ideas about the way to proceed in future.

In order to better understand the history of the compilation of French dictionaries and lexica in Gabon, it is important to go back to the roots of this dictionary production, especially with the arrival of the Europeans in Gabon.

The history of the French language in Gabon and bilingual dictionaries pairing French with indigenous languages

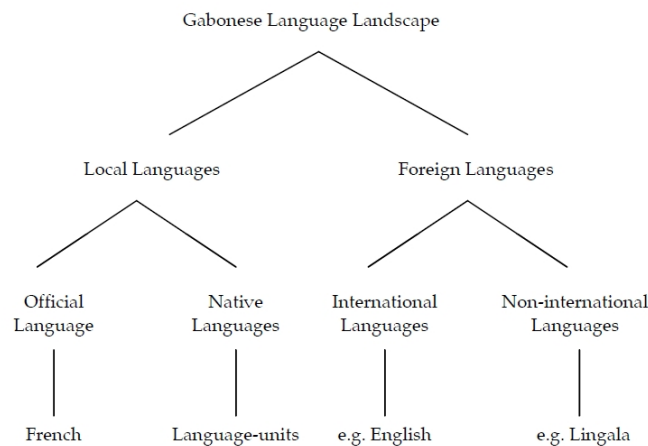
The first contact with Europeans was with the Portuguese in the fifteen century. During the following 350 years, first the Portuguese and later the French, Dutch, and the English carried on a lucrative trade in slaves from Gabon. The first permanent European settlement was made by the French, with the «agreement» of the Mpongwe ruler, in 1839. Libreville was founded a decade later by freed slaves. The French gradually extended their rule inland, and in 1866 appointed a governor to Gabon, which was then attached to the French Congo; it became part of French Equatorial Africa in 1910 (Encarta 1993-2000). After the attainment of independence, Gabon adopted the language of the former colonial master for functional purposes in official circles.

As far as lexicography is concerned, two eras of dictionary production in Gabon can be distinguished: the earlier era and the modern era. The early bilingual dictionaries in Gabon were compiled by missionaries and colonial officials and were to serve as reference works for European traders and French colonial administrators in their daily routine. These dictionaries can therefore be regarded, according to Gouws (2007: 314), as externally motivated products. Contrary to these dictionaries compiled by "foreigners", the modern era starts from 2002 with the publication of *Gedandedi sa geviya/Dictionnaire geviya-français* by Van der Veen and Bodinga-Bwa-Bodinga. The appearance of this dictionary opens a new era of dictionary publication in Gabon. As a matter of fact, it has

formed the basis for further lexicographic projects directed at the needs of the speech communities of Gabon (Mickala Manfoumbi 2004; UNESCO-LUTO 2006; Mavoungou and Plumel 2010; Idiata and De Nadaillac 2010). These new lexicographic products are intended to enhance the communicative abilities of the Gabonese speech communities and can therefore be regarded, according to Gouws (2007: 314-315), as internally motivated products. Moreover, Ella (2011, 2012) has discussed the compilation of Gabonese French dictionaries by showing that standard French dictionaries have to be adopted to fulfil the needs of the way French is spoken in Gabon.

Gabonese Language Landscape (GLL)

Ndinga-Koumba-Binza (2005a, 2005b, and 2007) indicated that the Gabonese language landscape refers to both foreign and local languages spoken in the country. The latter group comprises French, the sole official language, and the native languages.



Source: Ndinga-Koumba-Binza (2005a: 134)

With the exception of some languages of the pygmies, all Gabonese heritage languages are Bantu and they are to be found in zones A, B and H of Guthrie's referential classification (Guthrie 1953). Some of these languages are more dynamic than others. A sizeable number of Gabonese heritage languages falls within Fishman's (1991) stages 7 (i.e. they are used for cultural events and ceremonies), 6 (i.e. children are learning the language from their parents, neighbourhood, and communities) and 5 (i.e. there is local literacy in the community, and literacy programmes in the native languages), while others are unfortunately at stage 8 (i.e. they are close to extinction). So far no Gabonese heritage language falls within Fishman's (1991) stages 4 (i.e. the language is used in

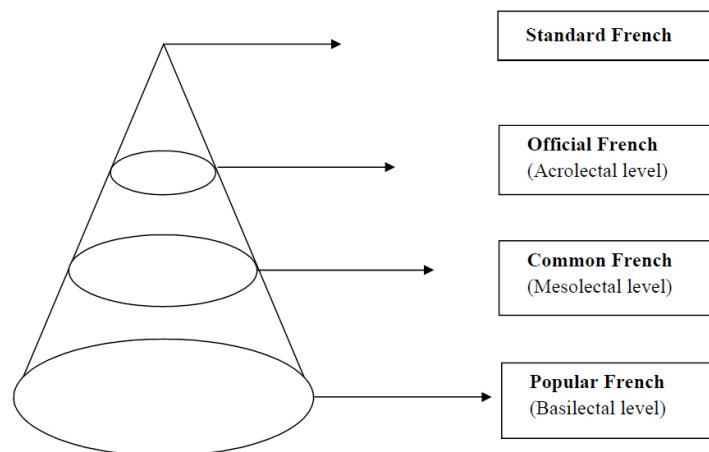
school), 3 (i.e. the language is used in the work sphere), 2 (the language is used in the local mass media, and local government) and 1 (i.e. the language exists at the highest levels in government, the universities, and national media). There is no standard language among the Gabonese heritage languages. "Standard language" is used here in the sense of a form or variety of any given language, "which is learned and accepted as correct across a community or set of communities in which others are also used" (Matthews 1997: 352). None has any constitutional status, nor is there any that is specifically mentioned in the Gabonese Constitution. The revised Constitution of 1994 (Art. 2, par. 8) simply committed the Republic to "endeavor to protect and promote national languages" without indicating which, and how, when, and in which order this should be done, nor to what extent (Emejulu 2000: 57-58 and 62). Finally, on the national level, there is no one dominant heritage language.

Varieties of French

One important dimension which determines language variety is the regional distribution of particular linguistic forms. In the case of French, a distinction can be made between:

- **National varieties** such as French spoken in France, Swiss French, Senegalese French, Gabonese French, etc., and
- **Regional dialects**, which are regional varieties within a national variety, such as Marseille French, Libreville French or Dakar French.

While there are many varieties of African French, common features include the use of an alveolar trill and the use of borrowed words from local languages. Four variants of French in Gabon can be distinguished, namely: standard French, official French (acrolectal level), common French (mesolectal level) and popular French (basilectal level).



Official French is the language of administration, schools, and all official circles. This variety is very close to Metropolitan French (spoken in Paris and considered standard). This French has been codified by centuries of dictionary-makers, grammar-book writers, printers and publishers, and educational systems.

Common French is very rich in Gabonisms. Gabonese people have named realities of their own environment with lexical items coined on their own soil or borrowed from other languages. They borrow words directly from local languages or they literally translate, often unknowingly, phrases or constructions from local languages.

Popular French, in which linguistic liberties are many, attests to the social life and the contemporary linguistic history of Gabon. This is a delightful language, full of different accents and mother-tongue interference.

Common French and popular French are not standardized or normalized and they are therefore excluded from the school and administrative spheres.

Gabonese French (cf. Boucher and Lafage 2000: XXIX-XXX) is a repertoire of a variety of lexical items and expressions. Idiosyncratic features of Gabonese French in vocabulary and syntax are presented below:

- Archaic, obsolete or old usage: *canguer* (to catch red-handed) rather than standard French *enchaîner* (to chain up);
- Abbreviation (a process of shortening a compound by using the initial letters of the words which are pronounced separately): *P.K.* (kilometric point) rather than standard French *point kilométrique* (kilometric point), *D.V.D.* ("Dos Ventre Dehors") and *V.C.D.* ("Ventre Cul Dehors") which are misappropriations of *Digital Versatile Disk* and *Video Compact Disk* pejoratively used to designate sexy young women and prostitutes, *clando* rather than standard French *taxi clandestine* or *taxi brousse* (bush taxi), *rapido* (sexual intercourse in a car with a prostitute);
- Modification of fixed expressions: *Quand les poules auront des dents et des cornes au menton* (and pigs might fly) rather than standard French *quand les poules auront des dents*;
- Usual graphic modification: *bonané* rather than standard French *bonne année* (happy New Year);
- Change of connotation: *charlatan* (ameliorative for "a traditional healer"), *cascadeur* (pejorative for "a young man whose carelessness can be life-threatening to others");
- Change of denotation: *bureau* (a woman who is kept by a married man), *caïman* (a crocodile), *sanglier* (a bush pig);
- Loan words from Gabonese languages: *musomfi* (a woman who has just given birth to a child), *nganga* (a traditional healer or sangoma), *safou* (the edible fruit produced by the *safoutier*);
- Anglicisms: *tchatcher* (to chat, to converse), *tchatcheur* (a good speaker);

- Neologisms: *buflon* (young buffalo), *dédeuillement* (the ending of the mourning period), *fétichisme* (fetishism), *tresseuse* (a female who plaits), *marabout* (a sangoma), *maraboutisme* (the craft of the sangoma), *marabouter* (to act as a sangoma), *se saper* (to dress elegantly), *sapeur* (an elegant person), *absenter* (to miss someone), *grever* (to be on strike), *siester* (to take a nap), *régler* (to have menstrual periods), *tôler* (to cover a house with corrugated iron), *têter* (to play a header with a soccer ball), etc.

Existing Gabonese French dictionaries

For many years, the general vocabulary of French as it is spoken in Gabon has been represented in the so-called bilingual externally motivated dictionaries that often had a Gabonese and a European language (French) as language pair. This situation began to change toward the end of the twentieth century. From 1972 to 1983, under the auspices of AUPELF (Association of Partially or Wholly French Language Universities, *Association des Universités Partiellement ou Entièrement de langue Française* in French), today called the AUF (Francophone University Agency, *Agence Universitaire de la Francophonie* in French), an extensive lexical research programme was conducted through a number of French-speaking countries of Black Africa. The project resulted in the publication in 1983 and in 1987 of a huge reference work, the *Inventaire des particularités lexicales du Français en Afrique Noire* (IFA). Unfortunately, Gabon did not participate in this research. To overcome this lack a number of students, French and Gabonese, embarked on this quest for their master's theses: Claudette Boutin-Dousset (1989), Caroline Thibaudier (1991), Maria Alves (1994), Marie Artigues, (1995), Karine Boucher (1997, 1998, 1999), Sandrine Ntsaga-Oyouuni (1998), Diane Bagouendi-Bagere (1999), Magali Italia (2000). In order to participate in the second phase of IFA, Karine Boucher wrote her M.A. dissertation (Boucher 1998) under the guidance of Suzanne Lafage. Different kinds of fieldwork were undertaken in Libreville, and collecting for this project was also done thanks to the participation of Gabonese students at the École Normale Supérieure (ENS) of Libreville. In 2000, the dissertation was published under the title: *Le lexique français du Gabon (entre tradition et modernité)*. For this version of the work released as a special issue of the journal *Le français en Afrique (Revue du Réseau des Observatoires du Français Contemporain en Afrique)*, issue 14, Suzanne Lafage joined in the writing. Boucher and Lafage's contribution is an alphabetically arranged lexicographic work of considerable size, consisting of 2500 articles covering 415 pages. Although the title of this publication modestly describes it as a "lexicon", it is in fact a comprehensive and thorough dictionary. The central list of Boucher and Lafage's work comprises a set of lexical items as used by the younger generation of Libreville inhabitants aged from 15 to 30. Fair attention is given to different style or normative levels, i.e. standard French, official French (acrolectal level), common French (mesolectal level) and popular French (basilectal level). Compare the following articles in this regard:

ABÈGNE, VAR. ABÈGN, ABÈ, n.m. Dispon. (du fang "maison commune des hommes"). Maison commune, placée au bout du village où les hommes ont coutume de se réunir. **V. CORPS* DE GARDE.** [...] *lorsque tout le monde est bien groupé dans l'abène (case* à palabre), on montre les douze images du catéchisme illustré des vérités nécessaires composé par Monseigneur Le Roy [...].* (Raponda-Walker, 1910 a: 6). *Autrefois, chez les Fang, l'abègne faisait surtout fonction de corps* de garde.* (Raponda-Walker/Sillans, 1983: 197). *La protection s'appuyait sur les maisons communes des hommes, tel que l' "abègn" des Fang — celles surtout qui étaient placées en bout de village — si bien que l'habitude s'est prise de les dénommer "corps* de garde".* (Pourtier, t.1, 1989: 159).

SYN.: abong, banza/banza*, case* à palabres, case* commune, case* de réunion, corps* de garde, ebanza*, mbandja*, temple du bwiti*.

BOUCHER and LAFAGE (2000: 2)

ADMINISTRATEUR, VAR. ADMINISTRATEUR DE BROUSSE, n.m. *Vx mais encore disponible à l'écrit.* **V. COMMANDANT*.** Fonctionnaire français de l'époque coloniale, placé à la tête d'un cercle* ou d'une subdivision. *L'administrateur, souvent seul dans son poste éloigné, représente tous les éléments de l'autorité. Il est pouvoir législatif: il essaie de faire concorder les lois françaises et le code de l'indigénat* avec les bonnes coutumes* du pays. [...] Il est pouvoir judiciaire. [...] Il est chef de l'état civil [...]. Enfin il est pouvoir exécutif et justicier [...].* (Grébert, 1928: 11). *Déjà je regardais en arrière le compte de mes réalisations, compte classique qui devait ressembler à celui de beaucoup d'administrateurs de brousse.* (Charnay, 1983: 157).

ENCYCL.: Les administrateurs étaient issus de l'Ecole coloniale, devenue plus tard Ecole Nationale de la France d'Outremer. **LOC.:** administrateur de brousse. **SYN.:** commandant*, commandant de cercle* (A.O.F. surtout), chef de subdivision* (A.E.F, Cameroun).

BOUCHER and LAFAGE (2000: 5)

After the publication of the *Lexique français du Gabon* in 2000, there was a long hiatus in French-language lexicography in Gabon. In 2008, Eric Dodo Bounguendza, a Gabonese linguist, published a Gabonese French dictionary entitled *Dictionnaire des Gabonismes*. This dictionary has a macrostructure consisting of 765 lemmata covering 142 pages which were identified specifically as Gabonese by the author.

Assamsè n.m. Boutique de fripier.

Il a acheté ce beau pantalon à assamsè.

Fréquent, oral, écrit, tous milieux.

Bana-bana n.m. Vendeur ambulancier de marchandises. *Devant les grands magasins de la capitale, il y a toujours des bana-bana en quête des clients.*

Fréquent, oral, écrit, tous milieux.

Eric DODO BOUNGUENDZA (2008: 23 and 29)

Just a year later, the dictionary of Dodo Bounguendza was followed by another lexicographic work. Although entitled *On est ensemble 852 mots pour comprendre le français du Gabon*, this publication is in fact a fairly small lexicon of Gabonese French consisting of 852 lemmata and covering 154 pages. It was published in 2009 by Lucien Ditougou, a Gabonese scholar lecturing in the Department of African Literature at Omar Bongo University.

142- Bouche: N. f. ① Avoir la ~: avoir la parole facile. *Il évite une confrontation avec Musunda parce qu'elle a la bouche.* ② Avoir la même ~: avancer les mêmes propos sans varier. ☞ Version. *Depuis qu'on l'interroge sur cette affaire, il a la même bouche.* ③ Avoir la même ~ que quelqu'un: être du même avis que lui. *J'ai la même bouche que Koumba.* ④ Être dans la ~ de quelqu'un: acquiescer ce qu'il dit ou faire sienne les paroles qu'il a prononcées. *Je suis dans la bouche de Ngondo. Oui, il a bien parlé.* ⑤ Faire la ~: narguer; s'enorgueillir. *A la tâche, Kumba ne fait plus la bouche.*

Lucien DITOUGOU (2009: 28)

A sample of 23 drawings by Lybek had been included as a bonus in the linguistic treatment of Gabonese French in the lexicon (see Fig. 1, 2 and 3).



Figure 1



Figure 2

The reason why Ditougou's contribution was well received by the public is that the pictorial illustrations included in the dictionary are masterpieces of Gabonese humour and self-portraiture. Lybek's drawings allowed Gabonese to recognize their own language specificity in a reference tool used not only to find the meaning or the orthography of a given lexical item but also for them to have a good laugh. As a matter of fact, many users have confessed that they bought the book in the first place for the humorous drawings. Of course, this is not part of the *genuine purpose* of a dictionary (cf. Wiegand 1998).



Figure 3

Dodo Bounguendza is also the author of another lexicographic work entitled *Diagnostic du Français du Gabon* published in 2010 as a normative product of the French language.

Cameraman D'abord il faut distinguer le «cameraman reporter» du «cameraman de plateau». Ce mot est un anglicisme et l'équivalent recommandé par la Commission de terminologie de l'O.R.T.F. est le «journaliste reporter d'images» qui est différent du «cameraman de plateau.» D'ailleurs, en français, on parle plus de cadreur. ↓*Pour couvrir cet évènement, je te suggère de prendre Olivier Antonio comme cameraman.* ↑*Pour couvrir cet évènement, je te suggère de prendre Olivier Antonio comme journaliste reporter d'images.*
Fréquent, citoyen ordinaire, journalistes, politiques, universitaires.

Eric DODO BOUNGUENDZA (2010: 28)

The work of Moussounda Ibouanga entitled *Français du Gabon: Approches socio-linguistique et lexicographique* was published in 2011. It is an alphabetically arranged monolingual lexicon consisting of 950 articles and covering 292 pages. The publication is based on several sources, its primary source being the lexicon of Karine Boucher and Suzanne Lafage (2000).

Combi: n. c. basilecte. ■ Personne avec qui on a une relation fraternelle. ▲ C'est une sorte de d'amitié réciproque. **Syn. Complice, pote, ami, frangin, dream team, type.** *Lui, c'est mon vrai combi.* (Herbert, 27 ans). Oral, fréquent, milieu des jeunes. (Ce terme est absent de Boucher, 2000). [kõmbi]
On est ensemble: locution du basilecte. (Gabonisme). ■ Se dit pour renforcer et retracer la solidarité et la fraternité des Gabonais. **Syn. On ne se perd pas de vue. Nous sommes ensemble.** *On est ensemble.* (Pony, 38 ans). Oral, courant, parler commun. (Cette phrase est absente de Boucher, 2000). [õneõsõblõ]

Firmin MOUSSOUNDA IBOUANGA (2011: 118 and 213)

To date the most thorough documentation of the Gabonese French vocabulary is in Boucher and Lafage's dictionary. Despite the fact that most Gabonese French lexicographic works are internally motivated products, they are still biased toward the central French of Paris. For example, when a dictionary of Gabonisms (defined as words or meanings not used in France) is produced in Gabon, the reality is that these Gabonisms are defined with regard to standard French and not with regard to French as it is spoken in Gabon. All existing Gabonese French dictionaries are implicitly regarding variant B as being inferior to variant A, that is standard French. As such, they do not express a high degree of emancipation. That is the reason why I plead for the production of a complete dictionary of variant B (the French language as it is spoken in Gabon).

Planning and compilation of a complete dictionary of variant B, namely Gabonese French

The idea is to produce a complete dictionary of variant B, namely Gabonese French. Such a dictionary will reflect a certain emancipation from the French language spoken in France (standard or central French, i.e. variant A). To date and as already said, all lexicographic referential work on the French language spoken in Gabon are differential lexica or dictionaries. In other words, Gabonese lexicographers and linguists working on Gabonese French do not need to define the lexicon of their language differentially. Gabonese French differs enough from standard French to be regarded as a new emerging language and not merely as a regional variety of French. Compared to French spoken in France, lexical particularities of African French-speaking countries should not be regarded as regionalisms but rather as national varieties. Moreover, a *local variety* of a given language is not synonymous with a *national variety* (Hausmann 1986: 4-5).

Gouws has rightly emphasized: "Successful joint ventures in the lexicographic practice need to be preceded by joint research in the field of theoretical lexicography, including a focus on the planning of dictionary projects." In this regard, I have discussed (cf. Mavoungou 2002) various metalexigraphic crite-

ria for the planning and compilation of the *Dictionnaire du Français du Gabon* (DFG). The DFG is meant to establish itself as a model for future general dictionaries of Gabonese French. It will be the first descriptive dictionary in a one volume edition to use the French spoken in Gabon as linguistic reference or norm, and will consider as "foreign" words and meanings used only in France. In what follows, I will briefly discuss various aspects regarding corpus building, national lexicographic units and their staff.

As far as the first issue is concerned, it should be mentioned that corpora used in the existing Gabonese French dictionaries are too small and crudely selected. Emejulu (2000: 62) has stressed:

Though dialectizing rapidly, Gabonese French is not standardized, nor normalized and is excluded from the school and administrative spheres. However, it strongly attests to the social life and the contemporary linguistic history of Gabon. A comprehensive, monolingual dictionary that should be periodically updated is very necessary to keep the trend of the linguistic and anthropological evolutions of our society; a lexicographic unit is therefore needed to embark on a large-scale nationwide corpus build-up and dictionary planning.

A very important demand made on a corpus today is that it should be balanced. In other words, before embarking on any lexicographic activity, the lexicographer has to make sure that he/she has a representative corpus of the language which is to be described. This means that it should contain texts from (a) different genres, (b) spoken and written language and (c) speakers of different regions, age groups, social backgrounds and genders (Herbst 2007: 30). So far only Gabonese French as it is spoken in Libreville has been systematically documented. A survey of French as it is spoken in Port-Gentil (the centre of Gabon's petroleum and timber industries, having recently become one of the richest cities in Gabon) can be found in a Ph.D. dissertation completed at the University of Stendhal-Grenoble III by Ompoussa (2008). Another notable Ph.D. dissertation, Moussounda Ibouanga (2006), written at the University of Aix-en-Provence gives an account of Gabonese French as it is spoken in Mouila (in the south-west of Gabon). Since common and popular French are not written, the bulk of data from towns and provinces not yet been surveyed will be obtained from interviews or will consist of answers to questionnaires completed by volunteers or students. Once the information is captured, transcripts of conversations, dialogues or interviews have to be computerized in the form of a database or in the form of a corpus. To these data, material taken from older investigations or from written sources such as newspapers produced in Gabon, novels, books, and especially existing Gabonese French monolingual dictionaries as well as bilingual dictionaries pairing French with indigenous languages will be added to the database or the corpus of Gabonese French. The future editorial board of the *Dictionnaire du Français du Gabon* could also rely on a long tradition of African French investigation at a number of universities and institutions in France (Aix-en-Provence University, the Paris III-Sorbonne Nou-

velle University, and so on). The idea is to document the basic vocabulary of the language (that is the core lexicon of French which is more or less identical for Gabonese and French people), the available common basic lexicon (which differs a lot between the Gabonese and the French), the scientific oriented general lexica and the highly specialized lexica (which are very much the same for specialists of the same discipline whether Gabonese or French). The pronunciation of Gabonese French lexical items will also be obtained through fieldwork. This is extremely important. Gabon has its own distinct idiosyncratic features of pronunciation, vocabulary and syntax, resulting from contact with indigenous languages. Pronunciation was omitted from Boucher and Lafage's work owing to insufficient data. This has to be corrected in the prospective *Dictionnaire du Français du Gabon*.

Across dictionaries, lexicographers usually experience similar problems which can be dealt with by means of similar solutions. On puristic grounds, some people are against slang words and expressions. Conscious of this fact, the future editorial board of the DFG will be well advised to carefully consider the inclusion of Gabonisms. For example, Gabrielle Saint-Yves (2001), reflecting on the reception of the *Dictionnaire québécois d'aujourd'hui* (DQA), wrote:

Many Quebec linguists, the *Office de la langue française* (OLF), and language teachers emphatically disagreed both with the selection of entries and its socio-linguistic labels. In reaction to media debates and social pressure, the Quebec Ministry of Education banned the pedagogical use of the DQA in the classroom.

If the future editorial board of the DFG feels that the public is not prepared to see slang words and expressions included in a General Monolingual Dictionary of Gabonese French, it must not be insisted that they should be included before a general consensus on a Gabonese norm could be reached. As far as this point is concerned, acceptance must be considered as most important.

With regard to the training of staff members of a dictionary project, the lack of trained lexicographers has often been a serious problem to overcome in a number of countries. In Central Africa, the Gabonese Republic is in the fortunate position that she has readily at hand quite a significant number of highly trained lexicographers. Ten Gabonese have received their doctoral degrees in lexicography at the University of Stellenbosch, South Africa. Despite the importance of this achievement, Gabon has not yet given full priority to the comprehensive planning of dictionary projects in the country. Gabon could utilize available expertise in order to produce more dictionaries for better communication. In this regard, National Lexicographic Units (NLUs) have to be established for all the ten language units of the country. In countries with established national lexicographic units, the former are generally hosted at tertiary institutions in the geolinguistic areas where the majority of first-language speakers live. Given that Gabon is a country with sixty-two indigenous languages but with French as the sole official language, a lexicographic unit is thus advisable to cater for the needs of the Gabonese regarding Gabonese French.

Prospects and challenges for French lexicography in Gabon

Although the Gabonese language landscape displays several native languages and a number of other foreign languages (Ndinga-Koumba-Binza 2005a: 134), none has ever been made official or been used for official purposes. The fact that French is increasingly becoming the mother tongue of younger generations gives it a promising future in Gabon. In the light of the previous section, Gabonese French lexicography may be seen as a long-term but worthwhile prospect. There is a need for Gabonese French dictionaries aimed at a very specific target user. For meeting the real needs as well as the reference skills of the intended target user, the focus should be on the compilation of pedagogic or teaching materials, such as school dictionaries, grammar books and syllabi in order to reflect the linguistic usage of the French language spoken in Gabon. Coming back to the organization of the National Lexicographic Units (NLUs) of the country, these units should be managed by a board of directors under the control of a language body of which the mandate will be to legislate over important linguistic matters including the issue regarding the extraordinary stock of gentiles, toponymic adjectives, derivatives of place names, proper nouns, and other neologisms in Gabonese French which are being created by journalists, writers, laypersons and singers to express the vivid national identity of the Gabonese (Mavoungou 2011).

Kwenzi-Mikala's (1998) internal classification offers an ideal theoretical platform for establishing National Lexicographic Units. As far as this point is concerned, the Gabonese French NLU should play a vital role. Employing the hub-and-spoke model (cf. Martin 1996), the NLU for Gabonese French will be in charge of the compilation of all the monolingual dictionaries dealing with Gabonese French. The NLUs of the other Gabonese languages will be faced with at least two challenges, namely the choice of the dialect as pre-selected language and the compilation of bridging dictionaries between the particular pre-selected Gabonese language (spoke) and Gabonese French, the official language (hub). In the light of Gallardo (1980: 61), the future editorial board of the specific unit should only consider the tremendous task of compiling a comprehensive monolingual dictionary of the particular pre-selected language later on. And the first step to do so is the compilation of a database. It is also well-attested that the preparation of a database may be costly and time-consuming. Owing to the fact that most existing dictionaries of Gabonese languages have been the input of single lexicographers, future lexicographic products should be the result of team efforts.

Mavoungou (2002) and Mavoungou et al. (2002) discussed various meta-lexicographic criteria for the planning and compilation of the *Dictionnaire du Français du Gabon* (DFG). As a result of these studies, a major project involving all Gabonese lexicographers is currently under way. Gabonese lexicographers, sociolinguists and educationists have also joined forces in order to produce the first idiomatic dictionary of Gabonese French (Mavoungou, Moussounda Ibouanga and Pambou, forthcoming).

Conclusion

The survey of Gabonese French dictionaries and lexica above shows that most dictionaries available for Gabonese French are differential products. Another feature is that corpora used in these dictionaries are too small and crudely selected. The only way to address this situation is twofold: (a) to collect spoken language at grassroots level through fieldwork, and (b) to obtain data from written sources such as newspapers produced in Gabon, novels, books, and especially existing Gabonese French monolingual dictionaries as well as bilingual dictionaries pairing French with indigenous languages in order to enlarge the database or the corpus of Gabonese French. The compilation of the DFG is answering the need for linguistic recognition and cultural autonomy. It is an attempt to go beyond all existing lexicographic products published in Gabonese French. It will represent all registers of French spoken in Gabon and will take account of all lexical items (including Anglicisms and socially banned, taboo or slang expressions) and their sociolinguistic labels. This dictionary project will only fulfil its purpose when completed. The different prospects raised above will be confronted with serious challenges ranging from financial, political to social support within and outside Gabon.

References

- Alves, M.** 1994. *La presse gabonaise en langue française: étude lexicale*. Mémoire sous la direction de S. Lafage. Paris: Université de Paris III-Sorbonne Nouvelle.
- Artigues, M.** 1995. *Participation à une étude des particularités lexicales du français parlé au Gabon*. Mémoire sous la direction de S. Lafage. Paris: Université de Paris III-Sorbonne Nouvelle.
- Bagouendi-Bagere, D.** 1999. *Etude des procédés de néologie lexicale dans le français parlé au Gabon*. Mémoire sous la direction de A. Quéffelec. Université de Provence-Aix-Marseille I.
- Boucher, K.** 1997. *Créativité lexicale et identité culturelle du français au Gabon*. Mémoire de maîtrise sous la direction de S. Lafage. Paris: Université de Paris III-Sorbonne Nouvelle.
- Boucher, K.** 1998. *Langues et identité culturelle des jeunes Librevillois de 15 à 30 ans: une enquête de terrain*. Mémoire de DEA sous la direction de S. Lafage. Paris: Université de la Sorbonne Nouvelle-Paris III.
- Boucher, K.** 1999. Approche des représentations sociolinguistiques dans un groupe de jeunes Librevillois. *Le français en Afrique* 13: 173-192.
- Boucher, K.** 2000. "Quand tu gorges, tu blazes!" *Contributions à la zè journée de l'école doctorale du 29 mai 1999, Langage et Langue, Paris III-Sorbonne Nouvelle*: 55-65.
- Boucher, K. and S. Lafage.** 2000. *Le lexique français du Gabon (entre tradition et modernité)*. Nice: Les presses U.F.R. Lettres, Art et Sciences Humaines de l'Université de Nice-Sophia Antipolis.
- Boulanger, J.-C. (Ed.)**. 1992. *Dictionnaire québécois d'aujourd'hui*. Saint-Laurent: DicoRobert.
- Boutin-Dousset, C.** 1989. *Matériaux pour un inventaire des particularités lexicales du français au Gabon*. Mémoire sous la direction de S. Lafage. Paris: Université de la Sorbonne Nouvelle.
- Charnay, R.** 1983. *La terre des adieux*. Paris: Olivier Orban.

- Ditougou, L.** 2009. *On est ensemble: 852 mots pour comprendre le français du Gabon*. Libreville: Éditions Raponda Walker.
- Dodo Bounguendza, E.** 2008. *Dictionnaire des gabonismes*. Paris: L'Harmattan.
- Dodo Bounguendza, E.** 2010. *Diagnostic du français au Gabon (Guide pratique destiné aux journalistes, aux politiques, aux administratifs et aux universitaires)*. Libreville: Éditions Ntsame.
- Ella, E.M.** 2011. Typologie hybride et modèles poly-fonctionnels dans la confection des dictionnaires au Gabon. *Mbaandza Revue d'étude et d'analyse francophones* 1: 113-126.
- Ella, E.M.** 2012. Adopter un dictionnaire bilingue pour un usage local. *Alac Revue Scientifique Internationale publiée par la Chaire UNESCO Interculturalité* 1: 265-287.
- Emejulu, J.D.** 2000. Lexicography, an Economic Asset in Multilingual Gabon. *Revue Gabonaise des Sciences du Langage / Gabonese Journal of Language Sciences* 1: 51-69.
- Encarta.** 1993–2000. Gabon. *Encarta Encyclopedia 2001*. Microsoft Corporation.
- Fishman, J.A.** 1991. *Reversing Language Shift: Theoretical and Empirical Foundations of Assistance to Threatened Languages*. Cleveland: Multilingual Matters.
- Gallardo, A.** 1980. Dictionaries and the Standardization Process. Zgusta, L. (Ed.). 1980. *Theory and Method in Lexicography*: 59-69. Columbia: Hornbeam Press.
- Gouws, R.H.** 2007. Dictionaries in South Africa: Aspects of Dictionary Culture and Government Policy. *International Journal of Lexicography* 20(3): 313-327.
- Grébert, F.** 1928. *Au Gabon (Afrique équatoriale française)*. Paris: Société des Missions Evangéliques.
- Guthrie, M.** 1953. *The Bantu Languages of Western Equatorial Africa*. Oxford: Oxford University Press.
- Hausmann, F.J.** 1986. Les dictionnaires du français hors de France. Boisvert, L., C. Poirier et C. Verreault (Eds.). 1986. *La lexicographie québécoise: bilan et perspectives*: 3-19. Laval: Les Presses de l'Université de Laval.
- Herbst, T.** 2007. *Linguistics: Concepts and Ideas for the Description of English*. Erlangen: Alleinvertrieb Mencke-Blaesing.
- Idiata, D.F. and L. de Nadaillac.** 2010. *Dictionnaire Isangu-Français*. Libreville: Les Éditions du CENAREST.
- Italia, M.** 2000. *Morphosyntaxe verbale dans des corpus de locuteurs basilectaux au Gabon*. Mémoire de maîtrise sous la direction de A. Quéffelec. Université de Provence-Aix-Marseille I.
- Kwenzi-Mikala, J.T.** 1990. Quel avenir pour les langues gabonaises? *Revue Gabonaise des Sciences de l'Homme* 2: 121-124.
- Martin, W.** 1996. Lexicographic Resources in a Multilingual Environment: An Orientation. *Lexicography as a Financial Asset in a Multilingual South Africa*: 9-24. Pretoria: Department of Arts, Culture, Science and Technology.
- Matthews, P.H.** 1997. *The Concise Oxford Dictionary of Linguistics*. Oxford: Oxford University Press.
- Mavoungou, P.A.** 2002. Vers un dictionnaire du français du Gabon. Emejulu, J.D. (Ed.). 2002: *Éléments de lexicographie gabonaise. Tome II*: 230-262. New York: Jimacs-Hillman Publishers.
- Mavoungou, P.A.** 2011. Regard sur les onomastismes dans le français de Libreville et leur traitement lexicographique. *Mbaandza Revue d'étude et d'analyse francophones* 1: 21-50.
- Mavoungou, P.A. et al.** 2002. *Vers un Dictionnaire du Français du Gabon (Premier dictionnaire parallèle contenant une centaine d'articles traités)*. Unpublished Document. Stellenbosch: University of Stellenbosch.
- Mavoungou, P.A. and B. Plumel.** 2010. *Dictionnaire Yilumbu-Français*. Libreville: Les Éditions Raponda Walker.

- Mavoungou, P.A., F. Mousounda Ibouanga and J.A. Pambou.** Forthcoming. *Dictionnaire des Collocations et Locutions figurées du Français du Gabon*.
- Mickala Manfoumbi, R.** 2004. *Lexique pove-français/français-pove*. Libreville: Éditions Raponda-Walker.
- Mousounda Ibouanga, F.** 2006. *Les Molvilois et leurs langues: dynamiques linguistiques à Mouila de 1900 à nos jours*. Thèse de Doctorat Nouveau Régime. Aix-en-Provence: Université Aix-Marseille I-Université de Provence.
- Mousounda Ibouanga, F.** 2011. *Français du Gabon: Approches sociolinguistique et lexicographique (le toli bangando)*. Paris: Éditions Universitaires Européennes (EUE).
- Ndinga-Koumba-Binza, Hugues Steve.** 2005a. Considering a Lexicographic Plan for Gabon within the Gabonese Language Landscape. *Lexikos* 15: 132-150.
- Ndinga-Koumba-Binza, Hugues Steve.** 2005b. Gabonese Language Landscape: Survey and Perspectives. *South African Journal of African Languages* 25(3): 97-116.
- Ndinga-Koumba-Binza, H.S.** 2007. Gabonese Language Landscape: Survey and Perspectives. *South African Journal of African Languages* 27(3): 97-116.
- Ntsaga-Oyouni, S.** 1998. *Approche de quelques traits du français parlé de jeunes Gabonais habitant la région parisienne*. Mémoire sous la direction de S. Lafage. Paris: Université de Paris III-Sorbonne Nouvelle.
- Ompoussa, V.** 2008. *Les particularités lexicales dans le français scolaire au Gabon: le cas de la ville de Port-Gentil*. Thèse de Doctorat Nouveau Régime. Grenoble: Université Stendhal-Grenoble III.
- Pourtier, R.** 1989. *Le Gabon. Tome I: Espace, histoire, société. Tome II: État et développement*. Paris: L'Harmattan.
- Raponda-Walker, A.** 1910. *Au pays des Ishogo, simple récit de voyage*. Publié Messenger du St-Esprit, repris Libreville: Fondation Raponda-Walker, Classiques Africains, 1993.
- Raponda-Walker, André and Roger Sillans.** 1996. *Plantes utiles du Gabon: essai d'inventaire et de concordance des noms vernaculaires et scientifiques des plantes spontanées et introduites du Gabon*. Paris: Lechevalier.
- Saint-Yves, Gabrielle.** 2001. *Lexicographic Revolution in Quebec*. [online]. Toronto: Lexperimenta. Available: <http://www.chass.utoronto.ca/~wulfric/articles/saint-yves2/>.
- Thibaudier, C.** 1991. *Contribution à une étude des néologismes dans la presse gabonaise*. Mémoire sous la direction de S. Lafage. Paris: Université de la Sorbonne Nouvelle.
- UNESCO-LUTO.** 2006. *Lexique du Kóya (langue des Pygmées du Nord-Est du Gabon)*. Libreville: Éditions UNESCO.
- Van der Veen, L.J. and S. Bodinga-bwa-Bodinga.** 2002. *Gedandedi sa geviya/Dictionnaire geviya-français*. Louvain/Paris/Sterling, Virginia: Peeters.
- Wiegand, Herbert Ernst.** 1998. *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie*. Volume 1. With 159 Illustrations in Text. Berlin/New York: Walter de Gruyter.

Addendum (Mavoungou et al., forthcoming)

bord de mer [bɔrdmɛr] *n.m.* ♦ Milieu géographique abritant nombre de quartiers huppés et sièges administratifs de la capitale, y compris la Présidence de la République ♦ *Aworet vit au bord de mer.* **palais du bord de mer** (on dit également *le bord de mer* = siège de la Présidence de la République). *Tous les leaders des partis politiques ont été reçus au palais du bord de mer afin de discuter des nouvelles modalités des prochaines échéances électorales. Les ambassadeurs accrédités au Gabon ont présenté leurs lettres de créances au palais du bord de mer.*

clando [klɔdo] *n.m.* PL. *clandos.* ♦ **1.** Moyen de transport utilisé illicitement pour l'acheminement des personnes et des biens, généralement, sur une distance assez longue ♦ *Makaya a emprunté un clando pour se rendre à Kango. Pour boucler ses fins de mois, Abessolo a décidé de faire du clando. Suite au mouvement de grève déclenché par les taximen de Libreville et de Port-Gentil, on pratique la débrouille: on prend le clando. Moussounda qui tient un commerce de friperie à Ntoun prend le clando régulièrement et paie 5000 Fcfa pour chaque ballot.* ♦ **2.** Personne qui vit de la profession de clando (*Voir clandoman*). ♦ *C'est un clando, il n'a rien à t'offrir.* ♦ **3. adv.** De façon clandestine. ♦ *Il est parti clando* (= filer à l'anglaise).

engong [ɛŋgɔŋ] *n.m.* ♦ (*dans le jargon footballistique des jeunes*) Jeu de vilain, mauvais jeu ♦ *Dépité par le but encaissé, les jeunes de l'équipe adverse ont décidé de jouer engong! Toutes les femmes d'Engong émancipées, il en fit toutes et sans exception ses femmes, ses femmes sans pour autant être mariées légalement avec l'une d'elles* (NDONG MBENG, 1992: 73-74). *Engong! Engong! Engong!* (tube populaire d'un groupe de la place: Raboon). ■ Engong est un terme populaire du français du Gabon tiré de l'épopée Mvet. Il désigne plus précisément le peuple et le pays des Immortels, redoutables et puissants guerriers. Dans le Nord-Gabon à l'origine et aujourd'hui dans la quasi-totalité du pays, ce terme est utilisé par les jeunes en particulier pour faire référence à un duel physique par analogie à l'esprit belliqueux qui caractérise nombre de guerriers d'Engong en tête desquels se trouve le célèbre Akoma Mba. *Voir mvet.*

Reference

- Ndong Mbeng, H.F. 1992. *Les matitis, mes pauvres.univers en contre-plaqué, en planche et en tôle.* St Maur: Éditions Sépia.

*Birds of a Feather Don't Always Flock Together: User Problems in Identifying Headwords in Online English Learner's Dictionaries**

Julia Miller, *School of Education, University of Adelaide, Australia* (julia.miller@adelaide.edu.au)

Abstract: Idioms, sayings and proverbs (referred to here as 'phrasemes'), are a central part of the English language. However, it is often difficult for learners of English as an Additional Language (EAL) to choose the correct headword when looking for such expressions in a dictionary. Learners may not recognise a word as belonging to a phraseme, and so may not look under a single, 'important' word. Moreover, their choice of a salient word may not accord with the lexicographer's. Thirdly, they may not recognise phraseme variants, such as *carry/take coals to Newcastle*. They may therefore often fail to find the phraseme altogether.

A study of 84 phrasemes in five online English learner's dictionaries (*Cambridge Advanced Learner's Dictionary*, *Collins COBUILD Advanced Dictionary*, *Longman Dictionary of Contemporary English*, *Macmillan English Dictionary for Advanced Learners* and *Oxford Advanced Learner's Dictionary*) revealed a lack of uniformity across and often within dictionaries. This paper is based on 14 of these phrasemes, which appear in one or more of these dictionaries and include proper nouns and/or variable words.

To make learner's dictionaries more user friendly (Zgusta 1971), it is argued that they need greater consistency in their choice of phraseme headwords, both within and between dictionaries, and that greater cross-referencing is necessary within a single dictionary. Five strategies are presented to help learners with their dictionary searches.

Keywords: ENGLISH AS AN ADDITIONAL LANGUAGE, DICTIONARY, ONLINE, HEADWORD, PHRASEME

Opsomming: *Birds of a Feather Don't Always Flock Together: Gebruikersprobleme by die identifikasie van trefwoorde in aanlyn Engelse aanleerderswoordeboeke.* Idiome, segswyses en spreekwoorde (hier na verwys as 'fraseme') vorm 'n sentrale deel van die Engelse taal. Dit is egter dikwels moeilik vir aanleerders van Engels as 'n Addisionele Taal (EAT) om die regte trefwoord te kies wanneer daar na uitdrukkings in 'n woordeboek gesoek word. Aanleerders kan dalk nie 'n woord herken as behorende tot 'n fraseem nie, en kan derhalwe nie onder een enkele 'belangrike' woord kyk nie. Meer nog, hulle keuse van 'n opvallende

* This article was presented as a paper at the Conference of the Australasian Association for Lexicography (AUSTRALEX), which was held at the Australian National University, Canberra, Australia, 28–29 November 2011.

woord kan nie ooreenstem met dié van die leksikograaf nie. Derdens, hulle kan dalk nie fraseem-variante, soos *carry/take coals to Newcastle*, herken nie. Hulle kan derhalwe dikwels glad nie daarin slaag om die fraseem te vind nie.

'n Studie van 84 fraseem in vyf aanlyn Engelse aanleerderswoordeboeke (*Cambridge Advanced Learner's Dictionary*, *Collins COBUILD Advanced Dictionary*, *Longman Dictionary of Contemporary English*, *Macmillan English Dictionary for Advanced Learners* and *Oxford Advanced Learner's Dictionary*) het 'n gebrek aan eenvormigheid tussen en dikwels binne woordeboeke getoon. Hierdie artikel is gebaseer op 14 van hierdie fraseem wat in een of meer van hierdie woordeboeke verskyn en sluit eiename en/of variante woorde in.

Om aanleerderswoordeboeke meer gebruikersvriendelik (Zgusta 1971) te maak, word daar bepleit dat hulle groter konsekwensie in hul keuse van fraseentrefwoorde nodig het, sowel binne as tussen woordeboeke, en dat groter kruisverwysing nodig is binne 'n enkele woordeboek. Vyf strategieë word aangebied om aanleerders te help met hul woordeboeksoektogte.

Sleutelwoorde: ENGELS AS ADDISIONELE TAAL, WOORDEBOEK, AANLYN, TREFWOORD, FRASEEM

Introduction

One of the main precepts for any dictionary is that it should be user friendly (Zgusta 1971), implying that someone can search a dictionary easily and with a successful outcome. In the case of multi-word expressions such as idioms, sayings and proverbs (henceforth referred to as 'phrasemes'), providing enough information to lead to a successful search is particularly problematic, since a variety of headwords could be chosen by both the lexicographer and the user, and the choices made by these two parties do not always coincide. Monolingual English learner's dictionaries (MELDs) include phrasemes, usually as sub-entries, in their paper dictionaries, and list them under headwords in their online dictionaries. There are dictionaries purely for idioms or proverbs, but learners of English as an additional language (EAL) may not know that a word is part of a phraseme and are therefore more likely to search for an individual headword in a MELD. All the major MELDs are now available online, and indeed Macmillan issued a press release in November 2012 to say that from 2013 their dictionary would only be available on the Internet, and they would no longer produce a print version (Macmillan Publishers Limited 2012). In theory, it should be easier to find a phraseme in an online dictionary.

The study which follows is based on the online versions of five British English MELDs, known as the 'Big 5', consulted online in November 2012: *Cambridge Advanced Learner's Dictionary (CALD)*; *Collins COBUILD Advanced Dictionary (COBUILD)*; *Longman Dictionary of Contemporary English (LDOCE)*; *Macmillan English Dictionary for Advanced Learners (MEDAL)*; and *Oxford Advanced Learner's Dictionary (OALD)*. There is a sixth online MELD, the *Merriam-Webster Learner's Dictionary*, but it was not included in the study, which addresses phrasemes used in Australia and which come predominantly from the UK,

using British English spelling (e.g. *ploughshare* rather than *plowshare*).

Correct understanding and use of phrasemes is especially important for learners of English as an additional language, because the meaning of phrasemes is often figurative and therefore hard to guess, and a misunderstanding can lead to confusion or even embarrassment. There are three particular problems which EAL learners face when searching for a phraseme in a dictionary. Firstly, they may not recognise a word as belonging to a phraseme, and so may fail to identify the headword. For example, *COBUILD* lists the phraseme *to beat/turn swords into ploughshares* at *ploughshare* but not at *sword*. Secondly, their choice of an 'important' or salient word may not be the same as the lexicographer's. Phrasemes containing proper nouns, such as *the man/woman on the Clapham omnibus*, offer particularly unusual, and therefore salient, nouns, but these may not be used as headwords in MELDs. For example, most learners could not be expected to be familiar with the words *Clapham* or *omnibus*, yet a search under either of these words in *LDOCE* would fail to find the phraseme, which lists it only under the first noun, *man* (but not under *woman*). Thirdly, users may not be aware of phraseme variants, such as *not to budge/give /move an inch*. This means that if a phraseme is listed under only one of the variants, a user may fail to find it. Only *CALD* and *OALD*, for instance, give the alternatives *budge/give/move* for this phraseme, with *LDOCE* and *MEDAL* only providing cross references at *budge* and *give* but not at *move*.

The Big 5 are becoming ever more sophisticated in their online search engines, and changes are being made daily. Such online dictionaries should theoretically offer the user a greater choice of search words. Consistency is necessary both within and between dictionaries, however, in order for users to conduct consistently successful searches. Learners too need to be given direction to help them use dictionaries efficiently. The study which follows compares the use of headwords for 14 phrasemes containing proper nouns or variable words in online versions of the Big 5 to see if these dictionaries are truly user friendly in regard to their search facilities for phrasemes, and suggests strategies for learners to improve their search experiences.

Background

The area of phraseology proliferates with terms used by different researchers to refer to the same thing; conversely, the same term may have different meanings for different researchers. The word 'idiom' is a prime example of this, as the concept embodied by the word is a notoriously difficult one to define. For this reason, the overall term 'phraseme', common in French and German phraseology (Dobrovolskij and Piirainen 2005: 30), has been adopted in this paper to refer to figurative expressions of more than one word, including idioms, proverbs and sayings. The MELDs themselves usually refer to these different expressions more loosely as 'idioms', or sometimes 'phrases'.

An understanding of English phrasemes allows EAL learners not only to communicate more freely with native speakers but also to orient themselves

and adapt to a culture. As Nation says (2001: 397), knowing a word means "being aware of restrictions on the word for cultural, geographical, stylistic or register reasons". As Fillmore, Kay and O'Connor (1988: 504) warn, however, "an idiomatic expression or construction is something a language user could fail to know while knowing everything else in the language". For this reason, it is important for EAL learners to understand phrasemes. It is also important for them to be aware of any cultural restrictions on their use, such as whether the term is humorous or old fashioned (Miller 2010), so that they can communicate effectively with different age groups in a range of situations.

It has been proposed that the teaching of phraseology "should occupy a central and uncontroversial position in instructed second language acquisition" (Granger and Meunier 2008: 247). Phrasemes are important for language learners because idioms and other fixed expressions come in chunks and thus provide syntactic patterns for encoding (producing language) and decoding (understanding language) (Nattinger and De Carrico 1992: 114), providing learners with "ready-made units appropriate for a particular situation" (Nattinger 1980: 341). Although phrasemes may be hard to learn (Kövecses and Szabó 1996: 326), they are nonetheless important. Decoding the meaning of a phraseme involves the use of various resources, and a dictionary can be paramount in this. However, finding a phraseme in a dictionary can be problematic, since the user has to know which headword to search under.

The use of headwords for listing certain types of phrasemes in a dictionary was first adopted in 1670 by John Ray, whose "material words" for proverbs were "'Printed in different Character, that so with the least cast of an eye any man may find any Proverb'" (Doyle 2007: 186). Alphabetical listing for proverb dictionaries was used first in the seventeenth century, based on the first word of the proverb (Doyle 2007: 183-185). This first word, of course, was variable, since many proverbs start with either *a*, *the* or a pronoun. Verbs too may vary, as in the example *take or carry coals to Newcastle*. Choosing the canonical form and showing variations is thus a major problem for lexicographers (Moon 1999: 273).

The use of headwords became more standardised in the twentieth century in separate works by Whiting and Tilley (Doyle 2007: 191), who both chose the first noun of a proverb as the alphabetical listing feature. If no noun was present, then the first finite verb was chosen, and if there was no verb then they used the "first important word" (Doyle 2007: 191). This method for listing proverbs and other phrasemes became the dominant method in the United States and was also adopted in general dictionaries, which usually include phrasemes as sub-entries under a headword chosen from the phraseme (Bergenholtz and Gouws 2007: 237-238), rather than as entries in their own right. In a study of six English dictionaries, however, Gates (1986: 102-103) found that five of them placed phrasemes at the first "major invariable word". For instance, *go/run to seed* was placed at *seed*. In general, though, the dictionaries in Gates' study had no consistency between them in their choice of headword, with some favouring nouns and others favouring verbs. This lack of consistency accords with Atkins and Rundell's (2008: 168) suggestion that phrasemes are "the most difficult

MWEs [multiword expressions] to handle in lexicography. In the absence of hard and fast criteria, it is well nigh impossible to be wholly consistent". The choice of a key word, or headword, is therefore a vexed matter, and is often "a largely subjective and complex decision" (Mulhall 2010: 1355) on the part of the lexicographer.

Online dictionaries should theoretically offer the user a greater choice of search words for phrasemes than paper dictionaries, but in practice this is not always the case. Although online dictionaries have the capacity to transcend the restrictions of the printed page, and can include hyperlinks, wild card searches and audio files, many online dictionaries are merely Internet versions of paper dictionaries (Nielsen and Mourier 2005: 110). They thus present the same problems as printed dictionaries in terms of phraseme searches, and may not always be user-friendly. Traditionally, entire phrasemes are not normally used as headwords in a dictionary, except in the case of compounds and phrasal verbs (Atkins and Rundell 2008: 181). Although online dictionary search engines are now making the entry of phrasemes in a search box more feasible, only three of the five MELDs in this study allow the user to find a phraseme by entering the entire phrase in the search box.

The strategies for choosing headwords mentioned by Doyle (2007) and Gates (1986) create several problems for learners of English as an additional language. If MELDs list phrasemes predominantly under nouns or verbs, this presupposes that the learner can recognise which part of speech a word belongs to. However, since learners often have trouble determining parts of speech (Nesi and Hail 2002: 282), the search process may be made more difficult for them. For example, in the saying *birds of a feather flock together*, the verbal form *flock* is less well known than its nominal equivalent, so that learners searching under the noun *flock* will fail to find the associated phraseme. Moreover, although surveys of dictionary use have shown that learners usually look first under nouns when searching for a phraseme (Moon 1998: 201), this strategy may not be an innately subconscious process, but may be the result of previous search experiences in other dictionaries. Those dictionaries that operate on a different principle will therefore present problems for learners.

Headwords used in the Big 5

Policies on headword use vary among the Big 5. In comparing their use of headwords for 84 phrasemes from five different categories (Biblical, literary or historical, UK in origin or reference, Australian in origin or reference, and older in origin or reference), there is a lack of consistency across, and sometimes within, dictionaries. *CALD*, *LDOCE* and *MEDAL* mostly use nouns and verbs as headwords, with a phraseme often cross referenced at both the noun and the verb. For example, the saying *birds of a feather flock together* is listed under *birds*, *feather*, *flock* and *together* in *CALD* and *MEDAL*, under *bird* in *LDOCE* and *OALD* (with a cross reference at *feather*), and not at all in *COBUILD*. *COBUILD*

and *OALD* prefer nouns, although *OALD* sometimes uses verbs or adjectives instead. In the paper versions of these dictionaries, *CALD* (2008: x) and *LDOCE* (2003: xiv) instruct their readers to look for idioms at the end of an entry under the "first important word"; *MEDAL* (2002: xi) directs readers to the "first main word"; and *OALD* (2010: R14) directs readers to the "first 'full' word". Instructions on idiom searches are not readily visible in *COBUILD* (2009). No advice is given in the online versions of these dictionaries, and this choice of "main", "full" or "important" word is of course a hard one for the user to make.

The following fourteen phrasemes appear in one or more of the Big 5 and include proper nouns and/or variable words. They provide a snapshot of the problems raised by many other phrasemes in these dictionaries. These phrasemes and their headwords in each MELD are listed in Table 1. The X symbol indicates that a phraseme does not appear in a particular dictionary and a word in brackets indicates that the phraseme is cross referenced from that word.

Table 1: Fourteen phrasemes and their headwords in the Big 5 (cross references in brackets)

Phraseme	CALD	COBUILD	LDOCE	MEDAL	OALD
To beat/turn swords into ploughshares	beat, turn, swords, ploughshares	ploughshare	sword, ploughshare	X	sword, (ploughshare)
A pretty/fine kettle of fish	pretty, fine, kettle, fish	X	X	X	X
To hide/cover a multitude of sins	hide, cover, multitude, sins	multitude (sin, sins)	multitude (sin)	hide, multitude	multitude (hide, cover, sin)
Give somebody an inch and they'll take a yard/mile	give, inch, take, mile	X	inch	inch (mile, yard)	inch (mile, yard)
Not to budge/give/move an inch	budge, give, move, inch	X	inch (and in an example at 'budge')	(budge) inch	inch (and in an example at 'budge')
To carry/take coals to Newcastle	carry, take, coals, Newcastle	X	coal, coals	coal, coals, Newcastle	coal, coals (Newcastle)
The man/woman on the Clapham omnibus	man, woman, Clapham, omnibus	X	man	X	Clapham, omnibus
Back of Bourke	X	X	X	Bourke	Bourke
To send someone to Coventry	send, Coventry	(send) Coventry	Coventry	send, Coventry	(send) Coventry
The road to Damascus	road, Damascus	X	Damascus	Damascus	Damascus

As old as Methuselah	as, old, Methuselah	X	Methuselah	(Methuselah — descrip- tion, no phraseme)	(Methuselah — example, no phraseme)
Not to know someone from Adam	not, know, Adam	X	Adam	Adam	know, (Adam)
In the land of Nod	land, nod	X	land, (nod)	(land, nod)	land, (nod)
To rob Peter to pay Paul	rob, Peter, pay, Paul	Peter	rob, (Peter)	rob, Paul	rob, (Peter)

Online searching has made it easier to find phrasemes in all the Big 5, but the process can still lead to frustration. The search process for each dictionary is outlined below, exemplified by the phraseme *to rob Peter to pay Paul*.

- *CALD*: After entering *rob* in the search box on the home page, the user is taken to the *rob* main definition page. A box on the right offers "More results", with choices for "All" or "Idioms". Since the user may not know that a word is part of a phraseme, the "All" box is used in this example, and in fact this phraseme appears at the top of the box. (For most other phrasemes, it may be necessary to click "See all results".) A search at *Peter*, *pay* and *Paul* also leads directly to the phraseme as a main entry. It is also possible to type the entire phraseme into the search box on the homepage. The *CALD* search process is thus very easy to use in this case. In other respects *CALD* is unusual, in that some phrasemes can only be found when the plural form of a noun is entered (*swords*, *ploughshares*, *sins*, *coals*) rather than the singular form of the noun, contrary to most other forms of dictionary search which require the use of the singular noun. This may confuse and frustrate some users.
- *COBUILD*: Only the headword *Peter* is used for the phraseme in this *MELD*. After entering *Peter* or *peter* in the search box, the user is taken to a list of entries. The phraseme appears as a phrase under the word *Peter*, rather than as an entry in its own right. The restricted number of headwords and the use of the phraseme in a definition rather than as a headword make this a less user-friendly process.
- *LDOCE*: On entering the word *rob* in the search box, the user is taken to a list of entries where they must decide which entry, including which part of speech, to click on. In the case of *rob* this presents no challenges. After clicking on the first entry, "**rob** verb", the user is taken to a list of entries which include the phraseme. A search at *Peter*, however, leads directly to a list of possibilities including *peter* (verb), *Peter*, *Peter Abelard*, *Blue Peter* and *Peter Pan*. The user must decide which is the most relevant. After clicking on plain *Peter* they are taken to a cross reference which directs them back to the phraseme at *rob*. This is partly user friendly, in that only one of the words in the phraseme is used as a headword and one is used

as a cross reference. However, the need to decide on a part of speech may complicate the process for many users.

- *MEDAL*: The search term *rob* leads directly to the phraseme towards the bottom of the page, under "Phrases". The phraseme is not listed or linked under *peter* or *pay*, but curiously the search word *Paul* leads to the definition under *rob*. It is also possible to type the entire phraseme into the search box on the homepage. The process is easy to follow, but it is a little confusing that only the entire phraseme, or the first and last words, produce results.
- *OALD*: The search term *rob* leads straight to the phraseme part way down the page under the heading "Idioms". Alternatively, at *peter* the user is directed back to the phraseme with a hyperlink to the entry at *rob*. The complete phraseme may also be entered in the search box on the homepage. This process is user friendly, but more links from the other words in the phraseme would improve it still further.

Obstacles to phraseme searches

There are thus three main obstacles which learners must negotiate if they are successfully to locate a phraseme in a MELD:

- **Identifying a word as belonging to a phraseme**
The first obstacle that users may face is that they may need to know that a word is part of a phraseme if they are to pursue their search past the first noun. Suppose, for example, that the learner is looking for the phraseme *to carry/take coals to Newcastle*. They may understand the individual meaning of each word, but will not necessarily realise that the words combine to give a figurative meaning. In that case, they may look under the first noun. Experience has taught them that nouns are usually listed in a dictionary in their singular form; they would therefore miss the phraseme in *CALD*, which uses the plural form *coals*. Similarly, with the phraseme *the road to Damascus*, only *Damascus* is used as the headword in *LDOCE*, *MEDAL* and *OALD*, so that users searching only under *road* will fail to find the phraseme. A wide choice of headwords is therefore necessary if users are to search successfully, and users should be encouraged to persist if their first search does not yield results. Again, this presupposes that users realise they are indeed searching for a phraseme.
- **Locating salient words**
Secondly, users must correctly identify the headword under which the phraseme is listed. Their choice of 'important' or salient word may not, however, accord with that of the lexicographer. Proper nouns in particular might be seen as particularly salient, yet in the nine examples of

phrasemes in this study containing proper nouns, the approach varies greatly among the Big 5 (see Table 1), and the proper noun is not always used as the headword.

Despite some consistency, therefore, there is also much variation between dictionaries. *Bourke*, *Coventry* and *Damascus* were easy to find, but not all the MELDs listed *Newcastle*, *Adam*, *Clapham*, *Nod* or *Peter*. In particular, the choice of *Paul* as a headword in *MEDAL* is most unusual, since it is neither the first noun nor the first salient word.

— **Recognising phraseme variants**

Thirdly, users may not be aware of phraseme variants, such as *to carry/take coals to Newcastle*. This means that if a phraseme is listed under only one of the variants, a user may fail to find it. Seven phrasemes in this study had possible variations, as demonstrated in Table 1 above. Treatment of variants again differed among the Big 5.

Discussion

The greatest consistency in search results within a dictionary was found in *CALD*, whose numerous headwords generally make searching easy without the user needing to know which part of speech the word belongs to, or whether a word is part of a phraseme. The only detracting factor was the need to enter plural noun forms in some cases, which is counter-intuitive for most dictionary users.

COBUILD tended to include phrasemes in its noun entry examples, rather than as individual entries. For example, under *multitude* we read, "If you say that something covers or hides **a multitude of sins**, you mean that it hides something unattractive or does not reveal the true nature of something". The phraseme is written in bold font, but the reader used to finding items as an entry on the left or in a vertical list might have trouble finding the phraseme. In the four examples in this paper, however, *COBUILD* is consistent in using unusual, or salient, nouns as its phraseme headwords, even when these are not the first noun in the phraseme. For example, *to beat/turn swords into ploughshares* is listed at *ploughshare* rather than at *sword*. *COBUILD*'s lack of inclusion of many phrasemes is a drawback for users, and the choice of phrasemes included does not always reflect those which are most commonly used. For example, Biblical phrasemes such as *to beat/turn swords into ploughshares* are less commonly used than phrasemes referring to older measurements, such as *not to budge an inch* (Miller Forthcoming), which does not appear in *COBUILD*.

The *LDOCE* user has first to negotiate the part of speech they need before they can find the phraseme. As Nesi and Haill (2002: 282) suggest, this may cause problems for some learners. The choice of possible topics under each headword further complicates the issue. For example, when searching for *the man/woman on the Clapham omnibus*, the user may also choose whether they are

looking for a particular "topic", such as "occupation", "board game", "military" or "family". Fortunately, it is also possible to click just on the headword itself. While *LDOCE* tends to prefer nouns to verbs as headwords, this policy is not always consistent. For example, *to rob Peter to pay Paul* is listed at the verb *rob*, but *not to know someone from Adam* is listed at the noun *Adam*. The choice of headword nouns is also not consistent, for while unusual words such as *multitude* are sometimes used (in *to cover a multitude of sins*), at other times more common nouns such as *man* are chosen (in *the man/woman on the Clapham omnibus*), although a more salient word such as *Clapham* or *omnibus* would have led to the phraseme more quickly.

MEDAL, like *CALD*, also includes phrasemes at main entries, but occasionally cross references them with hyperlinks at other entries. For example, at the end of the page under *budge*, we find "see also → *inch*". *MEDAL* is fairly consistent in its use of nouns as headwords, but at times the choice of salient noun is puzzling. Why, for instance, is the second noun *Paul* chosen instead of the first noun *Peter* as the headword for the phraseme *to rob Peter to pay Paul*?

In *OALD*, the headword is usually a noun, though not always the first noun in the phraseme; sometimes a verb is chosen, however, as in *not to know someone from Adam*. A welcome feature is the addition of idioms in small font underneath a headword entry. For example, at *Newcastle* there is a small entry on the main page directing the user to the phraseme *carry/take coals to Newcastle*.

The above discussion highlights the fact that one of the main problems for users is the lack of consistency between dictionaries, and sometimes within a single dictionary. If a user is familiar with a particular MELD, they may use a strategy which does not work in another MELD, leading to the frustration of an unsuccessful search. Of the five dictionaries in this study, *CALD* and *MEDAL* were the most user friendly; *COBUILD* included fewer entries and headwords but was consistent in its choice of headwords; *LDOCE* showed less consistency, and the lack of hyperlinks made any cross references less effective; and *OALD* tended to use nouns as headwords, but not always consistently.

For the user, only *CALD* comprehensively addresses the three main problems raised in this paper. First, almost every word in a phraseme is searchable as a headword, so that the EAL learner does not need to know that a word is part of a phraseme in order to find an expression. Secondly, the generous provision of headwords means that the learner does not have to second guess the lexicographer's choice. Thirdly, many phraseme variants are addressed, facilitating searches for variable phrasemes.

In general, however, the results of this study are in line with those of Gates (1986), in that there is still a lack of consistency between dictionaries. Although Atkins and Rundell (2008) highlight the difficulty of achieving total consistency in the treatment of multi-word expressions, online search facilities should make the search process simpler by providing wider choices. During the course of this study, from the months October 2011 to November 2012, it was found that *OALD* did indeed increase its range of cross references. It is

therefore likely, and of course desirable, that this trend continue across all the Big 5.

Five strategies are suggested to help EAL learners with phraseme searches. First, they need to be aware of the existence of phrasemes in English. When they are confronted by a group of words that seems to present an unfamiliar meaning (such as *carry/take coals to Newcastle*) they may then realise that the overall meaning of the phrase may be more than the sum of its parts, so that it might be necessary to use a dictionary even though they may understand the individual words. Secondly, they can be instructed in the ways in which dictionaries work generally with regard to phraseme headwords. Since, however, as shown above, there is a lack of consistency, they may try searching first under what appears to be a salient word. If unsuccessful, they should be encouraged to keep searching using different words in a sentence until their search yields results. Thirdly, EAL students can be made aware that phrasemes often have variants (such as *carry/take coals to Newcastle*); hence they might need to look under a different word if their first search is unsuccessful. Fourthly, since all the Big 5 are freely available online, and all vary in their coverage of phrasemes, learners may also be encouraged to try more than one dictionary. Finally, it would be helpful for EAL students to make their own lists of phrasemes and any variants, and to share such lists with their peers. The key word in all these strategies is 'persistence'. Teachers of EAL students can promote such strategies in class, but lecturers or teachers in any discipline can post simple instructions on dictionary use, and links to dictionary websites, in their course details on their Learning Management Systems.

Conclusion

It can be seen from this study that the choice of headwords varies greatly, both within and between dictionaries, making it difficult for students to conduct consistently successful searches and potentially leading to frustration (Gouws and Prinsloo 2005: 9). It seems axiomatic that dictionaries should be consistent within themselves when choosing headwords, but this appears not always to be the case. Nevertheless, greater consistency would benefit users, who could familiarise themselves with the style of a particular dictionary and even apply their techniques when searching in a different dictionary.

Those dictionaries which provided the greatest number of headwords and cross references were obviously the easiest to use, and *CALD* was the winner in this respect. The ability to find a phraseme at any of its constituent words is a great bonus for users who do not recognise a phraseme as a multi-word expression. It is therefore suggested that as wide a choice of headwords as possible be used. If this is not possible, then searching would be facilitated by a consistent use of salient words (such as proper nouns), first nouns, and first finite verbs, as initially proposed by Whiting and Tilley (Doyle 2007: 191).

Learners should be encouraged to persist when searching in a dictionary,

and to try different dictionaries until they are satisfied with their results. Although they may be unaware that a word forms part of a phraseme, it is still worthwhile for them to search under what may appear to be familiar words (such as *rob* in *rob Peter to pay Paul*) if they do not understand the meaning of a sentence. In this way, they may be led to a phraseme and will extend not only their decoding but also their encoding vocabulary. It is not necessary to be an English language teacher in order to promote such strategies to students in any discipline.

In terms of headword choices, it seems that while the Big 5 might be *birds of a feather* in one sense, they do not always *flock together*. A greater understanding of search strategies would therefore help EAL students to make better use of their dictionaries and improve their language skills.

References

- Atkins, B.T.S. and M. Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.
- Bergenholtz, H. and R.H. Gouws. 2007. The Access Process in Dictionaries for Fixed Expressions. *Lexicographica. International Annual for Lexicography* 23: 237-260.
- Cambridge University Press. 2011. *Cambridge Dictionaries Online*. Available: <http://dictionary.cambridge.org/> [7 November 2012].
- Dobrovolskij, D. and E. Piirainen. 2005. *Figurative Language: Cross-Cultural and Cross-Linguistic Perspectives*. Amsterdam: Elsevier.
- Doyle, C.C. 2007. Collections of Proverbs and Proverb Dictionaries: Some Historical Observations on What's in Them and What's not (with a Note on Current 'Gendered' Proverbs). Skandera, P. (Ed.). 2007. *Phraseology and Culture in English*: 181-203. Berlin: Mouton de Gruyter.
- Dykstra, A. and T. Schoonheim (Eds.). 2010. *Proceedings of the XIV Euralex International Congress, Leeuwarden, 6–10 July 2010*. Ljouwert: Fryske Akademy.
- Fillmore, C.J., P. Kay and M.C. O'Connor. 1988. Regularity and Idiomaticity in Grammatical Constructions: The Case of Let Alone. *Language* 64(3): 501-538.
- Gates, E. 1986. The Treatment of Multiword Lexemes in some Current Dictionaries of English. Snell-Hornby, M. (Ed.). 1988. *ZüriLEX '86 Proceedings. Papers Read at the EURALEX International Congress, University of Zürich, 9–14 September 1986*: 99-106. Tübingen: Francke Verlag.
- Gouws, R.H. and D.J. Prinsloo. 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.
- Granger, S. and F. Meunier. 2008. Phraseology in Language Learning and Teaching: Where to from here? Meunier, F. and S. Granger (Eds.). 2008. *Phraseology in Foreign Language Learning and Teaching*: 247-252. Amsterdam/Philadelphia: John Benjamins.
- HarperCollins. 2012. *myCOBUILD.com*. Available: <http://www.mycobuild.com/free-search.aspx> [7 November 2012].
- Kövecses, Z. and P. Szabó. 1996. Idioms: A View from Cognitive Semantics. *Applied Linguistics* 17(3): 326-355.
- Macmillan Publishers Limited. 2012. Stop the Press: Dictionary No Longer a Page-Turner. Available: <http://www.macmillaneducation.com/MediaArticle.aspx?id=1778> [7 November 2012].

- Macmillan Publishers Limited.** 2009–2012. *Macmillan Dictionary*. Available: <http://www.macmillandictionary.com/> [7 November 2012].
- Miller, J.** 2010. Coals to Newcastle or Glittering Gold? Which Idioms Need to be Included in a Learner's Dictionary for Use in Australia? Dykstra, A. and T. Schoonheim (Eds.). 2010: 1226-1233.
- Miller, J.** Forthcoming. Phraseology across Continents. Szerszunowicz, J. (Ed.). Forthcoming. *Intercontinental Dialogue on Phraseology. Volume 2*. Bialystok, Poland: University of Bialystok Publishing House.
- Moon, R.** 1998. *Fixed Expressions and Idioms in English: A Corpus-based Approach*. Oxford: Clarendon Press.
- Moon, R.** 1999. Needles and Haystacks, Idioms and Corpora: Gaining Insights into Idioms, Using Corpus Analysis. Herbst T. and K. Popp (Eds.). 1999. *The Perfect Learners' Dictionary (?)*: 265-281. Tübingen: Max Niemeyer.
- Mulhall, C.** 2010. A Semantic and Lexical-based Approach to the Lemmatisation of Idioms in Bilingual Italian–English Dictionaries. Dykstra, A. and T. Schoonheim (Eds.). 2010: 1355-1371.
- Nation, I.S.P.** 2001. *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Nattinger, J.R.** 1980. A Lexical Phrase Grammar for ESL. *TESOL Quarterly* 14(3): 337-344.
- Nattinger, J.R. and J.S. de Carrico.** 1992. *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.
- Nesi, H. and R. Haill.** 2002. A Study of Dictionary Use by International Students at a British University. *International Journal of Lexicography* 15(4): 277-305.
- Nielsen, S. and L. Mourier.** 2005. Internet Accounting Dictionaries: Present Solutions and Future Opportunities. *Hermes, Journal of Linguistics* 34: 83-116.
- Oxford University Press.** 2011. *Oxford Advanced Learner's Dictionary*. Available: <http://oald8.oxfordlearnersdictionaries.com/> [7 November 2012].
- Pearson Education Ltd.** 2012. *Longman Dictionary of Contemporary English*. Available: <http://www.ldoceonline.com/> [7 November 2012].
- Rundell, M. (Ed.).** 2002. *Macmillan English Dictionary for Advanced Learners*. Oxford: Macmillan Education.
- Summers, D. (Ed.).** 2003. *Longman Dictionary of Contemporary English*. Fourth edition. Harlow, UK: Pearson Education.
- Turnbull, J. (Ed.).** 2010. *Oxford Advanced Learner's Dictionary*. Eighth edition. Oxford: Oxford University Press.
- Walter, E. (Ed.).** 2008. *Cambridge Advanced Learner's Dictionary*. Third edition. Cambridge: Cambridge University Press.
- Zgusta, L.** 1971. *Manual of Lexicography*. The Hague/Paris: Mouton.

A Balanced and Representative Corpus: The Effects of Strict Corpus-based Dictionary Compilation in Sesotho sa Leboa*

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

Abstract: Theoretically the Northern Sotho language is made up of almost 30 dialects while practically it is not so, because the standard language was formed from very few of its dialects. As a result, even today the language has no corpus which is balanced or representative owing to the fact that almost all of the available corpora are compiled from the written standard language and the written dialects. The majority of the Northern Sotho dialects do not have written orthographies, and the few dialects which had written orthographies prior to standardization came to monopolize the standard language and the Northern Sotho corpora. Therefore, the compilation of a corpus-based dictionary in Northern Sotho is tantamount to a continuation of producing unbalanced and unrepresentative dictionaries, which continue to sideline and to marginalize the majority of the communities and the linguistic varieties which could potentially enrich both the Northern Sotho standard language and the Northern Sotho corpora. The main objective with this research is to analyze, to expose and to suggest ways of correcting these irregularities so that the marginalized Northern Sotho dialects can be accommodated in the standard language. This will obviously increase the size of the Northern Sotho standard language and the corpus by more than 50%.

Keywords: CORPUS, BALANCED CORPUS, REPRESENTATIVE CORPUS, STANDARDIZATION, DIALECT, ORTHOGRAPHY, MARGINALIZED DIALECTS, PRESTIGE DIALECTS, MISSIONARY ACTIVITIES

Opsomming: 'n Gebalanseerde en verteenwoordigende korpus: Die gevolge van streng korpusgebaseerde woordeboeksamstelling in Sesotho sa Leboa. Teoreties bestaan die Noord-Sotho taal uit byna 30 dialekte, terwyl dit prakties nie die geval is nie omdat die standaardtaal uit slegs 'n paar van sy dialekte gevorm is. Gevolglik het die taal selfs vandag nog geen korpus wat gebalanseerd of verteenwoordigend is nie as gevolg van die feit dat byna al die beskikbare korpusse saamgestel is uit die geskrewe standaardtaal en die geskrewe dialekte. Die meerderheid Noord-Sotho dialekte het nie geskrewe ortografieë nie, en die paar dialekte wat geskrewe ortografieë gehad het voor standaardisasie het begin om die standaard-

* This article is a revised version of a paper presented at the Seventeenth Annual International Conference of the African Association for Lexicography (AFRILEX), which was hosted by the Department of African Languages, University of Pretoria, Pretoria, South Africa, 2-5 July 2012.

taal en die Noord-Sotho korpusse te monopoliseer. Die samestelling van 'n korpusgebaseerde woordeboek kom gevolglik neer op 'n voortsetting van die totstandbrenging van ongebalanseerde en onverteengewordigende woordeboeke wat voortgaan om die meerderheid van die gemeenskappe en taalvariëteite opsy te skuif en te marginaliseer wat potensieel sowel die Noord-Sotho standaardtaal as die Noord-Sotho korpusse kan verryk. Die hoofdoel met hierdie navorsing is om maniere te ondersoek, uit te wys en voor te stel om hierdie ongelykhede reg te stel sodat die gemarginaliseerde Noord-Sotho dialekte in die standaardtaal ondergebring kan word. Dit sal vanselfsprekend die grootte van die Noord-Sotho standaardtaal en korpus met meer as 50% vermeerder.

Sleutelwoorde: KORPUS, GEBALANSEERDE KORPUS, VERTEENWOORDIGENDE KORPUS, STANDAARDISASIE, DIALEK, ORTOGRAFIE, GEMARGINALISEERDE DIALEKTE, PRESTIGEDIALEKTE, SENDELINGAKTIWITEITE

1. Introduction

Northern Sotho, or Sesotho sa Leboa, presently has corpora which were built entirely from published materials, and as such representing only the written and documented dialects. This is a major shortcoming, because the published documents in indigenous languages like Northern Sotho are usually based on the few dialects which are restricted to certain parts of society, while the majority of the undocumented dialects are sidelined. Northern Sotho is made up of approximately 30 dialects which are found in almost all five municipal districts of the Limpopo Province. Of all these dialects, almost half are marginalized 'languages'. This simply means that these dialects (or 'languages' as the communities themselves regard their dialects) are not included in written standard Northern Sotho.

This written Northern Sotho language, which is derived from the few documented dialects, i.e. the 'prestige' dialects which were fortunate to have written and published materials prior to standardization, are the ones which are represented in Northern Sotho corpora today. These irregularities are the subject of investigation in this research, whose objectives can be summarized as follows:

- (a) to discuss the shortcomings and disadvantages of relying solely on corpus-based dictionary compilations in indigenous languages like Northern Sotho,
- (b) to demonstrate that Northern Sotho does not have a 'balanced or representative' corpus,
- (c) to analyze factors leading to the marginalization of the majority of the Northern Sotho dialects, and
- (d) to show how purism and monopolies influenced the standardization of Northern Sotho, thereby leading to the marginalization of the majority of its dialects.

2. What is a corpus?

The term *corpus* is defined by several linguistic and lexicographic scholars, such as Watson (1976), Kennedy (1998), Gouws and Prinsloo (2005) among others. Watson (1976: 243) describes a corpus as:

a body of writings of a particular kind, or on a particular subject.

Kennedy's (1998: 1) definition of a corpus is as follows:

a corpus is a body of written text or transcribed speech which can serve as a basis for linguistic analysis and description.

The most direct and straightforward description of a corpus is given by Gouws and Prinsloo (2005: 21):

The collection of written and spoken material from the sources earmarked for the dictionary basis. Data is compiled and stored as a lexicographic data basis which should preferably be an electronic corpus. An electronic corpus can be defined in an oversimplified way as a computerized collection of texts, such a collection of texts can, for example, consist of tape recordings of conversations and written texts which have been typed into the computer.

These definitions show that corpora are supposed to be compiled from both written and oral materials. But, on the contrary, it is not always easy to compile a corpus for languages or dialects with no orthographies or written forms. Gouws and Prinsloo (2005: 21-22) say the following in this regard:

Unfortunately most corpora around the world lack sufficient data from spoken sources. The reason for this is that there are many logistical problems and ethical factors involved in the collection of spoken data. It is also much more expensive and time consuming to enlarge the corpus with spoken data compared to data available in electronic, printed or even handwritten format. Extending the corpus with data already in electronic format such as texts downloaded from the internet or texts already available on computer disk is relatively easy. Printed matter which is not available in electronic form can also relatively easily be computerized by means of Optical Character Recognition (OCR), commonly referred to as 'scanning'.

3. The issue of a 'balanced' and 'representative' corpus

A normal and appropriate general corpus for a language needs to be balanced and representative. According to Kennedy (1998: 20), "a general corpus is typically designed to be balanced, by containing texts from different genres — including spoken and written". Kennedy (1998: 52) further emphasizes that "for a corpus to be 'representative' there must be a clearly analysed and defined population to take the sample from".

Gouws and Prinsloo (2005: 25) re-emphasize these requirements with relevancy to the South African indigenous language situation as follows:

Important for lexicographic work in South Africa is that corpus compilers should be sensitive to all of these aspects, i.e. to build as far as possible, corpora that are big enough, well balanced and representative so that valid conclusions for lexicographic purposes can be drawn.

As far as the standard language is concerned, it could be argued that, although no spoken data has been included, the compilers of currently available corpora for Northern Sotho, did make a real effort to capture available written sources. Ideally the corpus should be extended to include large quantities of dialectal information.

3.1 The consequences of standardization on Northern Sotho

Standardization has prioritized certain Northern Sotho dialects, while it has marginalized others, as indicated by the following facts:

- Purism was used as pretext by most language committee members who dominated the official language bodies, for standardizing their own dialects to 'represent' all other dialects. The few dialects which have contributed to standardized Northern Sotho form less than half of all the Northern Sotho dialects.
- Of the approximately 30 Northern Sotho dialects, the only dialects which are represented in standard Northern Sotho and the available corpora are Sekone (which is spoken in the central and southern parts of the Waterberg district and a section of the Capricorn district), Sekopa and Sepedi, which are used in the Sekhukhune district and Sekgaga (dialect of the Gamphahlele and Gamothapo districts), as well as the few dialects spoken in the Mankweng and Mamabolo areas.
- Standardization has marginalized the majority of the 'potential' Northern Sotho dialects, i.e. the dialects which are grouped as dialects of Northern Sotho even though practically they differ by far from the Northern Sotho standard language. These include major 'languages' like Sepulana, Setlokwa, Khelobedu, Seroka, Sephalaborwa, Sehananwa, Sekgaga (of Maake and Mogoboya), Sekhutšwe, and others.

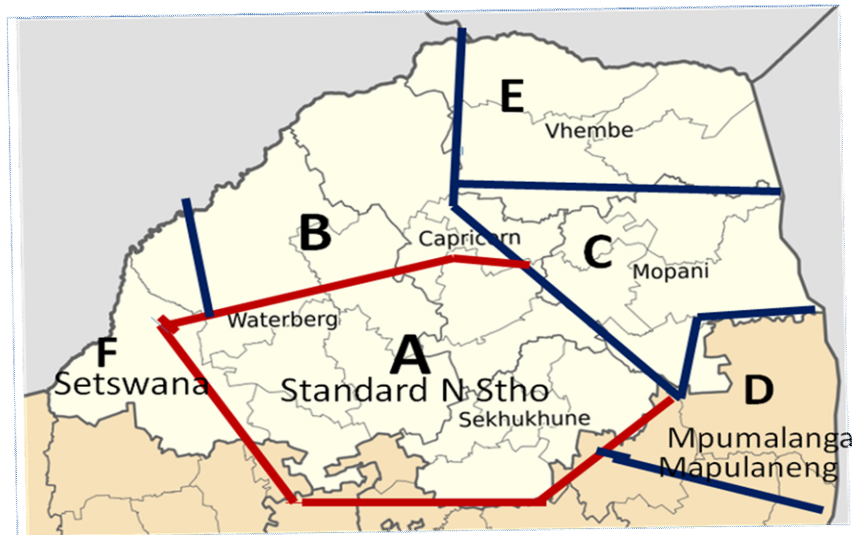
3.2 Demographic representation of the Northern Sotho dialects

Demographically the marginalized 'dialects' include all the linguistic areas of Botlokwa, Sekgosese–Lemondokop and the Senwabarwana areas in the Northern part of the Capricorn district and a small section of the Vhembe district; and the whole of the Mopani district, which encompasses, inter alia, the areas

of Bolobedu–Modjadjiskloof, Tzaneen, Trichardtsdal and Phalaborwa, as well as the Bushbuckridge–Mapulaneng area in the North Eastern part of Mpumalanga.

This means that Kennedy's principle, i.e. 'for a corpus to be 'representative' there must be a clearly analysed and defined population to take the sample from', was only considered with reference to the few dialects around Polokwane–Matlala, Lepelle–Nkumbi, Gasekhukhune and a section of the Waterberg district.

The following map explains the demography of Northern Sotho, showing dialectal distributions within the Limpopo and Mpumalanga regions:



Map showing the demography of the Northern Sotho dialectal regions

Area A: Standard Northern Sotho is formed on the basis of dialects spoken in this area. The area includes the central and south eastern part of the Waterberg district, the southern part of the Capricorn district and the whole of the Greater Sekhukhune municipal district. The dialects in this area include, inter alia, Sekopa, Sepedi, Sekgaga (of Mphahlele, Mothapo, etc.) and Sekone (of Moletši, Matlala, Bakenberg, Polokwane, Mothiba, Dikgale, etc.).

Areas B, C and D: All the Northern Sotho dialects in these areas are not represented in the standard language. These areas include:

the northern parts of the Waterberg and Capricorn districts (area B), inhabited by, inter alia, the Batlokwa and Bahananwa communities

the whole of the Mopani district, the north eastern part of the Capricorn district and a section of the southern part of the Vembe district (area C): These areas

are inhabited by the Batlokwa (in the Sekgosese–Lemondokop areas) and Baroka communities. These communities include, inter alia, the Balobedu, Bakgaga (ba Maake), Baphalaborwa, Banareng ba Sekororo, Batlokwa, etc.

the Bushbuckridge or Mapulaneng area, in the north eastern part of the Mpu-malanga Province (area D). This is the place of residence of the Mapulana communities.

Areas E and F: Even though there are a few Northern Sotho dialects in these areas, the overwhelming majority of the indigenous communities in area E are the Venda people, while area F is dominated by the Batswana communities.

3.3 The gap between the standard languages and the marginalized dialects

All the Northern Sotho dialects or 'languages' in areas B, C and D, and a few scattered remnants in area E (in the above map) are marginalized. The dialects in these areas differ considerably from standard Northern Sotho even though the communities are forced to use the standard language for official communications in education and all official correspondence. Sometimes the gap between the standard language and the marginalized dialects is so wide that most of the communities in area A (in the above map) need interpreters to engage in effective communication with the communities of areas B, C and D, especially with the Balobedu, Baphalaborwa and Mapulana. These marginalized dialects in areas B, C and D are not only excluded from the standard language, but also from the Northern Sotho official orthography and, eventually, from the official corpora.

3.4 The influence of standardization on the compilation of the corpora

The fact that no printed matter exists for the dialects left the corpus compilers no alternative than to concentrate on the already written and standardized language in compiling the Northern Sotho corpora. Ideally these corpora should be extended by the inclusion of dialectal data. Lexicographers should then be in a position to consider corpus-based dialectal data especially for lemma selection and translation equivalents. So, for example, a comprehensive dictionary should include lemmas such as *molema*, *khobe*, *lesalabu* and *kholophana* and also such words as translation equivalents for the lemmas *bat*, *fish*, *watermelon*, etc. respectively.

In assisting to increase and to improve the University of Pretoria corpus, the Sesotho sa Leboa National Lexicography Unit staff and the University of Pretoria lexicographers in the Department of African Languages, under the guidance of lexicographic scholars like Prof. D.J. Prinsloo and Prof. G.-M. de Schryver, much relied on what Gouws and Prinsloo (2005: 22) refer to as the OCR or Optical Character Recognition:

Printed matter which is not available in electronic form can also relatively easily be computerized by means of Optical Character Recognition (OCR), commonly referred to as 'scanning'.

This means that existing published or printed materials in Northern Sotho were used to develop the corpus. Unfortunately, these materials are only available in the 'one-sided' standard language because the marginalized dialects did not have any written or published materials, apparently because these dialects did not have any orthography. Even up to this moment, there are no available printed materials covering the marginalized dialects. Therefore, continuation of relying solely on the 'printed matter' will be tantamount to advancing and promoting the marginalization policy of the purists.

4. The role played by the other Sotho languages in the marginalization of the Northern-Sotho dialects

The gap which exists between the standard Northern Sotho language and some of its own dialects is much wider than that which exists between Northern Sotho and the other two Sotho languages, i.e. Setswana (Western Sotho) and Sesotho (Southern Sotho). The influence of these Sotho languages contributed much to the widening of the gap between the Northern Sotho standard language and its dialects. The process of the interaction and the relationships between the three Sotho languages can be divided into three phases, i.e. the missionary period, the unification or harmonization period and the separate development (apartheid) period.

4.1 The missionary period

This covers the period before 1929, when the role of developing and converting the Sotho oral languages into written 'languages' was effected only by the missionaries. The task of writing and compiling the first orthographies for the Northern Sotho language was started by the German missionaries, i.e. the Berlin Evangelical Missionary Society, during the 19th century. They established missionary stations under the Bapedi and Bakopa communities, whose dialects are much closer to Setswana and the Southern Sotho communities than the Northern Sotho dialects in the North and the Lowveld. That is why the first written so-called 'Sepedi' missionary orthography did not deviate much from the structures of the other two Sotho languages. This can be ascribed to the following factors:

- Historically the Bapedi and Bakopa communities are more aligned to the Batswana communities than to the other Northern Sotho dialectal groups.
- The Bapedi and Bakopa communities are in closer proximity to the Batswana and the Basotho communities when compared to the Northern

Sotho communities in the North and the Lowveld areas, who are, in turn, bordering on the Vatsonga and the Vhavenda communities.

4.2 The unification or harmonization period

This is the period between 1929 and 1961 when the Union Government took over the administration of the indigenous schools from the missionaries for the first time, and started regulating the development of the indigenous languages. As a result, when the Transvaal Education Department (TED) took over responsibility for organising the Sotho languages after 1929, the focus was more on uniting all the Sotho languages, i.e. Western Sotho (Setswana), Southern Sotho and Northern Sotho, into one standard language. These languages were, by then, regarded as the Sotho dialects. The Transvaal Sotho District Committee, which was formed by the TED, compiled and introduced its first Sotho orthography in 1930. After holding several meetings and conferences between 1930 and 1950, like the Somerset House Conference of 1947, the Orthography Sub-Committee of the Sotho Language Board revised and adopted the official Sotho orthography for all the Sotho 'languages' in 1950. After passing the Bantu Education Act in 1953, the Union Government took over the responsibility of running formal education from the missionaries, and the 1950 Sotho orthography became official in the Transvaal. Mojela (2008: 121) comments as follows in this regard:

It was only after 1929 that the Transvaal Education Department (TED) started making attempts at standardizing the Sotho languages in the former Transvaal which eventually led to the formation of the Language Boards (Mojela 2005: 46). In South Africa, for instance, it was only after the passing of the Bantu Education Act in October 1953 (Act No. 47 of 1953) that the South African government took over control of formal education from the missionaries.

The 1950 orthography brought the three Sotho 'languages' close together and this gave advantage to all the Northern Sotho dialects, like Sepedi, Sekopa, Sekone, etc., which are structured more closely to the Southern Sotho and Setswana languages, because the Northern Sotho standard language came to be based on this orthography. As a result, this compromised the Northern Sotho dialects, like Setlokwa, Selobedu, Sepulana, Seroka, etc. which are found in the far North and the Lowveld. All those dialects whose structures were too remote from Setswana and Southern Sotho were thus marginalized. Most of the early Northern Sotho publications and printed materials, which later came to be used as important sources in the standardization of the Northern Sotho language were written according to the 1950 orthography and the 1953 unified Sotho standard language. The following examples from Mojela (2008: 127) demonstrate the remoteness of the standard Northern Sotho language from its own dialects, when compared to its closeness to the Setswana language:

Sesotho sa Leboa	Setswana	Selobedu (NS dialect)	English
<i>mopani</i>	<i>nato/mopani</i>	<i>moṯhanare</i>	mopani tree
<i>leribiši</i>	<i>lerubisi</i>	<i>mmankhoṯo</i>	owl
<i>mmankgagane</i>	<i>mmamanthane</i>	<i>molema</i>	bat
<i>hlapi</i>	<i>tlhapi</i>	<i>khobe</i>	fish
<i>betha/itiya</i>	<i>betsa</i>	<i>moṯa/tiya</i>	hit
<i>legotlo</i>	<i>legotlo</i>	<i>lehoṯo/peba/manṯoro</i>	mouse
<i>legapu</i>	<i>legapu</i>	<i>lesalabu</i>	watermelon
<i>nona</i>	<i>nona</i>	<i>kholphana</i>	be fat/gain weight
<i>mokgaditswane</i>	<i>mogaditswane</i>	<i>mphekwa</i>	lizard

These examples demonstrate how the harmonization and unification of the Sotho languages brought the Northern Sotho standard language closer to Setswana and Southern Sotho, while at the same time distancing itself from its own dialects.

4.3 The separate development (apartheid) period

The attempt to unify the Sotho languages was destroyed by the policy of separate development, which aimed at developing all the South African societies separately. With this policy, the apartheid regime wanted to use a 'divide and rule' strategy to keep power in the hands of the white minority. As a result, the three subcommittees of the unified Sotho language, i.e. the subcommittees for Setswana, Southern Sotho and Northern Sotho were converted to fully-fledged autonomous Language Boards for the respective languages to develop each of them separately. Even though separate development gave the standardizing authorities enough chance to pay attention to the marginalized dialects in every language, in Northern Sotho too much damage had already been done because: (a) the written language and the available orthographies were dominated by the 1953 standard Sotho language, (b) the educated elite who came to dominate membership of the newly established Language Boards were educated on the basis of the 1953 Sotho standard language, (c) all the publications which were used in the standardization of Northern Sotho were totally foreign to the marginalized dialects, and (d) ultimately, the only available materials for the corpus compilers in the compilation of the Northern Sotho corpora were still those which excluded the marginalized dialects.

5. The repercussions of dialectal marginalization

The issue of the 'unbalanced' and 'unrepresentative' corpora, which resulted from a one-sided standardization has led to the emergence of a nationalistic spirit among the marginalized communities:

- The emerging elite groups and the *magoši* (traditional leaders) from areas

such as Bolobedu, Makhutšwe, Botlokwa, Senwabarwana, Mapulaneng, etc. have already started questioning the validity of incorporating their dialects in the Northern Sotho standard language, which is not only lexically and morphophonologically foreign to them, but in all practical respects, too different from their dialects.

- At the same time, most communities whose dialects were favoured by standardization started claiming ownership of the standard language, while those groups who were sidelined by the standardization started disowning and opting for withdrawal from the standard language, because they believe their 'languages' are misplaced.
- The Balobedu under the leadership of the 'self-imposed' Archbishop Prince Madlakadlaka, and influential activists such as Mr Phetole Mampule, under the influence of the philosophies of the Kara Heritage Institute of Dr Mathole Motshekga, started questioning the inclusion of Khelobedu into the Northern Sotho standard language. They have already submitted several petitions and requests to the Government and the Constitutional Court to have Khelobedu declared an official language. Some are even insisting that Khelobedu is more aligned to Tshivenda than to Northern Sotho, and as such, rather a Tshivenda dialect than a dialect of Northern Sotho.
- The Mapulana communities under the leadership of high profile personalities and their *magoši* are also demanding official status for Sepulana because they insist they are not Bapedi, but Basotho (in the east).

6. Conclusion

This research demonstrates that corpus lexicography in the South African indigenous languages, like Northern Sotho, is not always possible, because there are still many dialects within these languages which are not included in both the official orthographies and the standard languages. Since the corpora are compiled mostly from the written languages and from published materials, the dialects which did not have written forms or published materials will not always be included in the corpora. Even though sometimes oral materials were collected and included into the corpora, very few of these published materials were recorded in languages like Northern Sotho because, up to this moment, almost all the available Northern Sotho corpora do not have anything related to the marginalized dialects like Khelobedu, Sepulana, Seṭokwa, Sehananwa or Sephalaborwa. As a result, the available Northern Sotho corpora do not conform to the lexicographic principles of 'balance' and 'representativeness'. The standard language is neither balanced nor representative because it reflects less than half of its dialects.

In conclusion, this research recommends further research and a thorough

revision of the official orthography and the standard language to incorporate all the omitted dialects into the Northern Sotho language, before prescribing strict corpus-based lexicography. This will not only silence the emerging nationalistic spirit which threatens to divide the language, but will also double the size of the Northern Sotho corpus.

References

- Gouws, R.H. and D.J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.
- Kennedy, G.** 1998. *An Introduction to Corpus Linguistics*. London/New York: Longman.
- Mojela, V.M.** 1991. *Semantic Changes Accompanying Loan-words in the Northern Sotho Lexicon*. Unpublished M.A. Thesis. Pretoria: Vista University.
- Mojela, V.M.** 2005. Standardization and the Development of Orthography in Sesotho sa Leboa — A Historical Overview. Webb, V. 2005. *The Standardization of African Languages in South Africa*. Pretoria: University of Pretoria.
- Mojela, V.M.** 2008. Standardization or Stigmatization? Challenges Confronting Lexicography and Terminography in Sesotho sa Leboa. *Lexikos* 18: 119-130.
- Rooney, K. (Ed.-in-chief).** 2001. *Encarta Concise English Dictionary*. London: Bloomsbury.
- Watson, O. (Ed.).** 1976. *Longman Modern English Dictionary*. Harlow: Longman.

Digitizing the Monolingual Lusoga Dictionary: Challenges and Prospects

Minah Nabirye, *Department of Languages and Cultures, Ghent University,
Ghent, Belgium (minah.nabirye@UGent.be)*

and

Gilles-Maurice de Schryver, *KongoKing Research Group, Department of
Languages and Cultures, Ghent University, Ghent, Belgium; and
Department of African Languages, University of Pretoria, Pretoria, South
Africa (gillesmaurice.deschryver@UGent.be)*

Abstract: This article recounts the genesis, growth and turbulent events that accompanied the compilation of the *Eiwanika ly'Olusoga*, that is, the Monolingual Lusoga Dictionary. Contrasting academia with the trade, legacy with state-of-the-art dictionary compilation software, high praise and visibility with daily and down-to-earth drudgery, it recounts the events chronologically, from humble to grand, from paper to digital, from success to insignificance, and leads to a set of highly realistic proposals to be considered by all those involved in the compilation of explanatory dictionaries for the African languages.

Keywords: MONOLINGUAL LEXICOGRAPHY, DICTIONARY COMPILATION SOFTWARE, FUNDING, ACADEMIA, TRADE, DIGITIZATION, LUSOGA, UGANDA

Obufunze: Okuta Eiwanika ly'Olusoga mu mbeela y'omutegekowaziso ogusomwa ku kompyuta: Ebizibu n'ebiluubililwa. Olupapula luno lwandhula obuzibu obwekulungila mu kuwandiika *Eiwanika ly'Olusoga*. Ebizibu ebyalimu bigelaagelanhizibwa n'ebyo ebitele okwagwanibwa mu kweyunila ebyetaago by'emisomo, eby'obusuubuzi ni mu mitegeko egitela okukozesebwa mu kuwandiika amawanika. Ebyetaago ebili mu mitendela gino gyonsatule biweebwa okulaga nti, okutendelezebwa okwandibaile kugwaniile okuweebwa omulimu ogw'ekika kino tikufunibwa. Omutindo ogulwoozebwa nti guteebwawo abandi kwe bayinza okusenziila gubusibwabusibwa bw'ogugelaagelanhia n'ebizibu ebigwetooloile mu bulamu obwa buliidho. Olupapula luno lukulaga nti, okuwandiika kwa *Eiwanika ly'Olusoga* okw'atandiikila mu mbeela ennafu, kw'asobolwa okutumbulwa okutuusibwa ku mutindo gw'amawanika agandi mu nsi yoonayoon. Obuvumu bw'enkola eyo, bw'asobozesa eiwanika lino okuva mu mbeela y'ekitabo ekigemebwaku okwizibwa mu mbeela y'omutegekowaziso ogusomwa ku kompyuta. Ebyafaayo ebiweebwa mu lupapula luno bigendelela kuwa kyakubonelaku eli abo abandyenze okugelaagelanhia ebisoboka n'ebitasoboka mu kuwandiika amawanika mu nnimi dha Africa enzaalilanwa.

Ebigambo ebikulu: NAMAWIKA W'OLULIMI OLULALA, OMUTEGEKO GW'OKUWANDIIKA AMAWANIKA, OBUYAMBI MU BY'ENFUNA, EMISOMO, EBYOBUSUUBUZI, OMUTEGEKOWAZISO, LUSOGA, UGANDA

1. Monolingual Lusoga lexicography: spin vs. facts

The monolingual Lusoga dictionary has all the characteristics of a success story. Begun as a tiny addendum to an MA dissertation at Makerere University a decade ago, it gradually grew into a fully-fledged 700-page desktop dictionary which is currently available from every major bookshop in Uganda. Key funders included an Indian businessman from the Sugar Corporation of Uganda (who helped fund the fieldwork), His Excellency Muammar al-Gaddafi (who paid for the extended university studies), and the Chinese Embassy in Kampala (who provided a computer, printer and funded the entire print-run). Various family members also contributed amounts large and small. Conceived, researched, compiled, funded and printed in Africa, it was furthermore officially launched by Uganda's President, Y.K. Museveni, in October 2010.

In the process we also set up our own publishing company, *Menha Publishers* (which since then released a Festschrift for Patrick Hanks), made the data freely available as an online dictionary, and prepared a downloadable version of the dictionary for offline use on a PC. In addition to an MA and a dictionary in three media, this project introduced us to the world of computers, the Internet and web design, software such as MS Word, Shoebox, InDesign and TLex, but above all resulted in arguably the most advanced e-dictionary for any Bantu language currently available.

The monolingual Lusoga dictionary has all the characteristics of a long and painful struggle. With its roots in a mere MA study, it quickly outgrew the initial goal, requiring vast amounts of time and funding. Academically, the struggles with various university committees, who promised to upgrade the study but in the end failed to do so, were endless. It was a struggle to obtain the necessary hard- and software, a struggle to go fundraising while studying, a struggle to find the peace and quiet in order to compile a dictionary in a household averaging thirty people (parenting was out of necessity outsourced to family members), a daily struggle even to wait for electricity to come on, a struggle to find colleagues willing to contribute (to the fieldwork, the actual compilation, the proofreading), a struggle to confront the elders in the community (being a young, female, Muslim, in a male, Christian environment), a struggle to find a publisher (none was found), a struggle to set up a business in a country with very few formal businesses, a struggle to set up a company website, and a struggle to find a printing house (with trips all the way to India and South Africa in search of one). To date, less than 10% of the print-run has sold.

There are over two million Lusoga speakers, many of them in the diaspora. Only junk and F-words are being looked up in the online Lusoga dictionary, and exactly two copies of the e-dictionary were sold so far.

In a world in which novels, television series and even e-dictionaries provide different optional paths to their consumers (when it comes to story endings, actors being voted in or out, or layers of lexical information being presented), it seemed fitting to offer two paths to open this article with. Granted, they may seem one another's opposites, but they are in the end the two sides of the same coin. The left-hand column is the one typically adhered to in an academic register, but the right-hand column is the real world with which many lexicographers have to contend with. The story which follows will develop both columns, but without the clear division, as in reality each difficulty which leads to some kind of (academic) success provides for enough energy to keep on doing the right thing. What should be clear from the outset, however, is that from a purely business perspective, compiling a monolingual dictionary for what is and remains in the end a severely under-resourced language, even though spoken by over two million speakers (UBS 2006: 12), is often nothing but financial suicide. This is an important finding, one which should be kept in mind by all those who wish or have to compile monolingual dictionaries for minority languages, as is the case for several of the National Lexicography Units (NLUs) in South Africa. The NLUs are for example routinely disappointed with the local dictionary publishers when their products are not being considered for publication. Not even allowing for aspects such as inherent quality, one must remember that those publishers do have a point. A second premise which one can posit right away is that no amount of funding *on its own* will ever be enough in such a situation. One must be willing to devote one's own personal resources to the task at hand, in the conviction that future generations will end up appreciating the effort. Lip-service abounds, both from the community and the government, but in the end lexicographic activities in such an environment are a profoundly solitary undertaking. Realizing this is of paramount importance, and we again have the impression that not all NLUs in South Africa, for example, are fully aware of this. Being aware will help dissipate false expectations. Being a dictionary compiler in such an environment is not a job, it is a vocation, a calling.

The purpose of what follows, then, is two-fold. On the one hand it aims at describing the main steps that led to the monolingual Lusoga dictionary, in print, as a free online dictionary, and as a downloadable e-dictionary. Various levels of digitization have played a role throughout the gestation of these three different media. Although one could argue that a dictionary team beginning work today will (hopefully) do things differently — starting off with the very latest software within the very latest metalexicographical frameworks — the fact of the matter is that very many dictionary teams are still going through the convoluted processes described below. It is our hope that a brief description of such a "legacy approach" will convince all future dictionary compilers of the need to indeed do things differently from the start. On the other hand, important lessons may be drawn with regard to the "dream" that monolingual lexicography is a viable undertaking in any language under any circumstances.

Clearly, it is not. That doesn't mean one has to down tools right away. Several possibilities will be offered to capitalize on what is seemingly a lost undertaking. As it stands, the various monolingual Lusoga dictionaries plainly fail commercially, and the logs attached to the online version suggest that they also fail to satisfy the target user group. Conversely, the dictionaries and the meta-lexicographical underpinnings are an academic success, and they of course put the language itself on the map. For Busoga to be taken seriously, an important hurdle was cleared. In the words of the first editor of the *International Journal of Lexicography*, Busoga became "a truly enlightened nation":

[...] almost as much as national flags, national anthems, and national armies, national dictionaries [a]re icons of national pride and prestige: a truly enlightened nation is not only numerate and literate but also dictionarate. (Ilson 2012: 382)

2. Background to the compilation of the *Eiwanika ly'Olusoga* (WSG)

The compilation of the first monolingual Lusoga dictionary — *Eiwanika ly'Olusoga* (Nabirye 2009), henceforth abbreviated to WSG — began as one of the requirements for attaining a Master's Degree at Makerere University, now a decade ago. For this study only a dummy dictionary of about 500 entries was envisaged. The MA dissertation itself (Nabirye 2008) was supposed to provide the meta-lexicographical background, with the actual compilation merely an illustration.

At the time, therefore, no dedicated dictionary compilation software was sought, as it was assumed that the 500 entries could simply be written out using a word processor. Also, the use of computers in general, and the Internet (and search engines) in particular, were in their infancy in Uganda. Using and especially possessing a computer was a true luxury at the time. Not much was known about software to compile dictionaries with, and truth be told, generic off-the-shelf dictionary compilation software was only starting to be produced internationally (cf. De Schryver 2011a). For all these reasons, the use of Microsoft Word was thought to be sufficient.

3. The Microsoft Word version of the WSG

The words defined were based on a minimal corpus which was developed for this project. A number of Lusoga texts were scanned and the small resulting corpus was used to generate a word list, from which the 500 entries were randomly selected.¹ Entries were organized according to the style guide developed for the MA study. Dictionary formatting was manually inserted in MS Word. The draft was rather basic with nothing fancy to warrant any further inquest into any other types of software. The pages were of a manageable extent and relatively easy to correct in the MS Word document. The timeframe for this

study was one year from September 2003 onwards. The first draft of the dummy dictionary was completed in about three months' time.

On submission of the draft, however, the supervisor argued that if the compilation process was stretched for a little while longer, a full dictionary could be realized by the end of the MA study. The supervisor's challenge was phrased as follows:

Where do you think we are going to find another Musoga who will enrol for studies in linguistics, be able to pay all the university fees, pass all the necessary exams, specialize in lexicography, and complete the draft you have presented? (Kiingi 2004, pers. comm.)²

The guidance from then onwards was tilted from compiling just a dummy dictionary to a complete dictionary. This is how the undertaking to compile a fully-fledged monolingual Lusoga dictionary came about.

From an academic point of view, this research was not only aimed at producing the first-ever monolingual dictionary to be compiled in Lusoga, but also to lay down the metalexicographical foundations for Lusoga lexicography. At the time Lusoga did not have any official status and its documentation was absolutely minimal. The compilation of the dictionary envisaged would therefore help to document and preserve the language's most basic lexical information, an effort that could serve future studies of Lusoga. It would also aim at producing a complete product that could be readily accessed by the Busoga community.

In the second phase of the study, therefore, the compilation moved from aiming at only 500 dictionary entries, to a tenfold, namely 5 000 entries. The new demands led to a serious expansion of the research. The increase in the number of entries also forced us to seek a more professional piece of software to compile the dictionary with.

4. From Microsoft Word to Shoebox

A colleague at the department — Celestino Oriikiriza — who was also trying to compile a monolingual dictionary, in his case for Runyankore-Rukiga, managed to find a program for the manipulation of textual data: Shoebox version 5.0. He grappled with applying it to his research, knowledge of which he later shared with us. We also acquired a copy of the Shoebox software and started to manually transfer the draft of WSG from the MS Word document into Shoebox. The formatting specifications on how each entry should be organized were based on the same style guide used in the first draft. Although the transfer took some time, the advantage of the change of programs was that, this time, dictionary formatting was inbuilt. Unlike in MS Word where all formatting aspects were applied and checked manually, in Shoebox most of this was automated. Editing of the database from then on was more manageable. Although Shoebox is widely used by field linguists, who typically build a dic-

tionary as they analyse and interlinearise texts, we found that it was not well adjusted to catering for Bantu language features and some dictionary entry information had to be forced in unrelated fields to maintain the order of the entry as stipulated in the style guide.

At the time, no monolingual dictionaries for any of the Ugandan languages existed. All the projects that later gave rise to monolingual dictionaries, such as the ones for Luganda and Runyankore-Rukiga, were also in their infancy (cf. Kiingi et al. 2007 and Oriikiriza 2007). Of all the monolingual dictionary projects only the Lusoga dictionary project was undertaken with the aim of attaining an academic degree. The demands placed on the study were therefore exceptional because the focus had to be on meeting the goals of a scientific study within a specified timeframe.

5. Visits to dictionary centres in Africa

In the subsequent years a number of research visits were arranged to dictionary centres in Africa, in the hope of procuring references on dictionary compilation and additional assistance on the use of corpora in lexicography and the use of dictionary writing systems. The dictionary centres visited were the Institute of Kiswahili Research (IKR, abbreviated as TUKI in Swahili) at the University of Dar es Salaam, the corpus and dictionary units at the University of Pretoria, the Zulu NLU in Durban, and finally a private consultation with A.C. Nkabinde (the doyen of monolingual Zulu lexicography, cf. e.g. Nkabinde 1982 and 1985) in Pietermaritzburg. Unfortunately, TUKI dictionaries were not based on corpora and no dedicated dictionary compilation software was in use in Dar es Salaam. TUKI had however produced monolingual dictionaries for Swahili and these dictionaries informed the research. Quite surprisingly, no corpus query software or dictionary writing systems were introduced to us by our host at the University of Pretoria either.³ Literature on Zulu and Shona lexicography was consulted at the library. Nothing forthcoming was found at the Zulu NLU either. The consultations with A.C. Nkabinde, on the other hand, were found to be very informative, especially for comparative Bantu lexicographical research. Exposure to other monolingual Bantu dictionaries richly informed the arguments raised in the MA itself (Nabirye 2008) as well as the final rendition of dictionary data in the WSG (Nabirye 2009). However, since none of the visits advanced the know-how regarding the corpus and dictionary writing systems beyond what the study had already secured, no further inquest was sought on these aspects. Instead, the WSG was labelled as a non-corpus-based dictionary and no further discussion or inquiry into any other use of software was undertaken.

At the end of the visits, the study had established a stand on how to deal with Lusoga lexicography based on the literature reviewed on the compilation of dictionaries in cognate Bantu languages. This was the basis for describing the findings arrived at in the MA study. Since the WSG was the first monolin-

gual dictionary of Lusoga, most of the data specified for the dictionary proper was new with the WSG as the only record for it.

6. Megastructure of the WSG

We also came to the conclusion that the WSG which was initially envisaged as a chapter within the MA dissertation had to be considered as an independent result, appended to the dissertation. A detailed description of how all the components in the compilation stage were arrived at and brought together served as the study itself. Issues dealt with in the MA study include the orthography used, with a justification of how the writing system in general was specified and how the Lusoga grammar in particular was addressed, with an indication of all the new grammatical terminology. The WSG was conceived as a general-purpose dictionary, meant for mother-tongue speakers, with at least a minimum of primary seven (P7) education. The language used for the recording of dictionary data had to be simplified to ease access to the dictionary for its intended audience. Summaries of the most essential Lusoga language information, specified for the very first time in WSG, were given independent consideration with special treatment in the dictionary. The explanation and justification of this data was given in Nabirye (2008).

Given that the WSG was now an independent entity, it had to appear with all the information required to make it fitting as a complete dictionary. Being the first and only existing monolingual Lusoga dictionary, the WSG anticipated a lot of demands from the target audience. The dictionary content was as a result conceived to cater for general user needs by specifying information such as (1) a summary of the history of Lusoga, (2) a list of all the main abbreviations in general use in Uganda (English and otherwise), together with their Lusoga interpretations, (3) a language portrait detailing the language information specified in the dictionary, (4) pictures to enhance the definitions at about one hundred entries, and (5) an onomasiological section with different categories of things, such as birds, musical instruments, transportation mediums, gardening tools, etc. Lastly, since the morphology of Lusoga was found to be a challenge after testing pilot versions of the WSG (Nabirye 2008: 130), information considered to be of an irregular or unpredictable nature had to be prepared as "a list of sight words", to smoothen dictionary access. A section called "How to use your dictionary" was required to introduce the contents of the dictionary. This section was prepared in Lusoga and translated into English.

7. Shoebox runs its course

All in all, there were as many as eleven different parts that had to be put together in the final draft of the dictionary. First there were the front and back cover pages (in colour, to be printed on heavier paper), followed by the title

and imprint pages (2 pp.) as second component. The "How to use your dictionary" section in Lusoga (22 pp.) constituted the rest of the front matter, while the English translation thereof (21 pp.) constituted the second part of the back matter. The first part of the back matter was reserved for the list of sight words (63 pp.). The sixth component was the actual A-to-Z section, to be interspersed with the five additional sections mentioned under Section 6. Those interspersed additional sections had to have logical placements near related entries in the dictionary. Information on the history of Busoga consisted of 7 pages and had to be placed near the entry *ebyafaayo* 'historical issues', abbreviations consisted of 14 pages and had to be placed near the entry (*e*)*kifunze* 'abbreviation', language information consisted of 15 pages and had to be placed near the entry *gulaama* 'grammar', while the picture plates consisting of 8 pages could have been placed near the entry (*e*)*kifaanani* 'picture', but this placement was very close to the abbreviation section with only one page in-between so another placement was decided on, namely (*o*)*ku.faanana* 'to look like'. Pictures intended for inclusion at *particular entries* failed to export properly into the Shoebox generated file, so that the attribute for pictures in Shoebox was abandoned. All entries that needed pictures were listed and a plan was envisaged to have them inserted manually.

The best that Shoebox could do at this point was to give a full dictionary copy of the A-to-Z section, exported as an MS Word document. Shoebox was also used to automatically generate a list of all the irregular and unpredictable entries to be appended to the dictionary. This is as far as Shoebox could go. Inserting the other parts of the dictionary required another type of software in which all parts would be brought together and numbered accordingly. The final product of the dictionary had already been specified in the MA dissertation and it had to appear that way in the appended copy. The desktop publishing software called InDesign was found fitting to put all the different parts of the dictionary into one document that would then constitute the WSG to be appended to the MA dissertation for submission.

8. From Shoebox to InDesign

Shoebox was relatively compatible with InDesign and most of the data could directly be imported into InDesign with few alterations.⁴ Most of the inbuilt dictionary formats from Shoebox were maintained by InDesign. Information in the two programs was easy to correlate during the importation exercise. Among the problems that arose, however, was that InDesign did not have an automated application for headers and footers. This meant that all the headers on each page of the dictionary were lost in the process. We also lost the page numbering and the automated formatting of the dictionary which existed in Shoebox. We were back to something similar to the status of the draft generated in MS Word because most of the changes in the dictionary from here on were to be effected manually.

After setting the A-to-Z section, we moved on to place all the additional dictionary parts in the desired positions manually. Since the information in all the additional sections was independent, it was straightforward to intersperse those without affecting the formatting in the A-to-Z dictionary section.

The most trying part was to add to the InDesign file the pictures on selected entries. Since we were inserting them on entries already defined, the formatting of the respective pages changed with each insertion. Some pictures, such as the one for the skeleton, needed a full page immediately after the entry on which it was entered, here (*e*)*igumba* 'bone'. Other pictures, such as the one for measurements entered on (*e*)*kipimo* and bicycle entered on (*e*)*gaali*, needed half a page after the respective entries. We treated each problem as it came and did not really know the program well enough to anticipate future problems. The main advantage of InDesign was that, once the placements were made, it could hold the pictures in place while we re-arranged the altered formats.

It is only after grappling with the placement of all the parts of the dictionary that we had the opportunity to analyse the entire dictionary for the very first time. This is when final editing of the dictionary was begun. The alphabetical formatting inherited from Shoebox kept the majority of entries in their correct placements; however placement of all new and edited entries from here on was effected manually. And, as mentioned, all the new information added in the dictionary from here on also lacked the inbuilt automation of Shoebox. Changes made in the final draft thus only existed in the InDesign file and could not easily be tracked or automatically applied to related entries. Cross-checking updated information was difficult and of course inconsistencies were introduced.

As far as the demands of the study were concerned, the WSG had been compiled as specified in the dissertation chapters and evidence of this was given in the draft that was appended. The draft dictionary at this stage was an independent part of the research which was arguably a major contribution to the documentation of Lusoga. No prior standard record of this nature existed which is why the process to have it published was pursued. The dictionary which was originally envisaged to have only 500 entries, had by now grown to a massive 12 700 entries, equivalent to 552 printed pages just for the main text of the A-to-Z section.⁵

9. Setting up Menha Publishers

When the MA study was completed and submitted for examination at the end of 2007, we began the search for a publisher. We approached three publishing companies in Uganda. Both Longman and Macmillan were simply not interested in publishing our dictionary. Our contact person at Longman did send us to the liaison office of Oxford University Press in Kampala. At that office, we were informed that whatever is published in Uganda must be authorized by the Head Office in Nairobi, Kenya.⁶

The struggles to compile the dictionary and complete the MA study had served one lesson, in that we believed one must just keep searching for a way to obtain whatever needs to be done in the best way possible. We therefore went to buy books on publishing to get a clear picture of what a publishing company actually does once it receives a manuscript. Two of the books in particular proved informative enough and gave excellent guidance on how to set up one's own publishing company. After reading those books, we simply stopped looking for a publisher and boldly decided to set up our own publishing company rather.

In September 2007, we filed for setting up a publishing company and prepared the proper legal and financial requirements. Once the process was about three quarters underway, we set out looking for editors. These efforts to have native speakers edit the dictionary were unfortunately fruitless (Nabirye 2008: 148). The few 'specialists' who were approached all disqualified themselves. At the time, we could not find other speakers of Lusoga who had the necessary expertise to undertake this task. This was (and is) because the majority of Basoga have never learned to read or write Lusoga. Moreover, canvassing all the data and cross-checking their consistency was clearly too much a task to be completed manually.

Having been a single-handed study and compilation effort, all humanly possible resources had already been drawn and drained to bring both the study and the compilation to completion. For the published version of the dictionary, therefore, we decided to only address the major anomalies overlooked in the version appended to the MA. We reasoned that just like computer programs are often released with bugs that are only patched in subsequent updates, so could a dictionary be released with future editions to take care of the errors. This, of course, is in line with a move from an academic environment to the trade: whereas one can try to attain perfection in the first (spending endless amounts of time and money, when available), it is rarely a goal in the second.

Menha Publishers (U) Ltd. finally began official operations in June 2008. The project which had continuously been under financial constraints was salvaged by the Chinese Embassy in Uganda which offered substantial funds to finance the printing. We started hiring people to carry out tasks like designing the company logo. We paid for all the services as a company but the money was not sufficient to for example engage a professional website designer. A relative offered to help do what he could to have the website hosted and all the company did was to pay for the web hosting fees.⁷

To actually print their works, publishing companies join hands with printing houses. The search for one eventually led to visits to Kolkata and Bangalore in India, as well as Cape Town in South Africa. We settled for the latter, e-mailed the material for publication as a single PDF in March 2009, from where a few months later the dictionaries themselves were shipped to Mombasa, and then put on a train to Kampala. A sample page of the printed version of the WSG is shown in Addendum 1.

10. From InDesign to TLex

Following the submission of the MA, conference papers were presented about different aspects of the study and the compilation process. At those conferences we were able to meet with and talk to various dictionary publishers, to hear how they actualize their projects. We also had a chance to listen to presentations on the different software solutions available for dictionary compilation and that is how we ended up talking to the developers of TLex.

The initial interest at the time was to start work on a second edition of the WSG, using a better dictionary writing system than Shoebox. The developers of TLex indulged us in the advantages of their software and since the problems that arose in the compilation stage were still ripe in our minds, it was easy to ask relevant questions based on our compilation experience.

A lot has been written about TLex already in the scientific lexicographical literature, so we will limit ourselves to a single reference. The most concise overview of the various features of TLex can be found in De Schryver (2011a), which contains all the references to earlier publications for the reader who is interested in specific details of the software.⁸

We were quickly convinced and elected to have our data transferred to TLex. This in the understanding that TLex is primarily a powerful dictionary database which takes care of the A-to-Z section(s) of a dictionary, and that most extra-matter material is the domain of desktop publishing software, where a dictionary file exported from TLex may be joined to the extra texts that have been prepared in still other programs, like word processors and the like.

Although the WSG was initially rather well organised with explicit mark-up labels preceding each field as long as it resided in Shoebox, the move to InDesign, and the further compilation therein, meant that the programmers could only import the InDesign data into TLex. The InDesign data is inherently "flat" with the only remaining structure the formatting. The developers of TLex have developed in-house finite-state importers, which are able to analyse such features and differences like bold vs. italics. vs. small caps etc. in running text, and also take punctuation (, vs. ; vs. . vs. : etc. as well as various types of brackets) into account, in order to recreate or even simply to create a text that is properly structured. In simple terms, as the data is being transferred to TLex, a DTD (i.e. document type definition) needs to be built, which regulates the dictionary grammar. Such custom importers never import everything perfectly, mainly because the source files are rarely 100% consistent, especially those that involved a lot of manual intervention (as was the case for the WSG).

As expected, a number of problems arose during the importation stage of the WSG. The first basically revolved around the language barrier, given that the entire text (being a monolingual Lusoga dictionary) was literally foreign to the programmers. As a result, there were cases where some parts of the article information ended up being misplaced. The solution here was to continuously liaise with the programmers during the importation exercise, to check the draft

imports for any anomalies and to alert the programmers in time. This close cooperation enabled us to actually make substantial improvements to the internal structure of the dictionary articles, and to help design a solid DTD.

Another problem that arose during the importation was that we had added a new letter to the Lusoga alphabet — a velar nasal, η , which is not commonly used. When the importation was carried out, the default sorting did not know where to place entries containing the velar nasal. The placements of entries with the letter η were therefore found in all sorts of illogical positions. Examples of such entries include *(a)kaku η junta*, *bbili η gania*, *daja*, *η ganzaiza*. In TLex various sorting methods may be used and customised, and here it sufficed to add the velar nasal η in-between n and o in the four-pass table-based sorting which is based on ISO 14651.

The most annoying problems were those which resulted from inconsistencies in the InDesign file, inconsistencies either inherited from the MS Word document, the Shoebox database or inserted in the InDesign file itself. Solving those was a trying job for both us and the programmers, but of course a much better dictionary database was the result. For example, the derivation category and the consideration of unpredictable plural forms were two of the last-minute additions to the dictionary entry parts whose proper placements were problematic because they were not given in the information on parts of the entry in Shoebox.

One of the more interesting "clean-ups" was that TLex forced us to make a clear separation between actual dictionary contents (which are unique to each dictionary article) and all metalanguage (such as part of speech assignments or cross-reference texts, which are repetitive). All the metalanguage became part of the Style System, so easily changeable at any stage without the need to actually touch the dictionary contents. In this context, conditional metalanguage was also introduced on various levels. For example, to introduce run-ons, TLex will now automatically precede a single run-on with the meta-text "*bgz:*", but multiple run-ons with the meta-text "*bbgz:*".

In a first phase, the data in the TLex file was meant to mimic the layout seen in the InDesign file as much as possible, down to the fonts and abbreviations used, as that is what we were familiar with having worked with the data for so many years. Needless to say, all errors spotted during the conversion were of course corrected, so the data in the TLex file is now "the latest version". A screenshot of the imported WSG data into TLex is shown in Addendum 2.

11. The online version of the WSG

With dictionary data in TLex, it has always been trivial to export the material to any of the commonly-used formats, with the aim to produce a paper dictionary, an online dictionary, any other type of electronic dictionary, or even to reuse any (parts of the) data in another application. The current version of TLex (7.1), for example, has all of the following data export options:

- Comma Separated Values
- HTML (Web Page)
- RTF (Rich Text Format) (MS Word, LibreOffice et al.)
- XML (Data / Structured)
- XML (Formatted / Publisher-friendly)
- TLex Online Publishing [Advanced]
- Text
- Lemma signs
- Index
- ODBC database

In order to place a dictionary on a website, thus as an online dictionary in searchable format, one would typically choose the "TLex Online Publishing" option. Although doing so and preparing the website only takes a good day's work — provided one has a domain name and web space already, with database software installed on the server — it took us another two years before we took this step. The reasons for waiting so long are many, but basically we first wished to give the sales of the printed copies a chance, yet when seeing that those were not doing that great, we reversed the argument, now assuming that the free online version would help the sales of the paper copies.

The exact same contents from the TLex file were eventually placed online in June 2012. These are thus the contents from the WSG without the cross-referenced material (i.e. without (1) to (5) mentioned in Section 6). A screenshot of the online version of the dictionary, baptized *e-Eiwanika ly'Olusoga*, is shown in Addendum 3.

Given that there are far fewer space constraints on the Internet, the textual condensation may be lessened, by for instance starting each sense on a new line. Also, and in contrast to printing, colour may be used royally in an online environment, which helps to quickly navigate dictionary articles. Both of these were implemented for the *e-Eiwanika ly'Olusoga*. To further improve the usefulness of the dictionary, the symbols shown in Table 1 were also introduced.

Table 1: "Quick-help" for the online *e-Eiwanika ly'Olusoga*.

Symbol	Functionality	Example
&	AND	itaka & eitaka
	OR	itaka eitaka
"..."	Exact phrase	"kulondoola ensonga"
_	Single-character wildcard	_taka
%	Multi-character wildcard	%soga
/(1-9)	Within x words of one another, given order	"nga ni"/2
@(1-9)	Within x words of one another, any order	"ekigobeelwa ensonga"@4
#	XOR (find one or the other, but not both)	ekigobeelwa # ensonga
^	None of ...	^ekigobeelwa ensonga

With the symbols seen in Table 1, the user has been handed an extremely powerful search tool with which a dictionary may be searched in a manner unlike anything available before. One of the early "electronic dreams" has been implemented (cf. De Schryver 2003), and in this respect the *e-Eiwanika ly'Olusoga* may very well be the most advanced online dictionary currently available — not just for the Bantu languages, but for any language.⁹

12. The offline version of the WSG

The makers of TLex also have a module called the "TLex Electronic Dictionary System", with which downloadable e-dictionaries may be produced for offline use on a computer. The same contents can also be burned to a disc (CD-ROM, DVD, etc.) or written to a USB flash drive. Assuming that there would be a market for this type of dictionary as well, such an e-version of the WSG was also prepared, a screenshot of which is shown in Addendum 4.

In this offline version of the *e-Eiwanika ly'Olusoga*, the search options include:

- Match any of the given search terms
- Match all of the given search terms
- Case sensitive
- Match whole words only
- Search the full text

This e-dictionary further allows for automated Web searches, either for text or images, and has an MS Word plugin, which can automatically display, in a pop-up window, the dictionary contents of the words one is typing in or clicking on.

13. Marketing of the WSG

The three major bookshops in Uganda — Uganda Bookshop, Aristock Booklex and Makerere University Bookshop — are also the only true bookshops in the whole of Uganda, other books being sold from supermarkets. All three are located in Kampala. For each, the normal procedure for publishers is to approach them with copies of a book, with payments only forthcoming once (and months after) the books deposited have been sold. Since the end of 2009, payments for three consignments from Uganda Bookshop, two from Makerere University Bookshop, and just recently the first from Aristock Booklex were received. In addition, a few hardcopies were acquired by once-off customers directly from our warehouse in Kampala or through our website. All of this amounts to about 200 copies sold so far, which is less than 10% of the print-run. The great majority of the sales were local. There may have been a small uptake in the interest in the dictionary following the official launch of the WSG by

Uganda's President, on the 8th of October 2010, when about five articles in the main Ugandan newspapers also reported on the publication (cf. e.g. Jaramogi 2010, De Schryver 2011).

The efforts to promote the electronic versions of the dictionary were as follows: We sent several messages to the two main mailing lists that unite the electronically connected Basoga, i.e. BuSoga Yaife and Busoga Bulletin. Members of these mailing lists mostly constitute Basoga in the diaspora, whom we assumed to be our main target audience for an electronic product dealing with their language and culture, in their language. We also wrote targeted e-mails to various other interest groups and individuals to announce the release of the *e-Eiwanika ly'Olusoga*. At the same time, the company website of Menha Publishers was updated to include detailed information on both the paper and electronic versions of the dictionary. The online version of the *e-Eiwanika ly'Olusoga* is freely accessible from our website, while a fully automated system takes care of the purchases of the downloadable version. The electronic version of the dictionary was envisaged to serve a wider market of users who could order and pay online from any part of the world. The offline dictionary can even be downloaded as a trial version first.

In spite of all these marketing efforts — which, lest it be forgotten, come on top of nearly a decade of detailed research, painstaking dictionary compilation and inventive fund-raising running into the tens of thousands of Euros —, during the first fifteen months of *e-Eiwanika ly'Olusoga* being available, exactly two copies of the downloadable version were sold: one to a private user in Uganda, and one to a library in the US.

14. Actual use of the WSG

No studies have so far been undertaken of the actual use of the hardcopy version of the monolingual Lusoga dictionary. It may or may not be used successfully, it may or may not fill a lookup need — we simply don't know. With regard to the online version, however, we are in a position to look at how this product is used, given that we can study the log files attached to this Internet dictionary. Sadly, the findings are unsatisfactory.

For the first fifteen months that the dictionary has been online so far, just over 2 000 searches were made by about 1 000 different users. As a comparison, over the years, the *Online Swahili-English Dictionary* has attracted approximately 1 000 visitors a day, who perform about 2 000 searches every four and a half hours! As may be seen from Figure 1, the distribution of the number of searches per user in the *e-Eiwanika ly'Olusoga* is Zipfian, so most users actually only look up a single word and leave. There are a few return visitors, such as # 270, who looked up 59 words over a period of nearly 40 days, or # 62 who looked up 10 words over a period of nearly 9 days. Studying the actual searches for particular users (cf. De Schryver and Joffe 2004: 192) is not interesting, given the very large number of immaterial searches.

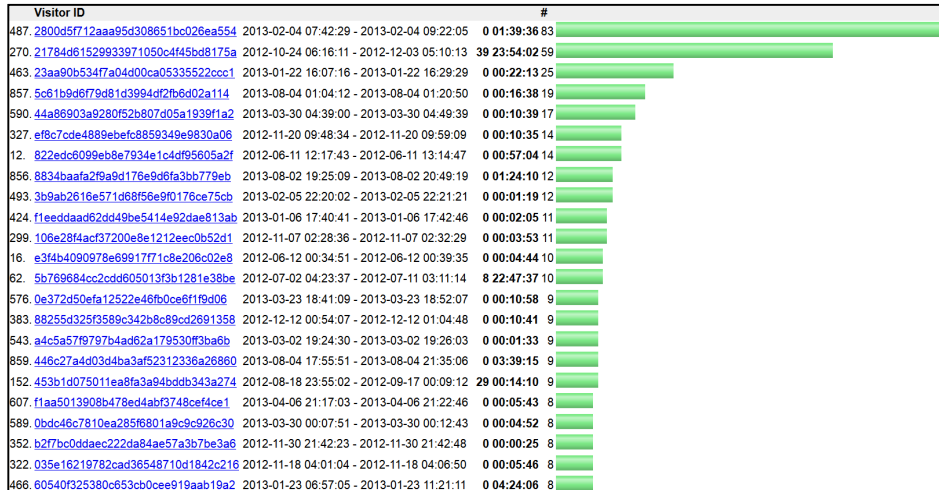


Figure 1: Number of searches per user of the *e-Eiwanika ly'Olusoga*.

Both the number of searches and the number of visitors has remained stable since the launch of the *e-Eiwanika ly'Olusoga*. This may be deduced from Figure 2, where the monthly number of searches and users are plotted.

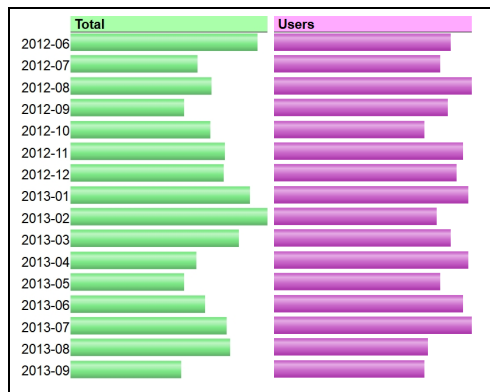


Figure 2: Monthly number of searches and users of the *e-Eiwanika ly'Olusoga*.

Most searches are immaterial because non-Lusoga words are being looked up in a monolingual Lusoga dictionary, and when Lusoga words are being looked up, they mostly belong to a limited number of registers. Just one quarter (25.02%) of the searches result in a "hit", meaning that the word or one of the words being looked up is/are found at least once in the full dictionary text. A massive three-quarter (74.98%) of the searches result in a "miss". The top-frequent "hits" and "misses" are reproduced in Tables 2 and 3 respectively.

Table 2: Most frequent "hits" in the *e-Eiwanika ly'Olusoga*.

Search	Freq.	Search	Freq.	Search	Freq.	Search	Freq.
hello	11	bi wanindi	4	bugisu	2	nze	2
		wan pod ...					
kiswahili	10	olusoga	4	akasolo	2	embooli	2
itaka	10	kuma	3	embolo	2	amaloboozi	2
eiwanika	8	omunie	3	omukyala	2	itaka & eitaka	2
tomba	8	diamond	3	kalenda	2	katonda	2
go	7	microscope	3	bye	2	okutomba	2
father	7	k	3	me	2	be healthy	2
i love you	6	school	3	wanzi	2	kuba	2
a	6	house	3	ighe	2	nkutu#	2
baba	6	_	3	bugiri	2	mudindo	2
mama	5	doctor	3	mapenzi	2	muna	2
eitaka	5	%	3	baaba	2	ensonga	2
taka	5	o	3	embwa	2	okwenda	2
boy	5	omudindo	3	mkeka	2	jambo	2
lusoga	4	se	2	okuluma	2	kale	2
ekinazi	4	amadhi	2	iganga	2	bantu	2
car	4	emmana	2	e	2	tai	2
enfuli	4	ekinhazi	2	me too	2	(hapaxes to follow)	

Table 3: Most frequent "misses" in the *e-Eiwanika ly'Olusoga*.

Search	Freq.	Search	Freq.	Search	Freq.	Search	Freq.
man	20	happy	5	together	3	bird	3
love	20	thank you	5	ly	3	omunege	3
water	12	home	5	look for	3	and	3
emana	11	spirit	5	blood	3	old	3
dog	11	cow	4	person	3	sitari	3
one	10	nyako	4	mana	3	tarihi	3
lion	8	god	4	how are you	3	moon	3
come	8	apple	4	omunye	3	genius	3
king	8	see	4	devil	3	tree	3
woman	8	atom	4	sun	3	yes	3
good	7	life	4	death	3	eat	3
morning							
fuck	7	child	4	install	3	my love	3
book	7	stone	4	girl	3	hand	3
mother	7	unite	4	stabalaize	3	sleep	3
table	6	fire	4	want	3	kodeyo	3
food	6	cat	4	building	3	hate	3
vagina	5	kikokotoo	4	snake	3	(searches with freq. 2 and hapaxes to follow)	
sex	5	omusadha	3	kuudhi	3		

A study of the most frequent "hits" in Table 2 reveals that 20% of the found material is actually English (typically mentioned in the etymology slots), that several of the Lusoga words are simply the words and symbols taken from the instructions to the dictionary (cf. Table 1, e.g. *itaka*, *eitaka*, *ensonga*, *taka*, *_*, *%*, ...), and that way too many of the other searches are F-words on the one hand: *tomba* 'fuck', *ekinazi* 'vagina', *enfuli* 'labia minora', *omunie* 'anus', *omudindo* 'anus', *emmana* 'vagina', *ekinhazi* 'vagina', *akasolo* 'penis', *embolo* 'penis', *okutomba* 'to fuck', *mudindo* 'anus', or basic vocabulary on the other: *baba* 'father', *mama* 'mother', *kuma* 'light', *omukyala* 'woman', *amadhi* 'water', ... Genuine searches include the words in the title of the dictionary, and words like *embooli* 'sweet potatoes', *amaloboozi* 'voices', ...

These were the hits; the picture for the misses is even more depressing. As many as 85% of the misses in Table 3 are simply English words, several of them again from the F-field: *fuck*, *vagina*, *sex*, or baby words: *man*, *love*, *water*, *dog*, *one*, *lion*, ... The few Lusoga misses include more (misspelled) F-words: *emana* 'vagina', *mana* 'vagina', *omunye* 'anus', *omunege* 'penis', misspellings of basic words: *omusadha* 'man', *kodeyo* 'hello', ... and foreign words: *nyako*, *sitiari*, *tarihi*, ...

Clearly, then, the use of Internet dictionaries remains biased towards prurient content and some high-frequency words (cf. De Schryver and Joffe 2004: 190). The type of words being looked up, as seen from both the hits and the misses, moreover indicates that the *e-Eiwanika ly'Olusoga* cannot be said to be used for any serious purposes. If ever there was a noble use for the expression cast pearls before swine, then this is it. This project is not only the adaptation of an academic study being fully misused by the community, it is also philanthropy gone very wrong.

15. What we can learn from all this

Wearing an academic hat, it is possible to explain away quite a number of the depressing findings. Some of the arguments could then go as follows. If we compare Lusoga to the neighbouring Luganda, for instance, one can state that Luganda has a longer tradition as a written language, dating back to at least a century ago (Meeuwis 1999). It has been a medium of instruction in Uganda for about half that time (Ladefoged et al. 1972: 87-99). To this date, Luganda is the language of the church and the media, both in Buganda and Busoga. When the monolingual Luganda dictionary was published in 2007 (Kiingi et al. 2007) all copies were sold within a year and they had to reprint soon after. For Lusoga, in contrast, a language that only received its first official recognition in Uganda and Busoga in 2005 (NCDC 2006: 5), it is still too early for a monolingual Lusoga dictionary to attract enough attention.

Also, Lusoga is not yet stable as a written language. One could hypothesize that most users will find it problematic to decide on the right spelling of the lemmas to be looked up, and after a few trials they may give up. That doesn't necessarily mean that such users do not want a monolingual Lusoga

dictionary per se; failure to figure out how the words of interest are written and listed in the dictionary simply drives away such potential users. Comparing Table 2 with Table 3 — where one notices that the same type of words and even the same words — are searched for in both correct and wrong spellings, actually gives weight to this argument.

Because Lusoga is only just beginning to have a presence in the written genre and in scholarly works, the majority of the academic papers written so far have been on problems that could help advance the description of Lusoga. Very few reference works exist on Lusoga, and fewer even have been written *in* Lusoga, which implies that the interest and need to use Lusoga in an advanced setting or in a way which requires one to check the proper form or the exact meaning(s) of a word in a dictionary, has not yet arisen.

Lastly, the WSG project was started and developed as an academic study. It was therefore designed and aimed to fulfil scholarly demands, not market-oriented demands. The need to market the dictionary arose after the project was passed by the academic bodies and therefore the way it is taken to the market and presented to this very niche market needs to be adjusted if it is to receive the attention we think it deserves.

Conversely, and wearing a business hat, one simply has to admit, based on the evidence seen in Tables 2 and 3, that what the Lusoga community needs first is a bilingual English–Lusoga dictionary. At a push, one could wish to conclude that they need a bilingualised dictionary, thus a monolingual Lusoga dictionary where English glosses are provided at each sense of each dictionary article.

Additionally, the material could have been made far more user-friendly in a digital environment. For one, the entire metalanguage could easily have been expanded: writing the parts of speech in full rather than use the current obscure abbreviations, or "in Luganda" rather than "Lg.", "example" rather than "gez.", and even "this word is a singular noun in class 7, with its plural in class 8" rather than "7/8", etc. One could also have decided to do away with orthographic conventions in the pronunciation field, such as those that regulate the compensatory lengthening of vowels. Using full words throughout rather than morphemes, could also have been considered. And so on.

Yet deep down the actual tension is actually one between a product that is needed to make a society dictionary, versus a product that is needed to make money, and must, by definition, be sellable and thus user-friendly. Monolingual dictionaries in a non-dictionary environment must therefore be facilitated by a *deus ex machina*. Even then, the battle remains an uphill one. The heavily government-funded and over-trained Northern Sotho NLU, for instance, has had their monolingual dictionary online for a number of years now, known as the *Pukuntšutlhaloši ya Sesotho sa Leboa ka Inthanete*. Several teams of lexicographers worked on the dictionary for well over a decade, a dictionary which potentially serves a community of over 4.6 million speakers in digitally advanced and well-connected South Africa. For the past 15 months,

about 1 400 visitors made use of this online monolingual Northern Sotho dictionary, searching for roughly 6 300 items, with a hit rate of 35%. While these figures are all higher than those for the monolingual Lusoga dictionary, the difference is clearly not as big as one would have hoped. Therefore, even though there may be little need for it in the present communities, monolingual dictionaries ought to be funded and the process guided by competent government bodies. Bringing it back to South Africa, the NLUs simply *must* focus on the production of monolingual dictionaries, as no one else will.

Endnotes

1. The first author would like to thank Brian Mugabi who helped scan the Lusoga texts, back in 2003, which served as the basis for the corpus of the MA study.
2. The first author would like to acknowledge the help and support of her supervisor, Dr. K.B. Kiingi, who ensured that both a worthy MA and a fully-fledged monolingual dictionary of Lusoga were eventually produced.
3. Ironically, back in October 2003 already, the first author of this article was in e-mail contact with the second author — then at the University of Pretoria. Both WordSmith Tools (for corpus querying) and TshwaneLex (for dictionary compilation) were discussed. The first author deemed both software programs too advanced or otherwise not suitable at the time. When the first author was at the University of Pretoria from August to November 2005, the second author had just left — about to relocate and be affiliated to the University of the Western Cape for a number of years. Both authors finally met in person at the Afrilex 2008 conference in Stellenbosch (and got married a year later). WordSmith Tools and TLex were taken up soon after, for all future work on Lusoga (cf. e.g. De Schryver and Nabirye 2010).
4. The first author would like to thank Hassan Wasswa Matovu who helped import the dictionary draft into InDesign and who was also responsible for the final dictionary typesetting.
5. The total number of printed pages is 704 (= 2 + 22 for the front matter, + 552 for the A-to-Z text, + 7 + 14 + 15 + 8 for the interspersed sections, + 63 + 21 for the back matter), which is exactly 22 quires of 32 pages each, the standard in bookbinding. Trying to fit one's contents into an exact multiple of 32 pages is always the cheapest option for printing. Some of the data that had been prepared was deleted to attain this multiple.
6. At the time, this very much felt like we were just being sent away. When, in early 2009, we checked with Oxford University Press Southern Africa, however, they also felt they could not take on a dictionary like ours. That said, OUPSA did help us find an excellent printing house in Cape Town.
7. When Menha Publishers worked on their next book, a Festschrift for Patrick Hanks (De Schryver 2010), new moneys were invested into the company and the website was updated.
8. All of these publications are also available from the company website of TshwaneDJe Human Language Technology, see the References for the URL.
9. The number of results shown per search has been limited to 5, however, this to make sure that the online dictionary contents cannot just be "stolen" in one go.

References

- Busoga Bulletin*. Available online at: <http://groups.yahoo.com/neo/groups/Busoga-Bulletin/info> [last accessed on 30 September 2013].
- BuSoga Yaife*. Available online at: <http://groups.yahoo.com/neo/groups/Busogayaife/info> [last accessed on 30 September 2013].
- De Schryver, G.-M.** 2003. Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography* 16(2): 143-199.
- De Schryver, G.-M. (Ed.)**. 2010. *A Way with Words: Recent Advances in Lexical Theory and Analysis. A Festschrift for Patrick Hanks*. Kampala: Menha Publishers.
- De Schryver, G.-M.** 2011. *AFRILEX President's Annual Report: 2010-2011*. Presented at the AGM of the AFRILEX 2011 Conference, 6 July 2011. Available online at: http://afrilex.africanlanguages.com/presreport1011_text.pdf (text), http://afrilex.africanlanguages.com/presreport1011_slides.pdf (photos) [last accessed on 30 September 2013].
- De Schryver, G.-M.** 2011a. Why Opting for a Dedicated, Professional, Off-the-shelf Dictionary Writing System Matters. Akasu, K. and S. Uchida (Eds.). 2011. *ASIALEX 2011 Proceedings. LEXICOGRAPHY: Theoretical and Practical Perspectives. Papers Submitted to the Seventh ASIALEX Biennial International Conference, Kyoto Terrsa, Kyoto, Japan, August 22-24, 2011: 647-656*. Kyoto: The Asian Association for Lexicography.
- De Schryver, G.-M. and D. Joffe.** 2004. On How Electronic Dictionaries are Really Used. Williams, G. and S. Vessier (Eds.). 2004. *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004, Lorient, France, July 6-10, 2004: 187-196*. Lorient: Faculté des Lettres et des Sciences Humaines, Université de Bretagne Sud.
- De Schryver, G.-M. and M. Nabirye.** 2010. A Quantitative Analysis of the Morphology, Morphophonology and Semantic Import of the Lusoga Noun. *Africana Linguistica* 16: 97-153.
- e-Eiwanika ly'Olusoga*. Available online at: <http://menhapublishers.com/dictionary/> [last accessed 30 September 2013].
- Iison, R.** 2012. IJL: The First Ten Years — and Beyond. *International Journal of Lexicography* 25(4): 381-385.
- InDesign*. Available online at: <http://www.adobe.com/products/indesign.html> [last accessed on 30 September 2013].
- Jaramogi, P.** 2010. *President Museveni Launches Lusoga Dictionary*. Available online at: <http://www.mediacentre.go.ug/details.php?catId=4&item=1036> [last accessed on 30 September 2013].
- Kiingi, K.B. (Chief Compiler) et al.** 2007. *Enkuluze y'Oluganda ey'e Makerere*. Kampala: Fountain Publishers.
- Ladefoged, P., R. Glick and C. Criper (with an introduction by C.H. Prator and additional material by L. Walusimbi).** 1972. *Language in Uganda*. Ford Foundation Language Surveys 1. London: Oxford University Press.
- Meeuwis, M.** 1999. The White Fathers and Luganda: To the Origins of French Missionary Linguistics in the Lake Victoria Region. *Annales Aequatoria* 20: 413-443.
- Menha Publishers*. Available online at: <http://menhapublishers.com/> [last accessed 30 September 2013].
- Microsoft Word*. Available online at: <http://office.microsoft.com/en-gb/word/> [last accessed 30 September 2013].

- Nabirye, M.** 2008. *Compilation of the Monolingual Lusoga Dictionary*. Unpublished M.A. Dissertation. Kampala: Makerere University.
- Nabirye, M.** 2009. *Eiwanika ly'Olusoga. Eiwanika ly'aboogezi b'Olusoga n'abo abenda okwegga Olusoga*. Kampala: Menha Publishers.
- NCDC. 2006. *THEMA. The Newsletter of the Thematic Primary Curriculum. Issue 1. August 2006*. Kampala: National Curriculum Development Centre.
- Nkabinde, A.C.** 1982. *Isichazamazwi 1*. Pietermaritzburg: Shuter and Shooter.
- Nkabinde, A.C.** 1985. *Isichazamazwi 2*. Cape Town: Oxford University Press Southern Africa.
- Online Swahili-English Dictionary*. Available online at: <http://africanlanguages.com/swahili/> [last accessed 30 September 2013].
- Oriikiriza, C.** 2007. *Kashoorozi y'Orunyankore-Rukiga*. Kampala: Fountain Publishers.
- Pukuntšutlhaloši ya Sesotho sa Leboa ka Inthanete*. Available online at: <http://africanlanguages.com/psl/> [last accessed 30 September 2013].
- Shoebox*. Available online at: <http://www-01.sil.org/computing/shoebox/> [last accessed 30 September 2013].
- TLex (aka TshwaneLex)*. Available online at: <http://tshwanedje.com/tshwanelex/> [last accessed 30 September 2013].
- TshwaneDje Human Language Technology — Publications*. Available online at: <http://tshwanedje.com/publications/> [last accessed 30 September 2013].
- UBS.** 2006. *The 2002 Uganda Population and Housing Census, Analytical Report, Population Composition*. Kampala: Uganda Bureau of Statistics.
- WordSmith Tools*. Available online at: <http://www.lexically.net/wordsmith/> [last accessed 30 September 2013].

Addendum 1: Sample page from the *Eiwanika ly'Olusoga* (WSG, Nabirye 2009: 238)

(o)ku.bunil.a

gez: *Omunaku abunga mu nsi muno!*
3) Okulambuula ebifo kamaala. gez: *Oyo abunze ensi eno okugimalaku.*
bbgz: Okubungila, Okubunza.

(o)ku.bunil.a [(o) kúbúnílá] kt. [T] [-buniile] [mbuniile] **bl: [Lg: okubunila]** 1) Okubwika omunhwa. gez: *Silika ela obunile mangu.* 2) (stl, ygt) Okusilika. gez: *Abaali baawakania baabawa sente baabunila.*

bbgz: Okubuniza.

(o)ku.butik.a [(o) kúbutíká] kt. [L] [-butiike] [mbutiike] **bl: [Lg: okubutika]** Okuteleka ekintu ng'emmele mu kanhwa yaalamu ekiseela nga togin-haanha oba kugimila. gez: *Abutiike embafu mu matama.*

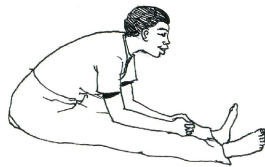
bbgz: Okubutikila, Okubutisa, Okwebutika.

(o)ku.butuk.a [(o) kúbutúká] kt. [T] [-butwike] [mbutwike] **bl: [Lg: okubutuka]** Okufuna obusundosundo ku mubili. gez: *Oluwusu lumbutwike twonal-wona.*

bbgz: Okubutukila, Okubutusa.

(o)ku.butul.a [(o) kúbutúlá] kt. [L] [-butwile] [mbutwile] **bl: [Lg: okubutula]** Okuleetela okufuna obusundosundo ku luwusu. gez: *Ebizigo ebyo bimbutula.*
bbgz: Okubutulibwa, Okubutulila.

(o)ku.buudhaal.a [(o) kúbúúdháálá] **tbk: (o)kú.búndáál.á** kt. [T] [-buudhaile] [mbuudhaile] **bl: [Lg: okubundaala]** 1) Okutyama weesiila ku magulu nga weefuniemu. 2) Okusilikilila oba okuwuubaala. gez: *Oidhie waatukya-mula tubaile tubuudhaaliile waka twenka.*
bbgz: Okubuudhaalila, Okubuudhaaza.



(o)ku.buusabuus.a

(o)ku.buuguuk.a [(o) kúbúúgúúká] kt. [T] [-bugwike] [mbuugwike]

1	Okuwulila
2	Okubugutana
3	Okuyiisa
4	Okutumbuka
5	Okwongela
6	Okuwaba
7	Okudobana

1) Okuwulila okwokelela. gez: *Ku mwoyo kuli kumbuuguuka.* 2) Okubugutana. 3) Okuyiisa einho ekintu ky'obaile ofumba. 4) Okutumbuka. gez: *Omulilo gubuugwike omulundi mulala gwagama ensiisila dhaakoleela.* 5) Okweyongela. 6) Okuwaba. 7) Okudobana.

bbgz: Okubuuguukila, Okubuuguusa,

(o)ku.buulil.a [(o) kúbúúlíílá] kt. [L] [-buuliile] [mbuliile] **bl: [Lg: okubuulila]** Okusomesa eidiini. gez: *Leelo babuuliliile ku kusonhiwagana.*
bbgz: Okubuulilwa, Okubuulilila, Okubuuliza, Okwebuulila.

(o)ku.buulilil.a [(o) kúbúúlílíílá] kt. [L] [-buuliliile] [mbuliliile] **bl: [Lg: okubuulilila]** 1) Okukobela omuwala ku by'obufumbo n'empisa ng'agya kufumbilwa. gez: *Basenga bo balina okukubuulilila ng'ogya kufumbilwa.* 2) Okuwabula omuntu abaile akezo enso-bi aleke kugilamu. 3) Okusomesa empisa.

bbgz: Okubuulililwa, Okubuulilika, Okubuuliliza, Okwebuulilila.

(o)ku.buuliliza.a [(o) kúbúúlílíízá] kt. [L] [-buuliliiza] [mbuliliiza] **bl: [Lg: okubuuliliza]** Okunoonheleza ku musango. gez: *Tulina okumaliliza okubuuliliza me tulyoke tuwawabe omusango.*

(o)ku.buusabuus.a [(o) kúbúúsábúúsá] kt. [L] [-buusabuusiiza] [mbuusabuusiiza] **bl: [Lg: okubuus-**

Addendum 2: Screenshot of the *Eiwanika ly'Olusoga* A-to-Z data in TLex
 (© Minah Nabirye 2003-2010)

The screenshot displays the TLex software interface for the entry 'Okubunila'. The main window shows the following text:

(o)ku.bunila [(o) kúbúnlá] *kt.[F] [-bunilé] [mbunilé] bi:* [Lg: okubunila] 1) Okubwika omunhwa. *gez: Siiika ela obunile mangu.*
 2) (stl, ygl) Okusiika. *gez: Abaali baavkama baabawa sente baabunila.*
bgz: Okubuniza.

>> Article is cross-referenced from

-buniile *weile. Bona: (o)ku.bunila*
mbunilile [mbúniilé] *kt. Bona: (o)ku.bunila*

(o)ku.butik.a [(o) kúbutíká] *kt.[L] [-butiike] [mbutiike] bi:* [Lg: okubutika] Okuteleka ekintu ng'emmele mu kanhwa yaamalamu ekiseela nga toginhaanha oba kugimila. *gez: Abutiike embafu mu matama.*
bbgz: Okubutikila, Okubutisa, Okwebutika.

(o)ku.butuk.a [(o) kúbutúká] *kt.[F] [-butwike] [mbutwike] bi:* [Lg: okubutuka] Okufuna obusundosundo ku mubili. *gez: Oltavusu lumbutwike lwonahwona.*
bbgz: Okubutukila, Okubutusa.

(o)ku.butul.a [(o) kúbutúlá] *kt.[L] [-butwile] [mbutwile] bi:* [Lg: okubutula] Okuleetela okufuna obusundosundo ku luwusu. *gez: Ebizigo ebyo bimbutula.*
bbgz: Okubutubwa, Okubututula.

(o)ku.buudhaal.a [(o) kúbútháálá] *kt.[F] tbk: (o) kú.búndáál.á [-buudhale] [mbuudhale] bi:* [Lg: okubundaala] 1) Okutyama weesila ku magulu nga weefuniemu. 2) Okusiikilila oba

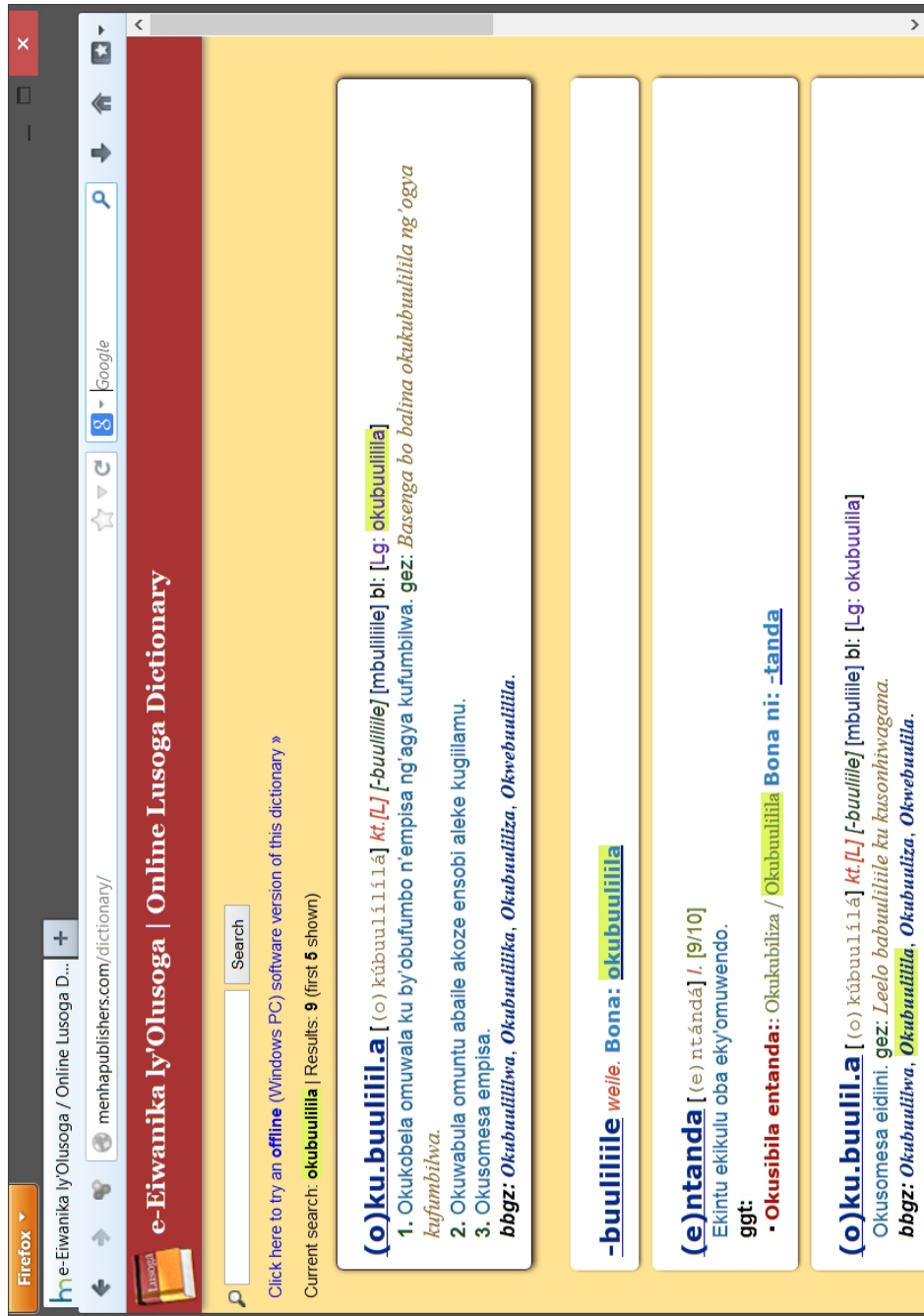
The interface also shows a list of related forms on the left, a detailed definition and example on the right, and a morphological table at the bottom.

Lemma: (o)ku.bunila. **LemmaSign:** (o)ku.bunila.a, PhoneticF: [Lusoga]

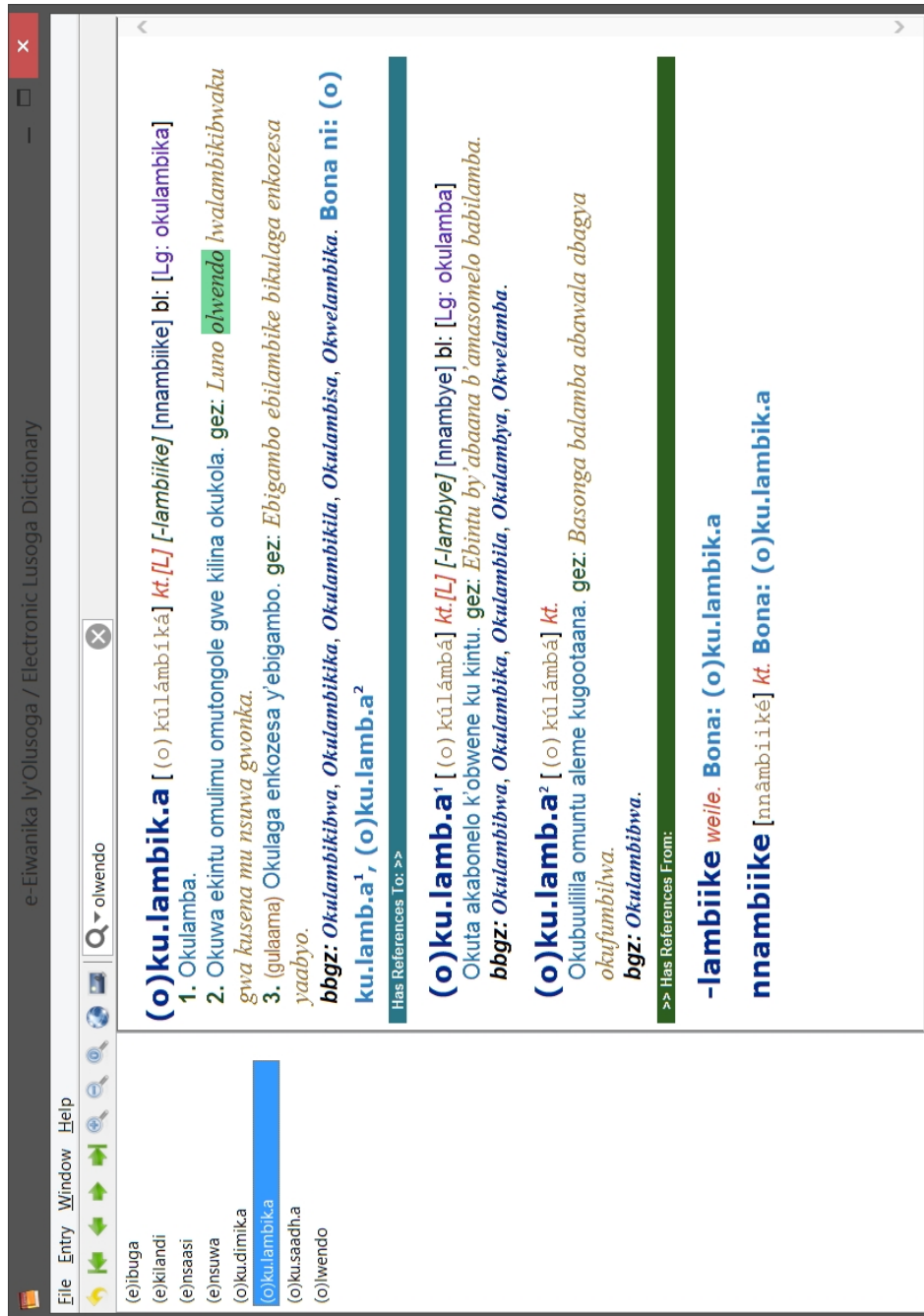
CompoundF: CompoundF=mbunilile
CognateLuganda: CognateLuganda=okubunila
Sense: 1 SenseNumber=1
 Definition: Okubwika omunhwa.
 Example: Siiika ela obunile mangu.
Sense: 2 SenseNumber=2, Label=stl ygl
 Definition: Okusiika.
 Example: Abaali baavkama baabawa sente
 Run-On: Run-On=Okubuniza

Attributes (F1)	Attributes (F2)	Sense (2): Label
Lemma: Label	Sense (1): Label	
bkl	bkl	nglwz
dhl	dhl	nhlz
dib	dib	saw
gulaama	gulaama	sbz
gwil	gwil	stl
ibul	ibul	tgl
idin	idin	totm
kaile	kaile	ttgl
kiina	kiina	vum
kyam	kyam	wl
Lgbl	Lgbl	waana
mat	mat	wk
mnhgn	mnhgn	yaaka
nglwz	nglwz	yaaye
nhlz	nhlz	ygl

Addendum 3: Screenshot of the online *e-Eiwanika ly'Olusoga* (© Minah Nabirye 2003-2012)



Addendum 4: Screenshot of the offline *e-Eiwanika ly'Olusoga* (© Minah Nabirye 2003-2012)



Development in Lexicography: From Polyfunctional to Monofunctional Accounting Dictionaries

Sandro Nielsen, *Centre for Lexicography, Aarhus University,
Aarhus, Denmark and International Centre for Lexicography,
Universidad de Valladolid, Valladolid, Spain*
(sn@ash.dk)

and

Pedro A. Fuertes-Olivera, *International Centre for Lexicography,
Universidad de Valladolid, Valladolid, Spain*
(pedro@emp.uva.es)

Abstract: This article describes the theoretical foundation of the accounting dictionaries as well as its practical results. Furthermore, the implementation of the project shows how the constant interaction of lexicographical theory with practical dictionary work over a period of 10 years has led to lexicographical development and adaptation resulting in ongoing adjustments to the dictionaries and their theoretical foundation. This is exemplified by studying the transition from polyfunctional to monofunctional dictionaries in an attempt to provide help in several types of usage situations, typically in communicative situations, e.g. when reading, writing or translating English, Danish or Spanish accounting texts, or cognitive situations, i.e. when users want to know more about accounting matters or accounting language. The article also indicates that the creation of this project connects e-lexicography with the knowledge-based economy. Lexicographers work in the cloud being physically thousands of kilometres away; they work with experts in accounting, databases and the Internet with the aim of constructing and updating a high-quality tool with relatively low information costs for users. Finally, this article shows that the future of lexicography rests on designing, constructing, and updating information tools that take into consideration not only the true nature of lexicography but also the possibilities of the Internet and its technologies, as advocates of the Function Theory of Lexicography do on a regular basis.

Keywords: SPECIALIZED LEXICOGRAPHY, FUNCTION THEORY, POLYFUNCTIONAL DICTIONARY, MONOFUNCTIONAL DICTIONARY, ACCOUNTING DICTIONARIES, INFORMATION TOOLS, LEXICOGRAPHICAL INFORMATION COSTS, ONLINE DICTIONARY, ELECTRONIC LEXICOGRAPHY, INTERNET, MEANING, WRITING, TRANSLATION, KNOWLEDGE, DANISH, ENGLISH, SPANISH

Opsomming: *Ontwikkeling in die leksikografie: Van polifunksionele tot monofunksionele rekenkundige woordeboeke.* Hierdie artikel beskryf sowel die teoretiese grondslag van die rekenkundige woordeboeke as die praktiese resultate daarvan. Verder

het die implementering van die projek getoon hoe die voortdurende wisselwerking van leksikografiese teorie met praktiese woordeboekwerk oor 'n tydperk van 10 jaar gelei het tot leksikografiese ontwikkeling en wysiging wat voortdurende aanpassings aan die woordeboeke en hul teoretiese grondslag tot gevolg gehad het. Dit word geïllustreer deur te let op die oorgang van polifunksionele na monofunksionele woordeboeke in 'n poging om hulp te verleen in verskeie soorte gebruikssituasies, gewoonlik in kommunikatiewe situasies, bv. wanneer Engelse, Deense of Spaanse rekenkundige tekste gelees, geskryf of vertaal word, of in kognitiewe situasies, d.i. wanneer gebruikers meer wil weet oor rekenkundige aangeleenthede of rekenkundige taal. Die artikel toon ook dat die skepping van hierdie projek e-leksikografie met die kennis-gebaseerde ekonomie verbind. Leksikograwe werk in die wolk deur fisies duisende kilometers van mekaar te wees; hulle werk met kundiges in die rekeningkunde, databasisse en die Internet met die doel om 'n hoëdrukgehaltewerktuig te bou en aan te pas met 'n relatief lae inligtingskoste vir gebruikers. Ten slotte toon hierdie artikel dat die toekoms van die leksikografie lê in die ontwerp, bou en aanpassing van inligtingswerktuie wat nie net die ware aard van die leksikografie in aanmerking neem nie, maar ook die moontlikhede van die Internet en sy tegnologieë, soos die voorstanders van die Funksieteorie van die Leksikografie op 'n gereelde basis doen.

Slutelwoorde: GESPECIALISEERDE LEKSIKOGRAFIE, FUNKSIETEORIE, POLIFUNKSIONELE WOORDEBOEK, MONOFUNKSIONELE WOORDEBOEK, REKENINGKUNDIGE WOORDEBOEKE, INLIGTINGSWERKTUIE, LEKSIKOGRAFIESE INLIGTINGSKOSTE, AANLYN WOORDEBOEK, ELEKTRONIESE LEKSIKOGRAFIE, INTERNET, BETEKENIS, SKRYF, VERTALING, KENNIS, DEENS, ENGELS, SPAANS

1. Introduction

Specialized dictionaries, i.e. dictionaries that cover specific domains and/or their languages, have been around for many years. They were primarily seen as practical tools of the trade, and little thought was given to a theoretical framework within which these dictionaries could be designed and produced. During the 19th and most of the 20th centuries, lexicography was considered a practical vocation based on linguistics and linguistic theories, but in the last half of the 20th century lexicographers began to study specialized dictionaries in order to analyze and describe these products as repositories of lexicon and knowledge as proposed in e.g. Frawley (1988) and Wiegand (1988). This theoretical framework applied to printed monolingual dictionaries and used primarily linguistic and text-linguistic principles. Around that time, lexicographers extended their interest to bilingual dictionaries by focusing mainly on equivalence and equivalents. Furthermore, during the 1990s, researchers introduced a new approach to lexicography and regarded general and specialized dictionaries as complex lexicographical tools that contain more than linguistic concepts and terms as described in e.g. Nielsen (1994), Schaefer and Bergenholtz (1994), Bergenholtz and Tarp (1995), and Bergenholtz (1996).

Lexicography, and by implication specialized lexicography, is thus a dynamic field that needs to keep up with and adapt to the development of

modern society, in particular the change from paper-based data to electronic data and the use of electronic information platforms, see e.g. Fuertes-Olivera (2013), Fuertes-Olivera and Tarp (2014), and Nielsen (2013). Bergenholtz (1996), Tarp (2000, 2001), Nielsen and Mourier (2005, 2007), Fuertes-Olivera (2009), Nielsen (2010), Fuertes-Olivera and Nielsen (2012), and Fuertes-Olivera and Tarp (2014) show that practical as well as theoretical lexicography cannot develop properly by focusing exclusively on documentation, description and analysis in a linguistic tradition but should turn its attention to the development of lexicographically relevant principles that can guide lexicographers in their quest for designing and producing dictionaries that help users solve specific types of problems in specific types of situations in the modern information society. Aspects of the recent development in lexicography will be described with particular reference to a dictionary project that involves a set of specialized online dictionaries within the field of accounting, collectively referred to below as the Accounting Dictionaries. The theoretical foundation of the dictionary project will be described as well as its practical results. Furthermore, the constant interaction of theory with practical dictionary work over a period of 10 years has led to lexicographical adaptation resulting in ongoing adjustments to the dictionaries and their theoretical foundation. This will be exemplified by studying the transition from polyfunctional to monofunctional dictionaries in an attempt to provide help that satisfies user needs.

2. Origins of the project

The Accounting Dictionaries, a set of 28 monolingual and bilingual dictionaries, are the result of a joint project involving lexicographers and subject-field specialists from a number of universities in Denmark and Spain. At present, the Danish participants come from the Centre for Lexicography affiliated with the Department of Business Communication, Aarhus University, and the Spanish participants are affiliated with the International Centre for Lexicography, University of Valladolid. In 2001, lexicographers at the Danish Centre started preparing the lexicographical basis for the compilation of a bilingual accounting dictionary covering the languages Danish and English. The planned dictionary was intended to be a practical tool designed to help Danish enterprises meet the increasing need to present their financial reporting in linguistically and terminologically correct English. From the start, the dictionary was designed as an electronic dictionary to be freely accessible on the Internet; see Nielsen (2002) for a detailed description of the theoretical foundation.

At the turn of the century, lexicography was slowly changing its focus though conventional wisdom was still oriented towards linguistic and text-linguistic approaches to dictionary making and research (see e.g. Schneider 1998, Lindemann 2000, and Landau 2001). The lexicographers felt that this focus on documentation, description and analysis based on a linguistic tradition was unsatisfactory; it did not provide the necessary help to establish a sound theo-

retical foundation embracing the types of problems involved in making the planned dictionary. This resembled a situation described by Chesterton (2001: 141): "It isn't that they can't see the solution. It's that they can't see the problem." As a result, the lexicographers had to look for another approach.

The problem was not to make a bilingual dictionary that would be a repository of various types of data representing linguistic concepts and principles documenting the lexicon of accounting. The real challenge was to make an online bilingual accounting dictionary that provided help to specific types of users to solve specific types of problems when translating accounting texts from Danish into English or producing accounting texts directly in English. In this light, the modern theory of lexicographical functions presented in e.g. Bergenholtz (1996), Bergenholtz and Kaufmann (1997) and Tarp (2000) indicated new possibilities for planning and compiling such dictionaries.

Any project based on a theoretical foundation need to have a clear understanding of what a dictionary actually is. A dictionary has traditionally been described as a reference work that presents the vocabulary of a language, usually in alphabetical order, with explanations of meanings in the same language or with equivalents in another language (see e.g. Hartmann and James 2001: 41; and Van Sterkenburg 2003: 396). However, the new lexicographical approach adopted regards the dictionary as an information tool with three significant features. The overriding feature is that the dictionary has one or more functions, e.g. communicative functions such as providing help to understand, translate and write texts, and cognitive functions such as providing help to acquire knowledge in communication-free contexts. This feature concerns the dictionary's lexicographical potential as an aid in particular types of usage situation. Secondly, the dictionary contains lexicographical data that support its function(s). This second feature interacts with the first one, because the potential of the dictionary can only be realized through the data presented. Finally, the lexicographical structures marshal the data into the task of fulfilling the dictionary function(s). This is the feature that makes it possible to unite the lexicographical data and the function(s) to produce a whole and to operationalize the relation and interaction between data based on specific lexicographical principles (see e.g. Nielsen 2003: 111-112; and Nielsen 2009b: 215). The point is that the intention behind the dictionary is not found in either the first, the second or the third feature, but in the aggregate of significant features, i.e. the electronic (or printed) dictionary in its entirety.

The adoption of the functional theory of lexicography means that the goal was to develop a lexicographical tool that would provide help to solve specific types of problems encountered by specific types of users in the context of translation into or production of texts in English. The next challenge was to identify and select the lexicographical data types that would support the intended lexicographical functions. Few reported studies deal with specialized translation dictionaries, e.g. Duvå and Laursen (1994), Nielsen (1994: 12-32), Wang (2001: 75-137) and Muráth (2002: 43-79), and only the first study concerns the lexico-

graphical needs of persons translating specialized texts into a foreign language. One drawback of the available user surveys is that the majority of informants are learners with limited experience and knowledge of translating specialized texts. Secondly, each set of findings relates to one particular text genre and therefore says nothing about translating specialized texts in general. Thirdly, the studies examine a limited number of subject fields so that they provide few new insights into the general activity of specialized translation. Fourthly, the number of informants in each study is so small that the results are not representative. Finally, the results reflect the subjective problems, needs, etc. recognized by the informants, whereas unrecognized needs go unnoticed. The findings therefore do not provide solid proof of actual usage situations or clear guidelines for designing specialized translation dictionaries (see Tarp 2009: 290-292 for a discussion of lexicographical user research).

In contrast to the above user studies, the functional theory allows lexicographers to work fast by consulting the research literature published in the field of translation studies. Several scholars have examined the elements of the translation process, and some of their findings are relevant for translating specialized texts and specialized lexicography. For instance, Nord (2005) demonstrates that the translation process contains several recursive steps that go beyond the level of terms and words, and Bell (2000: 211) focuses on translation units larger than words by identifying syntactic, semantic and pragmatic knowledge as necessary elements in translating texts. In particular, translators need syntactic knowledge about possible ways of combining words, and structuring word groups as well as restrictions in the foreign language in order to avoid linear dislocation. Here dictionaries can help by showing e.g. collocations and phrases that are relevant for the language of the domain. Semantic knowledge is relevant because translators need to know the exact meanings of domain-specific words and multi-word terms to properly understand differences in the structure of the domains in the two cultures involved, including hierarchical relationships such as superordination and homonymy. Clear definitions written in the appropriate language and style, synonyms and antonyms can help translators acquiring the necessary knowledge. Finally, pragmatic knowledge will tell translators how utterances are used in given contexts, i.e. how textual resources in the foreign language are conventionally used and what is textually appropriate or normal in communication within the field of accounting. Dictionaries can provide help through the presentation of collocations, phrases, example sentences and usage notes (See Sections 4 and 5 for a further discussion and examples).

In addition, some of the lexicographers in the project are experts in specialized translation and had, at the start of the project, more than 25 years' experience between them teaching specialized translation to students of English at undergraduate and postgraduate level; experience thus played a role in selecting the lexicographical data. In the field of epistemology, this type of experience is recognized as a source of knowledge and falls under the heading perception. It should be appreciated that the experience referred to is not the

same as introspection:

Through perception, we acquire (primarily) justified beliefs and knowledge about the external world; without these, we would be unlikely to survive. Through introspection, we acquire (primarily) justified beliefs and knowledge only about the mental world; with only this, our knowledge and justification would be sadly limited to our own minds. (Audi 2003: 91)

Compared to the method of surveying dictionary users, the use of findings by translation scholars combined with the relevant experience of translation experts has a number of benefits. Overall, the findings represent types of general elements of the translation process, and translation scholars and university teachers of specialized translation are experts in translating texts and have considerable practical as well as theoretical experience of the entire translation process. This may result in the inclusion of data types in the dictionary that can help users solve general types of problems, such as proscriptive presentation of equivalents, translation of collocations and phrases, as well as usage notes. Furthermore, specialized translation dictionaries that take these findings into consideration will address general user needs and not be biased towards subjective needs. Finally, the use of existing research results and perception allows lexicographers to design and complete a dictionary project within a relatively short time frame. The result was that, in order to fulfil the needs of translators and text writers, the lexicographers designed and made a syntagmatic specialized dictionary, i.e. a dictionary listing words and terms with definitions and inflectional paradigms, collocations, phrases as well as example sentences. The lexicographers also included various types of speech: nouns (single and multi-word units), adjectives, adverbs, prepositions, abbreviations, and verbs, including phrasal verbs. According to Lindemann (2000: 195), this distinguishes the accounting dictionary from existing ones covering the same field with only few collocations, phrases, sentences, and linguistic structures.

The three significant features of the dictionary must be realized as an online accounting dictionary proper. Many regard an electronic dictionary as a database that users access from an interface whose sole function is to give direct access to the dictionary. In other words, the database is identical with the dictionary. However, the Accounting Dictionaries can best be described as an electronic construction with three main components. Firstly, there is a database containing specially selected data that have been structured in a way that facilitates search and retrieval. Secondly, users will see an Accounting Dictionary in the form of a website that does not contain the lexicographical data as these are contained in the database and not in the user interface. Thirdly, in order to provide access to the lexicographical data, a search engine is introduced as an intermediary between the dictionary (user interface) and the database. This search engine allows users to search for data in the database and from there it retrieves the relevant data and presents the results of searches on the screen. In other words, the database is not identical with the dictionary.

The three-component structure has a number of practical and theoretical implications. Firstly, the database can be the source of several dictionaries as the search engine allows users to make structured searches in structured data through a specific set of online dictionaries. Secondly, the Accounting Dictionaries contain no macrostructures in the traditional text-linguistic sense of the word: a lexicographical structure that arranges lemmata in a specific order so that they can easily be found, as defined in e.g. Hausmann and Wiegand (1989: 336). The Accounting Dictionaries allow users to access data in the database and present the search results on the computer screen in a structured way, no matter where the data were actually located in the database. The macrostructure has been replaced by a data presentation structure that arranges the data retrieved from the database according to type, and presents these data in a predetermined order depending on user needs as identified by the type of help sought; see Nielsen and Almind (2011: 147-151) for a detailed discussion of the database.

The adoption of the functional theory of lexicography combined with the three-component electronic platform resulted in an extension of the project. It was now possible to have a large bilingual database in which a search engine could search for data in two languages and retrieve selected types of data irrespective of language. The lexicographers decided to produce a Danish-English accounting dictionary as well as a Danish accounting dictionary. One reason for this was that the Danish data types supporting the function translation from Danish into English could be used in a separate Danish accounting dictionary. The Danish definitions, inflectional paradigms, collocations, phrases, example sentences, cross-references and external links support communicative functions such as providing help to understand Danish accounting texts and to write accounting texts in Danish, as well as the cognitive functions of providing help to acquire knowledge about Danish accounting matters and accounting language. This strengthens the claim made earlier that the lexicographical database is not identical with the dictionary: one database and two dictionaries.

The two accounting dictionaries were made available to the public in 2003. They each contained about 4,200 headwords, or lemmas, about 15,400 collocations and phrases, and about 1,000 example sentences, providing users with guidance concerning correct language as well as subject-matter information. Being Internet dictionaries, the Danish accounting dictionary and the Danish-English accounting dictionary benefit from the advantage of open-ended electronic dictionaries: space. As accounting language develops rapidly in the modern world of financial regulation and crises, space is important, and the database allows lexicographers to add, revise and delete data, thereby enlarging the dictionaries on an ongoing basis (see Nielsen and Mourier 2007). In 2004, *Regnskabsordbogen dansk-engelsk* (Danish-English Accounting Dictionary) was published in print based on selected data found in the database and used for the Danish-English online accounting dictionary (Nielsen, Mourier and Bergholtz 2004).

The success with the two completed dictionaries led the lexicographers to design an English accounting dictionary and an English–Danish accounting dictionary. The theoretical basis as described above was retained in an adapted form and a new bilingual database was created partly based on the data from the first one. Work progressed along the same lines as before and the result was two online accounting dictionaries that provided help to specific types of users to solve specific types of problems when translating accounting texts from English into Danish, producing accounting texts directly in English, reading English accounting texts, and acquiring knowledge about English accounting matters and accounting language. The two new accounting dictionaries were made available on the Internet by end-2004, each containing about 6,000 headwords, about 20,000 collocations and phrases, and about 2,000 example sentences. In 2007, *Regnskabsordbogen engelsk–dansk* (English–Danish Accounting Dictionary) was published in print based on selected data found in the database and used for the English–Danish online accounting dictionary (Nielsen, Mourier and Bergenholtz 2007).

The change from linguistic to functional principles described so far illustrates important aspects of the development in lexicography in general and specialized lexicography in particular. A dictionary is seen as an information tool with three significant features that, combined with the theory of lexicographical functions, allow lexicographers to complete dictionary projects quickly. Moreover, the electronic platform with three components makes it possible to compile a database that can be used to make several electronic and printed dictionaries.

3. The inclusion of Spanish

In October 2008, a group of Spanish lexicographers, translators and accounting experts started the translation of the English lexicographical data stored in the accounting database. At that time, the database contained around 6,000 English dictionary articles, which were translated into or adapted to Spanish accounting in around 11 months (see Bergenholtz 2012, Fuertes-Olivera and Nielsen 2012, and Fuertes-Olivera and Tarp 2014 for a description of the data categories included in this database). The Spanish team worked as follows:

- The chief Spanish lexicographer explained the definition of the English lemma to the accounting experts. For around 95% of the English terms, the explanations given allowed the accounting experts to identify the corresponding Spanish equivalents.
- The chief Spanish lexicographer attached Spanish grammar and inflections to the Spanish equivalents.
- Two translators of English specialized texts translated the collocations and example sentences that contextualize each English headword (around 20,000 English collocations and 2,000 English example sentences).

- The chief Spanish lexicographer and the accounting experts included translations and/or adaptations of English synonyms and antonyms (where necessary). They also included contrastive notes, where necessary, and internal and external cross-references, i.e. links to other dictionary articles or to external web-pages, typically the homepage of the Instituto de Contabilidad y Auditoría de Cuentas, which is the official institution in charge of accounting standards, rules, and procedures in Spain.
- The chief Spanish lexicographer and the accounting experts coined the Spanish equivalents of the English terms that had not been identified in step 1 above; sometimes, this coinage consisted in accepting equivalents found in other dictionaries or in Internet homepages, or verbatim translations of English accounting lemmas (Fuertes-Olivera 2011).
- A Spanish native speaker trained in Spanish grammar and a Danish Professor of Spanish who teaches specialized translation proofread all the Spanish data present in the editors' layout "English-Spanish" of the accounting database (Figure 1). Their role was to check errors, mistakes, etc.

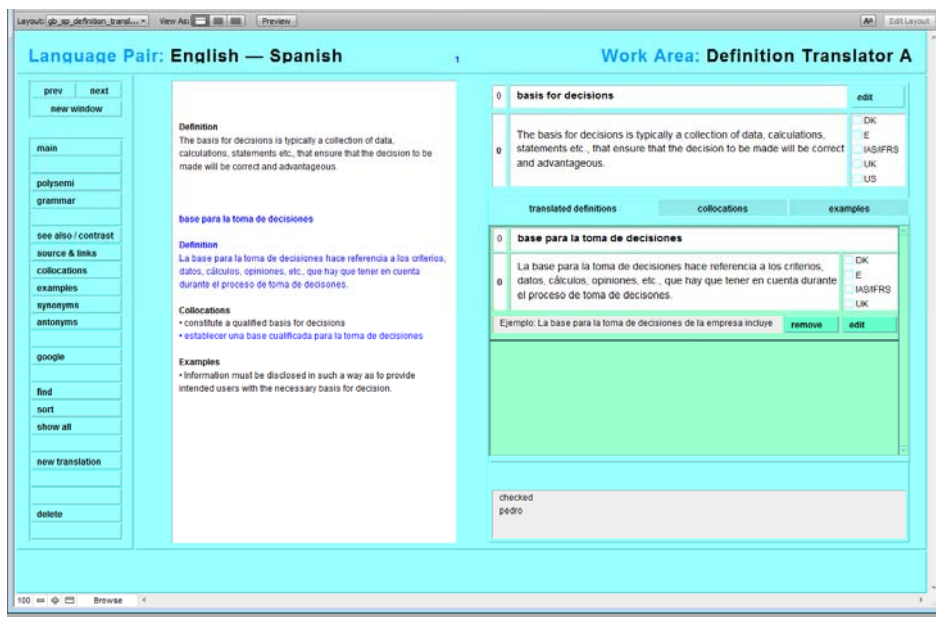


Figure 1: Screenshot of the English-Spanish accounting database (editors' layout: Definition)

By December 2009, IT experts based in the Centre for Lexicography, Aarhus University, converted the data into a polyfunctional online dictionary (Nielsen

et al. 2009), and a printed one (Fuertes-Olivera et al. 2010). There are some differences between them: for example, the printed dictionary does not contain external cross-references, i.e. links to web-pages. Examples 1 and 2 show the dictionary entry for **Consumer Price Index** in the online and printed dictionaries respectively: the grammar data and inflectional paradigms are only found in the online dictionary; the online dictionary does not use symbols but words, i.e. it uses "synonyms" instead of the sign "=":

Consumer Price Index US
<noun a, the -s>

Definition
The Consumer Price Index, CPI, is an index showing the rise and fall in prices of consumer goods and services sold in the retail market over a period of time.

Indice de Precios al Consumo

Synonyms
IPC

Collocations
• a change in the Consumer Price Index
un cambio en el Indice de Precios al Consumo

Synonyms
CPI US
Retail Price Index UK
RPI UK

Example 1: Consumer Price Index in the online dictionary (Nielsen et al. 2009)

Consumer Price Index US
= CPI US, Retail Price Index
UK, RPI UK
The Consumer Price Index,
CPI, is an index showing the
rise and fall in prices of consu-
mer goods and services sold in
the retail market over a period
of time.
Indice de Precios al Consumo
= IPC
▲ a change in the Consumer Pri-
ce Index un cambio en el Indice
de Precios al Consumo

Example 2: Consumer Price Index in the printed dictionary (Fuertes-Olivera et al. 2010)

The IT experts also prepared a new database for storing Spanish accounting data, e.g. the Spanish lemmas, their Spanish definitions, grammar, inflections, external and internal links, contrastive data, collocations, and example sentences.

This database is modeled as the English and Danish one, as shown in Figure 2:

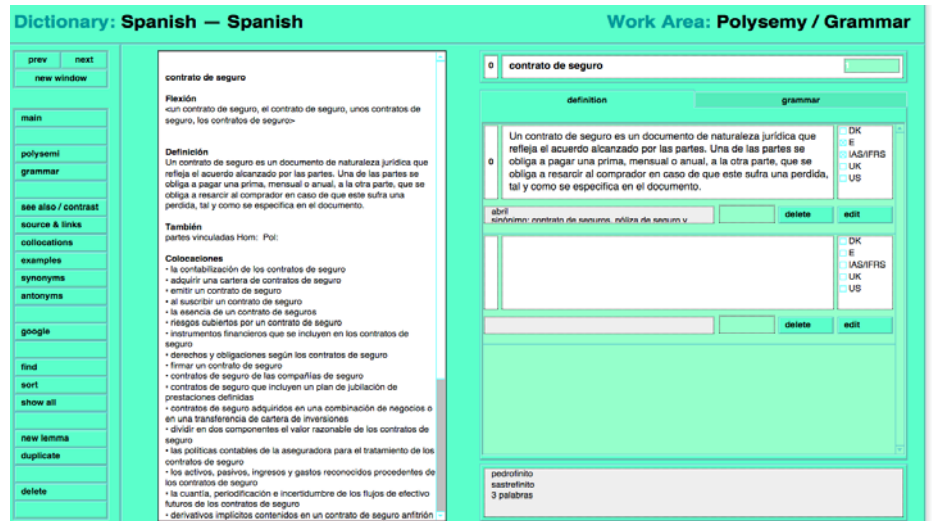


Figure 2: Screenshot of the Spanish accounting database (editors' layout: Polysemy and Grammar)

This new editors' layout was available in October 2011. At that time, the team initiated the storage of data as described below:

- The chief Spanish lexicographer and the accounting experts selected the Spanish lemmas, firstly from the Spanish equivalents included in the English-Spanish accounting dictionary, then from several accounting texts, and finally from an in-house corpus (see Fuertes-Olivera et al. 2013, and Fuertes-Olivera and Tarp 2014 for a description of the process).
- The chief Spanish lexicographer attached grammar and inflections to each Spanish headword.
- The chief Spanish lexicographer and the accounting experts crafted the Spanish definitions and decided on questions concerning homonymy and polysemy; this definition could be a translation or an adaptation of the English definition. It could also be new or containing cultural information connected with Spanish accounting rules and traditions (Fuertes-Olivera and Tarp 2014).
- The rest of the Spanish team completed the dictionary articles, e.g. collocations, example sentences, synonyms, usage and contrastive notes. All the data included were taken from accounting texts, most of which were downloaded from the Internet (see Fuertes-Olivera and Tarp 2014 for a description of the process).

By the end of 2012, the storage system contained around 6,500 Spanish dictionary articles. At that time, lexicographers and experts started working in a new layout of the accounting database, prepared by IT experts in Aarhus for including Spanish–English accounting data (Figure 3). At the time of writing this article, around 4,000 dictionary articles have been completed and it is expected that more than 6,000 dictionary articles will be completed by the end of 2013.

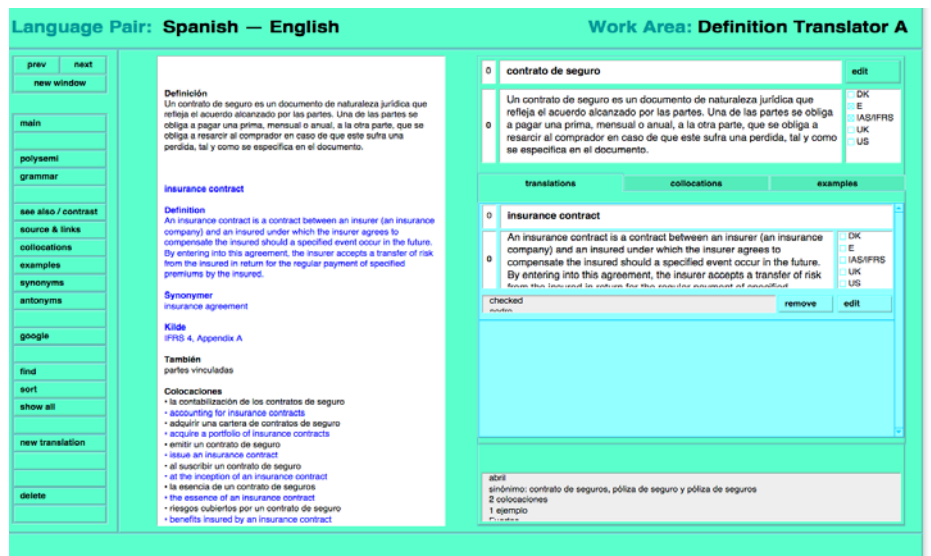


Figure 3: Screenshot of the Spanish–English accounting database (editors' layout: Definition)

As explained below, proponents of the function theory of lexicography are always open to include innovations, assuming that these agree with the conception of user needs in specific extra-lexicographical situations (Nielsen 2009a; and Tarp 2008). This intellectual situation resulted in replacing the polyfunctional dictionaries with innovative monofunctional dictionaries (Bergenholtz 2011, 2012), initially described as *Model T Ford* dictionaries, i.e. dictionaries whose articles and visualized lexicographical data are adapted to the various functions displayed by the dictionary, frequently assisted by different types of interactive options where users may define themselves and the activity for which they need information (Tarp 2011; and Fuertes-Olivera and Tarp 2014). In other words, the originally conceived and published polyfunctional accounting dictionaries are under a process of constant updating and conversion into monofunctional online dictionaries.

4. Conversion of the polyfunctional English and Danish dictionaries into usage-based monofunctional ones

The original polyfunctional nature of the dictionaries affected access to and presentation of data. When they searched for words, users would be presented with full articles resembling those in printed dictionaries with data types intended to support a plurality of usage situations. This meant that users had to read, or at least skim, entire articles to find the answers they were looking for because the text contained relatively little relevant data among a large collection of data. As a result the lexicographical information costs for users were higher than necessary (see Nielsen 2008 for a detailed discussion of information costs in lexicography). This is a common problem with general and specialized dictionaries, printed as well as electronic, that want to cater to everyone in every type of usage situation. However, various developments allow lexicographers to design dictionaries so that users can easily access the data and process these into useful information.

The ease with which users will be able to acquire the necessary information from the data in dictionaries can be examined in terms of lexicographical information costs. These costs may be defined as the efforts that users believe or feel are associated with consulting a dictionary, an article or any other text part of a dictionary. Search-related information costs are the efforts related to the look-up activities users have to perform when consulting a dictionary in order to access the relevant data, while comprehension-related information costs are the efforts related to users' ability to understand the data presented in a dictionary (Nielsen 2008: 173-174). The problem with polyfunctional dictionaries is that it may require relatively much effort to search for the answer users need in an article containing several data types of which most are irrelevant for a specific search. On top of that, it may require much effort to decode the meaning of the data in order to find the specific data that answer your question, because they will have to be interpreted in the light of the rest of the data to make sure that you have found the lexicographical help needed. In other words, lexicographical *Model T Fords* should be able to reduce search- and comprehension-related information costs by making access to and interpretation of data easy and quick.

The technological development of online resources can be combined with the development in the function theory of lexicography. By focusing on the needs of users in various types of usage situations, lexicographers can select electronic solutions that ensure retrieval of data that satisfy a specific type of need and can present the data in such a way that users can easily turn them into useful information. From end-2009 to mid-2011 the lexicographers at the Centre for Lexicography developed one single database that replaced the previous ones and now allows the lexicographers to use advanced electronic options. One result was that the polyfunctional monolingual and bilingual accounting dictionaries were substituted by monofunctional dictionaries. Per-

sons who find themselves in a particular type of communicative or cognitive usage situation may need help of a specific kind and consult the accounting dictionary most likely to provide assistance. Users can consult one of four Danish and four English accounting dictionaries for the following kinds of help:

- Meaning of accounting term (usage situation: I am reading an accounting text and want help to understand an accounting term)
- Use of known accounting term (usage situation: I am writing an accounting text and want help to produce the text with a known word)
- Use of accounting term whose meaning is known (usage situation: I am writing an accounting text and want help to find and use a term that has a specific meaning)
- Knowledge of accounting matters (usage situation: I want to know more about accounting matters or accounting language)

Users can consult one of four Danish-English and four English-Danish accounting dictionaries for the following kinds of help:

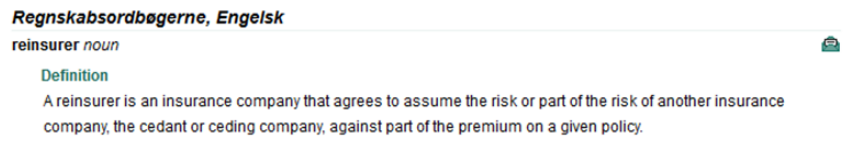
- Meaning of accounting term (usage situation: I am reading an accounting text and want help to understand an accounting term)
- Translation of known accounting term (usage situation: I am translating an accounting text and want help to translate a known word)
- Translation of collocation (usage situation: I am translating an accounting text and want help to translate a collocation)
- Knowledge of accounting matters (usage situation: I want to know more about (comparative) accounting matters or accounting language)

This means that each dictionary has different functions and search options specifically made for each function. There are now eight monolingual and eight bilingual accounting dictionaries with the languages Danish and English, all developed from a single database.

In order to keep information costs as low as possible, the search engine allows users to perform targeted searches in the database. Readers of accounting texts may want to know the meaning of the English accounting term *reinsurer* and consult the *English Accounting Dictionary: Meaning* whose function is to provide the meaning of accounting terms. The search engine makes a targeted search for the term and retrieves the definitional data addressed to the term as illustrated in Example 3.

Example 3 shows how the monofunctional dictionary helps users by presenting limited data that are specifically selected to support the communicative function: help users understand the meaning of a particular term. The relevant

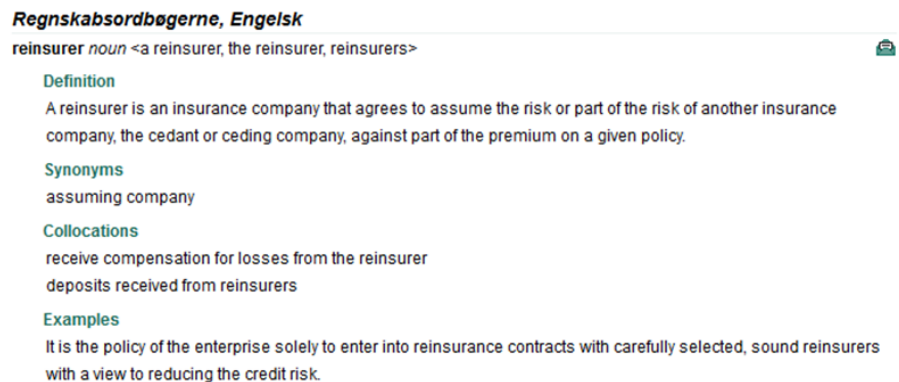
data are easy to find as they are not hidden among other, and for the function irrelevant, data; users only need a definition and nothing more in this type of usage situation.



Regnskabsordbøgerne, Engelsk
reinsurer *noun*
Definition
A reinsurer is an insurance company that agrees to assume the risk or part of the risk of another insurance company, the cedant or ceding company, against part of the premium on a given policy.

Example 3: Definition helping to understand the accounting term

Authors may want help to write texts in English in which the term *reinsurer* occurs. The search engine allows users to select the *English Accounting Dictionary: Text Production* whose function is to show how known accounting terms can be used. The result of the search is shown in Example 4.



Regnskabsordbøgerne, Engelsk
reinsurer *noun* <a reinsurer, the reinsurer, reinsurers>
Definition
A reinsurer is an insurance company that agrees to assume the risk or part of the risk of another insurance company, the cedant or ceding company, against part of the premium on a given policy.
Synonyms
assuming company
Collocations
receive compensation for losses from the reinsurer
deposits received from reinsurers
Examples
It is the policy of the enterprise solely to enter into reinsurance contracts with carefully selected, sound reinsurers with a view to reducing the credit risk.

Example 4: Help to write accounting texts where the expression is known

The definition in Example 4 enables users to check that the term found has the correct meaning, and the grammar data, synonyms, collocations and example sentences support the writing process. The writing process generally concerns terminology, phraseology, grammar, syntax, and pragmatics, hence the data types presented in the monofunctional accounting dictionary whose communicative function is to help users write accounting texts. In particular, writers prepare drafts, check the texts to ensure that generally acceptable grammar and spelling rules are complied with, that the appropriate terminology has been used consistently so that the texts are unambiguous, and that the texts are free from errors; see Nielsen (2006) for a discussion of monolingual accounting dictionaries for text production.

The *English Accounting Dictionary: Knowledge* provides help in cognitive usage situations. People may want to acquire general or specific knowledge

about the accounting concept *reinsurer* and the search engine retrieves and presents the data in a predetermined order illustrated in Example 5.

Regnskabsordbøgerne, Engelsk

reinsurer *noun* <a reinsurer, the reinsurer, reinsurers>

Definition
A reinsurer is an insurance company that agrees to assume the risk or part of the risk of another insurance company, the cedant or ceding company, against part of the premium on a given policy.

Synonyms
assuming company

Antonyms
cedant

Collocations
receive compensation for losses from the reinsurer
deposits received from reinsurers

Examples
It is the policy of the enterprise solely to enter into reinsurance contracts with carefully selected, sound reinsurers with a view to reducing the credit risk.

See also:
[reinsurance](#)

Sources
[IFRS 4, Appendix A](#)

Example 5: Help to acquire knowledge in cognitive usage situations

In Example 5, the definition explains the meaning of the term, which is supplemented by an example sentence. Synonyms and antonyms help users put the term *reinsurer* in its terminological hierarchy within the domain. In addition, users can click the underlined cross-reference (under See also) containing an embedded link and come to another article with relevant additional data; the item indicating the source of the definition also has an embedded link users can click to go to the website of the international financial reporting standard (IFRS) for more information.

Danes who translate accounting texts may have problems translating the English term *reinsurer*. They can find assistance in the *English–Danish Accounting Dictionary: Translation* whose function is to provide help to translate accounting terms, and Example 6 shows the search result for *reinsurer*.

In addition to presenting the meaning of the term in Example 6, the dictionary presents the recommended Danish equivalent: *reassurandør*. The alternative Danish term *genforsikringselskab* is presented as a synonym. English collocations and example sentences with their translations into Danish as well as the inflectional paradigm help users to translate accounting texts in which the search term occurs.

Examples 3 to 6 are examples of how the monofunctional accounting dictionaries covering the languages Danish and English work. Each example shows that the monofunctional approach keeps information costs low as the

dictionaries in each case only present data supporting communicative and cognitive functions respectively. In contrast, polyfunctional dictionaries would likely have presented all the data shown in Example 5 in all cases and left it to users to find the relevant data that could help them. Finally, it should be noted that searches are conducted in and results retrieved from one single database.

Regnskabsordbøgerne, Engelsk-Dansk

reinsurer *noun*

Definition
A reinsurer is an insurance company that agrees to assume the risk or part of the risk of another insurance company, the cedant or ceding company, against part of the premium on a given policy.

reassurandør *substantiv* <en reassurandør, reassurandøren, reassurandører, reassurandørerne>

Synonyms
genforsikringselskab

Collocations

- deposits received from reinsurers
depoter modtaget fra genforsikringselskaber
- receive compensation for losses from the reinsurer
modtage erstatning for tab fra reassurandøren

Examples

- It is the policy of the enterprise solely to enter into reinsurance contracts with carefully selected, sound reinsurers with a view to reducing the credit risk.
Det er virksomhedens politik, at genforsikringskontrakter udelukkende indgås med omhyggeligt udvalgte, solide reassurandører for at reducere kreditrisikoen.

Example 6: Data providing help to translate an accounting term

5. Conversion of the polyfunctional English and Spanish dictionaries into usage-based monofunctional ones

The conversion of the polyfunctional English and Spanish dictionaries into usage-based English and Spanish dictionaries followed the theoretical principles described in section 4 above. This process rested on the idea that lexicographical databases and dictionaries are different things. This has allowed the dictionary team to convert the originally conceived polyfunctional dictionaries into twelve accounting dictionaries, eight of which are already in operation: there are four English–Spanish accounting dictionaries and four Spanish accounting dictionaries that offer the same kind of help discussed in Section 4.

In addition, users of the English–Spanish and Spanish dictionary sets have two more innovations. One of them is that the English–Spanish accounting set contains English and Spanish definitions of the English lemmas. The Spanish definitions are included for two main reasons. The first is that some targeted users (Spanish students and interested laypeople) are not as proficient in English as their Danish counterparts and the Spanish definitions will help them to understand the accounting concepts (i.e. low information costs). The second reason is that some of the definitions also include cultural data, e.g. data for

explaining that *building societies* are English institutions that are different from Spanish *sociedades constructoras* (Eng.: *building firms* or *construction companies*); they are similar to Spanish co-operatives or mutual savings banks, two organizations that are easily identified in Spain and that have a similar function to the building society described in the dictionary article (Example 7).

(UK) **building society**

Definition

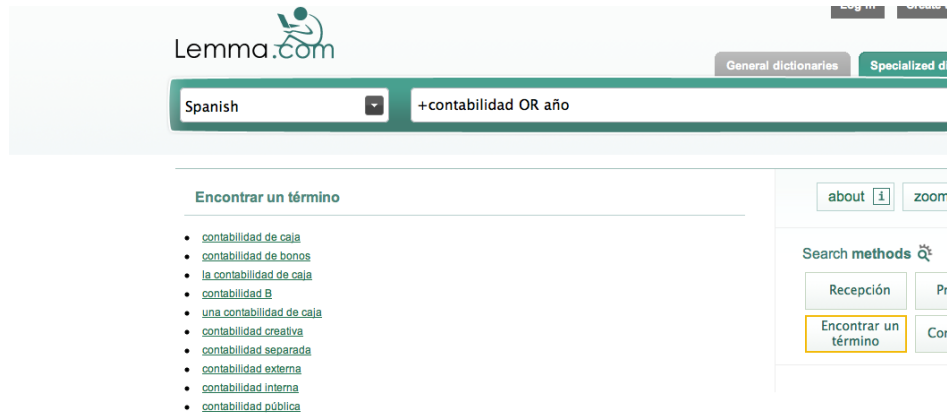
- Las sociedades de crédito hipotecario o inmobiliario son instituciones financieras que funcionan en el Reino Unido como cooperativa o mutua constructora. Sus objetivos incluyen el depósito de cuentas viviendas, la concesión de préstamos hipotecarios con avales y sin avales utilizados para la compra, construcción, o mejora de viviendas así como la oferta de varios tipos de servicios financieros.
- Building societies in the UK are financial institutions whose objectives include accepting deposits for saving, granting mortgage loans as well as unsecured loans to the depositors for the purchase, construction or improvement of owner-occupied dwellings as well as offering various financial services.

Example 7: Definitions of *building society* in the *English–Spanish accounting dictionary: Meaning*

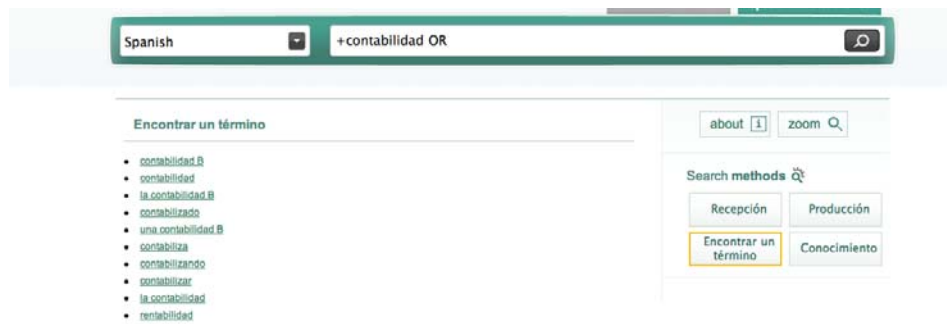
The second innovation is the inclusion of the search engine "encontrar un término" (Eng.: find a term) that allows searching with a term, part of it and three Boolean operators: "+", "-", and "or". For instance, a search for **+contabilidad +año** retrieves some terms (e.g. *contabilidad creativa*) and verb forms (*contabilizando*), which can be clicked (Example 8). A search for **+contabilidad OR año** retrieves different terms (e.g. *contabilidad de caja*) and some of its inflections (*la contabilidad de caja* and *una contabilidad de caja*) (Example 9). A search for **+contabilidad OR** retrieves several possible terms and inflections (Example 10). In other words, this search option offers users possible ways of finding a term they do not remember (perhaps, they only have the gist of the concept and need several searches before finding what they are looking for).



Example 8: Data providing help for finding an unknown accounting term



Example 9: Data providing help for finding an unknown accounting term



Example 10: Data providing help for finding an unknown accounting term

6. Conclusion and implication for the future

This article places the true nature of lexicography within the tenets of the *Function Theory of Lexicography*, i.e. the theoretical construction initiated in the Centre for Lexicography at the Aarhus School of Business (Bergenholtz and Tarp 2002, 2003, 2004; also Tarp 2008, and Fuertes-Olivera and Tarp 2014). It indicates that lexicography in general and specialized lexicography in particular deals with more than language and facts. It *also* keeps up with and adapts to the developments of a modern information society. For instance, the advent of the Internet allows lexicographers to move from constructing printed polyfunctional dictionaries to designing sets of electronic monofunctional ones. One example of this change is the conversion of the originally polyfunctional accounting dictionary into 28 monofunctional dictionaries, 24 of which are already in operation and can be consulted at Ordbogen.com or Lemma.com.

This article has traced the history of the lexicographical project since its inception in 2001. At that time, lexicographers at the Centre for Lexicography intended to construct a tool to help Danish enterprises meet the increasing need to present their financial reporting in linguistically and terminologically correct English. A few years later, the objectives of the project were enlarged, a new language – Spanish – was included, and new ideas and technologies came into operation, e.g. the distinction between the lexicographical database, the search engine, and the homepage. These new concepts integrated well within the general framework espoused in the function theory of lexicography: dictionaries are tools and are designed for satisfying specific types of needs in specific types of extra-lexicographical usage situations.

The making of the project also has implications for the future: this dictionary project is a typical product of the knowledge-based economy we have been constructing so far. Lexicographers work in the cloud being physically thousands of kilometres away; they work with experts in accounting, databases and the Internet with the aim of constructing and updating high-quality tools with low information costs, i.e. tools that are sold on a subscription basis. It may be envisaged that the future of lexicography is illustrated by the working of this project: It is based on sound research and is constantly transforming with the aim of producing sellable, innovative and useful lexicographical products. In other words, it is likely that the future of lexicography rests on designing, constructing, and updating information tools that take into consideration not only the true nature of lexicography but also the possibilities of the Internet and its technologies. This article has shown how this can be done within the general framework of the Function Theory of Lexicography.

Acknowledgments

Thanks are due to the Ministerio de Economía y Competitividad and to the Junta de Castilla y León for financial support (Grant FF12011-22885 and VA 067A12-1 respectively).

References

Accounting Dictionaries

- Fuertes-Olivera, P.A., H. Bergenholtz, S. Nielsen, P. Gordo Gómez, L. Mourier, M. Niño Amo, Á. de los Ríos Rodicio, Á. Sastre Ruano, S. Tarp and M. Velasco Sacristán. 2012. *Diccionario Inglés-Español de Contabilidad: Conocimiento*. Base de Datos y Diseño: R. Almind and J. Skovgård Nielsen. Odense: Lemma.com.
- Fuertes-Olivera, P.A., H. Bergenholtz, S. Nielsen, P. Gordo Gómez, L. Mourier, M. Niño Amo, Á. de los Ríos Rodicio, Á. Sastre Ruano, S. Tarp and M. Velasco Sacristán. 2012. *Diccionario Inglés-Español de Contabilidad: Recepción*. Base de Datos y Diseño: R. Almind and J. Skovgård Nielsen. Odense: Lemma.com.

- Fuertes-Olivera, P.A., H. Bergenholtz, S. Nielsen, P. Gordo Gómez, L. Mourier, M. Niño Amo, Á. de los Ríos Rodicio, Á. Sastre Ruano, S. Tarp and M. Velasco Sacristán.** 2012. *Diccionario Inglés-Español de Contabilidad: Traducción*. Base de Datos y Diseño: R. Almind and J. Skovgård Nielsen. Odense: Lemma.com.
- Fuertes-Olivera, P.A., H. Bergenholtz, S. Nielsen, P. Gordo Gómez, L. Mourier, M. Niño Amo, Á. de los Ríos Rodicio, Á. Sastre Ruano, S. Tarp and M. Velasco Sacristán.** 2012. *Diccionario Inglés-Español de Contabilidad: Traducción de Frases y Expresiones*. Base de Datos y Diseño: R. Almind and J. Skovgård Nielsen. Odense: Lemma.com.
- Fuertes-Olivera, P.A., H. Bergenholtz, S. Nielsen, P. Gordo Gómez, M. Niño Amo, Á. de los Ríos Rodicio, Á. Sastre Ruano, S. Tarp and M. Velasco Sacristán.** 2013. *Diccionario Español de Contabilidad. Conocimiento*. Base de Datos y Diseño: R. Almind and J. Skovgård Nielsen. Hamburg: Lemma.com.
- Fuertes-Olivera, P.A., H. Bergenholtz, S. Nielsen, P. Gordo Gómez, M. Niño Amo, Á. de los Ríos Rodicio, Á. Sastre Ruano, S. Tarp and M. Velasco Sacristán.** 2013. *Diccionario Español de Contabilidad: Encontrar un Término*. Base de Datos y Diseño: R. Almind and J. Skovgård Nielsen. Hamburg: Lemma.com.
- Fuertes-Olivera, P.A., H. Bergenholtz, S. Nielsen, P. Gordo Gómez, M. Niño Amo, Á. de los Ríos Rodicio, Á. Sastre Ruano, S. Tarp and M. Velasco Sacristán.** 2013. *Diccionario Español de Contabilidad: Producción*. Base de Datos y Diseño: R. Almind and J. Skovgård Nielsen. Hamburg: Lemma.com.
- Fuertes-Olivera, P.A., H. Bergenholtz, S. Nielsen, P. Gordo Gómez, M. Niño Amo, Á. de los Ríos Rodicio, Á. Sastre Ruano, S. Tarp and M. Velasco Sacristán.** 2013. *Diccionario Español de Contabilidad: Recepción*. Base de Datos y Diseño: R. Almind and J. Skovgård Nielsen. Hamburg: Lemma.com.
- Fuertes-Olivera, P.A., P. Gordo Gómez, M. Niño Amo, Á. de los Ríos Rodicio, Á. Sastre Ruano, S. Tarp, M. Velasco Sacristán, S. Nielsen, L. Mourier and H. Bergenholtz.** 2010. *Diccionario de Contabilidad Inglés-Español*. Navarra: Thomson Reuters-Aranzadi.
- Nielsen, S., L. Mourier and H. Bergenholtz.** 2003–2009. *Den Dansk-Engelske Regnskabsordbog*. Aarhus: Centre for Lexicography. Available at: <http://www.ordbogen.com>.
- Nielsen, S., L. Mourier and H. Bergenholtz.** 2003–2009. *Den Danske Regnskabsordbog*. Aarhus: Centre for Lexicography. Available at: <http://www.ordbogen.com>.
- Nielsen, S., L. Mourier and H. Bergenholtz.** 2004. *Regnskabsordbogen dansk-engelsk*. Copenhagen: Thomson.
- Nielsen, S., L. Mourier and H. Bergenholtz.** 2004–2009. *Den Engelsk-Danske Regnskabsordbog*. Aarhus: Centre for Lexicography. Available at: <http://www.ordbogen.com>.
- Nielsen, S., L. Mourier and H. Bergenholtz.** 2004–2009. *Den Engelske Regnskabsordbog*. Aarhus: Centre for Lexicography. Available at: <http://www.ordbogen.com>.
- Nielsen, S., L. Mourier and H. Bergenholtz.** 2007. *Regnskabsordbogen engelsk-dansk*. Copenhagen: Thomson.
- Nielsen, S., L. Mourier and H. Bergenholtz.** 2012. *Den Dansk-Engelske Regnskabsordbog: Betydning*. Odense: Ordbogen.com. Available at: <http://www.ordbogen.com>.
- Nielsen, S., L. Mourier and H. Bergenholtz.** 2012. *Den Dansk-Engelske Regnskabsordbog: Oversættelse*. Odense: Ordbogen.com. Available at: <http://www.ordbogen.com>.
- Nielsen, S., L. Mourier and H. Bergenholtz.** 2012. *Den Dansk-Engelske Regnskabsordbog: Oversættelse af ordforbindelse*. Odense: Ordbogen.com. Available at: <http://www.ordbogen.com>.

- Nielsen, S., L. Mourier and H. Bergenholtz. 2012. *Den Dansk-Engelske Regnskabsordbog: Viden*. Odense: Ordbogen.com. Available at: <http://www.ordbogen.com>.
- Nielsen, S., L. Mourier and H. Bergenholtz. 2012. *Den Danske Regnskabsordbog: Betydning*. Odense: Ordbogen.com. Available at: <http://www.ordbogen.com>.
- Nielsen, S., L. Mourier and H. Bergenholtz. 2012. *Den Danske Regnskabsordbog: Hvad hedder dette regnskabsudtryk?* Odense: Ordbogen.com. Available at: <http://www.ordbogen.com>.
- Nielsen, S., L. Mourier and H. Bergenholtz. 2012. *Den Danske Regnskabsordbog: Tekstproduktion*. Odense: Ordbogen.com. Available at: <http://www.ordbogen.com>.
- Nielsen, S., L. Mourier and H. Bergenholtz. 2012. *Den Danske Regnskabsordbog: Viden*. Odense: Ordbogen.com. Available at: <http://www.ordbogen.com>.
- Nielsen, S., L. Mourier and H. Bergenholtz. 2012. *Den Engelsk-Danske Regnskabsordbog: Betydning*. Odense: Ordbogen.com. Available at: <http://www.ordbogen.com>.
- Nielsen, S., L. Mourier and H. Bergenholtz. 2012. *Den Engelsk-Danske Regnskabsordbog: Oversættelse*. Odense: Ordbogen.com. Available at: <http://www.ordbogen.com>.
- Nielsen, S., L. Mourier and H. Bergenholtz. 2012. *Den Engelsk-Danske Regnskabsordbog: Oversættelse af ordforbindelser*. Odense: Ordbogen.com. Available at: <http://www.ordbogen.com>.
- Nielsen, S., L. Mourier and H. Bergenholtz. 2012. *Den Engelsk-Danske Regnskabsordbog: Viden*. Odense: Ordbogen.com. Available at: <http://www.ordbogen.com>.
- Nielsen, S., L. Mourier and H. Bergenholtz. 2012. *Den Engelske Regnskabsordbog: Betydning*. Odense: Ordbogen.com. Available at: <http://www.ordbogen.com>.
- Nielsen, S., L. Mourier and H. Bergenholtz. 2012. *Den Engelske Regnskabsordbog: Hvad hedder dette regnskabsudtryk?* Odense: Ordbogen.com. Available at: <http://www.ordbogen.com>.
- Nielsen, S., L. Mourier and H. Bergenholtz. 2012. *Den Engelske Regnskabsordbog: Tekstproduktion*. Odense: Ordbogen.com. Available at: <http://www.ordbogen.com>.
- Nielsen, S., L. Mourier and H. Bergenholtz. 2012. *Den Engelske Regnskabsordbog: Viden*. Odense: Ordbogen.com. Available at: <http://www.ordbogen.com>.
- Nielsen, S. L. Mourier, H. Bergenholtz, P.A. Fuertes-Olivera, P. Gordo Gómez, M. Niño Amo, A. de los Ríos Rodicio, A. Sastre Ruano, S. Tarp and M. Velasco Sacristán. 2009. *El Diccionario Inglés-Español de Contabilidad*. Available at: http://www.accountingdictionary.dk/regn/gbsp/regngbsp_index.php.

Other literature

- Audi, R. 2003. *Epistemology. A Contemporary Introduction to the Theory of Knowledge*. New York/London: Routledge.
- Bell, R.T. 2000. *Translation and Translating. Theory and Practice*. Eleventh printing. London/New York: Longman.
- Bergenholtz, H. 1996. Grundfragen der Fachlexikographie. Gellerstam, M., J. Järborg, S.-G. Malmgren, K. Norén, L. Rogström and C.R. Pappmehl (Eds.). 1996. *Euralex '96 Proceedings I-II, Papers Submitted to the Seventh EURALEX International Congress on Lexicography in Göteborg, Sweden*: 731-758. Göteborg: Department of Swedish, Göteborg University.
- Bergenholtz, H. 2011. Access to and Presentation of Needs-adapted Data in Monofunctional Internet Dictionaries. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011: 30-53.
- Bergenholtz, H. 2012. Concepts for Monofunctional Accounting Dictionaries. *Terminology* 18(2): 243-263.

- Bergenholtz, H. and U. Kaufmann.** 1997. Terminography and Lexicography. A Critical Survey of Dictionaries from a Single Specialised Field. Tarp, S. (Ed.). 1997. Specialised Lexicography and Terminology. *Hermes* 18: 91-125.
- Bergenholtz, H. and S. Tarp (Eds.).** 1995. *Manual of Specialised Lexicography. The Preparation of Specialised Dictionaries.* Amsterdam/Philadelphia: John Benjamins.
- Bergenholtz, H. and S. Tarp.** 2002. Die moderne lexikographische Funktionslehre. Diskussionsbeitrag zu neuen und alten Paradigmen, die Wörterbücher als Gebrauchsgegenstände verstehen. *Lexicographica. International Annual for Lexicography* 18: 253-263.
- Bergenholtz, H. and S. Tarp.** 2003. Two Opposing Theories: On H.E. Wiegand's Recent Discovery of Lexicographic Functions. *Hermes, Journal of Linguistics* 31: 171-196.
- Bergenholtz, H. and S. Tarp.** 2004. The Concept of "Dictionary Usage". Dollerup, Cai (Ed.). 2004. *Worlds of Words. A Tribute to Arne Zettersten. Nordic Journal of English Studies* 3(1): 23-36.
- Chesterton, G.K.** 2001. *The Scandal of Father Brown.* Kelly Bray: House of Stratus.
- Duvå, G. and A.L. Laursen.** 1994. Translation and LSP Lexicography: A User Survey. Schaefer, B. and H. Bergenholtz (Eds.). 1994. *Fachlexikographie: Fachwissen und seine Repräsentation in Wörterbüchern:* 247-267. Tübingen: Gunther Narr.
- Frawley, W.** 1988. New Forms of Specialized Dictionaries. *International Journal of Lexicography* 1(3): 189-213.
- Fuertes-Olivera, P.A.** 2009. Systematic Introductions in Specialised Dictionaries: Some Proposals in Relation to Accounting Dictionaries. Nielsen, S. and S. Tarp (Eds.). 2009: 161-178.
- Fuertes-Olivera, P.A.** 2011. Equivalent Selection in Specialized e-Lexicography: A Case Study with Spanish Accounting Terms. *Lexikos* 21: 95-119.
- Fuertes-Olivera, P.A.** 2013. E-lexicography: The Continuing Challenge of Applying New Technology to Dictionary Making. Jackson, H. (Ed.). 2013. *The Bloomsbury Companion to Lexicography.* London: Bloomsbury Publishing.
- 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., H. Bergenholtz, P. Gordo Gómez, S. Nielsen, M. Niño Amo, Á. de los Ríos Rodicio, Á. Sastre Ruano, and M. Velasco Sacristán.** 2013. The Construction of Online Specialized Dictionaries: Theory and Practice. *Fachsprache. International Journal of Specialized Communication* 35: 24-41.
- Fuertes-Olivera, P.A. and S. Nielsen.** 2012. Online Dictionaries for Assisting Translators of LSP Texts: The Accounting Dictionaries. *International Journal of Lexicography* 25(2): 191-215.
- Fuertes-Olivera, P.A. and S. Tarp.** 2014. *Theory and Practice of Specialised Online Dictionaries. Lexicography versus Terminography.* Berlin/New York: De Gruyter.
- Hartmann, R.R.K. and G. James.** 2001. *Dictionary of Lexicography.* London/New York: Routledge.
- Hausmann, F.J. and 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. *Wörterbücher. Ein internationales Handbuch zur Lexikographie/Dictionaries. An International Encyclopedia of Lexicography/Dictionnaires. Encyclopédie internationale de lexicographie. Volume 1:* 328-360. Berlin/New York: Walter de Gruyter.
- Landau, S.I.** 2001. *Dictionaries. The Art and Craft of Lexicography.* Second Edition. Cambridge/New York: Cambridge University Press.
- Lindemann, M.** 2000. Vorüberlegungen zur einem französisch-deutschen Wörterbuch der Fachsprache der Wirtschaft. Wiegand, H.E. (Ed.). 2000. *Germanistische Linguistik* 151-152: 173-214.

Hildesheim/New York: Georg Olms.

- Muráth, J.** 2002. *Zweisprachige Fachlexikographie*. Budapest: Nemzeti Tankönyvkiadó.
- Nielsen, S.** 1994. *The Bilingual LSP Dictionary. Principles and Practice for Legal language*. Tübingen: Gunter Narr.
- Nielsen, S.** 2002. *Lexicographical Basis for an Electronic Bilingual Accounting Dictionary: Theoretical Considerations*. <http://www.sprog.asb.dk/sn/lexicographicalbasis.htm> [accessed 31 July 2013].
- Nielsen, S.** 2003. Changes in Dictionary Subject Matter. Wiegand, H.E. (Ed.). 2003. *Untersuchungen zur kommerziellen Lexikographie der deutschen Gegenwartssprache I. Duden. Das grosse Wörterbuch der deutschen Sprache in zehn Bänden*: 109-115. Tübingen: Max Niemeyer.
- Nielsen, S.** 2006. Monolingual Accounting Dictionaries for EFL Text Production. *Ibérica* 12: 43-64.
- Nielsen, S.** 2008. The Effect of Lexicographical Information Costs on Dictionary Making and Use. *Lexikos* 18: 170-189.
- Nielsen, S.** 2009a. Reviewing Printed and Electronic Dictionaries. A Theoretical and Practical Framework. Nielsen, S. and S. Tarp (Eds.). 2009: 23-41.
- Nielsen, S.** 2009b. The Evaluation of the Outside Matter in Dictionary Reviews. *Lexikos* 19: 207-224.
- Nielsen, S.** 2010. Specialised Translation Dictionaries for Learners. Fuertes-Olivera, P.A. (Ed.). 2010. *Specialised Dictionaries for Learners*: 69-82. Berlin/New York: Walter de Gruyter.
- Nielsen, S.** 2013. The Future of Dictionaries, Dictionaries of the Future. Jackson, H. (Ed.). 2013. *The Bloomsbury Companion to Lexicography*. London: Bloomsbury Publishing.
- Nielsen, S. and R. Almind.** 2011. From Data to Dictionary. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011: 141-167.
- Nielsen, S. and L. Mourier.** 2005. Internet Accounting Dictionaries: Present Solutions and Future Possibilities. *Hermes, Journal of Linguistics* 34: 83-116.
- Nielsen, S. and L. Mourier.** 2007. Design of a Function-based Internet Accounting Dictionary. Gottlieb, H. and J.E. Mogensen (Eds.). 2007. *Dictionary Visions, Research and Practice: Selected Papers From the 12th International Symposium on Lexicography, Copenhagen, 2004*: 119-135. Terminology and Lexicography Research and Practice 10. Amsterdam/Philadelphia: John Benjamins.
- Nielsen, S. and S. Tarp (Eds.).** 2009. *Lexicography in the 21st Century. In Honour of Henning Bergenholtz*. Amsterdam/Philadelphia: John Benjamins.
- Nord, C.** 2005. *Text Analysis in Translation. Theory, Methodology, and Didactic Application of a Model for Translation-oriented Text Analysis*. Amsterdam/New York: Rodopi.
- Schaeder, B. and H. Bergenholtz (Eds.).** 1994. *Fachlexikographie. Fachwissen und seine Repräsentation in Wörterbüchern*. Forum für Fachsprachen-Forschung 23. Tübingen: Gunter Narr.
- Schneider, F.** 1998. *Studien zur kontextuellen Fachlexikographie. Das deutsch-französische Wörterbuch der Rechnungslegung*. Lexicographica Series Maior 83. Tübingen: Max Niemeyer.
- Tarp, S.** 2000. Theoretical Challenges to Practical Specialised Lexicography. *Lexikos* 10: 189-208.
- Tarp, S.** 2001. The Pedagogical Dimension of the Well-conceived Specialised Dictionary. *Ibérica* 10: 7-22.
- Tarp, S.** 2008. *Lexicography in the Borderland between Knowledge and Non-knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Tübingen: Max Niemeyer.
- Tarp, S.** 2009. Reflections on Lexicographical User Research. *Lexikos* 19: 275-296.
- 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.
- Van Sterkenburg, P.G.J. (Ed.).** 2003. *A Practical Guide to Lexicography. Terminology and Lexicog-*

raphy Research and Practice 6. Amsterdam/Philadelphia: John Benjamins.

Wang, W. 2001. *Zweisprachige Fachlexikographie. Benutzungsforschung, Typologie und mikrostrukturelle Konzeption*. Angewandte Sprachwissenschaft 8. Frankfurt am Main [etc.]: Peter Lang.

Wiegand, H.E. 1988. Was eigentlich ist Fachlexikographie? Mit Hinweisen zum Verhältnis von sprachlichem und enzyklopädischem Wissen. Munske, H.H., P. von Polenz, O. Reichmann and R. Hildebrandt (Eds.). 1988. *Deutscher Wortschatz. Lexikologische Studien. Ludwig Erich Schmitt zum 80. Geburtstag von seinen Marburger Schülern: 371-409*. Berlin/New York: Walter de Gruyter.

IsiXhosa Lexicography: Past, Present and Future*

Dion Nkomo, *School of Languages — African Language Studies Section,
Rhodes University, Grahamstown, South Africa*
(d.nkomo@ru.ac.za/deeouf@yahoo.co.uk)

and

Zola Wababa, *IsiXhosa National Lexicography Unit, University of Fort Hare,
Alice, South Africa; and School of Languages — African Language Studies
Section, Rhodes University, Grahamstown, South Africa*
(zwababa@ufh.ac.za)

Abstract: This article presents a panoramic and critical overview of isiXhosa lexicography and its impact on the intellectualisation of this indigenous South African language. The history of isiXhosa lexicography dates back more than two centuries. However, there still exists a need for dictionaries that serve the language-speaking community as practical tools for addressing diverse communication and learning-oriented needs in the current language policy dispensation. The IsiXhosa National Lexicography Unit (XNLU) is currently working on dictionary projects that attempt to address this situation while at the same time not losing sight of the mandate that the Pan South African National Language Board (PanSALB) placed on all the National Lexicography Units (NLU). For this to happen, the article argues that the NLU needs to put lexicographic practice into its historical perspective, i.e. conceiving dictionary projects in the light of existing dictionaries and lexicographic traditions in the language. Over and above that, there is a need to take into account the recent developments in lexicographic research, adopt co-operative lexicographic practice and develop a dictionary culture among the isiXhosa-speaking community.

Keywords: ISIXHOSA LEXICOGRAPHY, ISIXHOSA DICTIONARIES, XHOSA DICTIONARY PROJECT, NATIONAL LEXICOGRAPHY UNITS, ISIXHOSA NATIONAL LEXICOGRAPHY UNIT, NATIONAL DICTIONARY, SOUTH AFRICAN LEXICOGRAPHY, RESEARCH-BASED LEXICOGRAPHY, DICTIONARY CULTURE

Opsomming: IsiXhosaleksikografie: Verlede, hede en toekoms. Hierdie artikel bied 'n panoramiese en kritiese oorsig van die isiXhosaleksikografie en die invloed daarvan op die intellektualisering van hierdie inheemse Suid-Afrikaanse taal. Die geskiedenis van die isiXhosaleksikografie strek vir meer as twee eeue terug. Daar bestaan egter steeds 'n behoefte aan woordeboeke wat die taalsprekende gemeenskap dien as praktiese gereedskap om te voorsien in uiteenlopende kommunikasie- en leer-georiënteerde behoeftes in die huidige taalbeleidsbedeling. Die isiXhosa- Nasionale Leksikografie-eenheid (XNLE) werk tans aan projekte wat probeer om hierdie

* This article is a reworked version of a paper that was presented by the authors at the Eighteenth International Association of African Lexicography (AFRILEX) Conference held at the Nelson Mandela Metropolitan University, Port Elizabeth, South Africa, 2–5 July 2013.

situasie aan te spreek, maar terselfdertyd nie die mandaat uit die oog te verloor wat die Pan- Suid-Afrikaanse Taalraad (PanSAT) op al die Nasionale Leksikografie-eenhede (NLE's) geplaas het nie. Om dit te bewerkstellig, voer die artikel aan dat die NLE's hierdie leksikografiese praktyk in historiese perspektief moet plaas, d.w.s. deur woordeboekprojekte in die lig van bestaande woordeboeke en leksikografiese tradisies in die taal te beplan. Bo en behalwe dit is dit nodig om die onlangse ontwikkeling in leksikografiese navorsing in ag te neem, koöperatiewe leksikografiese praktyk te aanvaar en 'n woordeboekkultuur onder die isiXhosaspreekende gemeenskap te ontwikkel.

Sleutelwoorde: ISIXHOSALEKSIKOGRAFIE, ISIXHOSAWOORDEBOEKE, XHOSAWOORDEBOEKPROJEK, NASIONALE LEKSIKOGRAFIE-EENHEDE, ISIXHOSA- NASIONALE LEKSIKOGRAFIE-EENHEID, NASIONALE WOORDEBOEK, SUID-AFRIKAANSE LEKSIKOGRAFIE, NAVORSINGSGBASEERDE LEKSIKOGRAFIE, WOORDEBOEKKULTUUR

1. Introduction

This article presents a panoramic overview of isiXhosa lexicography up to the present. It offers a critical commentary of the major isiXhosa dictionaries and current dictionary products in the view of the line function of the National Lexicography Units (NLUs), of which the IsiXhosa National Lexicography Unit (XNLU) is part, in respect to isiXhosa as one of the official languages of South Africa. The main focus is on the impact of lexicography on this indigenous South African language and the lexicographic needs of the language-speaking community.

The history of isiXhosa lexicography, as briefly outlined in Section 2, dates back more than two centuries. The language currently boasts of a monumental three-volume dictionary, *The Greater Dictionary of isiXhosa* (GDX). This is in addition to the *Kafir-English Dictionary* (KED), the *Oxford English-Xhosa Dictionary* (OEXD), and a medium-sized monolingual dictionary, *Isichazi-magama SesiXhosa* (ISX), and many other dictionaries that are enumerated without detailed discussion in this article (cf. Table 1).

This article considers the strengths and limitations of the existing isiXhosa dictionaries. In the studied dictionaries, particular attention is paid to, among other aspects:

- Purpose of each dictionary (as enumerated in the blurb, preface and introductory text);
- Scope and size of the dictionary;
- Data categories included in relation to the purpose of the dictionary;
- User-friendliness of the dictionary.

Extending a similar approach to the current dictionary projects, this undertaking offers an up-to-date and futuristic view of isiXhosa lexicography, more

than two decades after Peter Mtuze conducted his own survey (cf. Mtuze 1992). Since then, a number of significant developments have taken place. Firstly, the Xhosa Dictionary Project (XDP), at the University of Fort Hare, has been reconstituted into the XNLU, thereby reconfiguring both the primary and secondary lexicographic processes which culminate in the production of isiXhosa dictionaries. Secondly, Volume 2 and Volume 1 of the GDG and the ISX have since been published, thereby rendering the inventory of isiXhosa lexicography richer than what it was in 1992. Finally, the currently on-going projects within the XNLU are being undertaken in a different context of language development, when isiXhosa is now one of South Africa's eleven official languages. This places big responsibilities on the lexicographers.

The survey of isiXhosa lexicography conducted in this article, therefore, is not undertaken for its own sake. It is a stock-taking exercise that is meant to inform the future activities of the XNLU.

The number and size of some of the available dictionaries in isiXhosa project a good image of the language, particularly when compared to the other African languages in general. However, a closer and critical analysis of the dictionaries themselves necessitates a distinction between dictionaries for the language on the one hand and dictionaries for the language users and language learners on the other. The line function of the NLUs which requires them to focus on a particular type of dictionary, if not adopted strategically, is likely to add to a long list of dictionaries that cannot be used by the majority of the language speakers who need them. A strategic approach is required that does not consider the line function of the NLUs as diametrically opposed to periodic and regular publication of other types of dictionaries which address some immediate lexicographic needs in the linguistic community, while ultimately contributing to the intellectualisation of a particular language in the long term. For this to happen, practical lexicography needs to catch up with recent theoretical developments in lexicography in order to address some challenges that are faced in the actual compilation of the dictionaries, thereby ultimately addressing the needs of isiXhosa speakers and learners in a user-friendly way.

2. A Brief History of isiXhosa Lexicography

The history of isiXhosa lexicography can be traced back to the 18th century (cf. Pahl 1989 and Mtuze 1992) and since then, the inventory of isiXhosa has grown as partly shown in Table 1 below. According to Pahl (1989) and Mtuze (1992), the pioneer of isiXhosa lexicography is Sparrman, a natural scientist who visited the Cape coast between 1772 and 1777 and is credited with compiling a short "list of Xhosa numerals, nouns, adjectives and verbs", differing little in form from the words as used today (Pahl 1989: xxxvii). The list was appended to his book entitled *A Voyage to the Cape of Good Hope Towards the Antarctic Polar Circle and Round the World but Chiefly into the Country of the Hottentots and Caffres from the Year 1772 to 1776*, indicated as Appendix to Sparrman (1776) in Table 1.

Dictionary Title	Editor(s)	Year of Publication
<i>Appendix to Sparrman (1776)</i>	Andrew Sparrman	1776
<i>Specimen of the Caffra Language</i>	Dr. J.T. van der Kemp	1803
<i>A Systematic Vocabulary of the Kaffrarian Language</i>	Rev. John Bennie	1826
<i>A Vocabulary of the Kafir Language</i>	John Aylif	1846
<i>A Zulu-Kafir Dictionary Etymologically Explained</i>	Jacob L. Döhne	1857
<i>A Dictionary of the Kaffir Language: Including the Xosa and Zulu Dialects</i>	Rev. Wm. J. Davis	1872
<i>A Kafir-English Dictionary</i>	Rev. Robert Godfrey / Dr. Albert Kropf	1899/1915
<i>A Concise Kafir-English Dictionary</i>	James McLaren	1915
<i>A Concise English-Kafir Dictionary</i>	James McLaren	1923
<i>The Concise Trilingual Dictionary in English, Xhosa, Afrikaans/Die Kort Drie-talige Woordeboek in Afrikaans, Xhosa, English</i>	Lionel E. Jennings	1961
<i>Xhosa Dictionary: English-Xhosa-Afrikaans, Xhosa-English-Afrikaans</i>	H. Nabe, Pw. Dreyer and G.L. Kakana	1976
<i>Oxford English-Xhosa Dictionary</i>	Arnold Fischer et al.	1985
<i>The Greater Dictionary of Xhosa</i>	H.W. Pahl et al.	1989-2006
<i>Isichazi-magama SesiXhosa</i>	Tshabe et al.	2008

Table 1: A partial inventory of isiXhosa dictionaries, their editors and dates of publication

With the exception of the two volumes of the GDX, Volume 1 and Volume 2, as well as the ISX, all the other listed dictionaries were published before the establishment of the XNLU. The establishment of the Unit with a specific line function, therefore, follows a very long history of lexicographic practice in isiXhosa.

However, despite a relatively long list of lexicographic works for an African language, there is a dearth of literature and research on isiXhosa lexicography. As Google search shows, only three hits for both *isiXhosa lexicography* and *Xhosa dictionaries* are academic works that specifically deal with isiXhosa dictionaries. The remainder of the hits only serve to confirm the existence of the dictionaries. Of the previous twenty-two volumes of *Lexikos*, the journal of AFRILEX, which has served to disseminate information about African lexicography, only five make reference to isiXhosa lexicography. Of easily accessible works on isiXhosa lexicography, Pahl (1989) and Mtuze (1992) offer the most comprehensive accounts. In Section 6 of the introduction of the third volume of the GDX, Pahl (1989: xxxvi-xxxix) offers a historical account of isiXhosa lexicography in order to provide the GDX with the necessary context. He benefits

largely from Rev. Robert Godfrey who, in the first edition of the KED, devoted some space to credit his predecessors and review their works. Mtuze (1992) adopts a similar approach in his survey of isiXhosa lexicography between 1972 and 1989, when the third volume of the GDX was published. Another publication worthy of mention because of its specific focus on isiXhosa lexicography is Moropa and Kruger (2000), which deals with the mistranslation of cultural terms in the KED.

The history of lexicography in general, or the historical perspective of dictionary research, in terms of Hartmann (2001), cannot be overemphasised. According to Wiegand (1984), it is important to study the history of lexicography in order to understand and learn from previous experiences and avoid repeating earlier mistakes. This point is also expressed by Hartmann (1983), who observes that every dictionary has a forerunner, as well as Landau (1984), who argues that dictionary making is a judicious act of piracy whereby lexicographers heavily rely on their predecessors. By revisiting the history of isiXhosa lexicography, the authors hope to chart the way forward for future works in this and other South African languages. This is particularly important at a time when both the NLUs and the Pan South African Language Board (PanSALB) are in need of a turn-around strategy for South African lexicography, more than ten years since the NLUs were established.

That isiXhosa lexicography is under-researched is unfitting, given such a long history of dictionary making. Not only lexicographers in the indigenous South African languages can learn from this history, but also lexicographers from other languages across the continent. However, for this to be possible, research on isiXhosa lexicography is still a long way behind. A student or researcher of isiXhosa lexicography still has to rely on primary sources such as the dictionaries themselves and some relevant documents which are in the possession of the XNLU, PanSALB and dictionary publishers. Furthermore, the researcher of isiXhosa lexicography still has to benefit from research on the lexicography of the other languages in streamlining research topics. To really make progress, the researcher may have to start from very general surveys before venturing into specific and detailed engagements with certain lexicographic issues. What this article hopes to achieve is to open up isiXhosa lexicography for metalexigraphic scrutiny by identifying, albeit in less detail, some key areas for research.

2.1 *A Kafir-English Dictionary*

The KED was first published under the editorship of Rev. Albert Kropf in 1899 before its second, and so far, the latest edition, edited by Rev. Robert Godfrey, in 1915. According to its original editor (Kropf 1899: iii), it was first proposed as "a memorial of the Kafir language which would soon be supplanted by English". In terms of typology, the KED is a mono-directional bilingual dictionary which describes isiXhosa using English. Beyond the fact that the language

lacked the relevant metalanguage for lexicographic description, the use of English as descriptive language is consistent with the view harboured by some missionaries when the dictionary project was first proposed. The language was being archived for the benefit of non-speakers of isiXhosa who had to study the language in order to understand the mother-tongue speakers (cf. Gallardo 1980).

Contrary to the original view among some missionaries, the KED had an unintended outcome. It has become an important element of the lexicographic inventory of the language, a real milestone in the intellectualisation of isiXhosa. Pahl (1989: xxxviii) regards Albert Kropf, as "the first lexicographer to have fathomed and thoroughly mastered the intricacies of Xhosa phonology ... so thorough and accurate that little fault can be found with it". He further suggests that "all future lexicographical work in isiXhosa must be based upon it". The usefulness of this dictionary has transcended the isiXhosa-speaking community of South Africa and is one of the South African dictionaries, the others being in Zulu, that have been used in Zimbabwe by Ndebele speakers and scholars (cf. Hadebe 2006). It has also transcended the period during which it was produced, with Pahl (1989: xxxviii) rightly observing that it "has stood the test of more than three quarters of a century". A number of isiXhosa academics and language practitioners continue to use it as an invaluable reference for different purposes, such as translation. Because of its enduring importance, the KED has been converted into a PDF document, thereby turning it into an electronic resource to increase its accessibility, portability and durability many years after it went out of print.

The article below is an example of the fairly comprehensive treatment of lemmata in the KED. As can be seen from this article, type of speech data, meaning, both literal and figurative, as well as usage examples are provided for the verb *ukutwala* (to carry).

uku-TW'ALA, *v. t. perf. twèle*. To carry, bear a burden: *bamtwála ofayo ngokúko*, they carried the sick person on a bed; of a tree, to bear fruit; fig. to endure, bear: *iábatáni idyokwe yam niyitwále*, take my yoke upon you. Phr. *wamtwála ngentshuntshe*, he stabbed him; *ukutwála ngamehlo*, to be fierce, wild, excited, incensed; *ukutwál' iyala*, to incur blame; *ukuyitwála intombi*, to abduct or carry off a girl. A girl who is being so abducted may yell piteously (*yo mama*) without attracting any notice, as the matter is too often done by arrangement with her friends. Often enough, however, the phrase simply means 'to elope with', as the young man and the girl have planned the matter between them, with the purpose of hastening the marriage or of avoiding the expense incurred by marriage.

Example of a Dictionary Article from the KED

The KED also has thirty-two pages of front-matter texts in which, among other issues, it describes the grammatical features of isiXhosa and enumerates their treatment in the dictionary. The dictionary, therefore, fits in with the earliest grammars of isiXhosa, but there would be need to improve or revise some of its grammatical representations.

Perhaps its major impediment in post-apartheid South Africa is the K word in its title, the main feature of many other isiXhosa and isiZulu dictionaries published during the colonial and apartheid periods (cf. Table 1). This word is offensive, its presence in the title remains a major put-off to some potential users of the dictionary today. Besides the title, the dictionary has also been criticised for the mistranslation of certain culture-specific terms (cf. Moropa and Kruger 2000). Associating a traditional marriage practice with abduction in the above KED article is a good example. In modern times, abduction would be associated with a criminal activity and this limits the dictionary in terms of its communicative and cognitive functions. Finally, with its second edition having been published in 1915, the old orthography becomes another impediment of the dictionary. This can be illustrated by the KED article where *uku-twala* (to carry) is entered instead of *ukuthwala* and other entries such as *tabatani* (take) instead of *thabathani* in the same article. Thus, notwithstanding its usefulness, modern lexicographers and language users will need to pay attention to such shortcomings of this otherwise timeless work of isiXhosa lexicography.

2.2 The Oxford English–Xhosa Dictionary

Based on the *Oxford Learner's Dictionary*, the OEXD is a mono-directional bilingual dictionary with as its primary purpose, the learning of isiXhosa as an additional language. In its preface, Fischer (1985) states "the idea of this *English–Xhosa Dictionary* was conceived during many years of studying and teaching Xhosa, and being confronted with the lack of an appropriate English–Xhosa Dictionary". The users are expected to know and use English as the point of departure when learning isiXhosa. Thus, English is the lemmatising language for which data categories meant to support mainly text production in isiXhosa are provided. The following is a typical article from this dictionary:

arm, n. ingalo (in- iin-); (sleeve) umkhono; (large branch) isebe elikhulu (ili- ama-); ~.ful, n. ubuninzi; ~-band, n. isacholo (isi- izi-); ~.chair, n. isitulo esinezixhaso-ngalo; ~-hole, n. umgxuma womkhono; ~.let, n. isacholo; ~.pit, n. ikhwapha (ili- ama-); the ~ of the law (fig.): umthetho; baby/child/ infant in ~.s : usana (ulu-iintsana); at ~'s length (fig.: avoid becoming familiar with somebody): -gcina kude with open ~.s (warmly, with enthusiasm): ngobubele; walk ~-in- ~ : -xhakana.

Example of a Dictionary Article from the OEXD

However, Fischer (1985) reports that the discussion of the original idea of this dictionary with teachers indicated "a long-felt need amongst Xhosa-speaking students as well", particularly when "confronted with essays and literature in the English curriculum". This resulted in the inclusion of literary English words which are explained in isiXhosa. Pahl (1989: xxxix) describes this dictionary as "a godsend to students of Xhosa" as it was produced in the then new orthography, compared for instance to the KED. Furthermore, the enriched dictionary articles also facilitate the comprehension of lemmata using isiXhosa, meaning that the dictionary also provides limited English-comprehension support to isiXhosa mother-tongue speakers. Finally, its size also makes the dictionary handy for Xhosa-English or English-Xhosa translators, interpreters and editors.

With the latest reprint of 2011 being the 21st impression of the OEXD, this dictionary is without doubt the most well-known and successful isiXhosa dictionary on the market. For example, in both 2012 and 2013, an annual pre-learning questionnaire survey that is conducted at the beginning of the undergraduate lexicography course at Rhodes University indicates that *isiXhosa dictionary* is synonymous with the OEXD. This knowledge is based on the students' previous but limited use of the dictionary at high school and its recommendation as an important reference work for isiXhosa students at the university, especially for translation studies. Despite this success, a close study of this dictionary is recommended to determine whether or not future reprinting should be accompanied by revision of certain areas. From a cognitive perspective, the dictionary is limited, like many other dictionaries in isiXhosa, as it displays cultural prejudices where the verb *abduct* is explained as forcefully dragging a girl into a marriage (cf. OEXD: 1). Yet *abduct* generally means much more than this.

2.3 *The Greater Dictionary of IsiXhosa*

The Greater Dictionary of isiXhosa (GDX) is a three-volume trilingual isiXhosa dictionary in which English and Afrikaans are the other featured languages. Its volumes were published separately and, curiously, in reverse order, i.e. Volume 3 (Q-Z), Volume 2 (K-P) and Volume 1 (A-J) in 1989, 2003 and 2006 respectively. Although the GDX was completed in the NLU dispensation, it is a product of the Xhosa Dictionary Project (XDP) which started in 1968. Having co-opted the XDP initiative into the NLU initiative which had a particular line function, the XNLU was allowed to complete this monumental lexicographic endeavour. The project was conceived on the backdrop of an acute need for "a modern, scientific standard Xhosa dictionary" (Pahl 1989: xxxix), the revision of isiXhosa orthography and a period of more than a century since the KED was out of print. Following the initiative of the rector of the University of Fort Hare and the endorsement by its Council, the project was to be housed at the University of Fort Hare.

According to Pahl (1989: xxxix), "the Fort Hare University Council had decided that the Xhosa entries were to be defined in English and Afrikaans, but

at a meeting of the Xhosa Dictionary Committee in 1968 it was decided that the definitions be given in Xhosa as well, thus making the dictionary fully trilingual". As a result, the GDX would appear as illustrated below:

úkù·thwálà nz/v *-túád- (mi/stat -thwélè, -thwelè; nzn/rec úkùthwálàná; nzk/met-pot úkùthwálèkà; nzl/ap úkùthwálèlá; nzs/caus úkùthwálísà; nzw/pass úkùthwálwà):

1 okomntu: ukuhamba ubeké into entloko okanye emagxeni okanye ezingalweni okanye kwinto enjengenqwanqwa uyisa endaweni; okwehashe okanye iqegu: ukuhambisa (umntu) okhwele okanye (into) ebekwe emhlana:
abafazi bazithwala entloko iinyanda zabo:

2 ukunxiba, ukubeka (isikhuselo okanye into yokuhamba entloko, njengomnqwazi, iqhiya, intshinga, isithsaba, njl):

3 okwesithuthi: ukulayisheka (umthwalo onga):

le lori ithwala ikhulu leengxowa:

4 okwesikhongozeli esikhulu, njengefatyi, itanka, ibhasi okanye iholo: ukungena (inani okanye umyinge othile):

le holo ithwala abantu abaliwaka:

5 okwemithi okanye izityalo: ukuxakatha, ukuba nezizhamo; ukwelisa izizhamo:

6 mi: okwaso nasiphi na isikhongozelo: -zele, -qulathe kanobom:
le tanka yamanzi ithwele:

7 okwento engesosithuthi enjengeentambo zombane, umbhobho, njl: ukuba nokuhambisa okanye ukusa (othile) kwenye indawo:

lo mbhobho ithwala amanzi athi anelise idolophu yonke:

8 ukuthabatha (ibhinqa) ulisa ekwendeni ngenkani; kgl Add 33:

9 ukuthi qhiwu (into enjengomsila, intonga, njl):

lakutyiwa sisibawu ithole lathwala ishoba lijikeleza umzi:

1 of a person: carry (a burden on the head or shoulders, in the arms or on a device like a stretcher); of an ox or horse: carry (a person or load); cf *ukubeleka, ukuphatha*:
the women carry their bundles of wood on their heads;

2 wear (something) on the head, eg a hat, head-cloth (SAE doek), head-scarf, ornamental headdress, crown, etc;

3 of a vehicle: be able to load (so much), have a carrying capacity of:
this lorry can carry a hundred bags;

4 of a large container, eg a barrel, tank, bus or hall: be able to contain, accommodate, hold (so much/many):

this hall holds a thousand people;

5 of trees or crops: yield, produce;

6 stat: of any container: be full, contain a considerable quantity:

this tank contains a good supply of water;

7 of a medium of conveyance other than a vehicle, eg an electric cable, water-furrow, water-pipes, etc: be able to convey or conduct (so much of a commodity):

this pipe carries enough water for the whole village;

8 abduct (a girl) as a wife for a man; see Add 33;

9 hold (something) up while walking or running, eg a horse its head, a calf its tail, a warrior a stick, etc:

when the calf was bitten by a gadfly, it ran around the homestead with its tail in the air;

1 van 'n persoon: ('n las of drag op die kop of skouers, in die arms of op 'n toestel, soos 'n draagbaar) dra; van 'n lasdier: ('n mens of vrag) dra; cf *ukubeleka* en *ukuphatha*:
die vroue dra hul dragte hout op hul kappe;

2 (iets, soos 'n hoed, doek of ornamentele band, kroon, ens) op of om die kop dra;

3 van 'n voertuig: ('n bepaalde massa of hoeveelheid) kan dra:
hierdie vragmotor kan 'n honderd sakke dra;

4 van 'n groot houder, bv vaatjie, bus, saal, ens: ('n bepaalde hoeveelheid) kan neem, hou, huisves:

hierdie saal huisves 'n duisend mense;

5 van 'n boom of oes: voortbring, dra;

6 stat: van 'n houer: vol wees, 'n goeie voorraad bevat:

die tenk het 'n goeie voorraad water;

7 van 'n geleidingsmedium, bv elektriese draad, watervoor, ens: in staat wees om ('n bepaalde hoeveelheid) te geleel:

hierdie waterpyp dra genoeg water vir die hele dorp;

8 ('n meisie) skaak as vrou vir 'n man; sien Add 33;

9 (iets) hoog hou terwyl geloop of gehardloop word, bv 'n perd sy kop, 'n kalf sy stert, 'n kryger 'n kierie, ens:

toe die kalf deur die steekvlieg gepla is, het hy stert in die lug om die kraal gehardloop;

Example of a Dictionary Article from the GDX

IsiXhosa is the lemmatising language but all three languages are used to provide explanations and other forms of lexicographic description. Because the dictionary also uses isiXhosa to provide information, it caters for the needs of both mother-tongue and non-mother-tongue speakers of isiXhosa. This was a major break from the tradition whereby all information about the language was hitherto given through another language, implying that those who did not know such a language, i.e. English or Afrikaans, could not benefit.

Besides the main text, the dictionary provides encyclopaedic information about isiXhosa and some culturally significant aspects of the language and its speakers. For example, between pages 685 and 754 of the third volume, the dictionary contains back matter texts which provide information on isiXhosa grammar and aspects of folklore in isiXhosa, e.g. *tikoloshe*, *mamlambo*, *ichanti*, *impundulu*, etc. as well as cultural practices such as circumcision, *ukuthwalwa kwentombi*, etc. As a container of knowledge (McArthur 1986), this dictionary provides information that non-mother-tongue speakers of isiXhosa and the

younger generation of isiXhosa speakers may find very valuable. Thus the dictionary is rich in both linguistic and extra-linguistic (encyclopaedic) information.

However, despite its stature and richness, this dictionary has probably made more impact on the language itself than on the speakers of the language. Hence the distinction made between dictionaries for the languages and dictionaries for the language speakers and learners. In other words, it has not proved to be a popular reference work, especially for students at both university and school levels. This has less to do with the quality of the dictionary, but more probably with its voluminous nature. Because of the amount of resources that were required to produce it, the dictionary is being sold at R425 per volume, which adds up to R1 275 for the three volumes. In addition to its price, the three-volume dictionary is not a portable reference work. As a result, over 4 300 of the 4 500 copies (1 500 per volume) are still with the publisher instead of dictionary users, which indicates that the dictionary has not sold very well. It represents archival material for isiXhosa, but is not a utility product as a dictionary should be.

2.4 *Isichazi-magama sesiXhosa*

Isichazi-magama sesiXhosa (ISX), published in 2008, is a medium-sized monolingual general-purpose dictionary. It is the first monolingual dictionary in isiXhosa, coming after a long tradition of the production of bi- and multilingual dictionaries. This dictionary was conceived and compiled within the NLU's dispensation, but it had to await the completion of the GDX. Its Editor-in-Chief confirms this in the acknowledgement section when he (Tshabe 2008: iv) writes:

Okokuqala mandiwenze umbulelo kuPanSALB ngenkxaso athe wasinika yona sileli Ziko lesiXhosa leSizwe loChazo-magama kuyo yonke le minyaka. Ndi-bamba ngazibini ngembono yakhe yokuba kuqalwe ngokuqulunqwa izichazi-magama ezilulwimi olunye. Usivumele ke thina kweli Ziko ngengebiso yeBhodi yabaLawuli beli Ziko okokuba sigqibezele umsebenzi owawusele uqaliwe kwangoo-1968 woqulunqo lwesichazimagama esiluwimi-ntathu esibizwa ngokuba yi *The Greater Dictionary of IsiXhosa*.

Firstly let me acknowledge PanSALB for supporting us as the IsiXhosa National Lexicography Unit for all these years. I appreciate the view that we should start to compile monolingual dictionaries. Yet PanSALB allowed our Unit, with the advice of the Board of Directors, to first complete the project of compiling the trilingual dictionary, namely *The Greater Dictionary of IsiXhosa*, which began in 1968.

This dictionary, therefore, may be regarded as a major step towards the accomplishment of the line function of the NLUs, i.e. to compile national dictionaries for the official languages (cf. Section 3). As a standard dictionary, the main components of the articles in the ISX are the lemma, word-class data, noun

classes for noun lemmata, explanations of meaning and illustrative examples, as can be seen from the example below. The dictionary also provides some data types characteristic of national dictionaries as linguistic and cultural reservoirs. For example, limited etymological data is given in the form of the source languages of borrowed words treated as lemmata, with abbreviations such as *Afr* (Afrikaans), *Eng* (English), *Jap* (Japanese) and *SAE* (South African English). *Isihlonipho* (euphemistic language of women), described in the front-matter, is also a key element of the cultural data that is included in the dictionary, but it is clear that its provision and treatment is even less comprehensive and less satisfactory compared to the GDX. Thus, there is still some way to go to accomplish the line function of the NLUs, for the national dictionary will have to be more comprehensive (cf. Section 3).

i-khwapha b 5/6:

1. igolonxa okanye indawo ethe
gongxo ngaphantsi ekudibaneni
kwengalo negxalaba:
2. imo yokunuka kwamakhwapha;
ivumba lokunuka kwamakhwapha,
umz kumntu ongakhange ahlambe,
okanye kumntu onuka amakhwapha
ngendalo: *ukho lomntu usiphethe
ngekhwapha kweli holo: lo
mntwana unekhwapha emncinci
kangaka.*

Example of a Dictionary Article from the ISX

Nevertheless, the landmark that this dictionary constitutes for isiXhosa has to be appreciated, in as much as it is a clear contribution towards the national dictionary. It is the first and so far the only dictionary that provides information on isiXhosa using the language itself. In an elated foreword, the XNLU Chairperson of the Board of Directors, Jadezweni (2008: ii), writes:

Baqondile ukuba mabakhe bahlukane nokuxhobisa ezinye izizwe ngolwimi lwabo — kaloku isiXhosa esi seside ngathi yinto emayifundwe zezinye izizwe ukuze zizazi iingqondwana "zala maXhosa". Kungoko le ncwadi yesigama ichaziweyo iza kunika isidima abanini-ntetho kuba ulwimi lwabo nalo lunakho ukwenza zonke eziya zinto zenziwa zezinye iilwimi zehlabathi. Okunene nasi isigama setekhnoloji sesichazwa ngesiXhosa njengoko nesiNgesi sisichaza.

(They (Xhosa speakers) are correct to stop developing other 'nations' using their language — it has been long since isiXhosa was treated as something that had to be learnt by other 'nations' in order to understand the mindset of "these Xhosas". As a result, this book that is being discussed will give dignity and confidence to the speakers of the language, that their language can also serve functions that have been served by other languages of the world. Surely, even the terminology of technology is also being defined using isiXhosa in the same way as English.)

Clearly, there is resonance to the sentiments expressed by Béjoint (2000) that monolingual general-purpose dictionaries are national emblems and expressions just like national flags and national anthems, a view that informed Samuel Johnson's approach to the mission of compiling *The Dictionary of the English Language* (cf. Johnson 1747). It is clear, therefore, that the future of the language and its ability to fulfil the ideals that saw isiXhosa being accorded its status as one of the country's official languages is felt in the dictionary. The mandate that PanSALB placed on the NLUs, therefore, is conceived in such a nationalistic spirit. However, it is critical to consider this mandate in the light of language planning in South Africa, of which lexicography is part, and the lexicographic needs of the speakers and learners of the respective languages.

3. Towards the National Dictionary: Current Dictionary Projects in isiXhosa

After this brief review of isiXhosa dictionaries, the future of isiXhosa lexicography is now considered in the light of current dictionary projects. Perhaps the most crucial question is whether the XNLU is on course to accomplish its line function of producing a national dictionary for isiXhosa. Again, it would be important to discuss the reason for the current state and possible direction for isiXhosa lexicography.

As noted in the previous section, lexicographic practice in isiXhosa predates the establishment of the NLUs in South Africa by over two centuries. Through the XDP, an organised structure for lexicographic work already existed in the language. This structure was considered ideal to be co-opted into the new lexicographic dispensation. Yet it was equally vital that the line function of the NLUs had to be co-opted into an established tradition for the language to develop and for both its speakers and learners to obtain dependable reference tools.

A closer look at the dictionaries produced and some current projects by a number of NLUs suggests that there is no consensus with regard to the line function of the NLUs and how this function could be accomplished. Consequently, the optimism that accompanied the establishment of the NLUs has turned into frustration as their mandate seems to be difficult to execute. Whereas different NLUs are engaged in projects which seem to be distanced from their line function, the NLUs do not find it difficult to justify their projects in view of the lexicographic needs of the linguistic communities they seek to serve and the developmental stages of the respective languages. The XNLU is no exception.

The terms of reference for the establishment of the NLUs were first outlined in the National Lexicography Units Bill (B 103-96), which is discussed in detail by Kumalo (1999), and further encapsulated in the Articles of Association of the NLUs. According to the Bill, the objectives of the NLUs "shall be to initiate, maintain, continue, complete and from time to time improve the compila-

tion of *the dictionary and other products*" (B 103-96: 4). The definite article *the* underlines the line function of the NLUs as focusing on a particular type of dictionary. The definitions section of the Bill describes this dictionary as "the definitive product, on the general vocabulary of an official South African language, as produced by the unit for the language concerned" (B 103-96: 3). This dictionary is the national monolingual general dictionary (Alberts 2011), or, simply put, the national dictionary. The national dictionary is defined in Hartmann and James (1998: 98) as "[a] cultural or historical dictionary of significance to the language of a country, often financed by central government or an academy and compiled by a team of lexicographers". To the best of our knowledge, none of the NLUs for the nine indigenous South African languages are currently working on such a dictionary. Instead they are busy with various types of dictionaries which, in terms of the Bill would be accounted for as the "other products" (B 103-96: 4). It is perhaps due to the failure of the NLUs to make the anticipated progress concerning their line function, and the uncertainty that they can realistically achieve this feat in the foreseeable future, that Alberts (2011: 51) argues:

The NLUs need to be given the resources that are necessary to revert to the original idea of producing properly researched, definitive, monolingual dictionaries for each of the official languages instead of compiling all kinds of dictionary products to generate funds (dictionaries that commercial publishers are able to compile much better, and that they are already publishing successfully).

The XNLU is currently working on three projects, namely, the Mathematics and Science Pedagogic Dictionary, the Concise Trilingual Dictionary and the IsiXhosa Monolingual Dictionary. A brief discussion of these projects poses a number of questions in relation to the progress of isiXhosa lexicography and the line function of the NLUs.

The product of the first project is a specialised dictionary intended for teachers and learners in the intermediate and senior phase of primary schooling. It represents the Unit's endeavour to promote mother-tongue education through a dictionary that facilitates access to, and communication in, mathematical and scientific content by learners and teachers in isiXhosa. The project, therefore, addresses communication and cognitive problems in the respective subject fields of Mathematics and Science by describing specialised concepts, documenting, explaining and illustrating the use of the relevant terminology using isiXhosa which is one of the main languages of the Eastern and Western Cape Provinces. The first draft of this dictionary has been reviewed for publication and the Unit is currently incorporating the suggestions of the reviewers who included subject specialists (Mathematics and Science), a mother-tongue isiXhosa linguist and a lexicographer.

The second project is the Concise Trilingual Dictionary Project. Its main intention is to abridge the bulky GDX into a concise, portable and user-friendly dictionary that can be sold at an affordable price and be easily used by school

and university students, both mother-tongue isiXhosa speakers and non-mother tongue speakers whose mother tongue is either Afrikaans or English. The GDX is, thus, the dictionary basis for the current project. The Unit contracted a Cape Town-based company to convert the hard copy of the GDX into an electronic rich text format (RTF) document. GDX entries were then copied into a Word document which can be edited, with editing focusing on abridging the GDX by eliminating certain types of entries while adding others, e.g. curriculum words collected from the Department of Education. The Unit intends to make use of TshwaneLex when the actual dictionary writing stage commences after the establishment of the dictionary basis and dictionary files.

The last of the current projects, i.e. the IsiXhosa Monolingual Dictionary Project, also emanates from the GDX. Its rationale is based on the size of the GDX, but unlike the concise trilingual dictionary project, it eliminates all the Afrikaans and English entries to result in a monolingual dictionary. Again, not only eliminating the entries in the other languages is the main focus, but the intended dictionary will also be updated by adding baseline curriculum terms from the Eastern and Western Cape Departments of Education and other new words. The target users of the dictionary are school and university mother tongue students of isiXhosa. The envisaged dictionary should be an updated and a more comprehensive version of the ISX.

Although only brief insights have been provided with regard to the current XNLU projects, it is clear that the Unit does not currently have the national monolingual general isiXhosa dictionary as one of its projects. Looking backwards and forwards, the following questions may therefore be posed concerning the current dictionary projects:

- How different is the mathematical section of the Mathematics and Science Pedagogic Dictionary from *Isichazi-magama Sezibalo* published by Cambridge University Press in 2010?
- How different is the scientific section of the Mathematics and Science Pedagogic Dictionary from *Isichazi-magama sezeNzululwazi neTeknoloji*, a Science and Technology Dictionary published by New Africa Books?
- Is size the only real problem with the GDX such that the Concise Trilingual Dictionary Project will result in a user-friendly dictionary?
- Will one concise version of the GDX be suitable for use by school (taking into account the different school levels) and university students?
- Will the IsiXhosa Monolingual Dictionary Project produce a dictionary that would be fundamentally different from the ISX that was published in 2008?
- Overall, how do the current dictionary projects advance isiXhosa lexicography cognisant of the line function of the XNLU?

These questions remain crucial for the XNLU to produce new dictionaries that

will not only be snippets and combos of previous dictionaries, but tools that will contribute towards the intellectualisation of isiXhosa. No concerted attempt will be made to address each of the questions here, for equally elaborate discussions pertaining to the individual dictionary projects would be required to address these questions. Such discussions will constitute the research-based lexicographic practice that the XNLU aspires to adopt as it moves forward. Here it will suffice to focus on the last two questions in as much as they resonate with the argument of Alberts (2011) cited above.

There is no disputing that isiXhosa, like all the other indigenous African languages, needs a comprehensive monolingual general dictionary in the mould of the *Woordeboek van die Afrikaanse Taal* in Afrikaans and other such dictionaries in more developed languages. Obviously, the NLUs will need more resources to conduct thorough research towards such dictionaries. The current budget for the NLUs, which is not even enough for the salaries of the staff, is a clear under-investment in this area. However, channelling all the available resources towards such a dictionary may not be ideal for the indigenous African languages in which other small dictionaries are needed to assert the official functions that the democratic constitution bestowed on these languages (cf. Mini 1996). Expecting commercial publishers to fill this big void by producing such dictionaries may also be asking for too much in the current situation where the implementation of the multilingual policy has been half-hearted. The success with which the commercial publishers have been producing other dictionaries for all the indigenous languages remains limited and questionable but understandable. Since commercial publishers are profit-driven, they will always be reluctant to produce books in African languages, unless the government commits funds to subsidise their costs. Strategies are needed that will, for instance, enable the NLUs to focus on the national monolingual general dictionary projects while producing lexicographic spin-offs in the form of other dictionaries that can make an immediate impact in the education sector, for instance. In any case, the material that is being used for current projects may also be employed for the comprehensive national monolingual dictionary and the XNLU may be considered to be on the right track, especially with the current IsiXhosa Monolingual Dictionary Project.

4. Other Issues for IsiXhosa Lexicography Moving into the Future

There are other crucial issues that should be considered for isiXhosa lexicography to move forward into the future. Some of these are discussed in the subsections that follow below.

4.1 Research-Based Lexicography

The establishment of the NLUs emerged as a response to the democratic language policy that sought to elevate South Africa's indigenous languages

(Gouws 2003) and their successful establishment may be attributable to this enabling environment. However, the major role played by lexicographic research should never be underestimated. It was crucial for the lexicographically informed to make research-based submissions and presentations to the policy-makers, and lexicographers such as Professors Danie Prinsloo and Rufus Gouws, as well as Dr Mariëtta Alberts, deserve special mention. It was also these lexicographers who organised and ran workshops on various lexicographic issues before and after the establishment of the NLUs, thereby ensuring that dictionary work would not be conducted by unskilled people (cf. Gouws 2001).

During the early years of the NLUs, the lexicographers working on different languages would also present papers at AFRILEX and other conferences on the challenges encountered in their work and the procedures they adopted to address such challenges. This provided an opportunity where lexicographic practice and theory would interact and enrich each other. Such enthusiasm later diminished. However, the Dictionary Roundtable session at the Eighteenth International AFRILEX Conference, held at the Nelson Mandela Metropolitan University, brought a revival of interest as the NLUs presented status and progress reports on their projects, in addition to showcasing some of their products. It became clear that despite some concerns with their achievements, they were, however, striving to contribute towards the development of their respective official languages, albeit under difficult circumstances.

Moving forward, the XNLU intends to engage in research-based lexicography for a number of reasons. Firstly, with the user-perspective having gained prominence in modern-day lexicography (Hartmann and James 1998), the Unit needs to be fully aware of the user needs and reference skills. In this way, the lexicographers will be able to conceive dictionary projects that will be of functional value. Secondly, the XNLU should open itself for metalexicographic research. This may entail publicising its products and projects in order to generate feedback and ideas from dictionary users and metalexicographers that may improve their products. In this way, they may also learn from the experience of their counterparts working on other languages locally and abroad. Finally, the concept of research-based lexicography will involve the lexicographers from the Unit in investigating language-specific issues, using linguistic, educational and lexicographic theories to arrive at well-informed decisions regarding their practices. Research-based lexicographic practice will not only improve the quality of isiXhosa dictionaries, it will also contribute to the available body of knowledge about South African lexicography, which is currently lacking when it comes to certain languages. It will contribute towards developing and entrenching a beneficial dictionary culture and an informed dictionary criticism.

4.2 Developing a Dictionary Culture

The practical work of compiling dictionaries at the XNLU needs to be accom-

panied by the development of a dictionary culture among the potential dictionary users. Whereas lexicographic practice on its own will contribute towards the development of the language, developing a dictionary culture will go a long way in empowering the isiXhosa speech community through the products of that lexicographic practice. As Gouws (1996: 97) argues, dictionary-making is a cultural practice that should lead to another cultural practice, i.e. dictionary use. If the NLU's continue to produce dictionaries that nobody uses, then it would be difficult for them to get feedback that will improve their practice. Above all, they would have laboured in vain. The little resources that are allocated to the NLU's would turn out to be a waste. Yet in the South African context, it was mainly the need to develop the languages so that they could be used for expanded social spaces, that placed emphasis on the dictionaries (Gouws 2003) and other corpus development outputs. Language users and language learners should venture into those spaces that were previously reserved for English with, among other resources, dictionaries as their communication and learning instruments. Not only do they need dictionaries, they also need the reference skills that will enable them to gain optimally from using dictionaries. At the moment, isiXhosa dictionaries are generally not well-known. Dictionaries are still less used at schools and students proceed to university with little or no knowledge of the available dictionaries especially in the African languages. The XNLU has a part to play in eradicating this problem. They need to go to schools to promote their dictionaries and demonstrate their importance in the real situations of the potential users. A project is in the pipeline in this regard (cf. Section 4.5).

4.3 Lemmatisation in isiXhosa Dictionaries

As in Bantu languages in general (Gouws and Prinsloo 2005a; 2005b), lemmatisation remains another critical issue that should be given attention by both practising and theoretical lexicographers in isiXhosa. This pertains to the choice between stem-based and word-based lemmatisation approaches in entering and ordering lemmata. The last time that the issue seems to have been given serious thought was when, in the second edition of *A Kafir-English Dictionary*, Godfrey (1915: ix) stated:

The nature of the structure of the prefix-using Kafir language, differing absolutely from that of the suffix-using English, renders the construction of a Kafir dictionary on English lines an impracticability. ... Were we to follow the alphabetical order of Kafir words, as we do in English, we should find the great bulk of the dictionary entered under the vowels *i* and *u* and we should have very little use for any other letter. For this reason Kafir scholars have practically agreed that a Kafir dictionary should follow the alphabetical order of stems rather than words. Kropf's dictionary was constructed on this plan.

Since then, all dictionaries in which isiXhosa is the lemmatising language have

followed suit. The advantages of both approaches have been discussed before (cf. Gouws and Prinsloo 2005a, 2005b). On the stem-based approach, Godfrey (1915: ix) cautiously concluded:

Although the greatest care has been exercised to discover the stems and to place under them only such words as are actually derivatives, further study of the language will demand a certain amount of re-arrangement.

It is such a further study that is advised in this article to determine whether the stem-based approach should continue to be adopted for all dictionaries in which isiXhosa is the lemmatising language, especially at a time when a shift has started in its sister languages such as isiZulu. Gouws and Prinsloo (2005b) observe that if applied wisely in the form of left-expanded article structures, the approach offers the best of both approaches. However, apart from further research that can ascertain user competences when it comes to dissecting words into prefixes and stems, this article also advises that in dictionaries targeted at young users and those dealing with specialised languages, where providing information regarding the structure of the language is not important, a word-based approach should be adopted.

4.4 Electronic Lexicography

Computer assistance was identified as a potentially important element in the formulation of the general theory of lexicography (Wiegand 1984). There are several ways in which computer technology has since transformed and improved practical lexicography (cf. De Schryver 2003, Sinclair 1987). This includes the use of electronic corpora for a variety of lexicographic tasks and dictionary writing systems which permit simultaneous participation of different lexicographers in one project even from different workstations. The XNLU and other NLUs need to exploit the opportunities of electronic lexicography in their practice. The national monolingual dictionary projects would benefit much from the use of corpora which would store large masses of linguistic data and display the various linguistic nuances that such dictionaries should capture. Electronic databases may also enable the lexicographers to enter all the data that will eventually be included in the national monolingual dictionary project but suppress it for the purposes of deriving smaller dictionaries from time to time. Even more, some dictionaries, including the national monolingual ones, may be published as electronic dictionaries to avoid publishing costs and render the dictionaries organic in the sense of being continuously updateable without devaluing previous publications.

The lexicographers working for the NLUs need to be equipped with the skills that will ensure that their computers are not nice-to-have gadgets or mere versions of type-writers but tools that will enable them to store data, process it and address language-specific issues such as lemmatising and defining in Bantu languages.

4.5 Co-operative Lexicography

Given that the NLUs constitute the core of South African lexicography and serve as a testimony of the government's commitment towards lexicography, it should always be borne in mind that they are a vital part of the official language planning efforts in the country. The concept of co-operative lexicography was proposed from the outset of the establishment of the NLUs (Kumalo 1999: 214) and came up again in their recent review (Alberts 2011: 35-37). Co-operative lexicography should be part of strategic operations which can maximise the limited resources that the NLUs have to work with while at the same time maximising on positive impacts of lexicographic projects and products on the development, learning and use of all official South African languages. A few cases of co-operative lexicography are under way or in the pipeline for isiXhosa lexicography undertaken at the XNLU.

Firstly, co-operative lexicography is essential in the XNLU's bi- or multilingual dictionary projects, given that its main focus and linguistic expertise is mainly in one language. The Unit would have to collaborate with its peers working on the respective languages covered in particular projects. So far, the Dictionary Unit for South African English (DSAE) has supplied the XNLU with its English wordlist that it compiled for its own projects and this will consolidate the dimension of the latter's projects covering English. The Bureau of the WAT may, for instance, also have to be consulted for assistance with the Afrikaans section in the Trilingual Concise Dictionary Project. In this way, it may be possible to optimise the assistance that different sets of language speakers and learners may obtain from one dictionary.

Secondly, the XNLU needs to strengthen collaboration with other stakeholders such as the Department of Education, schools and universities whose key activities constitute dictionary user situations and determine lexicographic needs. Such stakeholders need to be involved as advisors and participants in some lexicographic projects, and most crucially as potential users, future compilers and critics of prospective products. In this way, the Unit will be able to create public awareness of its activities, develop a dictionary culture among the potential dictionary users and generate constructive feedback on its products. That the University of Fort Hare library did not have copies of the GDX until 2011 is an illustration of how weak collaboration has been between the NLU and its host institution, a case of a poor tenant and landlord relationship. Currently the Unit, working with the African Language Studies Section of Rhodes University's School of Languages, have put together a joint-community engagement plan through which dictionary skills workshops will be conducted in Eastern Cape schools.

Lastly, the XNLU and all the NLUs need to find ways of working closely with commercial publishers. There is much that can be gained by co-operating rather than competing with commercial publishers who have much experience and resources required for successful lexicography. The DSAE and WAT collaboration with publishers such as Longman and Pharos have proved to be

effective in facilitating the production of dictionaries that fall outside the line function of the NLUs but serve important immediate functions, especially in education. Hall et al. (2013) utilised the publishers slot in the Dictionary Roundtable at the Eighteenth International AFRILEX Conference to highlight some key considerations for the NLUs seeking publication deals with commercial publishers. In order for the NLUs to find some common grounds with publishers, the importance of a business plan would be reiterated, whereby consultations are made at the dictionary conceptualisation stage instead of a situation whereby the NLUs offer finished dictionary projects to different publishers. Finding themselves stranded with their completed projects and resorting to less well-known and less strategic publishers has not helped some NLUs. The main reason, in our view, why the GDX and ISX are not available in bookshops and less well-known among the isiXhosa-speaking community is working with a publisher outside the isiXhosa-speaking provinces, thereby making the marketing of the dictionaries quite difficult. If the GDX had been published by OUP, for instance, it is unthinkable that it would be difficult to acquire it in bookshops.

4.6 A Review of the Current Structures of South African Lexicography

While there has been a concern that the NLUs have not lived up to their expectations (Alberts 2011), the PanSALB Session at the Eighteenth International AFRILEX Conference confirmed that the current structures of South African lexicography, the core of which is PanSALB and the NLUs, have struggled to function. Resource constraints, which have implications on the staffing of the NLUs and the scope of lexicographic work that the NLUs could realistically undertake, emerged as a major issue. The PanSALB presentation also fell short in that, having provided the now well-known background to the establishment of both PanSALB and the NLUs, it had a blank slide when it came to the Board's vision for the NLUs. That the NLUs have collectively made a submission of revisions of the Articles of Association demonstrate the strain that the lexicographers have taken not only to produce dictionaries but also to survive under the current situation. The XNLU, like the other NLUs, faces an uncertain future and since it has been the main role player in isiXhosa lexicography, it is also the future of isiXhosa lexicography that is at stake. However, one has to be optimistic that, rather than judging South African lexicography to have failed, the challenges faced by NLUs compel the stakeholders to go back to the drawing board and find new solutions that will ensure that South Africa maintains its place on the global lexicographic map.

5. Conclusion

This article dealt with the past, present and future of isiXhosa lexicography. However, some issues discussed are reflective of the lexicography of African

languages and South Africa in general. The article established that isiXhosa has a long and rich history of lexicographic practice. However, it noted that although there is much that can be learnt from that history, the existing dictionaries need to be either updated or supplemented by new dictionaries in order to meet current lexicographic needs among the isiXhosa-speaking community. Perhaps this explains the current projects being undertaken by the XNLU.

This article does not proclaim to be a closed case. More definite conclusions could only be made with regard to the past and the present. We can only plan, but not always predict successfully what will happen in the future. Accordingly, the article only sought to initiate discussions about South African lexicography more than a decade since the establishment of the NLUs, which came after more than a century of lexicographic practice in African languages. Some potentially productive proposals have been made in this article. The article therefore hopes to have opened up possibilities for research in and debate on isiXhosa lexicography, with the major objective being to contribute to lexicographic practice and dictionary use. At the same time, it highlighted that South African lexicography is at a cross-roads, e.g. the choice between a national dictionary and other projects. As such, there is an urgent need to plan anew to ensure that the optimism of the past decade does not vanish into a false promise.

References

- Alberts, M.** 2011. National Lexicography Units — Past, Present, Future. *Lexikos* 21: 23-52.
- Articles of Association of a Company Act of 1973 (Section 60 (1); regulation 18).*
- Aylif, J.** 1846. *A Vocabulary of the Kafir Language*. London: Wesleyan Mission-House.
- Béjoint, H.** 2000. *Modern Lexicography*. Oxford: Oxford University Press.
- Bennie, J.** 1826. *A Systematic Vocabulary of the Kaffrarian Language in Two Parts, to which is Prefaced an Introduction to Kaffrarian Grammar*. Lovedale: Glasgow Mission Press.
- Davis, Wm. J.** 1872. *A Dictionary of the Kaffir Language: Including the Xosa and Zulu Dialects*. London: Wesleyan Mission-House.
- De Schryver, G.-M.** 2003. Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography* 16(2): 143-199.
- Döhne, J.L.** 1857. *A Zulu-Kafir Dictionary Etymologically Explained*. Harvard: G.J. Pike.
- Fischer, A. et al.** 1985. *Oxford English-Xhosa Dictionary*. Cape Town: Oxford University Press.
- Gallardo, A.** 1980. Dictionaries and the Standardization Process. Zgusta, L. (Ed.). 1980. *Theory and Method in Lexicography*: 59-69. Columbia: Hornbeam Press.
- Godfrey, R.** 1899. *A Kafir-English Dictionary*. Alice: Lovedale Mission Press.
- Gouws, R.H.** 1996. A Sequence for Meeting Lexicographic Needs. *Lexicography as a Financial Asset in a Multilingual South Africa*: 97-110. Pretoria: Department of Arts, Culture, Science and Technology.
- Gouws, R.H.** 2001. Lexicographic Training: Approaches and Topics/Formation lexicographique: approches et thèmes. Emejulu, J.D. (Ed.). 2001. *Éléments de lexicographie gabonaise. Tome I*: 58-94. New York: Jimacs-Hillman.

- Gouws, R.H.** 2003. Towards the Formulation of a Theoretically Motivated Model for the National Lexicography Units in South Africa. Hartmann, R.R.K. (Ed.). 2003. *Lexicography: Critical Concepts*: 218-246. London/New York: Routledge.
- Gouws, R.H. and D.J. Prinsloo.** 2005a. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.
- Gouws, R.H. and D.J. Prinsloo.** 2005b. Left-expanded Article Structures in Bantu with Special Reference to isiZulu and Sepedi. *International Journal of Lexicography* 18(1): 25-46.
- Hadebe, S.** 2006. *The Standardisation of the Ndebele Language through Dictionary-making*. Harare/Oslo: The ALLEX Project.
- Hall, M. et al.** 2013. *New Dictionaries and Trends*. Presentation at the Eighteenth International Association of African Lexicography (AFRILEX) Conference held at the Nelson Mandela Metropolitan University, Port Elizabeth, South Africa, 2-5 July 2013.
- Hartmann, R.R.K.** 2001. *Teaching and Researching Lexicography*. Harlow: Longman/Essex: Pearson Education.
- Hartmann, R.R.K. and G. James.** 1998. *Dictionary of Lexicography*. London/New York: Routledge.
- Jennings, L.E.** 1961. *The Concise Trilingual Dictionary in English, Xhosa, Afrikaans / Die Kort Drietalige Woordeboek in Afrikaans, Xhosa, English*. Johannesburg: Ad Donkers Publishing.
- Johnson, S.** 1747. *The Plan of a Dictionary of the English Language*. Facsimile edition 1970. Menston: The Scholar Press.
- Kropf, A.** 1915. *A Kafir-English Dictionary*. Second edition. Alice: Lovedale.
- Kumalo, M.B.** 1999. The National Lexicography Units — Existing and Prospective. *Lexikos* 9: 211-216.
- Landau, S.I.** 1984. *Dictionaries. The Art and Craft of Lexicography*. New York: Charles Scribner's Sons (of The Scribner Press)/Cambridge: Cambridge University Press.
- Mbude-Shale, N. et al.** 2008. *Science and Technology Dictionary / Isichazi-magama SezeNzuluwazi neTeknoloji*. Cape Town: New Africa Books.
- McArthur, T.** 1986. *Worlds of Reference. Lexicography, Learning and Language from the Clay Tablets to the Computer*. Cambridge: Cambridge University Press.
- McLaren, J.** 1915. *A Concise Kafir-English Dictionary*. London: Longmans, Green and Company.
- McLaren, J.** 1923. *A Concise English-Kafir Dictionary*. London: Longmans, Green and Company.
- Mini, B.M.** 1996. The Lexicographic Needs of Xhosa. *Lexicography as a Financial Asset in a Multilingual South Africa*: 67-74. Pretoria: Department of Arts, Culture, Science and Technology.
- Ministry of Arts, Culture and Technology.** 1996. The National Lexicography Units Bill (B 103-96). Pretoria: Ministry of Arts, Culture and Technology.
- Moropa, K. and A. Kruger.** 2000. Mistranslation of Culture-specific Terms in Kropf's *Kafir-English Dictionary*. *South African Journal of African Languages* 20(1): 70-79.
- Mtuze, P.T.** 1992. A Critical Survey of Xhosa Lexicography 1772-1989. *Lexikos* 2: 165-177.
- Nabe, H., Pw. Dreyer and G.L. Kakana.** 1976. *Xhosa Dictionary: English-Xhosa-Afrikaans, Xhosa-English-Afrikaans*. Johannesburg: Educum Publishers.
- Pahl, H.W. et al.** 1989-2006. *The Greater Dictionary of Xhosa*. Pietermaritzburg: Nutrend Publishers.
- Sinclair, J.M. (Ed.).** 1987. *Looking Up. An Account of the COBUILD Project in Lexical Computing and the Development of the Collins COBUILD English Language Dictionary*. London/Glasgow: Collins ELT.
- Sparman, A.** 1776. *A Voyage to the Cape of Good Hope Towards the Antarctic Polar Circle and Round the World but Chiefly into the Country of the Hottentots and Caffres from the Year 1772 to 1776*. Lon-

don: G.G.J. and J. Robinson.

Tshabe, S.L. et al. 2008. *Isichazi-magama SesiXhosa*. Pietermaritzburg: Nutrend Publishers.

Van der Kemp, J.T. 1803. Specimen of the Caffra Language. *Transactions of the Missionary Society* 1: 442-458.

Wababa, Z. and K. Welman (Eds.). 2010. *Isichazi-magama seZibalo seZikolo*. Cape Town: Cambridge University Press.

Wiegand, H.E. 1984. On the Structure and Contents of a General Theory of Lexicography. Hartmann, R.R.K. (Ed.). 1984. *LEXeter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9-12 September 1983*: 13-30. Tübingen: Max Niemeyer.

'n Kritiese beskouing van woordeboeke met geamalgameerde lemmalyste

D.J. Prinsloo, *Departement Afrikatale, Universiteit van Pretoria,
Pretoria, Suid-Afrika (danie.prinsloo@up.ac.za)*

Opsomming: Die onlangs gepubliseerde Groot Woordeboek (Afrikaans en Nederlands), ook bekend as ANNA, is die eerste woordeboek met 'n geamalgameerde lemmalys gebaseer op die model van Martin en Gouws (2000). ANNA baan ook die weg vir 'n soortgelyke benadering vir ander nouverwante tale soos die Sothotale en die Ngunitale van Suid-Afrika. Die voor- en nadele van vyf aspekte te wete (a) vergelyking en kontras, (b) gebruikersvriendelikheid, (c) ruimtebesparing, (d) ordening van betekenisonderskeidings en (e) die aanbod van 'n afsonderlike grammatikale kompendium word krities beskou. Die beginsels waarop amalgamerings berus, die tersaaklike lemmatipes, asook enkele kenmerke van die model word vooraf kortliks bespreek.

Sleutelwoorde: GEAMALGAMEERDE LEMMALYS, VERWANTE TALE, OORTOLLIGHEID, ANNA, TWEETALIGE WOORDEBOEK, AFRIKAANS, NEDERLANDS, INLIGTINGSKOSTE, SISTEEMKOSTE

Abstract: A Critical Examination of Dictionaries with Amalgamated Lemmalists. The recently published Groot Woordeboek (Afrikaans en Nederlands), also known as ANNA, is the first dictionary with an amalgamated lemmalist based on the model of Martin and Gouws (2000). ANNA also paves the way for a similar approach for other closely related languages such as the Sotho languages and the Nguni languages of South Africa. The advantages and disadvantages of five aspects namely (a) comparison and contrast, (b) user-friendliness, (c) space saving, (d) ordering of senses and (e) provision of a separate grammatical compendium are critically evaluated. The principles of amalgamation, the relevant lemma types, as well as certain characteristics of the model are briefly discussed beforehand.

Keywords: AMALGAMATED LEMMALIST, RELATED LANGUAGES, REDUNDANCY, ANNA, BILINGUAL DICTIONARY, AFRIKAANS, DUTCH, INFORMATION COSTS, SYSTEM COSTS

Inleiding

Die doel van hierdie artikel is 'n kritiese evaluering van woordeboeke wat vir naverwante tale soos Afrikaans en Nederlands, die Sothotale, en die Ngunitale saamgestel word met 'n enkele (geamalgameerde) lemmalys vir die verwante tale. Die onlangs gepubliseerde Groot Woordeboek (Afrikaans en Nederlands),

voortaan benoem volgens die projek se eie identifikasie as ANNA, sal ontleed word as die hoofverwysingsbron. Die fokus is op die voor- en nadele van só 'n woordeboek met Afrikaanse en Nederlandse lemmas in 'n geamalgameerde lemmalys in vergelyking met twee afsonderlike woordeboeke of woordeboekedele Afrikaans–Nederlands en Nederlands–Afrikaans.

Die aard en samestelling van ANNA is omvangryk en dit is nie moontlik om binne die bestek van 'n artikel aan al die tersaaklike aspekte aandag te gee nie. In hierdie artikel word die fokus derhalwe beperk tot vyf aspekte, naamlik (a) vergelyking en kontras, (b) gebruikersvriendelikheid, (c) redundansie, (d) ordening van betekenisonderskeidings en (e) die aanbod van 'n afsonderlike grammatikale kompendium. (Uitvoerige besprekings van die aard en eienskappe van ANNA kan onder meer gevind word in Martin (2012a en 2012b), Martin en Gouws (2000), Marais (2011), Bosman (2013) en in die gebruiksaanwysings van ANNA.) Die artikel het ook ten doel om deur die analise van (a) tot (e) betreffende inligtingskoste en die model-/sisteenkoste van 'n geamalgameerde benadering, voorstelle aan die hand te doen hoe geamalgameerde woordeboeke meer gebruikersvriendelik gemaak kan word. Ten einde die argumente ten opsigte van (a) tot (e) in perspektief te plaas, word die beginsels waarop amalgameering berus, die tersaaklike lemmatipes, asook enkele kenmerke van die model vooraf kortliks bespreek.

Die geamalgameerde model

ANNA maak daarop aanspraak dat dit die eerste woordeboek is wat saamgestel is met 'n geamalgameerde lemmalys, dit wil sê een enkele lemmalys vir beide Afrikaans en Nederlands. 'n Woordeboek met 'n geamalgameerde lemmalys is in die eerste plek 'n vertaalwoordeboek wat vertaalekwivalente in die doeltaal vir woorde in die brontaal gee en omgekeerd. 'n Fundamentele verskil met tradisionele vertaalwoordeboeke is egter dat dit as gevolg van die geamalgameerde benadering 'n kontrasterende woordeboek is wat die ooreenkomste en verskille tussen die betrokke tale op 'n meer direkte manier as afsonderlike woordeboeke, bv. Afrikaans–Nederlands en Nederlands–Afrikaans uitlig. Martin en Gouws (2000) en Martin (2012b) beklemtoon dit dan ook dat beide ooreenkomste en verskille op 'n doeltreffende en kontrastiewe manier in ANNA uitgelig word.

The amalgamation model is not an explanatory model but a descriptive one. It aims to describe the lexemes of two languages in a bilingual dictionary in a *directly contrastive way*, contrary to the *indirect way* found in a traditional bilingual dictionary. (Martin 2012b: 413)

Die artikels van *coffee*, *rain* en *motho* in (1) vir 'n beplande geamalgameerde woordeboek Engels–Sotho/Sotho–Engels illustreer die twee aspekte van vergelyking/kontras en ruimtebesparing vir albei dele op 'n eenvoudige manier.

(1)

coffee *n.* kofi

rain *n.* pula

motho *n.* a person, a human being

Vergelyking/kontras blyk uit die feit dat *koffie*, *reën* en *mens* in al drie die Sothotale dieselfde vertaalekwivalent het. Groot ruimtebesparing vind plaas deurdat elke artikel slegs een keer gelys word en nie drie keer soos wat dit in afsonderlike woordeboeke vir Sepedi, Setswana en Sesotho die geval sou wees nie. In gevalle soos (2) waar oorvleueling slegs ten opsigte van twee van die Sothotale voorkom waar die model steeds goed ten opsigte van vergelyking/kontras, maar die wins ten opsigte van ruimtebesparing is kleiner omdat verskillende vertaalekwivalente ter sake is en as sodanig gemarkeer moet word.

(2)

his *pp.* gagwe [Sep/Set], hae [Ses]

verb *n.* lediri [Sep/Set], leetsi [Ses]

In (2) word die lemmas *his* en *verb* steeds slegs een keer gelys maar verskil die vertaalekwivalent vir Sesotho (Ses) van dié van Sepedi (Sep) en Setswana (Set). In (3) lê die wins slegs op die vlak van vergelyking/kontras en die feit dat die lemma nie drie keer herhaal word nie. Daar is geen besparing ten opsigte van vertaalekwivalente nie deurdat die lemma in elkeen van die Sothotale anders vertaal word.

(3)

page *n.* letlakala [Sep], leqephe [Ses], tsebe [Set]

In al hierdie voorbeelde, ongeag die beperkte omvang van die bewerking van die onderskeie lemmas, word die ooreenkomste en verskille, asook die besparing duidelik uitgebeeld. Dit is ook voor die hand liggend dat dit 'n basiese vereiste en vertrekpunt vir die samestelling van woordeboeke met geamalgameerde lemmalyste is dat die tale aansienlik oorvleuel met betrekking tot die vorm en betekenis van die woorde. Bloot op sigwaarde ten opsigte van tekens (tokens) geneem, wys Gouws et al. (2004) op 'n groot oorvleueling met 'n hoë gebruiksfrekwensie tussen Afrikaans en Nederlands, soos in tabel 1 aangedui word.

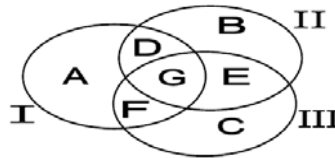
Tabel 1: Afrikaans in vergelyking met Nederlands met frekwensies per miljoen woorde

Item	Afrikaans	Dutch	Item	Afrikaans	Dutch	Item	Afrikaans	Dutch
<i>van</i>	19 688	18 091	<i>niet</i>	46	7 358	<i>als</i>	26	4 011
<i>het</i>	15 111	16 655	<i>te</i>	8 876	7 025	<i>aan</i>	4 338	3 846
<i>een</i>	1 892	16 223	<i>die</i>	58 733	6 520	<i>of</i>	6 653	3 811

<i>en</i>	22 516	13 663	<i>met</i>	7 139	6 268	<i>nog</i>	1 930	3 507
<i>in</i>	15 619	13 483	<i>voor</i>	1 316	6 220	<i>wel</i>	276	3 146
<i>is</i>	15 628	11 539	<i>maar</i>	3 862	5 121	<i>om</i>	7 323	3 074
<i>dat</i>	4 841	9 565	<i>ook</i>	3 068	4 414	<i>wat</i>	5 268	2 918
<i>op</i>	7 132	8 125	<i>dan</i>	907	4 032	<i>kan</i>	2 069	2 488

(Gouws et al. 2004: 798)

Vir die Sothotale het Prinsloo (2006) die 10,000 mees frekwente woorde in Sesotho, Setswana en Sepedi met mekaar vergelyk en bevind dat die woorde in dié tale tot 'n hoë mate oorvleuel. Die drie Sothotale het 19,4% woorde gemeen, Sepedi en Setswana deel 32,7%, Sepedi en Sesotho 26,9% en Setswana en Sesotho 34,4%, figuur 1.



Figuur 1: Oorvleueling van Sepedi, Setswana en Sesotho ten opsigte van die 10,000 mees frekwente woorde
 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 2006: 199)

Dit lei tot 'n enkele geamalgameerde lemmalys van 22,537 in vergelyking met 'n lys van 30,000 lemmas in drie afsonderlike woordeboeke, dus 'n besparing van byna 30%.

Lemmatipes in ANNA

ANNA onderskei vyf tipes lemmas, naamlik (a) absolute kognate, (b) absolute kognate met verskillende vorme, (c) gedeeltelike kognate, (d) niekognate en (e) valse vriende, en hierdie vyf tipes lemmas vorm vervolgens die basis vir besinning in hierdie artikel.

Absolute kognate is Nederlandse en Afrikaanse woorde wat na vorm en betekenis identies is soos *testament/testament*, *leemte/leemte* en *riool/riool*. *Absolute kognate met 'n vormverskil* het in albei tale dieselfde betekenis en die vormverskil is sodanig dat dit nie herkenbaarheid belemmer nie, byvoorbeeld *gelyknamig/gelyknamig*, *leeglopen/leegloop* en *schoen/skoen*. *Gedeeltelike kognate* staan teenoor absolute kognate deurdat hulle in minstens een betekenisonderskeiding verskil. So, byvoorbeeld, deel *blijven* en *bly* die betekenisonderskeiding 'nie weggaan nie' maar nie die Afrikaanse betekenis 'êrens tuisgaan' nie.

Niekognate het dieselfde betekenis maar daar is 'n duidelike vormverskil soos *keuken/kombuis*, *sinaasappel/lemoen*, ens. *Valse vriende* is woorde wat vormlik ooreenstem maar verskillende betekenis het, soos *amper*, *verskoning* en *verskoon*.

Ooreenkomste en verskille: vergelyking en kontras

Daar kan met die intrapslag onomwonde verklaar word dat ten opsigte van vergelyking en kontrastering van die naverwante tale, geamalgameerde woordeboeke, en ANNA in die besonder, volpunte kry ongeag die omvang van die artikels. Die uitwys van ooreenkomste en verskille lê die model ten grondslag en manifesteer op indrukwekkende wyses in die bewerking van beide Nederlandse en Afrikaanse lemmas, onder meer ten opsigte van vertaalekwivalente, betekenisonderskeidings, kollokasies en gebruiksvoorbeelde. Vergelyk byvoorbeeld die bewerking van die lemma *ouderwets* as 'n eenvoudige maar tipiese voorbeeld in (4).

(4)

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 deeglikheid*; 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*

In die eerste plek behels die bewerking deurlopende vergelyking en kontras ten opsigte van vertaalekwivalente, betekenisonderskeidings en voorbeelde in drie gedeeltes, naamlik wat (a) Nederlands en Afrikaans gemeen het (A/N), (b) uniek is aan Nederlands (N) en (c) eie is aan Afrikaans (A). Die konvensies '=' (is gelyk aan) en '≠' (is nie gelyk aan nie) is kripties en funksioneel om ooreenkomste en verskille te merk. So, byvoorbeeld, wys die is-gelyk-aan-teken dat die betekenisonderskeiding 'in werking stel' in (9) in beide Nederlands en Afrikaans voorkom in voorbeelde soos *de wasmachine aanzetten* en die *wasmasjien aansit*, maar dat mens iemand in Afrikaans tot spoed *aanspoor* en nie soos in Nederlands *aansit* nie. Uitvoerige besprekings van die bewerking van die vyf lemmatipes kan gevind word in die voortekste van ANNA, Martin (2012a en 2012b), en Martin en Gouws (2000).

Gebruikersvriendelikheid betreffende inligtingskoste en sisteemkoste

Die geamalgameerde model is 'n nuwe woordeboekkonsep wat verskil van die tradisionele tweetalige woordeboeke waaraan die gebruiker gewoond is. Hoewel daar uiteraard baie ooreenkomste met tweetalige woordeboeke is, moet die

gebruiker besef dat dit 'n unieke sisteem met sy eie kenmerke en voor- en nadele is. Wat betref gebruikersvriendelikheid kan gevra word wat die gebruiks- en sisteemkoste vir die gebruikers is. Die term sisteemkoste word in hierdie artikel gebruik om te verwys na die eienskappe, noodwendige vereistes en tekortkominge van die geamalgameerde model wat verband hou met wat Martin (2012b) "system side effects" en selfs "system pitfalls" noem. Sisteemkoste dui ook op wat David Lieb "cognitive overhead" noem en deur David Demaree gedefinieer word as "how many logical connections or jumps your brain has to make in order to contextualise the thing you are looking at" (Cognitive Overhead, Or Why Your Product Isn't As Simple As You Think (<http://techcrunch.com/2013/04/20/cognitive-overhead/>)). Ter wille van goeie balans moet dit duidelik gestel word dat die leksikografiese benadering in sogenaamde tradisionele woordeboeke ook gebruiks- en sisteemkoste vir die gebruikers inhou. So, byvoorbeeld, vereis 'n alfabetiese ordeningsbenadering kennis van die alfabet, en uitspraakleiding met behulp van die IPA-ortografie kennis van fonetiese simbole. Die verskil is egter dat die meeste woordeboekgebruikers reeds met die tradisionele sisteem vertrou is, terwyl die geamalgameerde model van nuuts af bemeester moet word. Optimale gebruik van geamalgameerde woordeboeke veronderstel 'n groot mate van voorafkennis van die model/sisteem asook van addisionele woordeboekkonvensies.

Die toegangstruktuur van ANNA verskil in verskeie opsigte van ander woordeboeke en 'n voorvereiste vir optimale gebruik is dat die gebruiker eers deeglike voorafstudie sal moet doen van die uitleg, konvensies, teksgedeeltes, kruisverwysingstelsel, ens. wat in die woordeboek gebruik word. Hierdie vereiste is in beginsel 'n nadeel omdat dit bekend is dat gebruikers oor die algemeen nalaat om die gebruikersgids van 'n woordeboek te raadpleeg, om nie van intensiewe bestudering daarvan te praat nie. So, byvoorbeeld, meld Bosman (2013: 48) dat geeneen van die studente wat aan 'n gebruikerstudie van ANNA deelgeneem het, die voortekste gelees het nie.

Die vooraforiëntering vir die gebruik van ANNA lê ook op meer as een vlak. Die belangrikste is begrip van die gebruik van simbole soos A (Afrikaans), N (Nederlands), N/A (Nederlands en Afrikaans), = (Nederlands dieselfde as Afrikaans), ≠ (Nederlands verskillend van Afrikaans), dat alle teksgedeeltes wat op Afrikaans betrekking het, kursief gedruk is en alle Nederlandse teks in Romeinse letters aangebied word, ens. Dit is dan ook belangrik dat die samestellers geen foute begaan in die rigtinggewende voorbeelde in die gebruiksaanwysings soos die foutiewe Nederlandse spelling (p. 27) binne die Afrikaanse frase "*bek van 'n trechter trechtermond*" in plaas van "*bek van 'n tregter trechtermond*" nie.

'n Tweede vlak behels studie van die gebruiksaanwysings ten opsigte van (a) die algemene aard van die woordeboek, (b) amalgamasiemodel en die tipes trefwoorde, (c) bou en struktuur van die artikels, (d) behandeling van voorbeelde en verbindings, (e) verwysings, (f) agtertekste en (g) ander aspekte soos

metataal, volgordekweessies, ens. wat uitvoerig in die gebruiksaanwysings beskryf word.

'n Derde vlak behels kennis van inherente kenmerke van die sisteem soos die aanbod en ordening van ekwivalente en betekenisonderskeidings wat hieronder in meer detail bespreek word.

Die gebruiksaanwysings is omvangryk en beslaan 18 bladsye en eindig met baie belangrike gebruiksinligting ten opsigte van artikeluitleg in figuur 2.

Die gebruik van ANNA vereis dus veel meer van die gebruiker as bloot die kennis van standaardwoordeboekkonvensies waarvoor die gemiddelde gebruiker beskik ten einde die inligting te vind en maksimaal daarby te kan baat. Bosman (2013) gebruik die terme *toegangsgemak* en *leksikografiese inligtingskoste* in dié verband. Wat inligtingskoste betref, bespreek sy die aangeleentheid uitvoerig aan die hand van gebruikerstudies en bevind dat die respondente in die algemeen nie daarin geslaag het om die gevraagde inligting maklik en vinnig te kon vind nie. Sy kom tot die gevolgtrekking dat die "inligtingskoste nogal hoog is" en dat studente 'n volledige voorafdemonstrasie benodig:

Gemeet aan Hartmann en James (1998) se definisie ("the relative ease with which information can be located in a reference work") dui die bogenoemde ondersoekdata en ontleding daarop dat die inligtingontsluitingsproses nóg vinnig nóg maklik was. ... Sowel soekverwante as begripsverwante koste is aan die hoë kant. (Bosman 2013: 51)

Dit is te verstane dat 'n nuwe woordeboekmodel eiesoortige toegangsvereistes en -konvensies sal hê en verder dat daar volgens Martin (2012b) ook sekere noodwendige "system side effects", oftewel sisteemkoste sal wees wat meer voorafkennis, voorafstudie en begrip van die teikengebruiker sal verlang. Dit is belangrik dat onderskeid gemaak moet word tussen sisteemkoste as eienskappe en vereistes wat inherent aan 'n woordeboekmodel is en waarvoor die leksikograaf eintlik geen beheer het nie, en sisteemkoste as wyses waarop die leksikograaf wel die gebruiker ter wille kan wees om die woordeboek meer gebruikersvriendelik te maak. So, byvoorbeeld, moet die samestellers dit oorweeg om die artikeluitleg in figuur 2 op 'n baie meer prominente plek aan te bied – verkieslik heel voor in die woordeboek – en die voorteks te stratifiseer ten opsigte van byvoorbeeld noodsaaklike gebruiksinligting versus kenmerke van en verdere inligting oor die model. Kennis van die gebruiksinligting in figuur 2 moet genoegsaam wees vir aanvanklike gebruik van die woordeboek tot tyd en wyl die gebruiker meer vertrouwd raak met die model.

Ruimtebesparing en redundansie

Soos wat hierbo ten opsigte van 'n geamalgameerde benadering vir die Sothotale opgemerk is, is die veronderstelling van die geamalgameerde benadering dat dit redundansie sal verminder op sowel makro- as mikrostruktuurvlak.

Artikeluitleg (Afrikaans)

Algemeen: Nederlandse inligting (trefwoorde, grammatikale gegewens, betekenisonderskeidings, voorbeelde, vertalings) is deurgaans romeins gedruk, Afrikaanse inligting *kursief*. Bykomende inligting oor inhoudelike aangeleenthede (= kommentaar) word altyd in Afrikaans gegee en is dus *kursief* gedruk.

Vir meer inligting kan ook die Gebruiksaanwysings en die lys Gebruikte afkortings, merkers en simbole geraadpleeg word.

Nederlandse trefwoorde word in vet romeinse druk en Afrikaanse trefwoorde in vet kursiewe druk aangegee.

Boskrifnummers word gebruik om te onderskei tussen woorde wat eenders gespel word maar tot verskillende woordsoortkategorieë behoort en/of sterk verskil t.o.v. betekenis.

A/N in die kantlyn dui aan dat die daaropvolgende betekenis(se) in Afrikaans sowel as in Nederlands voorkom. Waar meerdere betekenisonderskeidings voorkom, word dié in sirkeltjies genommer.

A of N in die kantlyn dui aan dat die daaropvolgende betekenis(se) uitsluitlik in Afrikaans of Nederlands onderskeidelik voorkom.

Valse vriende, d.w.s woorde wat in Afrikaans en Nederlands (nagenoeg) dieselfde vorm het maar sterk van betekenis verskil, word deur 'n spesiale simbool II aangedui.

Trefwoorde word gevolg deur woordsoortinligting, met direk daarna bykomende grammatikale inligting tussen vierkantige hakies.

Kort betekenisomskrywings en vertaalekwivalente word in vet druk gegee, Nederlands in romeinse en Afrikaans in kursiewe druk.

'n Isgelykaanteken (=) dui aan dat die daaropvolgende voorbeeld(e) in die twee tale ooreenstem.

'n Isongelykteken dui aan dat er 'n relevante verskil is tussen voorbeeld(e) en vertaling(e).

Punthakies word gebruik vir etikette wat bykomende inligting verskaf.

By sommige trefwoorde staan 'n verwysing, óf na 'n ander trefwoord (deurgaans van Afrikaans na Nederlands), óf na die grammatikale kompendium. 'n Verwysing wat op 'n koppelteken eindig, verwys na 'n woorddeel. Derglike verwysings wat in 'n grys balk gedruk staan, dien as bykomende hulpmiddel by die soek van woorde waarna gesoek word.

borg nw. [de; mv: -en], **borg** nw. [mv: -e]
 (borsom) **borg** = borg betalen **borg betaal**
 (die beloof te schulden te betalen) **borg** = de borg aanspreken *die borg aanspreek*; ergens borg voor staan *borg staan vir iets* ≠ zich borg stellen voor iets/iemand *borg staan vir iets/iemand*
 (besermingsmiddel) **beveiliging**
 (iemand wat vir finansiering instaan) **sponsor**
borg ^{ww.tr. (II)}
 (finansieel ondersteun) **sponsore**, **adopteren** ≠ 'n (glans)geleentheid **borg** een evenement sponsore; sy wil 'n Soedanese kindjie **borg** ze wil foster parent worden van een Soedanees kindjie
borgen ^{ww.tr. (II)}
 (beveiligen tegen losraken) **vassluit**
boony nw. [de; mv: bomen], **boom** nw. [mv: bome]
 (grote plant met stevige stam) **boom** = de boom van kennis/van goed en kwaad) *die boom van kennis (van goed en kwaad)* <fig.> de kat uit de boom kijken *die kat uit die boom kyk*; <fig.> boomplje groot, plantertje dood *boomplje groot, plantertje dood* ≠ <fig.> een boom van een vent 'n berg van 'n kere!, 'n blok van 'n man, 'n yslike kere!; <fig.> je kunt de boom in! *hy kan gaan boomklim!*; <fig.> hoge bomen vangen veel wind *hoë bome vang die meeste wind*; <fig.> door de bomen het bos niet zien *deur/vanweë die bome die bos nie meer sien nie*; <fig.> aan de vruchten kent men de boom 'n mens ken die boom aan sy vrugte; <fig.> oude bomen verplant men niet 'n ou boom verplant nie maklik nie/kan nie maklik verplant word nie <fig.> *die appel val nie ver van die boom nie* zo vader, zo zoon, de appel valt niet ver van de stam/boom; <fig.> *tussen die boom en die bas wees weifelachtig zijn*, noch het een noch het ander zijn; <fig.> *buig die boomplje terwyl hy nog jonk is* jonge rijzen kan men buigen (maar oude bomen niet); <fig.> *dit gaat maar tussen die boom en die bas* het gaat maar matig
batter ^{nw. (zie boter) nw.}
potter - zie ook **boter**

Figuur 2: Gebruiksinligting ten opsigte van artikeluitleg in ANNA

Op die vlak van die makrostruktuur word veronderstel dat die redusering van

twee of meer lemmalyste na een noodwendig tot ruimtebesparing sal lei en nie gebruikersonvriendelik sal wees nie. Hierdie veronderstelling vorm die belangrikste motivering in Prinsloo (2006) vir die gebruik van die model:

A bidirectional English → {siSwati, isiXhosa, isiNdebele and isiZulu}, {siSwati, isiXhosa, isiNdebele and IsiZulu} → English dictionary is comparable to *four* bidirectional bilingual dictionaries, English–siSwati, siSwati–English, English–isiXhosa, isiXhosa–English, English–isiZulu, isiZulu–English, and English–isiNdebele, isiNdebele–English. Thus two directions for the envisaged model versus eight sides for separately bridging English and a Nguni language. (Prinsloo 2006: 195)

Prinsloo (2006) aanvaar ook sonder meer dat die omvang van ruimtebesparing groot is in gevalle soos (1) waar dieselfde lemma in al drie die Sothotale gebruik word, en steeds bydra tot besparing waar 'n lemma minstens twee tale verteenwoordig soos in (2).

Dieselfde veronderstelde ruimtebesparing geld ook vir 'n Afrikaans–Nederlands geamalgameerde lemmalys. Martin (2012b: 418) stel dit onomwonde: "amalgamation *reduces redundancy*" (sy kursivering). Daar kan egter geargumenteer word dat faktore soos die noodwendige afsonderlike lysing van sekere Afrikaanse lemmas asook eise aan die voorafkennis van die gebruiker, óf redundansie in die hand werk óf gebruikersonvriendelik is. Hierdie faktore moet derhalwe as debiete teen ruimtebesparing in die model verreken word.

Die omvang van ruimtebesparing wissel uiteraard vir die verskillende lemmatipes (a) tot (e) wat in ANNA onderskei word. In die geval van absolute kognate wat vormlik identies is en gedeeltelike kognate, is ruimtebesparing 'n gegewe deurdat slegs een lemma gelys hoef te word. Vir niekognate noodsaak die vormverskil afsonderlike lemmatisering en is daar nie ruimtebesparing nie. In ANNA word valse vriende afsonderlik gelemmatiseer en gevolglik word ruimte ook nie bespaar nie. Die laaste kategorie, absolute kognate met 'n vormverskil, lewer gemengde resultate. Martin (2012b: 417) gaan van die standpunt uit dat die vormverskille wat hulle vertoon klein genoeg is sodat hulle maklik herkenbaar is en noem hartinfarkt/*hartinfark*, stikken/*stik* as voorbeelde.

Die redaksionele beleid is om nie die Afrikaanse lemma op te neem indien die Afrikaanse lemma die Nederlandse lemma alfabeties direk sou volg of voorafgaan, of hoogstens deur sewe items van mekaar geskei is nie. So, byvoorbeeld, word die Afrikaanse lemmas *erupsie* en *versmal* nie afsonderlik gelemmatiseer nie omdat dit die Nederlandse lemmas *eruptie* en *versmallen* respektiewelik alfabeties direk sou voorafgaan. Dieselfde geld vir die Afrikaanse lemmas *holoog* en *onfatsoenlik* wat alfabeties op enkele lemmas na, die Nederlandse lemmas *hologig* en *onfatsoenlijk* volg. Indien die betrokke lemmas alfabeties verder as sewe items van mekaar voorkom, word die Afrikaanse lemma ooreenkomsdig ANNA se redaksionele beleid wel in die toepaslike alfabetiese posisie gelys met 'n kruisverwysing na die Nederlandse lemma soos in (5) en (6)(a).

(5)

skrei zie **schreien**

skriba zie **scriba**

skrif zie **schrift**

skrif- zie ook **schrift-**

...

*skrik*¹ nw. zie **schrik** nw.

*skrik*² ww. zie **schrikken** ww.

skrik- zie (ook) **schrik-**

(6)

(a) *neerskryf* zie **neerschrijven**

(b) **neerschrijven** ww.tr., *neerskryf* ww.tr.

A|N (**opschrijven**) *neerskryf* ≠ je belevenissen in een schrift neerschrijven *jou belewenisse in 'n notaboek opteken*

Die redaksionele beleid word egter nie streng gevolg nie. In die geval van (6)(a) word die Afrikaanse woord wel as lemma opgeneem hoewel dit slegs vier alfabetiese posisies na (6)(b) verskyn. Hierteenoor is die Afrikaanse lemma *bekyk* wat alfabeties meer as 70 lemmas verwyderd van *bekijken* sou wees, nie opgeneem nie. Dieselfde geld vir *geskei* wat, indien dit gelemmatiseer sou wees, 65 lemmas verwyderd van *gescheiden* in die woordeboek sal wees. Omdat *geskei* nie opgeneem is met 'n kruisverwysing na *gescheiden* nie, sal die gebruiker nie weet om onder *gescheiden* te soek nie. Dieselfde geld vir *krygsraad* wat ook etlike bladsye verwyderd is van die alfabetiese posisie vir *krijgsraad*. Die probleem by laasgenoemde twee voorbeelde lê onder meer daarin dat die onderskeid in spelling "vroeg" in die woord d.w.s binne die eerste 3–5 karakters voorkom. In ander gevalle soos byvoorbeeld *brandstichting* versus *brandstiging*, *geruststellen* versus *gerusstel* en *ommuur* versus *ommuur* waar die verskille later in die woord voorkom, is die gebruiker se kansse veel beter om wel die lemma te vind.

Die aanbod van 'n afsonderlike grammatiese kompendium noodsaak verdere kruisverwysings wat ook redundansie in die hand werk.

Die vraag is dus of die regte balans hier gevind is ten opsigte van ruimtebesparing en gebruikersvriendelikheid.

Op die vlak van die mikrostruktuur word eweneens veronderstel dat die redusering van twee lemmalyste na een, noodwendig tot ruimtebesparing sal lei. Martin (2012b: 407) bied gepas die bewerkings van *robot* ter illustrasie in (7) en (8).

(7)

In the Dutch–Afrikaans part:

robot

(automaat die arbeid verricht) [automaton carrying out work] *robot, blikman*

In the Afrikaans–Dutch part:

robot

(paal met lig wat verkeer reël) [pole with light regulating traffic] *stoplicht, verkeerslicht*
(outomaat wat werk verrig) [automaton carrying out work] *robot*

(8)

In ANNA the information from the two parts is brought together, *amalgamated*, resulting in an entry like this:

robot/*robot*

A/N (outomaat die arbeid verricht) *robot, blikman*

A (*paal met lig wat verkeer reël*) *stoplicht, verkeerslicht*

Indien die Engelse vertalings buite rekening gelaat word, beteken dit dat ANNA in (8) in 17 woorde vermag wat in 21 woorde in tradisionele tweetalige woordeboekformaat nodig sou wees. In hierdie spesifieke geval is dit 'n 19% besparing.

Ten opsigte van die A/N artikels kan 'n sterk saak uitgemaak word ten gunste van besparing deurdat inligting slegs een keer gegee word waar dit in die geval van 'n tweerigtingwoordeboek (N–A, A–N) twee keer aangebied sou word.

Die vraag is egter tot watter mate die afsonderlike N- en A-gedeeltes ruimtebesparing in die hand werk omdat die tale dan wel afsonderlik bewerk word. So, byvoorbeeld, sou die N- en A-gedeeltes in (4) in die N–A- en A–N-kant van 'n tweerigtingwoordeboek aangebied word en derhalwe nie tot ruimtebesparing in die geamalgameerde model lei nie. Vergelyk ook die artikels in (9) met dié in (10) en (11) ten opsigte van die aanbod en ordening van betekenisonderskeidings in 'n geamalgameerde model versus afsonderlike bewerkings in twee verklarende woordeboeke vir Afrikaans en Nederlands ten opsigte van die omvang van die artikels.

(9)

aanzetten¹ ww.intr., **aansit** ww.intr.

A|N ● **(dik maken) aansit** ≠ pinda's zetten erg aan *grondboontjies laat jou baie gewig aansit* ≠ *vyf kilogram (gewig) aansit* vijf kilo aankomen

● **(ongewenst naderbij komen) aansit** ≠ <inf.> daar kwam de ME aanzetten *daar kom die onlustepolisie toe aangesit*; laat/vroeg komen aanzetten *laat/vroeg aangesit kom* ≠ <fig., inf.> daar kom je nu mee aanzetten *vir wat kom jy nou daarmee aangesit?* ≠ <fig.> *met iets aangesit kom* met iets komen aandragen, met iets komen aandraven

aanzetten² ww.tr., **aansit** ww.tr.

A|N ● **(vastmaken aan iets) aansit, vaswerk, aanwerk, insit** = knopen aan een jas zetten *knoppe aan 'n jas aanwerk* ≠ mouwen aan een jas aanzetten *moue aan 'n jas/baadjie insit, moue aan 'n jas aansit* ≠ <fig.> iemand de duimschroeven aanzetten *iemand laat bontstaan*

● **(in werking stellen) aansit** ≠ de wasmachine aanzetten *die wasmasjien aansit* ≠ die lig *aansit* het licht aandraaien; *die enjin/motor aansit* de motor/auto starten

N ● **(aansporen tot (iets)) aanspoor, opstook** = hij zette mij aan tot spoed *hy het my tot spoed aangespoor* ≠ iemand ergens toe aanzetten *iemand opstook om iets te doen*

● **(nadruk geven) beklemtoon** ≠ lippen aanzetten *lipstiffie aansmeer*

- (scherper maken) *slyp* = een mes aanzetten 'n mes slyp
- (beginnen iets te doen) *begin* ≠ de sprint aanzetten *die pas begin versnel* (in 'n fietswedren)

A (aan die liggaam sit) **omdoen, aanleggen** = 'n verband aansit een verband aanleggen; 'n ander das aansit een andere das omdoen ≠ 'n skoon doek vir die baba aansit de baby een schoner luiers omdoen; 'n kind aan die bors sit een kind (aan de borst) aanleggen; vir die kind 'n borslap(pie) aansit het kind een slabbetje omdoen, het kind een slabbetje voordoen

(10) e-HAT

aan'sit ww. [aangesit]

- 1 Aan, by iets sit: *Al die gaste moet nou aansit aan die groot tafel.*
 - 2 Aanwerk, vasheg, voorsien van: *Nuwe knope aan 'n hemp aansit. 'n Nuwe steel aan 'n graaf aansit.*
 - 3 Aan die gang sit: *Die motor aansit.*
 - 4 Aan die werk sit: *Die span werkers aansit om te skoffel.*
 - 5 Opwerk, aanhits: *Mense tot opstand aansit.*
 - 6 Bykry: *Gewig, vet aansit.*
 - 7 (<Eng.) Jou baie verbeel; spog, voorgee.
- aansit:** ~motor, ~sleutel. aansitterig.

(11) Van Dale Groot Woordenboek

<p>aan·zet·ten¹ zie ook: aanzetten tot onvergankelik werkwoord</p> <p>1 zette aan, heeft aangezet dik maken ▶ betekenisverwante terme</p> <p>2 zette aan, is aangezet licht aanbranden ▶ betekenisverwante terme</p> <p>¶ idioom komen aanzetten met iets aankomen met een verzoek, een verhaal enz.; een voorwerp aanbrengen sonder dat dit gewens of verwag word</p>	<p>aan·zet·ten² zie ook: aanzetten tot vergankelik werkwoord; zette aan, heeft aangezet</p> <p>1 ook absoluut beginn (iets te doen) ▶ betekenisverwante terme</p> <p>2 (iets) aan een voorwerp vastmaken ▶ betekenisverwante terme ▶ konteks</p> <p>3 inschakelen ▶ betekenisverwante terme ▶ konteks</p> <p>4 benadrukken ▶ betekenisverwante terme ▶ konteks</p> <p>5 (iets) zo plaatsen dat het bijna aan iets anders raak ▶ betekenisverwante terme</p> <p>6 skerp, wetten ▶ betekenisverwante terme ▶ konteks</p>
--	---

Die aanbod en ordening van ekwivalente en betekenisonderskeidings betrefende sisteemkoste

Martin (2012b: 420) stel dit duidelik dat die geamalgameerde benadering sekere nuweaspekte het wat eie is aan die sisteem. Hy verwys daarna as "system side

effects". Só word die aanbod van vertaalekwivalente en die ordening van betekenisonderskeidings deur die sisteem self bepaal en nie deur byvoorbeeld frekwensie van voorkoms van betekenisse nie. In gevalle waar Afrikaans en Nederlands dieselfde vertaalekwivalent het, kry dié paar in ANNA voorkeur, al sou daar 'n meer gebruiklike woord in Afrikaans bestaan. Martin (2012b: 420) noem *taxateur/taksateur* as voorbeeld waar *taksateur* as vertaalekwivalent bo die meer gebruiklike *waardeerder* verkies word.

(12)

taxateur nw. [de; mv: -s], **taksateur** nw. [mv: -s]

A|N (iemand die waarden van zaken bepaalt) <form..> *taksateur, waardeerder* = een taxateur benoemen 'n taksateur benoem, 'n waardeerder benoem; een beëdigd taxateur 'n beëdigde taksateur

Alhoewel *waardeerder* in die A/N-gedeelte teregkom, is daar geen aanduiding vir die gebruiker dat *waardeerder* by verre die gebruiklikste vorm in Afrikaans is nie. So, byvoorbeeld, dui 'n soektog in die Media24 Argief vir die koerant *Beeld* daarop dat *taksateur* slegs in een berig voorgekom het teenoor 264 berigte waarin *waardeerder* voorkom. Die feit dat *waardeerder* nie in die sisteem as hoofbetekenis uitstaan nie, kan as 'n debiet betreffende sisteemkoste beskou word.

Die ordening van betekenisonderskeidings word ook deur die geamalgameerde sisteem dikteer. Die betekenisonderskeidings wat Afrikaans en Nederlands gemeen het, word eerste gegee, gevolg deur dié wat eksklusief is vir Nederlands en dan dié wat eksklusief Afrikaans is. So, byvoorbeeld, word in die bewerking van *ei/eier*, 'vrouwelijke geslachtscel' as eerste betekenisonderskeiding aangebied terwyl die tweede betekenis 'legsels v. bepaalde dieren, bv. vogels' by verre die hoofbetekenis vir *eier* in Afrikaans is. Bestudering van korpusdata dui daarop dat 93% van die gebruik van *eier* die tweede betekenisonderskeiding in (13) is en *eier* as 'vroulike geslagsel' slegs 7% van die gebruik uitmaak.

(13)

ei nw. [het; mv: -eren], **eier** nw. [mv: -s]

A|N • (vrouwelijke geslachtscel) *eier* = een bevrucht eitje 'n bevrugte eiertjie; het eitje in die baarmoeder die eiertjie in die baarmoeder

• (legsels v. bepaalde dieren, bv. vogels) {Opm.: In Afr. kom die woord 'ei' ook in dié betekenis voor, maar slegs in die uitdrukking 'vir 'n appel en 'n ei'.} *eier* = een ei leggen 'n eier lê; een ei uitbroeden 'n eier uitbroei; de eitjes van insecten die eiertjies van insekte = <fig.> voor een appel en een ei vir 'n appel en 'n ei

Dus kan ook ten opsigte van die ordening van betekenisonderskeidings beweer word dat die sisteemkoste hoog is met betrekking tot gebruikersinligting omdat die gebruiker nie kan sien watter betekenisonderskeiding die mees frekwente een is nie.

Meer problematies is gevalle waar die hoofbetekenis van bepaalde Afrikaanse lemmas nie aangebied word nie, soos byvoorbeeld by *kaal/kaal* en *aansit/aansit* in (14) en (9).

(14)

kaal bnw., **kaal** b.nw.

A|N ● (**arm(oedig)**) **kaal** ≠ een kale boel/bedoening/ontvangst 'n *power spul/toestand/overwelkoming* ≠ <fig.> hoe kaler, hoe royaler *hoe kaler, hoe rojaler, hoe kaler jonker, hoe groter pronker, hoe kaler jakkals, hoe groter stert*; <fig.> een kale neet 'n *kaal jakkals*; <fig.> zo kaal als een luis so *kaal soos 'n luis*; <fig.> zo kaal als een kerkrat/neet so *kaal soos 'n kerkmuis/neet*; <fig., vero.> een kale juffer 'n *onbemiddelde dame wat haar deffig voordoet*; <fig.> kale kak *pure aanstellery/snobisme*; <fig.> kale drukte 'n *ophef oor niks* ≠ **kaal uittrek** *agterbly/terugkom* berooid *achterblyven/terugkeren* ≠ <fig.> *iemand kaal uittrek* iemand *kaalplukken*; <fig.> *iemand rot en kaal steel* iemand *uitschudden*

● (**onbegroeid, onbedekt**) **kaal, bles, bleskop, haarloos** = kale bomen/muren *kaal bome/mure*; iets kaal vreten *iets kaal vreet* ≠ een kale deur 'n *ongeverfde deur*; je kaal laten knippen/scheren *jou kop kaal laat knip/skeer*; een kaal hoofd 'n *kaal kop, 'n bleskop*; een kale man 'n *man met 'n bles, 'n kaalkopman*; een kale plek 'n *kaal kol, 'n bles* ≠ <fig.> de kale huur *die basiese huur*; <fig.> kaal worden/zijn *bles word/wees, deur jou hare groei, kaal word/wees* (bv. *deur blare te verloor*); <fig.> zo kaal als een biljartbal/knikker zijn *totaal kaalkop wees, heeltemal bles wees* ≠ met 'n *kaal bolyf* met ontbloot bovenlijf; *kaal sonaanbidders* ontblote zoonaanbidders; *kaal arms/bene* blote armen/benen; *sonder sy serpie voel hy so kaal* zonder zijn sjaltje voelt hij zich zo naakt; *kaal rotse* naakte rotsen; *kaal nael flitsen, streaken*; *kaal rondloop* naakt rondlopen; *kaal swem* naaktzwemmen ≠ <fig.> *iemand met jou oë kaal uittrek* iemand de kleren van het lijf kijken

In die geval van *kaal/kaal* sorg die ordening van die betekenisonderskeidings 1. "arm(oedig)" en 2. "onbegroeid, onbedekt" dat daar geen prominensie aan die Afrikaanse hoofbetekenis "sonder klere" gegee word nie. Die eerste vermelding van *kaal* in dié betekenis word aan die einde van die tweede betekenisonderskeiding gemaak. Aangesien die artikel redelik lank is, dra dit daartoe by dat die Afrikaanse hoofbetekenis maklik misgekyk kan word. Die gebruiker wat Nederlands ken, en *kaal* opsoek, vind dus nie sonder meer die hoofbetekenis van *kaal* in Afrikaans as "sonder klere" nie. Volgens korpusdata verteenwoordig laasgenoemde betekenis ruim 50% van die gebruik van *kaal*. eHat gee wel die nodige prominensie deur dit as die eerste betekenisonderskeiding aan te bied. Vergelyk (15).

(15) eHat

kaal b.nw., bw.

1(a) Sonder klere aan die lyf; naak ... (b) Sonder klere aan 'n gedeelte van die liggaam ...

Net so word die hoofbetekenis van *aansit*, te wete 'by iets, bv. 'n tafel aansit', nie aangegee nie, selfs nie eers in die A-gedeelte (sien (9) hierbo) wat spesifiek op Afrikaans afgestem is nie. Ook in hierdie geval dui korpusanalise op 'n 28% gebruik van *aansit* as "by iets, bv. 'n tafel aansit" gevolg deur 20% "aanwerk/aanheg/vasheg", 17% "aanskakel", 15% "gewig optel/aansit" met 'n opmerklike 12% van die gebruik vir die uitdrukking *ore aansit*. (Laasgenoemde verdien sekerlik vermelding as gebruiksvoorbeeld in Afrikaanse woordeboeke.) eHat se betekenisonderskeiding 1 en 2 in (10) is dus nommerpas, en by 1 word ook die

mees tipiese gebruiksvoorbeeld, naamlik om by 'n tafel aan te sit, gegee.

Benewens bogenoemde gevalle waar die mees frekwente betekenis nie eerste gelys of as sodanig markeer word nie, en gevalle waar die hoofbetekenis glad nie gelys word nie, moet die sisteemkoste ook bereken word vir die gevalle soos *vaak/vaak*, *bestellen/bestel*, en *bakkie/bakkie* waar duidelike betekenisoorvleueling maar ook belangrike betekenisverskille voorkom. Dit behels die lemmatisering en bewerking van polisemiese lemmas wat vormlik ooreenkom maar in een of meer betekenisonderskeidings verskil. Dit kan ook 'n valse-vriende-verhouding impliseer. Gedeeltelike kognate is problematies en vra om 'n komplekse hantering.

ANNA lemmatiseer en bewerk wel Nederlandse en Afrikaanse woorde wat slegs vormlik ooreenstem en geen enkele betekenisonderskeiding in gemeen het nie, as valse vriende. Gevalle soos *verskoon*, *verskoning* en *amper/amper* word in ANNA as valse vriende hanteer. Vergelyk (16) en (17) vir *verskoning* en *amper*.

(16)

verskoning nw. [mv: *verskonings*] {!!}

- A ● (*apologie*) **excuus** ≠ *verskoning maak/aanbied* je excuus/excuses maken/aanbieden, je verontschuldigen aanbieden/maken/uiten; *by jou vriende verskoning maak* je bij/tegenover je vrienden verontschuldigen; (*om*) *verskoning vra vir jou gedrag* je verontschuldigen voor je gedrag, je excuseren voor je gedrag
- (*ekskuus, uitvlug*) **smoes, uitvlucht, excuus** = *verskonings soek/vind* uitvluchten zoeken/vinden; 'n *geldige/goeie verskoning* hê een geldig/goed excuus hebben ≠ 'n *swak/flou verskoning* een slap/mager excuus; *iets as verskoning gebruik* iets als excuus aanvoeren/gebruiken; 'n *verskoning om na die bottel te gryp* een alibi om naar de fles te grijpen; 'n *verskoning versin/uitdink* een uitvlucht verzinnen; *as verskoning (aanvoer dat ...)* ter verontschuldiging (aanvoeren dat ...); *verskonings! verskonings!* allemaal smoesjes!
- (*kennisgewing v. afwesigheid*) **afzegging** ≠ *verskoning aanteken* verstek aantekenen; *ondanks die baie uitnodigings is daar nog net drie verskonings* ondanks de vele uitnodigingen zijn er nog maar drie afzeggingen

(17)

amper¹ bw. {!!Opm.: In Afr. is 'amper' en 'byna' sinoniem.}

- A (*net nie/byna*) **bijna, haast** ≠ *dit het amper/haas onbetaalbaar geword* het is haast onbetaalbaar geworden; 'n *mens kan amper(s) nie jou oë glo nie* je gelooft haast niet wat je ziet; *so ampertjies in/net-net oor centimeters naast/over* ≠ <fig.> *amper, maar nog nie stamper nie* bijna maar niet helemaal

amper² bw. {!!Opm.: In Ned. is 'amper' en 'nauwelijks' sinoniem.}

- N (*nauwelijks*) **net-net, nouliks** ≠ *hy verdient amper genoeg om rond te komen hy verdien nouliks genoeg om uit te kom*; *hy kan amper bij de bovenste plank hy kan net-net by die boonste rak bykom*

Eenvoudig gestel, kan gesê word dat iemand wat *amper* 'n doel bereik het in Nederlands net-net daarin geslaag het, maar in Afrikaans net-net misluk het. Dit is die taak van die leksikograaf om duidelike gebruikersleiding ten opsigte van valse vriende te gee en die keuse van dubbele uitroepingstekens "!!" as merker in (16) en (17) suggereer die belangrikheid van die verskille. In (16) is

Afrikaans en Nederlands valse vriende met betrekking tot al drie die betekenisonderskeidings. ANNA se benadering tot valse vriende verskil van die een wat deur Gouws et al. (2004) voorgehou word, naamlik dat valsevriende-verhoudings ook as sodanig bewerk moet word in gevalle waar polisemiese woorde in minstens een betekenisonderskeiding in 'n valsevriende-relasie tot mekaar staan. Hulle is ook van mening dat gebruikers spesifieke leiding in dié verband moet kry:

When speakers of either Afrikaans or Dutch consult a bilingual Afrikaans/Dutch//Dutch/Afrikaans dictionary to retrieve information regarding the respective Dutch or Afrikaans lexical items *bestellen* and *bestel* they need guidance to warn them that although the items in both languages have the sense 'to order', only the Dutch item has the sense 'to deliver'. (Gouws et al. 2004: 802)

Vergelyk (18) en (19).

(18)

bakkie s.nw. (-s) **1 v.v.!** Ligte vragmotor: *Ons laai die tuinvullis sommer self op die ~ en neem dit na die stortingsterrein.* **2** 'n Klein houer: *Ma sit my toebroodjies soggens in 'n plastiese ~.* Verkleinwoord van **bak**.

Gouws et al. (2004: 798) gebruik die merker "v.v." en in gevalle waar valse vriende op grond van hoë gebruiksfrekwensie as "gevaarlik" beskou word, word 'n uitroepteken bygevoeg soos vir bakkie in (18).

(19)

DUTCH	AFRIKAANS	DUTCH	AFRIKAANS
vaak ¹ 'often'	dikwels	<i>bestellen</i>	
vaak ² 'sleepy'	vaak	1 'to order'	bestel
		2 'to deliver'	aflewer

Gouws et al. (2004: 800, 802)

Gevalle waar sowel betekenisoooreenkomste as -verskille vir 'n bepaalde lemma voorkom, word nie in ANNA as valse vriende hanteer nie. Martin (2012b: 421) hou tergen/*terg* as tipiese voorbeeld voor waar belangrike verskille slegs deur 'n opmerking ondersteun word.

(20)

tergen ww.tr., **terg** ww.tr. {Opm.: Hoewel 'terg' in Afr. ook wel 'treiter' kan beteken, word dit oor die algemeen minder sterk gebruik as in Ned. en beteken dit eerder skertsend as gemeen pla.}

A|N (**sarren**) **terg**, **tart**, **treiter** = iemand op alle moegelijk manieren tergen iemand op alle moontlike maniere tart ≠ graag terg graag plagen; veral Amsterdammers kan ongenadig terg

vooral Amsterdammers kunnen geweldig zuigen; *iemand terg omdat hy vet is* iemand plagen omdat hij dik is

Eerder as om op die verskille te konsentreer met betrekking tot valse vriende, verwys hy na hierdie benadering as 'n poging om weg te beweeg van verskille en beskou dit as deel van die nuwe-effekte van die geamalgameerde sisteem. Trouens, Martin (2012b: 420-421) verwys daarna met die sterk term "slaggate" in die sisteem:

More serious perhaps than the system side effects ... which ... can be overcome by the system, are, what I call, pitfalls. ... cases where there is a meaning overlap or minor meaning/usage differences between Afrikaans and Dutch. As a rule, in ANNA, we have chosen to abstract away from these differences, in other words, to 'lump', not to 'split', quite in line with the amalgamation approach itself.

Al drie die gevalle **bakkie/bakkie**, **bestellen/bestel** en **vaak/vaak** wat in Gouws et al. (2004) voorgehou word, en eersgenoemde selfs as gevaarlike valse vriend, word in ANNA gevolglik nie as valse vriende hanteer nie en resorteer in dieselfde kategorie as *tergen/terg*, naamlik die benadering waarvolgens daar "weg van die verskille geabstraheer word" en die verskille as gewone betekenisse bewerk word. Vergelyk byvoorbeeld die artikels van *bakkie/bakkie*, *vaak/vaak* en *bestellen/bestel* in (21).

(21)

bakkie nw. [het; mv: -s], **bakkie** nw. [mv: -s]

A|N (verkleinwoord v. 'bak') **bakkie**

N ● (27mc-zender) **radiosender**

● (aanhangwagentjie) **oop sleepwaentjie agter 'n motor**

● (kopje koffie) **koppie koffie**

A ● (ligte vrugmotor) **pick-up, kleine open bestelauto**

● (soort roeibootjie) **roeibootjie**

vaak¹ b.nw.

A (*slaperig*) **slaperig** ≠ *vaak* word slaap krijgen; *vaak wees* slaap hebben

vaak² nw. [de; mv: -], **vaak** nw. [mv: -]

A|N <BN> (*slaperig gevoel*) **vaak** = <fig.> *praatjies voor de vaak* *praatjies vir die vaak* ≠ *vaak* hebben *vaak wees* ≠ <fig.> *Klaas Vaak Klaas Vakie, sandmannetjie* ≠ *deur die vaak oorval word* door de slaap overvallen worden

vaak³ bw.

N (*veelal, in de meeste gevallen*) **dikwels, baiekeer, baiemaal** = we sijn vaak te laat voor de opening *ons is dikwels te laat vir die opening* ≠ het gebeurt steeds vaker *dit gebeur steeds meer dikwels, dit gebeur al hoe meer*; let wel op want de ruit is al vaak genoeg gebroken *wees versigtig, want die ruit is al dikwels genoeg gebreek*; hij ging vaak bij haar op bezoek *hy het baiekeer by haar gaan kuier*

bestel nw. [het; mv: -], **bestel** nw. [mv: -]

A|N (*organisasie*) **bestel, bedeling** = het publieke bestel *die publieke bestel*

bestellen ww.tr., **bestel** ww.tr.

A|N ● (*laten komen*) **bestel** = een boek/trui/biertje bestellen 'n *boek/trui/biertje bestel*; een

taxi bestellen 'n taxi bestel ≠ kos/drank/bestel voedsel/ drank laten aanrukken
● ((van post aan huis bezorgen) bestel = een telegram/een brief bestellen 'n telegram/
'n brief bestel = <fig.> een overledene ter aarde bestellen 'n oorledende ter aarde bestel

Die vraag is dus of *bakkie/bakkie*, *vaak/vaak* en *bestellen/bestel* sonder meer as valse vriende getipeer en bewerk moet word of alternatiewelik soos enige ander polisemiese lemma behandel moet word met byvoeging van addisionele gebruikersleiding soos in (20).

Daar kan geargumenteer word dat in geval van enige vorm van valse vriende dit belangrik is om die gebruiker bedag te maak op die verskille sodat die woorde nie verkeerd gebruik word nie.

Dit moet egter duidelik gestel word dat die bewerking van sulke gevalle in ANNA voldoende inligting aan die gebruiker bied al word hulle nie eksplisiet as valse vriende gemerk deur die woordeboek se konvensie "!!" nie. Die vraag is net of die gebruiker vaardig genoeg sal wees om uit (20) te konkludeer dat die hoofbetekenis van *tergen* deur byvoorbeeld *provokeer* weergegee kan word, terwyl *terg* met *pla* gedefinieer kan word. Wat gevalle soos *bakkie/bakkie*, *vaak/vaak* en *bestellen/bestel* betref, is die vraag of geamalgameerde woordeboeke die gebruiker nie meer eksplisiet bedag moet maak op veral "gevaarlike" valsevriende-verhoudings nie.

'n Afsonderlike grammatikale kompendium

Daar kan met die intrapslag gevra word of dit werklik nodig is om 'n sogenaamde grammatikale kompendium as 'n afsonderlike teks aan te bied. Uit 'n sekere hoek beskou, is dit nie in die gees van amalgamerings waar die onderliggende filosofie juis die reduksie van twee (of meer) lemmalyste tot 'n enkele lemmalys is nie. Tweedens is dit op verskeie vlakke gebruikersonvriendelik. Gebruikers wat normaalweg nie aandag aan die agtertekste in woordeboeke gee of die gebruiksinligting noukeurig lees nie, sal nie eers besef dat daar so 'n gedeelte bestaan nie. Dit kan vererger word deur die feit dat die bestaan van die kompendium nie prominent genoeg bekendgestel word nie of selfs moontlik dat gebruikers die begrip *kompendium* self nie verstaan nie, soos inderdaad blyk uit die gebruikerstudie van Bosman (2013).

Die Nederlandse frase wat 'n kruisverwysing na die agtertekste bevat, **zie kompendium**, is nie deur die studente verstaan nie of bloot geïgnoreer. Respondent B het byvoorbeeld gedink dit is 'n vertaalekwivalent vir die Afrikaanse woord *is*. ... 'n Ernstiger probleem is dat die frase **zie kompendium** wat gebruikers na hierdie agtertekste verwys, hoegenaamd nie in die gebruiksaanwysings verduidelik word nie. (Bosman 2013: 49)

Dit is onwaarskynlik dat veral onervare teikengebruikers sal weet dat voorsetsets, telwoorde, voegwoorde, voornaamwoorde en lidwoorde funksiewoorde is. Selfs al sou die teikengebruikers dit weet, is dit onwaarskynlik dat hulle

lemmas soos *diegene*, *genoeg*, *iemand* en *minder* as funksiewoorde sal herken wat in 'n afsonderlike lys opgesoek moet word. Dit geld veral dié woorde wat, benewens as funksiewoorde, ook as ander woordsoorte optree, en wat, soos wat ANNA daarna verwys, in die "eintlike woordeboekgedeelte" aangebied word. So, byvoorbeeld, is die kans gering dat die gebruiker sal weet om *genoeg* as bywoord in die hoofteks op te soek maar as voornaamwoord in die kompendium. Net so is dit onwaarskynlik dat hy/sy die kennis het om *minder* as byvoeglike naamwoord en ook as bywoord in die hoofteks op te soek maar as telwoord in die kompendium. Die gebruiker sal gevolglik eers via kruisverwysings by die lemmas uitkom, wat verhoging in inligtingskoste impliseer. Die afsonderlike kompendium verswaar dus die mediostruktuur en veroorsaak dat gebruikers meer in die woordeboek moet rondblaai. In die geval van *genoeg* en *minder*, byvoorbeeld, is die kruisverwysing duidelik: **genoeg**² vnw. zie **compendium**. Die kruisverwysing is minder gebruikersvriendelik as die gebruiker byvoorbeeld *by* wil opsoek, vergelyk (22):

(22)

by¹ nw. zie **bij** nw.

by² voors. zie **bij**

by- zie (ook) **bij-**

Die gebruiker wat nie weet dat voorsetsels in die kompendium opgeneem word nie, word nie ingelig dat **by**² in die *compendium* opgeneem en bewerk is nie. Hy/sy sal stellig eers na **bij** in die hoofteks gaan waar **bij** as naamwoord en werkwoord bewerk word met 'n kruisverwysing: **bij**³ voorz. zie **compendium**. Hoewel die soektog uiteindelik suksesvol is, is die inligtingskoste hoog.

(23)

A	1	P	934
B	45	PH	997
BJ	143	PS	1075
D	146	PSH	1076
E	174	PŠ	1078
F	195	PŠH	1079
FS	267	R	1082
FŠ	268	S	1153
G	269	Š	1199

Die leksikograaf het dikwels 'n moeilike keuse tussen 'n leksikografiese benadering wat "linguisties meer korrek" of die "beste opsie" is en 'n benadering wat minder akkuraat, maar meer gebruikersvriendelik is. 'n Tipiese voorbeeld wat groot gebruikersfrustrasie tot gevolg het, is die Groot Noord-Sotho Woordeboek (GNSW) waar die samestellers afgewyk het van 'n normale alfabetiese ordening ten gunste van 'n fonetiese ordening omdat laasgenoemde na hulle mening grammaties meer korrek is. Dit het onder meer tot gevolg dat lemmas

wat met *bu-* begin voor lemmas wat met *bj-* begin in die alfabetiese lys aan-gegee word. Hierdie benadering noodsaak dat 'n spesiale blad voor in die woordeboek geraadpleeg moet word alvorens die gemiddelde en selfs gevorderde gebruiker woorde kan opsoek. Vergelyk (23) wat 'n uittreksel uit die alfabetiese gids is.

Samevatting en aanbevelings

Dit kan onomwonde gestel word dat die geamalgameerde model, en ANNA in besonder, ongeëwenaard is wat betref die inligtingswaarde ten opsigte van vergelyking en kontrastering van twee of meer verwante tale. Die woordeboek slaag uitnemend in die sistematiese en konsekwente aanbieding van sodanige ooreenkomste en verskille in al die kognaatkategorieë wat vir ANNA onderskei word. Dit is in dié opsig 'n uitstekende inligtingsbron.

Die voordele van 'n geamalgameerde woordeboek moet egter opgeweeg word teen 'n aantal moontlike negatiewe aspekte. Eerstens is 'n mate van redundansie onvermydelik as gevolg van die behoefte aan kruisverwysings ten opsigte van die kompendium, asook minstens in kategorieë (b) en (d), waar 'n aparte lemma vir die ander lid van die taalpaar gegee moet word in die toepaslike alfabetiese posisie ten einde die gebruiker te lei na die lemma waar die nodige inligting gevind kan word.

Tweedens lei 'n geamalgameerde benadering tot relatief lang(er) artikels bloot omdat behandeling in dieselfde artikel aangebied word vir twee of meer tale (drie in die geval van die Sothotale en selfs vier vir die Ngunitale). Dit beteken dat die gebruiker langer inskrywings moet deurwerk om die inligting te bekom waarna hy/sy op soek is. Die deurlees van langer artikels staan in teenstelling tot huidige tendense in die leksikografie om te waak teen oormatige inligtingsaanbod en die gebruiker meer direk te lei na die presiese verlangde inligting. Vergelyk byvoorbeeld Prinsloo et al. (2011) waar gebruikers deur keuse-boomstrukture direk na die juiste inligting gelei word.

Derdens kan geredeneer word dat in teenstelling met die A/N-gedeelte in 'n gegewe woordeboekartikel, die afsonderlike A- en N-gedeeltes nie bydra tot ruimtebesparing in vergelyking met afsonderlike Afrikaans-Nederlandse en Nederlands-Afrikaanse woordeboeke of woordeboekdele nie.

'n Vierde veronderstelde voordeel van ANNA is dat dit geskik is as 'n leeswoordeboek, dit wil sê dat gebruikers dit sal lees vir ontspanning. Dit is te betwyfel en sal deur gebruikerstudies bevestig moet word.

Die geamalgameerde benadering tot die samestelling van woordeboeke het besliste voordele maar ook sekere beperkinge wat in hierdie artikel as *inligtingskoste* en *sisteemkoste* voorgehou is. Die leksikograaf moet deurgaans probeer om die voordele van die model maksimaal te benut en geen steen onaangeroer laat om die beperkinge van die benadering te minimaliseer nie. Dit is veral aspekte van gebruikersvriendelikheid wat besondere aandag moet geniet. Beperkinge wat toegeskryf kan word aan inherente eienskappe van die model

of, volgens Martin (2012b), gegewe newe-effekte of slaggate van die sisteem, kan nie almal sonder meer besweer word nie. Die leksikograaf moet egter daarna streef om dié tipes inligtings- en sisteemkoste uit te weer wat onnodig negatief op gebruikersvriendelikheid inwerk. In beginsel moet hy/sy dus 'n onderskeid maak tussen inherente sisteemeienskappe en -beperkings versus dit waaraan die leksikograaf wel iets kan doen.

'n Eerste stap kan wees om omvattende gebruikersinligting saam te stel, soos wat dit voortreflik in ANNA gedoen is en dan die inligting te stratifiseer deur middel van 'n enkelblad-gebruikersgids heel voor in die woordeboek. 'n Tweede stap is die oordrag van inligting aan die gebruiker ten opsigte van die algemene aard van die woordeboek, die amalgamasiemodel, die tipes trefwoorde, bou en struktuur van die artikels, behandeling van voorbeelde en verbindings, verwysings, agtertekste, ens. 'n Derde vlak kan meer inligting rakende die inherente kenmerke van die sisteem soos die aanbod en ordening van ekwivalente en betekenisonderskeidings verskaf. Samestellers van geamalgameerde woordeboeke kan dit selfs oorweeg om 'n demonstrasie van die gebruik in een of ander elektroniese formaat by die papier- sowel as elektroniese weergawe in te sluit. Soos wat dit vir die samestelling van gebruikersgidse in enige woordeboek geld, moet die leksikograaf seker maak dat daar veral nie foute in rigtinggewende voorbeelde in die gebruikersgids is nie.

Geamalgameerde lemmalyste noodsaak wel afsonderlike lemmatisering van sekere kategorieë van woorde soos niekognate en kognate met 'n vormverskil. Hier moet die mediostruktuur die gebruiker lei na die korrekte lemma waar 'n bewerking aangebied word. Dit is belangrik dat die leksikograaf konsekwent sal hou by die redaksionele beleid soos in die gebruikersleiding aangedui is, byvoorbeeld die verstekreël "groter as sewe" wat in ANNA gebruik word. Hierbenewens moet versigtige oorweging gegee word aan die teiken-gebruiker se vermoë om die verband tussen vormlik verwante lemmas te kan insien. Indien die leksikograaf enigsins twyfel, kan 'n ad hoc-besluit geneem word om sekere woorde in die teikentaal of -tale te lemmatiseer.

Ten opsigte van vertaalekwivalente en die ordening van betekenis behoort samestellers dit sterk te oorweeg om vertaalekwivalente en betekenisonderskeidings wat die hoofbetekenis of mees frekwente betekenisonderskeiding in die teikentaal verwoord, maar nie eerste aangebied kan word nie, op een of ander wyse as sodanig te markeer.

Samestellers kan verder meer aandag gee aan valse vriende-verhoudings by polisemiese lemmas en dit oorweeg om dié verhoudings baie duidelik te markeer of die betrokke lemmas formeel as valse vriende te bewerk.

Die leksikograaf moet die voor- en nadele van 'n afsonderlike kompendium vir funksiewoorde in geamalgameerde woordeboeke krities teen mekaar opweeg en besluit of dit nie meer gebruikersvriendelik sal wees om met 'n enkele geamalgameerde lemmalys te volstaan nie. Dit sou dan steeds moontlik wees om, indien nodig, funksiewoorde op een of ander gepaste wyse te markeer.

Ten slotte kan gesê word dat ten spyte van sekere tekortkominge inherent aan die amalgamasiebenadering, ANNA 'n woordeboek van hoë leksikografiese prestasie is en 'n *monument* vir die Afrikaanse en Nederlandse leksikografie!

Acknowledgement

This work is based on the research 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 those of the author, and that the NRF accepts no liability whatsoever in this regard.

Verwysings

Woordeboeke

- ANNA = Martin, Willy (Hoofred.). 2011. *Pharos Groot Woordeboek. Afrikaans en Nederlands (Prisma Groot Woordenboek Afrikaans en Nederlands)*. Kaapstad: Pharos.
- e-HAT = Odendal, F.F. en R.H. Gouws. *Verklarende Handwoordeboek van die Afrikaanse Taal. Vyfde uitgawe op CD ROM*. 2005. Kaapstad: Pearson.
- 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.
- Hartmann, R.R.K. en G. James.** 1998. *Dictionary of Lexicography*. Londen/New York: Routledge.
- Van Dale Groot Woordenboek op CD ROM = *Van Dale Groot Woordenboek Hedendaags Nederlands. Versie 2.0*. 2002. Van Dale Lexicografie bv. Utrecht/Antwerpen. Zocksoftware 2002. C-Content, Rosmalen.

Ander literatuur

- 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.
- Cognitive Overhead, Or Why Your Product Isn't As Simple As You Think*. Lieb, David = <http://techcrunch.com/2013/04/20/cognitive-overhead/>. (Geraadpleeg 22 Mei 2013).
- Gouws, R.H., D.J. Prinsloo en G.-M. de Schryver.** 2004. Friends will be Friends – True or False. Lexicographic Approaches to the Treatment of False Friends. Williams, G. en S. Vessier (Reds.). 2004. *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004, Lorient, France, July 6-10, 2004*: 797-805. Lorient: Faculté des Lettres et des Sciences Humaines, Université de Bretagne Sud.

- Marais, R.** 2011. Een mooie dikke dame. "ANNA", het Groot Woordenboek Afrikaans en Nederlands. *Ons Erfdeel* 1: 190-192.
- Martin, Willy en 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 (Reds.). 2000. *Proceedings of the Ninth EURALEX International Congress, 8-12 August 2000*: 783-792. Stuttgart: Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.
- Martin, W.** 2012a. Amalgamated Bilingual Dictionaries. Genis, René et al. (Reds.). 2012. *Between East and West. Festschrift for Wim Honselaar on the Occasion of his 65th Birthday. Pegasus Oost-Europese 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. *Media24 Argief: Beeld* = <http://152.111.1.88/cgi-bin/beeld.cgi> (Geraadpleeg 6 Mei 2013).
- Prinsloo, D.J.** 2006. Compiling a Bidirectional Dictionary Bridging English and the Sotho Languages: A Viability Study. *Lexikos* 16: 193-204.
- 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 (Reds.). 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.

Old Wisdom: The Highly Relevant Lexicographical Knowledge Obtainable from a Specialized Dictionary from 1774*

Sven Tarp, *Department of Afrikaans and Dutch, University of Stellenbosch, Stellenbosch, South Africa and Centre for Lexicography, Aarhus University, Aarhus, Denmark (st@asb.dk)*

Abstract: There is little doubt that lexicography has experienced important progress during the years, especially when one compares the lexicographical works from the various centuries. However, history should not be viewed as an ever-growing progress but as a process with its ups and downs. In this respect, some old lexicographical works, especially from the Age of Enlightenment, are in some aspects extremely advanced even compared with present-day dictionaries. Malachy Postlethwayt's *Universal Dictionary of Trade and Commerce* from 1774 is one such example. The article provides a general presentation of this rare and interesting dictionary which was published in various editions in the 18th century. The presentation includes a detailed description of the objective, functions, structure, and content of this impressive work. The article then focuses on its cognitive and operative functions, and the way they are integrated into one and the same reference work. Upon this basis, the article shows how modern lexicography can use past experiences in order to meet the real and practical needs of present-day users.

Keywords: LEXICOGRAPHY, HISTORICAL LEXICOGRAPHY, FUNCTION THEORY, OPERATIVE FUNCTIONS, COGNITIVE FUNCTIONS, DICTIONARIES OF ECONOMICS, INTERDISCIPLINARY VOCATION OF LEXICOGRAPHY, INFORMATION TOOLS

Opsomming: Ou wysheid: Die hoogs tersaaklike leksikografiese kennis bekombaar uit 'n gespesialiseerde woordeboek uit 1774. Daar is min twyfel dat die leksikografie belangrike vooruitgang deur die jare ondergaan het, veral wanneer 'n mens leksikografiese werke uit die verskillende eeue vergelyk. Geskiedenis moet egter nie gesien word as 'n steeds groeiende vooruitgang nie, maar as 'n proses met sy wisselvallighede. In dié opsig is sommige ou leksikografiese werke, veral uit die Eeu van Verligting, in sommige aspekte uiters gevorderd, selfs vergeleke met die hedendaagse woordeboeke. Malachy Postlethwayt se *Universal Dictionary of Trade and Commerce* van 1774 is een so 'n voorbeeld. Die artikel verskaf 'n algemene

* This article was presented as a paper at the Seventeenth Annual International Conference of the African Association for Lexicography (AFRILEX), which was hosted by the Department of African Languages, University of Pretoria, Pretoria, South Africa, 2–5 July 2012.

bekendstelling van hierdie seldsame en interessante woordeboek wat in verskillende uitgawes in die 18de eeu gepubliseer is. Die bekendstelling sluit 'n uitvoerige beskrywing van die doel, funksies, struktuur en inhoud van hierdie indrukwekkende werk in. Die artikel fokus daarna op sy kognitiewe en operatiewe funksies, en die manier waarop hulle in een en dieselfde naslaanwerk geïntegreer is. Op grond hiervan toon die artikel hoe die moderne leksikografie vroeëre ervaring kan gebruik om aan die werklike en praktiese behoeftes van hedendaagse gebruikers te voldoen.

Sleutelwoorde: LEKSIKOGRAFIE, HISTORIESE LEKSIKOGRAFIE, FUNKSIETEORIE, OPERATIEWE FUNKSIES, KOGNITIEWE FUNKSIES, EKONOMIEWOORDEBOEKE, INTERDISIPLINÊRE ROEPING VAN DIE LEKSIKOGRAFIE, INLIGTINGSGEREEDSKAP

1. Introduction

There is little doubt that lexicography, both as a practice and a theory, has experienced important progress during the years, especially when one compares the lexicographical works from the various centuries. However, history — including lexicographical history — should not be viewed in a perspective of ever-growing progress but as a process with its ups and downs. The above statement should therefore not be regarded as something absolute and rigid, especially not because there are important exceptions to this general rule to be found in various lexicographical works, especially those under influence of the Age of Enlightenment which, in some aspects, are extremely advanced even compared with present-day dictionaries.

In the Age of Enlightenment, i.e. only a few hundred years ago, lexicography was much more a vocation than a profession. Many dictionaries from that period, both famous ones and long forgotten ones, were compiled by people who were paid nothing or very little for their efforts; they simply felt the compulsion to make them. They were people who did not need to carry out any sophisticated user research; in most cases they knew their future users personally, they knew their specific problems and needs, and often they had experienced the same kinds of problems and needs themselves. They were often practical people: the captain who compiled a dictionary in order to allow his officials, and often also himself, to communicate and solve technical and commercial problems when they entered a harbour in a distant country; the priest who wrote a small glossary to assist his simple-minded confirmation students when they had problems understanding difficult words and expressions used in the catechism; the missionary who produced bilingual dictionaries in order to spread the evangelism and preach it in "exotic" countries and languages, etc. There were also great thinkers like Diderot and d'Alembert, who were convinced that science and enlightenment communicated in an easily accessible way would change the world for the better; there were scholars like Samuel Johnson who simply wanted to ascertain the purity and standard of English and assist his fellow countrymen in improving their English skills.

There were a lot of people, many of whom were industrious, visionary, and innovative authors who did not pay much attention to established con-

ventions within a field that was still not generally known as lexicography. They were not afraid of experiments and innovations. They took practical decisions and opted for practical solutions to the problems they knew so well. When they compiled dictionaries for the people whom they knew personally, they simply used the forms, methods, and techniques which they considered most appropriate to solve problems which they frequently had experienced in their own life. Sadly enough, this lexicographical spring was followed, not by a warm and sunny summer, but by a cold winter called romanticism which killed many of the seeds planted by the previous generations and left them, so to say, to "the gnawing criticism of the mice". In this way, many valuable experiences and innovations were condemned to an anonymous existence on the dusty shelves of old libraries.

Malachy Postlethwayt's *The Universal Dictionary of Trade and Commerce* is one such example. The full title of this specialized dictionary — which is in fact more a narrative than a title (see the literature list) — provides immediate information about its content and its usefulness for the potential users. It was first published in London in instalments between 1751 and 1755 and later as a two-volume book in three other editions, in 1757, 1766, and 1774. As far as it has been possible to investigate, the dictionary is not mentioned in any recent literature on lexicography, not even in the volume on *Specialized Dictionaries* in the *History of English Lexicography* (Cowie 2009), nor in any of the relatively few contributions on historical lexicography published in *Lexikos*, *Lexicographica*, *International Journal of Lexicography* and other journals of lexicography.

By contrast, Postlethwayt has been briefly presented in various British bibliographies, e.g. Chalmers (1816) and Hewins (1895). And to this should be added that Postlethwayt and his work have been discussed several times by a number of experts in economics and economic history, among them Marx (1867), Johnson (1965), Groenewegen (2004), Mankin (2008), and Besomi (2011). These researchers, however, have obviously not analysed the dictionary from a lexicographical perspective but from the point of view of economics, just as various linguists frequently analyse some specific types of dictionaries from the point of view of their own discipline.

The following contribution will try to remedy this omission, not because Postlethwayt's dictionary surpasses other economic dictionaries from the same period, but because Postlethwayt — just like Samuel Johnson seven years before the publication of *A Dictionary of the English Language* — also published a dissertation about the dictionary he was working on, cf. Postlethwayt (1749). This dissertation — together with the *Introduction* in the dictionary — provides a rare insight into the author's motivations and own ideas about the content and structure of his brainchild.

2. Who was Malachy Postlethwayt?

Malachy Postlethwayt (1707–1767) was a British economist and publicist of cer-

tain reputation in his epoch; in fact, his writings are claimed to "have exerted a good deal of influence on the trend of British economic thought", cf. Johnson (1965: 185). He was the author of a number of essays on trade and commerce, but his most noted work is undoubtedly *The Universal Dictionary of Trade and Commerce*, to which "he devoted twenty years", cf. Hewins (1895: 219).

Apart from his publications, little is known about Postlethwayt's life, family, and social position. In his biography, Chalmers (1816: 219) provides the following details:

He was born about the year 1707 but where, of what parents, or how educated, we have not discovered. In the introductory discourse to his work entitled *Great Britain's True System*, he informs us, that nature having given him a very tender and weak constitution, he studiously declined and avoided, as much as he could, every degree of public life, as being inconsistent with, and indeed destructive of, that small snare of health which he had several years enjoyed, and which his studies had not mended and yet he preferred the studious life, as being more independent. He complains, however, of want of encouragement; and "humbly hopes that some people will be candid and ingenuous enough to think that he has a right to be treated upon a footing something different from that of an upstart idle chemist or projector, who has never given proof of any talents that might deserve the public regard and attention." Whether this complaint was redressed, we know not. He died Sept. 17, 1767, and probably not in very opulent circumstances, as he was buried in Old-street church-yard. The coffin, at his own request, was filled with unslacked lime.

If this is true, Postlethwayt seems to be one of the many lexicographers who got more fame than money for their efforts. In addition, the British economist has been criticised for being a plagiarist. According to Johnson (1965: 205), his dictionary included "ideas taken from fifty other past or contemporary writers" as well as "practically all" of the *Essai sur la nature de commerce en général* (Essay on the Nature of Commerce in General) published by Richard Cantillon in 1755. However, in the opinion of Groenewegen (2004: 999), "this accusation has been greatly exaggerated".

The discussion of plagiarism will be left to the experts in the history of economics. Apart from the fact that no copyright regulations existed in the 18th century, in a lexicographical perspective it is not so relevant if a dictionary is the brainchild of the author himself, or if it has been copied from other sources. Most important for a lexicographical work is that it serves its purpose and meets the needs of its genuine end users. It should never be forgotten that the world's greatest dictionary ever, the Chinese *Yongle Dadian* of a total of 11,095 volumes, was partially copied from a large number of already existing books whose content, in this way, was made more easily accessible.

Also important to note about Postlethwayt is "his greater interest in political problems; his more intense economic nationalism", cf. Johnson (1965: 402). This political interest and "economic nationalism" were expressed in various ways, first of all in the fact that Postlethwayt's major objective with his diction-

ary was to strengthen and raise the British Empire, an objective that he openly stated both in his plan for the dictionary, cf. Postlethwayt (1749), and in his *Introduction* to its fourth edition, cf. Postlethwayt (1774: v):

The great end aimed at by the author is to promote and advance, to the utmost of his abilities, the commercial prosperity and happiness of the British empire, and her colonies and plantations in America.

Postlethwayt's political and economic nationalism — today it would be called colonialist and imperialist ideas — also had other serious implications. During his life he was considered by many to be a paid agent of the Royal Africa Company. Nothing was ever proven but there is no doubt that he was a dedicated propagandist for the Company's mercantilist interests, which he defended and promoted in some of his publications, among them *The African Trade, the Great Pillar and Supporter of the British Plantation Trade in North America* (Postlethwayt 1745). In this essay he supported British expansion through trade with Africa and the colonies, and promoted the importance of slavery for British commerce and industry.

Whether an agent or not, the conclusion is that Postlethwayt was a child of his time, an intellectual who, like many other British, each of them in their own way, contributed to the colonialist destruction of Africa. In this respect, it is important to emphasize that the author's political, economic, commercial, and other ideas reflected in the dictionary are of purely historical interest; it is not from these outdated ideas that present-day lexicography may learn something from Postlethwayt, but from the way he prepared his data and transmitted his message by means of a dictionary.

3. Objectives of the dictionary

The Universal Dictionary of Trade and Commerce was partially a translation of a similar French work, Jacques Savary des Bruslons' *Dictionnaire universel de commerce*, first published in 1723 and subsequently in various editions and translated, apart from the English version, into German as *Allgemeine Schatz-Kammer der Kaufmannschaft* (1741–1743), and Italian as *Dizionario di commercio dei signori Fratelli Savary* (1770), cf. Besomi (2011: 28).

In 1749, i.e. two years before the appearance of the first instalment of his dictionary, Postlethwayt published a short dissertation on "the Plan, Use, and Importance" of the dictionary. In the title of this 52-page booklet, the author openly states that it will be translated from the French original, compiled by "the late Celebrated Mons. Savary, Inspector-General of the Manufactures of the Custom-House of Paris". However, he also adds that it will contain "considerable Additions and Improvements" from his own hand, and that it is "Addressed to the Nobility, Gentry, Merchants, and Traders of Great Britain" with the objective to "more particularly accommodate the Fame to the Trade and Navigation of the British Empire".

As to the important question of the sources to the wealth of a nation, later treated with much more scientific rigour by Adam Smith (1776), Postlethwayt (1749: 1) writes:

It is allowed on all hands, that agriculture and commerce nourish and enrich a nation.

In this respect, the author deplors that the British Empire suffered from some serious problems. He explains how the various people engaged in trade and commerce in one way or another — i.e. "Statesman, the Senator, the private Gentleman, the Trader, or the Manufacturer" — frequently do not have a "satisfactory knowledge of Facts in complicated matters of a commercial nature", and that the acquisition of this knowledge is not an easy matter, cf. Postlethwayt (1749: 2):

Foreign and domestic trade admitting of so infinite variety of matter, and the knowledge communicated to the world, by those skilled and experienced therein, being scattered in an infinity of volumes, it is no easy matter to have immediate recourse to what may be occasionally requisite.

For Postlethwayt (1749: 2), the solution to this serious problem for the British trade and commerce, and eventually for the British Empire, has to be found in lexicography:

A subject of this extensive nature therefore being reduced to the form of a Dictionary, for alphabetical reference, seems the most naturally adapted to answer these desirable purposes, and especially so, as the compilers can have no motive to deceive.

As an example of the practicability of this idea, he refers explicitly to Savary des Bruslons' *Dictionnaire universel de commerce*, a "celebrated work" which has proven "how far an universal knowledge of commerce is capable of being reduced into the like form". Postlethwayt (1749: 3) then lets this work "speak for itself" in order to:

communicate, not only the *original Plan*, but to give a succinct account of such additional matter, as is proposed to be incorporated throughout the whole, in an english dress, to the end that the fame may be more peculiarly adapted to the state of the trade and navigation of the british empire.

In these few lines we acquire a fairly good idea of the principal intention of Postlethwayt's dictionary:

- His main objective is to strengthen the British Empire which, according to him, is nourished by its agriculture and commerce.
- However, he has observed a serious problem, namely that the relevant people do not possess satisfactory knowledge of facts regarding trade and commerce, and that these people, in addition, have neither time nor

the possibility to obtain this knowledge because it is scattered in an infinity of volumes.

- For the benefit of the Empire and its commerce, he therefore sets himself the task of solving the problem by means of an alphabetically structured dictionary providing access to the required information.

In this way, the objective of the dictionary, its target user group, the users' needs, and the context where the needs may occur, are defined by the author much better than in many modern dictionaries where publicity has taken over; or they seem to be so defined, because Postlethwayt has more surprises up his sleeve.

4. The fourth edition

In 1774, seven years after Postlethwayt's death, the fourth edition of his *Universal Dictionary* appeared. Nothing is said about the author and his fate, or why it was not published before, but both the title and the content suggest that Postlethwayt worked on the dictionary after the Peace Treaty in 1763, probably until 1767 when he suddenly died "as he had often wished", cf. Chalmers (1816: 220). The fourth edition is the most developed and the one to which Postlethwayt contributed most of his own material; it is generally this edition that is discussed by economists and historians, and it will therefore also be the one to be dealt with in what follows.

The fourth edition of *The Universal Dictionary of Trade of Commerce* consists, like the previous editions, of two volumes, each of which having the considerable size of 44 x 29 x 17 cm. It comprises five main sections:

- Dedication
- Introduction, including "a General Idea of the Content"
- Preliminary Discourse the First
- Preliminary Discourse the Second
- The Dictionary as such (without pagination)

The dedication, dated February 10, 1766, i.e. a year and a half before the author's death, is addressed to the Mayor of the City of London as well as to the city's merchants and tradesmen who, according to the author, "constitute the very active Soul of the Commerce of the whole British State". The entire tone of this dedication reflects the "want of encouragement" mentioned by Chalmers (1816).

The first part of the *Preliminary Discourse* is Postlethwayt's personal introduction to the subject where he mainly explains "the Present Commercial Grievances" and the deplorable situation in which Britain finds itself according

to his understanding. The second part of the *Preliminary Discourse* contains the author's own visions and recommendations on how Great Britain could use the Peace Treaty from 1763 with Spain "to redress the National Grievances represented in the preceding Preliminary Discourse", and make "the Peace of Europe more lasting, and to save Great Britain the future Expence of engaging in the Continental Wars of Europe", a fact showing how he also intended to use the dictionary as a medium to communicate his personal ideas for exerting influence on political and economic affairs. From a lexicographical perspective, the two parts of the *Preliminary Discourse* represent, in an embryonic form, a prelude to the systematic introductions included in a number of modern specialized dictionaries, e.g. the *Encyclopedic Dictionary of Gene Technology*.

It is, however, the *Introduction* that contains the most interesting comments. Here the author initially presents his dictionary as "the first of its kind that was ever published in Great Britain", and then continues in a far from modest way saying that the previous editions have given it a reputation that makes recommendations almost superfluous (page iii):

Having obtained a reputation for above fourteen years, with those esteemed the best judges in the kingdom, amongst the nobility and gentry; and its credit and character also amongst commercial people of the best understanding, as well merchants as tradesmen, manufacturers and artisans of every respectable rank, render it the less necessary to urge any thing now in its recommendation.

In the *Introduction*, the author — apart from stating the objective of the dictionary already quoted above — also touches on a topic which permeates the whole dictionary, i.e. the relation to and envy of France; this country is considered not so much an enemy as a competitor who in many respects (law, trade, manufacture, etc.) is doing far better than Britain, and from whom the country and its traders and manufacturers can learn much (page iii):

But the difference between us and France conflicts chiefly in this; that they take no less care in the execution of their laws of trade, than in making them: We are remarkable for good laws, but are shamefully neglectful in their execution. It is not upon this account, that our bakers are suffered to adulterate their bread, to the injury of the public health; and to daily trick the poor in the weight, as well as the rich and poor in the quality of their bread; In France these things are better guarded against.

Anybody who has read novels and social reports from London in the 18th and 19th century will know that the city's bakers continued to adulterate their bread for a long period, and that Postlethwayt's recommendations did not seem to have any notable effect on this regrettable practice. However, from a modern lexicographical point of view, probably the most intriguing particular in the *Introduction* is that Postlethwayt clearly explains that the dictionary does not only have a cognitive function providing knowledge to its users (page v):

The whole containing a greater fund of commercial knowledge, relating to the

universal trade and navigation, both practical and political, than ever appeared in any dictionary of commerce, in this, or any other state or empire.

What is really surprising is that the author also conceived the dictionary with an operative function in order to provide its intended user group with various sorts of proposals, recommendations and suggestions to be used in different contexts (page iii):

This work abounds with no small variety of suggestions and intimations for the advancement of numerous branches of the trade and commerce of these kingdoms; which every discerning reader will make his proper use of, either for his own private benefit and advantage, or those of the public. Many manufacturers have wrote letters to the author, returning him their grateful thanks for the services he has laboured to do them; and he has been happily instrumental to set several new manufactures on foot for the public interest.

Operative functions are defined as the lexicographical assistance to users who need "advice and instructions in order to perform any kind of mental or manual action"; cf. Tarp (2008: 185).

It is only during the last five or six years that modern lexicography — and especially the supporters of the function theory — has started discussing these functions and the need to have them covered by various types of dictionaries and other information tools, an idea which has still not been generally accepted by the lexicographical community. For instance, at the Seventeenth International Conference of AFRILEX, held at the University of Pretoria in July 2012, an interesting discussion developed when the French-Danish lexicographer and supporter of the function theory, Patrick Leroyer (2012), presented his vision for a wine dictionary with both cognitive and operative functions, and was then contested by the President of AFRILEX, Gilles-Maurice de Schryver, who claimed that this was not lexicography.

However, returning to Postlethwayt — and the Age of Enlightenment — it will be found that operative functions as defined by the function theory were already an integrated part of lexicography in that period. And the reason for this was that the authors knew that their users demanded this sort of information as it is clearly expressed in the letters from the manufacturers, to which Postlethwayt refers in the above quotation.

5. The content

In a spirit typical for the 18th century, Postlethwayt (1749) writes that he has read everything that is worth reading about trade and commerce. In a certain manner it seems that he has not only read "everything", but also put "everything" into his *Universal Dictionary*, described by Johnson (1965: 188) as a "huge storehouse of economic facts, laws and theory". It is therefore not easy to give a short description of its rich content and structure.

The dictionary is structured in alphabetically organised articles, some of which contain very short definitions of only a few lines whereas others consist of complex texts of many pages. Many of the articles include long extracts from laws, letters, books, newspapers, and other documents. As a rule these long articles start with a definition, or explanation in a broader sense of the word, then incorporates some extracts from already existing documents, a section called Remarks and sometimes even Additional Remarks added by the author to the various editions. The description of the phenomena is often mixed with Postlethwayt's personal opinions, a detailed explanation of how they are produced and used, as well as recommendations on how to improve their production. The articles frequently contain hand-drawn illustrations, tables, and examples of accounting covering several pages.

The topics treated in the dictionary are of a large variety, in some cases only marginally related to trade and commerce as it is conceived today:

- Trade, commerce, law, and accounting
- Economic theory and practice
- Banking system, money, gold, and silver
- Navigation and naval affairs
- Measures, mathematics, and arithmetic
- Manufacture: silk-worms, brewery, malt, bricklaying, and ship building
- Cities, regions, and countries
- Anecdotes of distinguished citizens of London
- Etc, etc., etc.

The above list could easily be extended to cover several pages. Hewins (1895: 219) considered that Postlethwayt "presented his results without method or conciseness", and in a certain sense he is right. If his dictionary is compared to Savary des Bruslons' *Dictionnaire universel de commerce*, then it is evident that the latter is better structured and therefore easier to access. For instance, in the French dictionary, the article *commerce* covers a total of 250 pages and is internally structured with many sublemmata organised in a systematic way. In Postlethwayt's dictionary, the same article covers only about 10 pages, mainly owing to the different language structure, but this and many other articles are either very extensive or constantly interrupted by remarks, additional remarks, and long quotations from laws and other documents as already explained.

If the lexicographical data collected by the compiler are presented "without method or conciseness", as Hewins (1895) claims, this will almost inevitably make it more difficult, or even impossible, for the users to access and find the required data if they do not want to read the whole article from end to end. However, in defence of Postlethwayt could perhaps be argued that he had to

systematize his data — which go far beyond those incorporated in Savary des Bruslons' dictionary — in a period when economics had still not developed into a scientific discipline in the modern sense of the word. Adam Smith had not yet published his *Wealth of Nations*, and David Ricardo and Karl Marx were not yet born or even thought of.

6. Some articles

It is impossible to do justice to Postlethwayt's dictionary in the framework of this contribution, but a few articles selected from among the shorter ones may give an idea of its rich content. On the first page under the letter A, we find the article *aaggi-doggii*:

AAGGI-DOGGII, a Persian word, signifying 'The bitter mountain.' It is so called, on account of being a very dangerous passage for the trading caravans, which travel into Persia, and take their route of Constantinople to Ispahan. It is situated a day's journey from Louri, on the frontiers of Persia, near Chaouqueu. When these caravans are arrived at the pass of this mountain, all the camels and horses are numbered, for each of which the caravan-bachi takes a duty, which he employs partly for the pay of soldiers that guard the caravans, and partly for other small charges: but he detains the greatest part for himself.

It goes without saying that this article is not included in the dictionary in order to inform about geography in general. It has a very practical purpose and is highly relevant for those traders who plan to travel from Constantinople to Ispahan with a caravan, although the information about the caravan-bachi's duty seems more anecdotic than useful for anything but envy. In fact, sometimes Postlethwayt seems to enjoy writing in an anecdotic or even humoristic style as can be seen in the following article:

AMPHISCII. In geography and astronomy, the people who inhabit the Torrid Zone. They are thus denominated, as having their shadow turned sometimes one way, and sometimes another, i.e. at one time of the year to the north, and at another to the south.

As already mentioned, space does not permit the reproduction of the long articles which would illustrate the real complexity and richness of the dictionary; however, a look at the article *Lubecker* will nevertheless give a rather good idea of the composition of many articles:

LUBECKER.
As of old time the common councils of the free and imperial city of Lubeck, and several other towns upon the North and East Seas, have been in union with divers towns of the United Provinces, for defence of the liberty of trade, and of the rights belonging thereunto: and as now the States-General of the United Provinces, together with the burgomasters and common council of Lubeck, have found it convenient, in the present conjuncture, to treat about

renewing the aforesaid union and amity, by their respective deputies, who, after several conferences, and reports made, agreed upon the articles following:

I. That this union shall not be offensive, but only for preserving the freedom of commerce, &c. that the respective citizens and subjects of the union may enjoy the same without molestation, and mutually protect each other; so that the manifold difficulties detrimental to their common trade may be removed.

II. (...)

Postlethwayt then reproduces the remaining 15 articles composing the whole text agreed upon, and at the bottom of the document he also provides the final statements and signatures:

In witness whereof, two copies of these presents have been made, and confirmed by oath, and sealed with the seals of the States-General, and town of Lubeck, and signed by their respective clerk and syndic, each party retaining one copy. Done in May 1613, and underwritten,

By order of the States-General

Signed AERSSSEN,

And by order of the honourable the common council of the town of Lubeck,

Signed MARTIN NORDAMUS

Syndic of the Republic of Lubeck.

One may discuss the value of this information, but it could probably have been useful for the British merchants and traders doing business with the Lubeckers on foreign ground. As such, it is very similar to the kind of external data to which the users are frequently referred in modern specialized lexicography, in web-based dictionaries often with direct linking to the respective home pages.

Another article showing how Postlethwayt incorporated various sorts of text and document in his dictionary is *letter of credit*:

LETTER of CREDIT, is where a merchant, of correspondent, writes a letter to another, requesting him to credit the bearer with a certain sum of money.

Form of a Letter of Credit

Mr. E.F.

SIR,

London, May 10, 1753

My last to you of the 15th of March, wherein I wrote what was needful, in answer to your's of the 10th of the same month; and this serves chiefly to desire you to furnish and pay unto Mr. C. D. English gentleman, to the value of two thousand crowns, at one or more times, according as he shall have occasion for it, and request the same of you, taking his receipt, of bills of exchange, for the monies which you shall so furnish him with, and put it to my account; and this my letter of credit shall be your sufficient warrant for so doing.

To Mr. E.F. merchant

at Madrid Your's, &c. A. B.

In this rather typical article, the author first provides a small and easily under-

standable explanation of the term *letter of credit*, and instead of going into complex details he then furnishes a form which can be directly used by his readers if and when they need to write a letter of credit. This is one of many examples showing that Postlethwayt did not compile his dictionary merely to "describe the words", as a modern linguist would claim, but rather to assist his users and solve their real and practical problems which he undoubtedly knew very well. These users would surely find the above information highly relevant and useful in their daily work.

The article immediately after *letter of credit* provides another interesting example of how the dictionary was conceived:

LETTER-FOUNDER, He casts types, of letters, for printers: there are but two in London; 'till very lately we had types from Holland; but that excellent artist Mr. Caslon, having excelled all foreign founders, not only furnishes us at home, but sends great quantities abroad.

This is the kind of information which we nowadays would expect to find on the "yellow pages", in a telephone guide, a "who's who", an Internet directory, or a similar printed or electronic information tool conceived for consultation purposes. If the telephone had been invented in Postlethwayt's lifetime, he would probably have furnished the "excellent artist" Mr. Caslon's telephone number as well. When the function theory postulates that the very essence of lexicography is to provide quick and easy access to data from which punctual information can be retrieved, cf. Tarp (2009), its opponents frequently put forward as an argument that the theory does not make a sharp distinction between dictionaries and telephone guides. In fact, they are right, because there is no reason why lexicographical works should not provide this sort of information too. Postlethwayt gave his users the information which he knew that they needed and demanded, and nothing prevents modern lexicography from doing the same and designing integrated information tools, based upon advanced computer and information technologies, with a view to meeting the real and complex needs of the present-day users.

7. Operative functions

It has already been stated above that *The Universal Dictionary of Trade and Commerce* also had an operative function where the author provided "suggestions and intimations", as he called them. There are plenty of examples illustrating how this function was implemented in practice, but only three of them will be discussed here. For instance, under the lemma *lead*, one will find various sublemmata — or subtitles — such as *minium*, *litharge*, *burnt lead*, *white lead* or *cerusse*, *saccharum saturni*, *Remarks*, *The method of smelting lead ore at the great works with pit-coal*, and *The method of smelting lead ore with wood fuel*, some of which actually contain a sort of receipt as can be seen in the following extract from the article:

BURNT LEAD

Cut a quantity of this thinnest milled lead that can be got, into small plates, with which fill an earthen vessel that will bear the fire, and, with powder of brimstone, laid layer upon layer; when the sulphur is burnt away, the lead will be in a blackish powder: five ounces of brimstone suffice for half a pound of lead. Stir the matter while on the fire, and, when cold, wash the powder in three or four waters. It is used externally as the litharge: mixed with lard alone, it makes a good unguent for the piles.

Another similar example can be found in the *Remarks* addressed to the lemma *lint* where Postlethwayt very politely calls upon "any gentleman who has the conveniency of water for erecting a lint-mill" to go into the lint business:

REMARKS

The woollen manufacture of England never came to it's perfection 'till the business of the woolstapler came to be a trade; and it must be the same with the linen. — Here is a fine branch of business for any gentleman who has the conveniency of water for erecting a lint-mill; he may either raise lint himself, or buy it in the boon from others, and dress it by his mill for the heckle; he may have his heckles in a convenient house near to it, for dressing and stapling the lint by proper heckles, into fine dressed flax, fine dressed tow, common tow, backings and beards, for the service of the country spinners.

The last illustration of an operative function is taken from the article *malt* which covers several pages and includes many reflections and recommendations on the making of this important product. The article starts with a traditional definition:

MALT, is barley prepared, to fit it for making a potable liquor called beer or ale.

After this short definition, the article comprises various sections such as *Observation on the manner of Malting*, *Remarks*, and *An Abstract of the acts of parliament which have been made in relation to Malt*. In the first of these sections, and after describing the process of malting, Postlethwayt comments:

This is the general process of malting, wherein almost every maltster has his secret, or particular way of working. But, to render the operation perfect, the following cautions must be observed: (1.) That the barley be newly thrashed, or at least newly winnowed. (2.) That it is not mixed, or made up of different sorts. (3.) That it be not over-steeped in the cistern, or so long as to make it soft. (4.) That it be well drained. (5.) That it is carefully looked after in the wet couch, so as to stop the first tendency of the blade to shooting. (6.) Another caution is, to turn the wet couch inside outmost, if the barley comes, that it shoots more in the middle than on the sides. (7.) To keep it duly turning after it is out of the wet couch. (8.) To give it the proper heating in the dry heap. (9.) To dry and crisp it thoroughly upon the kiln, but without a fierce fire, so as to be several days in drying a kiln of pale malt. And, if these directions be carefully observed, the malt will always be good.

As it can be clearly seen, the character and structure of this last extract from the *Universal Dictionary* are very similar to what can be found in modern user guides, instruction manuals, how-tos, etc. It is a proof of the close relationship existing between dictionaries and these types of reference work, a relationship based upon the fact that they are all tools designed to be consulted in order to retrieve punctual information which can subsequently be used for many different purposes. In fact, Postlethwayt's dictionary documents that in the Age of Enlightenment no sharp boundaries existed between the various types of reference works, and that these were occasionally even integrated into one and the same tool.

8. A living and discussing dictionary

Reading *The Universal Dictionary of Trade and Commerce* is far from boring. It was compiled long before the introduction of a stereotype lexicographical style with rigid style sheets and models for the writing of definitions and explanations. In fact, Postlethwayt's dictionary appears as a living and discussing work where the author frequently "interrupts" the presentation in order to provide his own personal opinion on the matter treated, or even to discuss and change his previous opinion. For instance, under the lemma *lives [annuities upon lives]*, an article of several pages, he suddenly considers it necessary to comment on some ideas quoted from another author:

Here I have given the author's proposition and instance in his own words; but, if I may beg leave to express my meaning of them in my way, which, perhaps, may tend to explain and illustrate them, I should say: ...

Although always ready to give his personal opinion, Postlethwayt generally tries not to be unfair to other opinions. For example, a little further on in the above-mentioned article, under the *Remarks*, he refers to the different opinions expressed by various scholars and provides cross-references to a number of articles where the "substance" of the respective opinions can be found:

REMARKS

We have judged it necessary to enter so far into this delicate subject, and the objections made to the long embraced hypothesis of the learned Dr. Halley, and the foundations built thereupon by Mr. De Moivre and others (for the objections against the latter are not less weighty than those against the former, though we have not room fully to consider them here) in order to shew the reader, who is desirous to go to the root of this matter, that the works of the learned Weyman Lee, Esq; are well deserving his attentive perusal. There have, indeed, been some objections made to Mr. Lee's essay, but whoever compares them candidly and impartially with the reply that he has made to them, will hardly think, I am at present inclined to believe, that the fundamentals of his reasoning are at all invalidated. However, those who may still be of opinion, that what Dr. Halley, Mr. De Moivre, and Mr. Richards have said, will stand the test, will find the substance thereof in this work, under the articles Annuities and Leases; and the chief

objections that have been made against them, under the articles Lives, and Mortality [Bill of Mortality].

For the reader's further satisfaction, the following is the connected substance that Mr. Lee endeavours to prove, in opposition to Dr. Halley, Mr. De Moivre, Mr. Richards, and others, on LEASES and LIFE-ANNUITIES.

Postlethwayt is not afraid of admitting that he sometimes changes his opinion on certain questions. As mentioned above, in his dissertation on the plan of the dictionary, cf. Postlethwayt (1749), he expressed the idea that "agriculture and commerce nourish and enrich a nation". This idea is repeated in the fourth edition of his dictionary from 1774. However, if one consults the article *manufacture*, one will find the following highly interesting comment:

We begin to be now convinced, that we are nearly as much enriched by the labours of our fellow-creatures, as by the productions of the earth; and, if we have reason to rejoice at the abundance which nature, from year to year, produces for us, we may reap no less reasonable satisfaction from all the variety of employments in human society, and especially by means of our manufactural arts.

Although it may seem a little eclectic, with this comment Postlethwayt is actually taking a big step away from the physiocrats and towards a more modern understanding of the real sources of a country's wealth. The same moderation or change of ideas can be found in the article *Africa*. Here he still speaks about the "very beneficial traffic" of "gold, teeth, wax, and negroes", but at the same time he puts forward a long-term vision of "laying absolutely aside the slave-trade, and cultivating a fair, friendly, human, and civilized commerce with the Africans", a vision based upon Christian and "civilized" principles and, not to forget, the benefits to trade:

In short, Africa, though a full quarter of the globe, stored with an inexhaustible treasure, and capable, under proper improvements, of producing so many things delightful, as well as convenient within itself, seems utterly neglected by those who are civilized themselves, and its own inhabitants quite unsolicitous of reaping the benefits which nature has provided them. What it affords in its present rude, unimproved state, is solely given up to the gain of others, as if not the people only were to be sold for slaves to their fellow-creatures, but the whole country was captive, and produces its treasures, merely for the use and benefit of the rest of the world, and not at all for their own.

Whether, instead of making slaves of these people, it would not rather become such nations that assume to themselves the name and character of Christians, to give them a relish for the blessing of life, by extending traffic into their country in the largest degree it will admit of, and introducing among them the more civilized arts and customs, may be submitted to consideration.

(...)

But it is to be feared that, while the slaving trade with these people continues to be the great object of the Europeans, it will ever spirit up wars and hostilities among the negro princes and chiefs, for the sake of making captives of each other for sale. This, therefore, will ever obstruct the civilizing of these people, and

extending of the trade into the bowels of Africa, which, by the contrary means, might be easily practicable.

Although somewhat patronizing, in this text Postlethwayt proves that he has a fairly good idea of the very negative consequences which the European colonialist policy and interference in Africa had for this continent. It seems that the author has come under the influence of the epoch's growing ideas of freedom and human rights which a few decades later were expressed in the American *Declaration of Independence* and the French *Declaration of the Rights of Man and of the Citizen*. However, in spite of the great advance that these general declarations represented at the ideological level, strong economic interest in both countries prevented them from leading to the immediate abolishment of slavery. Against this background, Postlethwayt once more stands before us as a child of the contradictory ideas and practice existing in his lifetime.

9. Conclusions

Within present-day lexicographical circles it is frequently claimed that lexicography should be considered a sub-discipline of linguistics, a sort of "applied linguistics". Although it cannot be denied that linguistics has played an important role in the compilation of a large number of dictionaries, Postlethwayt's *Universal Dictionary of Trade and Commerce* is an example of a dictionary to the production of which no specialized linguistic knowledge was needed or used. The author himself was an economist, just like the authors of a whole range of economic dictionaries produced in Europe during the last three centuries, cf. Besomi (2011). However, until now none of these authors have claimed that lexicography should be considered a sub-discipline of economics, a sort of "applied economics"; and it is not difficult to predict how they would react if they were told that they had been "applying" linguistics theories. What Postlethwayt's *Universal Dictionary* shows, is, on the one hand, that lexicography has a great interdisciplinary vocation, and, on the other hand, that the *essence of lexicography* should be found in the elements that are common to all lexicographical works, i.e. the core elements remaining when an abstraction is made from all the specific types of knowledge needed to compile specific dictionaries, cf. Tarp (2009).

As already stated, the specific knowledge reflected in the dictionary is, as a rule, completely outdated and only interesting from a historical perspective. The really intriguing aspect is how the author tries to meet the users' needs with a great variety of lexicographical data. In the fourth edition he informs us that he has received many letters with positive response to the previous editions. This suggests that he had a fairly good knowledge of his target users and their needs, and that it is exactly these complex needs which he tried to meet with the various types of data and the various "styles" and structures found in the dictionary. In a certain sense, Postlethwayt's dictionary is a structural and stylistic mess. Although modern lexicography could certainly improve the

access to the relevant data, there seems to be a reason for this mess. The dictionary does not only provide information for cognitive needs, but also for operative ones. In some aspects it appears like a text book, at other times it is discussing, even putting in doubt former statements. It contains a big variety of up-to-date knowledge of that epoch, a "huge storehouse", not only of "economic facts, laws and theory", but also of practical "suggestions and intimations", all of these designed to meet the real needs and demands of his mid-eighteenth century's audience. In this respect, the Dutch specialist in the history of economics, Peter Groenewegen (2004: 1000), writes in his biographical article on Postlethwayt that the dictionary:

therefore contained many practical articles on inventions and improvements, as well as on commercial practice such as banking, commercial bills, and customs house business, hence catering for the considerable interest in compendia of knowledge of the mid-eighteenth century. The last enthusiasm was also evident in Postlethwayt's proposal for a remodelled Royal Society to enable it to collect information useful to British trade as well as to recommend new manufactures and new avenues of trade for official encouragement and support.

At a moment when traditional dictionaries are experiencing the start of a crisis owing to the fact that a growing number of potential users tend to satisfy their information needs by other means, most frequently on the Internet, modern lexicography can without any doubt learn much from dictionaries published in the Age of Enlightenment where the authors, like Postlethwayt, were not afraid of putting into their works exactly the types of data demanded and needed by their potential users. Today, with the advent of the computer and information technologies, especially the Internet, lexicographers should let themselves be inspired by past experiences and develop integrated information tools which may meet the complex needs of modern society whether these needs are of a communicative, cognitive, operative or interpretive art.

Acknowledgements

Thanks to the University of Valladolid, Spain, for allowing me to do research at the *Biblioteca Histórica de Santa Cruz*. Special thanks to Professor Pedro A. Fuertes-Olivera for arranging my visit to this marvellous library which contains about 18,000 books written between 970 and 1815, among them various old and rare dictionaries like the ones compiled by Postlethwayt and Savary des Bruslon.

Literature

Besomi, Daniele (Ed.). 2011. *Crises and Cycles in Economic Dictionaries and Encyclopaedias*. London/New York: Routledge.

Cantillon, Richard. 1755. *Essai sur la nature de commerce en général*. London: Fletcher Gyles.

- Chalmers, Alexander (Ed.).** 1816. *The General Biographical Dictionary: Containing an Historical and Critical Account of the Lives and Writings of Most Eminent Persons in Every Nation; Particularly the British and the Irish; from the Earliest Accounts to the Present Time.* Vol. XXV. London: J. Nichols and Son.
- Cowie, A.P. (Ed.).** 2009. *The Oxford History of English Lexicography. Volume II: Specialized Dictionaries.* Oxford: Oxford University Press.
- Groenewegen, Peter.** 2004. Postlethwayt, Malachy (c. 1707–1767). Mathew, H.C.G. and Brian Harrison (Eds.). 2004. *Oxford Dictionary of National Biography.* Vol. 44: 999-1000. New York: Oxford University Press.
- Hewins, William Albert Samuel.** 1895. Postlethwayt, Malachy. *Dictionary of National Biography.* Vol. 46. London: Smith, Elder & Co.
- Johnson, Edgar Augustus Jerome.** 1965. *Predecessors of Adam Smith: The Growth of British Economic Thought.* New York: A.M. Kelley.
- Kaufmann, Uwe and Henning Bergenholtz (Eds.).** 1998. *Encyclopedic Dictionary of Gene Technology. English–Spanish.* Toronto: Lugus Libros LatinAmerica Inc.
- Leroyer, Patrick.** 2012. *From Subject-Based to User-Based Dictionaries: Rejecting the Cultural Opposition Thesis in Specialized Lexicography.* Paper presented at the Seventeenth International Conference of the African Association for Lexicography, held at the University of Pretoria, Pretoria, 2–5 July 2012.
- Mankin, Robert.** 2008. Pins and Needles. Adam Smith and the Sources of the Encyclopédie. Brown, Vivienne (Ed.). 2008. *The Adam Smith Review* 4: 181-205. Oxford: Routledge.
- Marx, Karl.** 1867. *Das Kapital. Kritik der politischen Oekonomie. Erster Band.* Hamburg: Verlag von Otto Meissner.
- Postlethwayt, Malachy.** 1745. *The African Trade, the Great Pillar and Support of the British Plantation Trade in America.* London: J. Robinson.
- Postlethwayt, Malachy.** 1749. *A Dissertation on the Plan, Use, and Importance, of the Universal Dictionary of Trade and Commerce; Translated from the French of the Late Celebrated Mons. Savary, Inspector-General of the Manufactures of the Custom-House of Paris: With such Considerable Additions and Improvements, As will Appear at Large Throughout this Dissertation; And which More Particularly Accommodate the Fame to the Trade and Navigation of the British Empire, Addressed to the Nobility, Gentry, Merchants, and Traders of Great Britain.* London: John and Paul Knapton.
- Postlethwayt, Malachy.** 1751–1755. *The Universal Dictionary of Trade and Commerce: Translated from the French of the Late Celebrated Mons. Savary, Inspector-General of the Manufactures of the Custom-House of Paris: With Large Additions and Improvements Incorporated throughout the Whole Work; Which More Particularly Accommodate the Fame to the Trade and Navigation of these Kingdoms; and the Laws, Customs, and Usages to which All Traders are Subject.* London: John and Paul Knapton.
- Postlethwayt, Malachy.** 1774. *The Universal Dictionary of Trade and Commerce: With Large Additions and Improvements, Adapting the Same to the Present State of British Affairs in America, Since the Last Treaty of Peace Made in the Year 1763. With Great Variety of New Remarks and Illustrations Incorporated throughout the Whole: Together with Everything Essential that is Contained in Savary's Dictionary: Also, All the Material Laws of Trade and Navigation Relating to these Kingdoms, and the Customs and Usages to which All Traders are Subject.* The Fourth Edition. London: W. Strahan, J. and F. Rivington, J. Hinton.

- Savary des Bruslons, Jacques.** 1723. *Dictionnaire universel de commerce: contenant tout ce qui concerne le commerce qui se fait dans les quatre parties du monde, par terre, par mer, de proche en proche, et par des voyages de long cours, tant en gros qu'en détail. Ouvrage posthume du Sieur Jacques Savary Des Bruslons. Continué sur les Memoires de l'auteur, et donné au public par M. Philemon Louis Savary.* Paris: Chez J. Estienne.
- Smith, Adam** 1776. *An Inquiry into the Nature and Causes of the Wealth of Nations.* London: W. Strahan and T. Cadell.
- Tarp, Sven.** 2008. Revival of a Dusty Old Profession. *Hermes. Journal of Linguistics* 41: 175-188.
- Tarp, Sven.** 2009. Beyond Lexicography: New Visions and Challenges in the Information Age. Bergenholtz, Henning, Sandro Nielsen and Sven Tarp (Eds.). 2009. *Lexicography at a Crossroads: Dictionaries and Encyclopedias Today, Lexicographical Tools Tomorrow:* 17-32. Linguistic Insights. Studies in Language and Communication 90. Bern: Peter Lang.
- Xie, Jin (Ed.).** 1408. *Yongle Dadian. Vol. 1-11,095.* China.

Cuban School Dictionaries for First-Language Learners: A Shared Experience*

Sven Tarp, *Department of Afrikaans and Dutch, University of Stellenbosch,
South Africa; International Centre for Lexicography,
University of Valladolid, Spain; and Centre for Lexicography,
Aarhus University, Denmark (st@asb.dk)*

and

Leonel Ruiz Miyares, *Centre for Applied Linguistics, Santiago de Cuba,
Cuba (leonel@lingapli.ciges.inf.cu)*

Abstract: One of the reasons why many publishing houses hesitate when allocating resources to the development of more advanced concepts of school dictionaries for first-language learners is the relatively poor sales and corresponding profits coming from this category of dictionaries. A major challenge is therefore how to change the situation, stimulate the use and sales of school dictionaries, and guarantee the necessary investment in product development. This contribution will take up this challenge. It will give no final answer, but it will discuss an experience with a Cuban school dictionary which has reached its target users in large numbers. The contribution will particularly look at the multi-faceted relation between lexicographers and users which characterises the Cuban dictionary project. The objective is not to copy this or any other experience but to inspire lexicographers looking for alternative solutions to a problem which is affecting both school children and lexicographers in many countries and language communities.

Keywords: PEDAGOGICAL LEXICOGRAPHY, PEDAGOGICAL DICTIONARIES, SCHOOL DICTIONARIES, FIRST LANGUAGE LEARNERS, MOTHER-TONGUE LEARNERS

Opsomming: Kubaanse skoolwoordeboeke vir eerstetaalleerders: 'n Gedeelde ervaring. Een van die redes waarom baie uitgewerye huiwer wanneer middele toegewys word aan die ontwikkeling van meer gevorderde konsepte van skoolwoordeboeke vir eerstetaalleerders, is die relatiewe swak verkope en ooreenstemmende profyte wat van hierdie kategorie woordeboeke afkomstig is. 'n Belangrike uitdaging is dus hoe om die situasie te verander, die gebruik en verkope van skoolwoordeboeke te stimuleer en die nodige belegging in produkontwikkeling te waarborg. Hierdie bydrae sal dié uitdaging aanvaar. Dit sal geen finale antwoord gee nie, maar dit sal 'n ervaring met 'n Kubaanse skoolwoordeboek bespreek wat sy teikengebruikers in groot getalle bereik het. Die bydrae sal veral kyk na die veelfasettige verhouding tussen leksikograwe en gebruikers wat die Kubaanse woordeboekprojek kenmerk. Die doel is nie om hierdie of enige ander erva-

* This article was presented as a paper by Sven Tarp at the Eighteenth Annual International Conference of the African Association for Lexicography (AFRILEX) at the Nelson Mandela Metropolitan University, Port Elizabeth, South Africa, 2–5 July 2013.

ring na te boots nie, maar om leksikograwe te inspireer wat vir alternatiewe oplossings van die probleem soek wat sowel skoolkinders as leksikograwe in baie lande en taalgemeenskappe raak.

Sleutelwoorde: OPVOEDKUNDIGE LEKSIKOGRAFIE, OPVOEDKUNDIGE WOORDEBOEKE, SKOOLWOORDEBOEKE, EERSTETAALLEERDERS, MOEDERTAALLEERDERS

1. Introduction

During November 2011, a workshop on school dictionaries for first-language speakers was organized at the University of Pretoria (cf. Gouws and Tarp 2012). The informative discussions showed, among others, that a major reason why most South African publishing houses hesitate when they have to allocate resources to the development of more advanced concepts of school dictionaries is the relatively poor sales and corresponding profits coming from this category of dictionaries. This is a general problem in South Africa with the possible exception of Oxford's monolingual school dictionaries of English. Low use of school dictionaries, low sales, low profits, and low interest in the development of more user-friendly and attractive school dictionaries: all of this undoubtedly constitutes a classical vicious circle! A major challenge for these engaged in pedagogical lexicography is therefore how to break the circle:

What can be done to change the situation, stimulate the use and sales of school dictionaries, and guarantee the necessary investment in product development?

There is certainly no simple answer to this question but it may be worth studying positive experiences from other countries. In this respect, it may be interesting to look at a Cuban experience. Although having a different social system, Cuba is a developing country like many African countries; the available resources are limited and have to be used carefully according to priorities. On this island, a school dictionary, *Diccionario Básico Escolar*, has in a few years been distributed in a total of 126.800 copies among a population only 20 per cent the size of the South African population and with an even lower percentage of children of school-going age. In addition, the *Diccionario Básico Escolar* has been published on CD and has been made freely available on the Internet from a Basque server as well as on the Cuban national Intranet which can be accessed from all Cuban schools, Joven Clubs (Internet Cafés) and a small but growing number of private homes — a fact which does not seem to discourage the interest in the printed version, perhaps quite the opposite.

What is the "secret" behind this notable achievement?

In order to answer this question, one of the organizers of the Pretoria 2011 workshop has joined forces with one of the collaborators in the Cuban dictionary project. Our intention is not to tell others how to do their work. We do not propose that the Cuban experience should be copied by anybody. We believe that each dictionary project must have its individual characteristics adapted to the specific context and tradition in the respective country and language com-

munity. But we hope that this experience may inspire other lexicographers engaged in the production of school dictionaries to find alternative ways to improve their products and reach the school children in need of such tools when learning their first language. With this in view, in the following, we will describe the *Diccionario Básico Escolar* with focus on the close and peculiar contact between lexicographers and users in the pre- and post-compilation phases.

2. The preparatory work

Edited by Eloína Miyares Bermúdez, the *Diccionario Básico Escolar* has been planned, prepared, compiled and actively promoted by collaborators from the Centro de Lingüística Aplicada (Centre for Applied Linguistics), located in the eastern Cuban City of Santiago. The preparatory work started back in the early 1990s. Between 1990 and 1995 researchers from the Centre carried out a study of the active-functional vocabulary of Cuban primary-school children between the 2nd and 6th grade. The study was based on a total of 7.001 (5.873 written and 1.128 oral) compositions collected in both rural and urban areas in nine of Cuba's then fourteen provinces.

The result of this detailed study, which was later published in Miyares Bermúdez et al. (2006), did not only allow the researchers to establish the urban and rural school children's growing — and varying — active vocabulary over five grades (8 to 12 years). They also retrieved valuable information about the pupils' specific problems in terms of oral and written text production. All this allowed the researchers to compose a corpus of 700.000 codified words which constituted the central (but not only) part of the corpus which eventually assisted in the compilation of the *Diccionario Básico Escolar* and its twin sister, the *Diccionario Escolar Ilustrado*. The latter is designed to initiate the pupils in the world of dictionaries and attend to their lexicographical needs in the first four grades of Primary School.

In his analysis of Spanish school dictionaries, Hernández (2003: 24) writes that "there is not a single school dictionary which clearly informs about the sources of its nomenclature". The Santiago-based *Centro de Lingüística Aplicada* has acted differently. From the very beginning, it has been very informative about its sources and methodology which have been explained in great detail in various publications, for instance Ruiz Miyares and Miyares Bermúdez (1992), Ruiz Miyares (1994), Ruiz Miyares (1997), Miyares Bermúdez and Ruiz Miyares (2006), and Miyares Bermúdez et al. (2012).

In this respect, it is worth noting that what distinguishes the lexicographical basis of the *Diccionario Básico Escolar* (and its twin sister) is not so much the existence of a corpus. Corpora composed of textbooks, exercises, etc., are generally used in the compilation of school dictionaries, cf. Michael Rundell's (2004) Introduction to the *Macmillan School Dictionary*. The real interesting thing is the fact that the corpus in this case includes texts composed by the target user group itself, allowing the lexicographers to obtain a much more profound knowledge of the real problems and needs of this group which Wiegand (1977: 59)

once called the "well-known unknown".

The authors of the Cuban school dictionary have a very profound knowledge of the users and their real problems and needs. The 7.001 written and oral compositions collected among the foreseen target users represent a statistically reliable basis for determining these needs and the corresponding lexicographical solutions. In addition, the main editor, Eloína Miyares Bermúdez, has an academic background first as a teacher and later as a researcher in linguistics and first-language learning, while other collaborators have similar experience but also other types of expert knowledge.

The team working on the *Diccionario Básico Escolar* is interdisciplinary; it includes experts in linguistics and computer science. This kind of team work is very important for lexicographical research. The informatics experts developed three important systems: the computer system to analyze and process the 7.001 written and oral compositions, the editor of the *Diccionario Básico Escolar* as well as its digital version (for CD and the Internet), cf. Ruiz Miyares and Miyares Bermúdez (1992), Ruiz Miyares (1994), Ruiz Miyares (1997), Alegría, Iñaki et al. (2006a), Alegría, Iñaki et al. (2006b), and Alegría, Iñaki et al. (2006c). Both the editor and the digital version were prepared in collaboration with researchers and specialists from the University of the Basque Country.

3. Short presentation of the dictionary

The first edition of the *Diccionario Básico Escolar* was published in 2003; subsequently it appeared (until now) in two other editions (2008, 2009) with two reprints of the third edition (2010, 2012) for a total of 126.800 copies. As explained by Miyares Bermúdez and Ruiz Miyares (2006: 104), it is designed to meet the lexicographical needs of school children attending the "second circle" of Primary School (Grades 5 to 6), the Basic Secondary School (Grades 7 to 9) and High School (Grades 10 to 12), i.e. a total of eight grades with children from about ten to eighteen years old. According to the Introduction to the third edition, also students of "polytechnic centres and the users of the Joven Clubs", i.e. students within the same age group, should be added to the expected user group. As to the learning situations and activities in which the defined target users can use it, Miyares Bermúdez (2009: ix) writes:

This third edition comprises more than 10 500 entries and about 20 000 senses, all of them basic and fundamental for the *knowledge and normative use of the language*. The majority of the entries express different spheres of reality, especially aspects of social life, nature, science, culture, sport, and other novel elements that allow *the actualization of the knowledge* of the young and also the adult people.

Miyares Bermúdez and Ruiz Miyares (2006: 104, 110) provide further details:

Its main objective is that the school children achieve an *enlargement* and *better knowledge* of the Spanish vocabulary as well as an *improved orthography* and *proper use of the words* according to their function.

The major impact of the *Diccionario Básico Escolar* is of a social and educative character because the pupils of the second circle of Primary School, Basic Secondary School and High School in Cuba can count on this work that will assist them in their *linguistic, cultural, scientific, ethic and patriotic formation*.

From the above it results that the dictionary is, in the first place, designed to assist school children of Grades 5 to 12 having lexicographically relevant information needs related to text reception, text production and their cognitive formation. The more than 10.500 lemmata have been selected from various sources: the corpus described in the previous section, textbooks from various grades, Cuban newspapers and magazines for young people, as well as a special corpus based on a study of the school children's use of affixes in Grades 7 to 9. The selection represents modern Spanish language including a number of Cubanisms, Americanisms and, to a less extent, Anglicisms and Gallicisms as well as hundreds of terms taken from scientific disciplines studied in the various grades. The individual articles always contain:

- part of speech,
- explanations,
- example sentences,
- hyphenation,
- and inflection (of nouns, adjectives and verbs, in this last case with indication of past participle and reference to a list of 83 model verbs in the back matter of Volume II).

To this should be added the frequent inclusion of:

- underlining of certain letters in the lemma in order to warn against typical orthographic mistakes,
- gender,
- explicit grammatical data,
- diatechnical, diatopic, diaintegrative and diastratic marking,
- pronunciation (when it is a foreign word, e.g. *iceberg*),
- synonyms, antonyms, and cognate words,
- locutions, phraseologisms, and proverbs,
- explanation of locutions, phraseologisms, and proverbs,
- example sentences with locutions and phraseologisms,
- diminutives and augmentatives, and
- references to 19 thematic tables with a total of 584 illustrations.

The following four articles provide an idea of the design and content of the articles:

electricidad sf. Fenómeno físico producido por el movimiento de electrones de los átomos, el cual se manifiesta en la atracción y repulsión que aparece entre ellos. La electricidad no se crea ni se destruye. Es una forma de energía de empleo cómodo por lo fácil de su conducción; y muy útil, pues tiene muchos usos: se puede transformar en mecánica en los motores; térmica, en las resistencias de calefacción; luminosa, en el alumbrado eléctrico, etc. *El descubrimiento de la **electricidad** fue un hecho de gran importancia para el hombre.*

e-lec-tri-ci-dad

gota sf. **1** Porción pequeña de un líquido que al caer adopta la forma de una esfera. *Noté que el tiempo cambiaba cuando sentí caer de repente dos **gotas** de lluvia sobre mi brazo.* **2** Poca cantidad de algo. *Me ha servido una **gota** de comida.* Sin. pizca. **3** Enfermedad que produce dolor en las articulaciones y a veces afecciones viscerales. *Los enfermos de **gota** tienen un tratamiento médico intenso.* // loc. adv. **ni gota**. Nada, ninguna cosa. *No encontré **ni gota** de lo que me pediste.* // fras. **sudar la gota gorda**. Esforzarse mucho para conseguir algo. *Sudó la **gota gorda** para llegar hasta aquel lugar.*

go-ta; gotas (pl.); gotica (dim.)

gracioso, a adj. **1** Se dice de lo que tiene cierto atractivo en su aspecto. *Las muchachas celebran a los jóvenes **graciosos**.* Sin. bonito, atractivo, simpático. Ant. feo. **2** Se aplica a aquello que resulta chistoso, cómico o que nos hace reír. *Nos pareció **gracioso** el espectáculo.* Sin. divertido, agradable, cómico. **3** Se atribuye a la persona que hace chistes, cuentos, etc. *Jorge era el actor más **gracioso** de ese elenco.* Sin. ocurrente.

gra-cio-so, a; graciosos, as (pl.)

parecer vintr. (27) **1** Tener algo determinado aspecto físico o apariencia. *El trabajo que le encomendaron **parecía** fácil. Su padre nos **pareció** muy amable.* **2** vpr. Tener alguien o algo un aspecto o apariencia semejante o común a otra persona o cosa. *Odilia se **parece** mucho a su hermana.* Sin. semejarse. **3** vintr. Haber o existir razones para creer una cosa. ***Parece** que va a salir el Sol.* (Solamente se emplea en tercera persona). // loc. adv. **al parecer**, **según parece**. Como se observa o por lo que se sabe. ***Al parecer**, a Carlos se le hizo tarde. El profesor, **según parece**, está enfermo y no vendrá hoy a clases.*

pa-re-cer; perecido (p.p.)

As to the three functions discussed above, the *Diccionario Básico Escolar* complies with the requirements in this regard, at least in terms of the data catego-

ries necessary to meet the school children's lexicographical needs. As to *text reception*, the short explanation of the lemmata as well as of the locutions, phraseologisms and proverbs, written in a simple and clear language, and always followed by an example sentence, is virtually what is needed to help the children understand the words — and their various senses — included in the dictionary. As to *text production*, the *Diccionario Básico Escolar* contains a variety of data categories providing information about orthography, pronunciation, part of speech, inflection, locutions, phraseologisms, proverbs, synonyms, antonyms, cognate words, diminutives, augmentatives, and hyphenation which are all highly relevant for text production. Although some of these data categories are seldom found in Spanish school dictionaries, they are nevertheless absolutely necessary not only in order to assist the children in composing texts in correct Spanish, but also to develop and produce a rich and varying language in terms of words and expressions. Finally, as to its *cognitive function* in its various aspects, a look at the article *electricidad* shows how a relatively complex concept is explained with plain words.

4. Reactions to the dictionary

The *Diccionario Básico Escolar* has been reviewed various times by international scholars, both in Spanish (e.g. Forgas and Bargalló 2006, and Nomdedeu Rull 2012), and in English (e.g. Corpas Pastor 2005, and Tarp 2012). The reviews have generally been positive, although it has been remarked that the dictionary pretends to embrace too many grades (and years of age), "a maybe too excessive ambition" because of the "big cognitive and academic difference" between the users of the upper and lower grades covered by the dictionary (Forgas and Bargalló 2006: 254). Apart from this, almost all reviewers have coincided in appreciating the meticulous work performed by the authors and their efforts to adapt the dictionary to the school children's linguistic, mental, cultural, and encyclopaedic development.

In Cuba, the feedback has also been positive; this is, among other things, reflected in the many copies sold as well as in the new editions where the authors have made improvements and modifications taking into account the many opinions expressed by the public. In this respect, the opinion of parents, teachers and students has exceedingly contributed to the development of the *Basic School Dictionary* thanks to a series of initiatives which we will describe in the following.

5. Secrets

A dictionary is the brainchild of its authors. The major desire of any lexicographer, who has dedicated much work and love to a specific dictionary, is that the expected target users utilise and appreciate it. This is even more so when the dictionary in question is a school dictionary and the target users are

children being initiated in the marvellous world of their first language. This is true in Cuba, Africa, and elsewhere. However, too often the lexicographers' brainchild is taken out of their hands when they have finished the last article of the last letter. The promotion and distribution is put in the hands of marketing experts and other professionals from the publishing houses — well-prepared people who may do their job well but probably without the same feelings and personal involvement as those who produced the dictionary in the first place. This may lead to a certain degree of "business as usual".

Eloína Miyares Bermúdez and the other lexicographers at the Centro de Lingüística Aplicada cherish their school dictionary. They have therefore decided to play an active role in its promotion and distribution. This distribution follows, on the one hand, the normal channels for books and school material in Cuba. On the other hand, however, the authors have taken a number of initiatives which, it could be said, represent a *reunion with their users*. We will here briefly mention three of these.

The first initiative is not completely unknown to lexicographers in other countries. In Cuba, there is an annual FERIA del Libro (Book Fair) starting in Havana, the capital, with subsequent sessions in provinces all over the island during the months of February and March. Hundred thousands of Cubans participate in this very popular event, among them many parents together with their children looking for "normal" books as well as educative material — and the authors of the *Diccionario Básico Escolar* do the same. They have made use of the FERIA del Libro to present their dictionary. The result is impressive: apart from the opportunity to talk directly to their end users and their parents, during February 2011 they sold 50.000 dictionaries in one week in Havana only. If the number of printed copies had been larger, even more dictionaries could probably have been sold.

The second initiative is unique. During the past years, the Centro de Lingüística Aplicada has organised "caravans" which until now have visited 36 of Cuba's 168 municipalities and sponsored meetings where the authors themselves have introduced — and donated — the dictionary to teachers and other relevant people interested in it. This has allowed the establishing of direct links to the schools and, ultimately, to close the circle that started several years before with the research into the school children's active-functional vocabulary.

Since 1992, the Centro de Lingüística Aplicada has organized these Scientific National Caravans on initiative of the two founders of the Centre, Dr Vitelio Ruiz Hernández and Dra Eloína Miyares Bermúdez. The fundamental aim of these caravans, which are also supported by the Cuban Ministry of Education, is to spread the results of the Centre's scientific research across the whole country to raise the Spanish language culture of all students of elementary, secondary and high schools. Within the framework of these caravans, workshops and meetings are held where the researchers expound these scientific results and explain to the participants how to use them. In the most recent visit of the caravans (2008–2009), the printed and digital (CD) versions of the

Diccionario Básico Escolar were distributed freely in 3.800 and 3.000 copies respectively. Simultaneously, the books *Ortografía Integral* (Integral Spelling) and *Léxico activo-funcional del escolar cubano* (a study of the active-functional vocabulary of Cuban primary-school children) were handed out among the participants in 2.000 and 1.000 copies, respectively. In Figure 1, the route and stages of the caravan (2008–2009) are indicated.

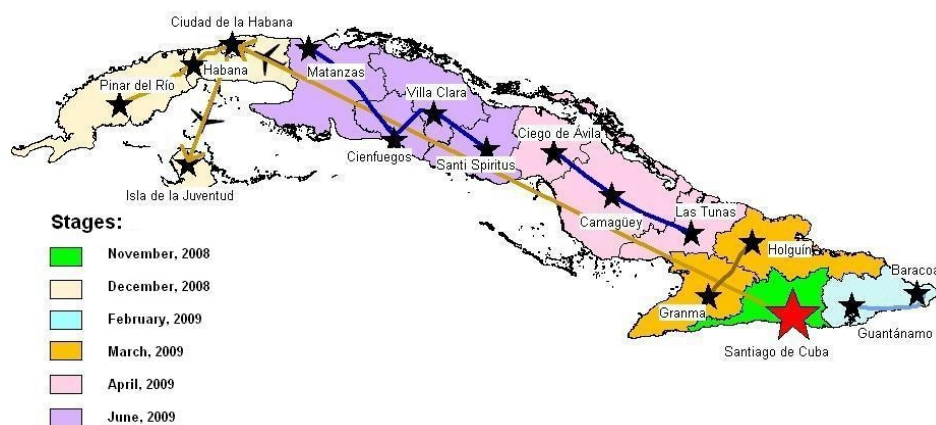


Figure 1: The route taken by the National Scientific Caravan in 2008–2009

In the various workshops and meetings organised by the caravan all over the country, more than a 1.000 persons participated, among them teachers, parents, students and executives who subsequently played the role of multipliers of this linguistic-pedagogic experience when they took it to the Cuban classrooms. Additionally, in these meetings, the scholars from the Centro de Lingüística Aplicada received many suggestions and ideas aimed at improving and extending their scientific work. One of these was to produce a digital version of the *Diccionario Escolar Ilustrado*, i.e. the dictionary for the youngest school children. The corresponding software has now been developed and it will be presented at the meetings planned to take place in 2013 and 2014 during the next caravan. Information about the caravans can be found on the website of the Centro de Lingüística Aplicada (www.linguistica.santiago.cu).

The third initiative is also related to the *Diccionario Básico Escolar*. It consists of the annual organisation of lexicographical "marathons" at school, municipal and provincial level (still not at a national level), where school children are stimulated to consult dictionaries and compete in terms of their lexicographical skills. These marathons are a new initiative taken by the founders of the Centro de Lingüística Aplicada and have been held in the eastern province of Santiago to stimulate a general dictionary culture among Cuban school children. The basic idea is to organise a competition where the time the partici-

pants use to search for words in the dictionary is measured, and where the one who uses the least time is declared the winner of the marathon. The participants in these competitions are mainly children from secondary school. Since 2009, tens of thousands of children at school level and more than 2.000 at municipality and province level have participated in the first five occurrences of the marathon using the *Diccionario Básico Escolar* as instrument of competition; as such, it also has an important impact on the use of this dictionary. The marathons have until now only been organised in the province of Santiago de Cuba but it is now planned to extend the competition to the rest of the country. Some of the results of the marathons can also be seen at www.linguistica.santiago.cu.

6. Accessible price

There are several additional factors contributing to the success experienced by the authors of *Diccionario Básico Escolar*. Since 1959, the Cuban government has regarded education as one of its most important priorities and has invested considerable amounts of money in this sector. As a result of this policy, the educational system in Cuba today is among the most esteemed in the world. This also means that present-day parents are much more conscious of their children's needs in this respect. This positive attitude may partially explain their willingness to buy school dictionaries for their children.

However, it must not be ignored that salaries in Cuba are not particularly high. As most families in Africa, Cuban families have to mind their money and prioritise on what it is spent. They will not be able to afford any price — even for the benefit of their children.

In this respect, two measures have been taken to ensure an accessible price of the *Diccionario Básico Escolar*. First, the authors received a donation from the Basque government to print and distribute the second edition (3.800 copies) at a reasonable price. Later they received funding from the Instituto Cubano del Libro (Cuban Book Institute) which permitted the printing of 120.000 dictionaries. Together with the use of cheap paper and printing techniques, the funding has made it possible to offer the third edition of the *Diccionario Básico Escolar* — 1.200 pages in two volumes — at the very low price of 30 Cuban Pesos (about 11 South African Rand) thus making it accessible to any Cuban family with school children.

7. Conclusion

Each new edition and impression of the *Diccionario Básico Escolar* has been quickly sold out in spite of the considerable number of printed copies. Various factors may explain this remarkable success. Government funding and the low distribution prices are undoubtedly important factors. But just as important is the quality of the dictionary and the efforts made by the lexicographers to

make their product known among the school children and their parents. The caravans to various municipalities, the meetings with local educational authorities and first-language teachers, the direct encounters with school children and their parents — all of these contribute to the promotion of the dictionary. In addition, the stimulation of a dictionary culture in the classrooms by means of lexicographical marathons should not be ignored either.

The Cuban experience shows above all that a multi-faceted relation between lexicographers and users — not only in the preparation and compilation phases, but also in the post-publication phase — is mutually inspiring for both parts and, hence, highly recommendable. As we wrote in the beginning of this contribution, we are very much aware that the specific experience described is embedded in a particular cultural tradition and context, and that it therefore cannot be copied. We are convinced that there are also other roads leading to Rome. But we hope at least to have provided some material for reflection and inspiration — for the benefit of school children, parents, and lexicographers.

Bibliography

Dictionaries

- Miyares Bermúdez, Eloína (Ed.).** 1998. *Diccionario Escolar Ilustrado*. Santiago de Cuba: Editorial Oriente/Madrid: Ediciones Libertarias.
- Miyares Bermúdez, Eloína (Ed.).** 2003. *Diccionario Básico Escolar*. First edition. Santiago de Cuba: Centro de Lingüística Aplicada.
- Miyares Bermúdez, Eloína (Ed.).** 2008. *Diccionario Básico Escolar*. Second edition. Santiago de Cuba: Centro de Lingüística Aplicada.
- Miyares Bermúdez, Eloína (Ed.).** 2009. *Diccionario Básico Escolar*. Third edition. Santiago de Cuba: Editorial Oriente.
- Miyares Bermúdez, Eloína (Ed.).** 2009. *Diccionario Básico Escolar*. Electronic edition, version 3.0. Santiago de Cuba: Centro de Lingüística Aplicada/Donostia: IXA Taldea, Universidad del País Vasco. [<http://ixa2.si.ehu.es/dbe/index.html>]
- Rundell, Michael (Ed.).** 2004. *Macmillan School Dictionary*. Oxford: Macmillan.

Other literature

- Alegría, Iñaki, Xabier Arregi, Xabier Artola, Mikel Astiz and Leonel Ruiz Miyares.** 2006a. A Dictionary Content Management System. Corino, E., C. Marelló and C. Onesti (Eds.). 2006. *Proceedings XII EURALEX International Congress, Turin, Italy, September 6th–9th, 2006. Vol. I:* 105–109. Alessandria: Edizioni dell'Orso.
- Alegría, Iñaki, Xabier Arregi, Xabier Artola, Mikel Astiz and Leonel Ruiz Miyares.** 2006b. Building an Electronic Version of the Cuban Basic School Dictionary. Corino, E., C. Marelló and C. Onesti (Eds.). 2006. *Proceedings XII EURALEX International Congress, Turin, Italy, September 6th–9th, 2006. Vol. I:* 243–250. Alessandria: Edizioni dell'Orso.

- Alegría, Iñaki, Xabier Arregi, Xabier Artola, Mikel Astiz and Leonel Ruiz Miyares.** 2006c. Different Issues in the Design and Development of the Electronic Cuban Basic School Dictionary in Linguistics in the Twenty First Century. Miyares Bermúdez, Eloína and Leonel Ruiz Miyares (Eds.). 2006. *Linguistics in the Twenty First Century*: 273-288. Cambridge: Cambridge Scholars Publishing.
- Corpas Pastor, Gloria.** 2005. [Review of] Eloína Miyares Bermúdez (Ed.). *Diccionario Básico Escolar*. Santiago de Cuba: Centro de Lingüística Aplicada. 2003. *International Journal of Lexicography* 18 (2): 267-272.
- Forgas, Esther and María Bargalló.** 2006. Los diccionarios del Centro de Lingüística Aplicada de Santiago de Cuba. *Revista de Lexicografía* 12: 249-266.
- Gouws, Rufus H. and Sven Tarp.** 2012. School Dictionaries for First-Language Learners. *Lexikos* 22: 333-351.
- Hernández Hernández, Humberto.** 2003. La lexicografía escolar del español: desde sus orígenes hasta su consolidación. *Revista Internacional de Lingüística Iberoamericana* 1: 7-25.
- Miyares Bermúdez, Eloína (Ed.).** 2009. Introducción a la tercera edición. *Diccionario Básico Escolar*. Third edition. Santiago de Cuba: Editorial Oriente.
- Miyares Bermúdez, Eloína and Leonel Ruiz Miyares.** 2006. Dos ejemplos de la lexicografía pedagógica cubana: el Diccionario Escolar Ilustrado y el Diccionario Básico Escolar. Campus Soutu, Mar, Felix Córdoba Rodríguez and José Ignacio Pérez Pascual (Eds.). 2006. *América y el diccionario*: 197-211. A Coruña: Universidade da Coruña.
- Miyares Bermúdez, Eloína (Ed.). In collaboration with Nancy C. Álamo Suárez, Celia Pérez Márqués, Katia González Heredia and Leonel Ruiz Miyares.** 2006. *Léxico activo-funcional del escolar cubano*. Santiago de Cuba: Centro de Lingüística Aplicada.
- Miyares Bermúdez, Eloína, Xabier Artola Zubillaga, Iñaki Alegría Loinaz, Xabier Arregi Iparagirre, Leonel Ruiz Miyares, Cristina Álamo Suárez and Celia Pérez Márqués.** 2012. Las últimas ediciones del Diccionario Básico Escolar de Cuba. Nomdedeu Rull, Antoni, Esther Forgas Berdet and Maria Bargalló Escrivà (Eds.). 2012. *Avances de lexicografía hispánica I*: 201-213. Tarragona: Universidad Rovira i Virgili.
- Nomdedeu Rull, Antoni.** 2012. [Review of] Miyares Bermúdez, Eloína (dir.) (2009): *Diccionario básico escolar*. Santiago de Cuba, Centro de Lingüística Aplicada. *Revista de Lexicografía* 17: 243-256.
- Ruiz Miyares, Leonel.** 1994. Aplicación de la computación al Estudio del Vocabulario Básico del Escolar Cubano. *Estudios de Comunicación Social*: 96-105. La Habana: Editorial Academia.
- Ruiz Miyares, Leonel.** 1997. Versión avanzada de un sistema computacional aplicado a una investigación lexicológica. *Estudios de Comunicación Social*: 85-113. La Habana: Editorial Academia.
- Ruiz Miyares, Leonel and Eloína Miyares Bermúdez.** 1992. Ampliación del Sistema de Computación VEC. *Selección de ponencias del Congreso Internacional Expolingua Habana '92*: 599-607. La Habana: Expolingua.
- Tarp, Sven.** 2012. New Experiences in Pedagogical Lexicography: Two Cuban School Dictionaries. *Hermes, Journal of Linguistics* 48: 95-107.
- Wiegand, Herbert Ernst.** 1977. Nachdenken über Wörterbücher: Aktuelle Probleme. Drosdowski, Günther, Helmut Henne and Herbert E. Wiegand (Eds.). 1977. *Nachdenken über Wörterbücher*: 51-102. Mannheim/Vienna/Zurich: Bibliographisches Institut.

Die rol van taalleeronderrig in Afrikaanse woordeboekwerkboeke

Michele F. van der Merwe, *Departement Kurrikulumstudies: Afrikaans, Fakulteit Opvoedkunde, Universiteit Stellenbosch, Stellenbosch*
(michelevdm@sun.ac.za)

Opsomming: Hierdie artikel beskou die rol van woordeboekwerkboeke in die vaardighedsopvoeding van gebruikers in Afrikaans. Woordeboekwerkboeke word gewoonlik op 'n spesifieke skool- of aanleerderwoordeboek gerig en sluit sodoende by die opvoedkundige leksikografie aan. Die hoofdoel van werkboeke behoort dus 'n opvoedkundige een te wees. Alhoewel die opvoedkundige leksikografie meer belangstelling in Suid-Afrika wek, word weinig aandag gegee aan die belangrike rol van woordeboekwerkboeke. Woordeboekwerkboeke behoort baie gebruikersgerig te wees, aangesien dit vir spesifieke teikengebruikers saam met die gebruik van 'n spesifieke woordeboek opgestel word. Die doel van werkboeke word in hierdie artikel geïdentifiseer deur verskeie werkboeke in die Afrikaanse konteks te analiseer. By nadere ondersoek blyk dit dat werkboeke diverse doelstellings het en ook diverse inhouds bevat. Werkboeke is daarop gemik om teikengebruikers meer woordeboekgeletterd te maak, d.w.s. om woordeboeke sinvol te kan gebruik en inligting daaruit suksesvol te ontsluit. Die wyse waarop inligting in werkboeke aangebied en dus oorgedra word, reflekteer sekere onderrigbenaderings in taal en hoe dit gestalte kan vind in werkboeke. Wanneer 'n onderrigbenadering suksesvol toegepas word, is die verwagting dat gebruikers se vaardighede sal verbeter. Gevolglik word daar verwys na die ontwikkeling van sekere vaardighede.

Sleutelwoorde: AFRIKAANSE OPVOEDKUNDIGE LEKSIKOGRAFIE, SKOOLWOORDEBOEKE, AANLEERDERSWOORDEBOEKE, WOORDEBOEKWERKBOEKE, TAALLEERONDERRIG, TAALLEERONDERRIGBENADERINGS, TEKSGBASEERDE BENADERING, KOMMUNIKATIEWE BENADERING, LEERDERGERIGTE ONDERRIGBENADERING, INDUKTIEWE ONDERRIGMETODE, KONSTRUKTIVISME, ONDERRIGLEERAKTIWITEITE, LEERDERS SE WOORDEBOEKVAARDIGHEDE, ONDERWYSER, LEKSIKOGRAAF, UITGEWER

Abstract: Criteria for Dictionary Workbooks in Afrikaans. This article reflects on the role of dictionary workbooks in the skills training of users in Afrikaans. Dictionary workbooks are usually directed at a specific school or learners' dictionary and in this way links up with pedagogical lexicography. The main purpose of workbooks should thus be an educational one. Although pedagogical lexicography is generating more interest in South Africa, very little attention has been paid to the important role of dictionary workbooks. Dictionary workbooks should be very user oriented, since they are compiled for specific target users using a specific dictionary. The purpose of dictionary workbooks is identified in this article by analysing various workbooks in the Afrikaans context. On closer investigation it was found that workbooks could have diverse purposes and could contain diverse contents. Workbooks are aimed at making target users more dictionary literate, i.e. to be able to use dictionaries meaningfully and to unlock information suc-

cessfully. The manner, in which information in workbooks is presented and thus conveyed, reflects certain educational approaches in language and how they can find shape in workbooks. When an educational approach is successfully applied, the expectations are that user skills should be improved. Hence reference is also made to the development of certain skills.

Keywords: AFRIKAANS PEDAGOGICAL LEXICOGRAPHY, SCHOOL DICTIONARIES, LEARNER'S DICTIONARIES, DICTIONARY WORKBOOKS, LANGUAGE TEACHING, LANGUAGE TEACHING APPROACHES, TEXT-BASED APPROACH, COMMUNICATIVE APPROACH, LEARNER-CENTRED TEACHING APPROACH, INDUCTIVE TEACHING METHOD, CONSTRUCTIVISM, LANGUAGE TEACHING ACTIVITIES, DICTIONARY SKILLS OF LEARNERS, TEACHER, LEXICOGRAPHER, PUBLISHER

1. Inleiding

Opvoedkundige leksikografie (met verwysing na skool- en aanleerderwoordeboeke) geniet die afgelope tyd meer belangstelling in Suid-Afrika, soos geïllustreer deur onlangse artikels oor die onderwerp. Vergelyk onder andere Tarp en Gouws (2004), Gouws (2006), Gouws en Tarp (2008) en Beyer (2010) in hierdie verband. Daar is egter weinig aandag aan die rol van woordeboekwerkboeke en die inhoud daarvan aandag gegee. Vergelyk Van der Merwe (2013). Woordeboekwerkboeke het 'n groot rol te speel in sowel skole as in die samelewing, aangesien dit die probleemoplossende potensiaal van woordeboeke belig.

In hierdie artikel word die bydrae van werkboeke tot die vaardigheidsopvoeding van woordeboekgebruikers ondersoek. Die vraag word gevra wat die doel van woordeboekwerkboeke is en of die werkboeke slaag in die doel wat gestel word. Die rol van woordeboekwerkboeke in die taalonderrigsituasie word beskryf. Benaderings tot taalonderrig in skole word verduidelik en 'n spesifieke metode vir taalonderrig met behulp van werkboeke word voorgestel. Watter vaardighede word aangeleer deur die gebruik van sulke werkboeke en watter vaardighede behoort aangeleer te word? Hoe moet die inligting aangebied en oorgedra word? Die artikel word vanuit die oogpunt van 'n kurrikulumstudie benader en probeer om 'n bydrae op teoretiese gebied te maak.

2. Die doel van woordeboekwerkboeke

Volgens Chi (1998: 565) geniet naslaanwerke hoë status in die Chinese kultuur, aangesien dit beskou word as "'n onderwyser wat nie kan praat nie". Woordeboeke word naas die onderwyser in die klas as die betroubaarste alternatief vir leer beskou. Chi wys daarop dat moderne leksikograwe en uitgewers 'n verskeidenheid tegnieke of maniere toepas om te verseker dat naslaanwerke en spesifiek woordeboeke, toegankliker, deursigtiger en makliker is om te verstaan as in die verlede.

Woordeboekwerkboeke word ook ontwerp om gebruikers te leer om woor-

deboeke te gebruik met behulp van verskillende soorte oefeninge. 'n Verskeidenheid gegradeerde oefeninge word aangebied om gebruikers van hulp te wees, ook in die klaskamer.

Stark (1990: 30) beskou 'n woordeboekwerkboek as 'n tekstipe met unieke status. Dit is 'n teks georiënteer ten opsigte van gebruikers om hulle te help om optimale voordeel uit 'n spesifieke naslaanwerk, naamlik 'n woordeboek te kan trek.

Die *Macmillan English Dictionary for Advanced Learners Workbook* se doel word duidelik in die voorwoord aan die gebruiker gestel, naamlik: "This workbook will help you to get the best out of your dictionary, and by working through these activities you will be able to find words quickly, learn how to find pronunciation and stress, use grammatical information, find and explore meanings, learn which words are used together and choose the right word for the context. Explanations and examples are followed by activities to practise your dictionary and language skills." Die volgende motivering word ook aan gebruikers gestel: "Being able to use a dictionary well is a great pleasure and will bring independence and confidence to your language studies."

Volgens die *Nuwe Woordeboek sonder grense Werkboek* (WSGW) sal die werkboek "leerders help om die vaardighede aan te leer om *Nuwe Woordeboek sonder grense* doeltreffend te gebruik". Leerders sal volgens die aannames op die agterblad van die werkboek gelei word om "hul woordeskat te verbreed, alfabetiese volgorde te leer ken, die struktuur van woordeboeke te verstaan en daarmee te werk, asook om woordsoorte op 'n eenvoudige vlak te onderskei en te gebruik".

Die *Oxford Tweetalige Werkboek* (OTW) belooft op die agterblad dat die werkboek "maklik is om te gebruik en vol nuttige aktiwiteite is. Dit help leerders om algemene struikelblokke te oorkom, bv. om die regte vertaling te kies. Dit bou leerders se woordeskat uit met behulp van oefeninge met voorvoegsels, meervoude, verkleinwoorde, woordsoorte, ens. Die OTW dien as gids vir leerders om die meeste uit hulle woordeboeke te kry, bv. trappe van vergelyking, onreëlmatige werkwoorde, hulp met die skryf van 'n e-pos of brief. Dit sluit antwoorde vir al die aktiwiteite, asook bondige aantekeninge vir onderwysers of ouers op elke bladsy in." Beide werkboeke se fokus is dus op taal- en woordeboekonderrig.

Volgens *Pharos Leerderboek by die Skoolwoordeboek* (PLS) (voorwoord, p. vi) "fokus die aktiwiteite daarop om woordeboekvaardighede by gebruikers in te skerp. Hulle word ook toegerus met die nodige vaardighede om suksesvolle briewe te skryf en projekte met gemak aan te pak." Die term *woordeboekvaardighede* word nie verduidelik nie. Die opsteller van die werkboek verwys ook na die sogenaamde buitetekste, wat nie noodwendig deel uitmaak van die opdrag van 'n skoolwoordeboek nie en leerders word vaardighede daarvoor aangeleer, bv. die skryf van briewe. Die fokus is op die aanleer van woordeboekvaardighede en die ontwikkeling van skryfvaardighede vir transaksionele tekste.

In die *Pharos Leerderboek by die Aanleerderswoordeboek* (PLA) (voorwoord)

word die doel van die werkboek soos volg geformuleer: "Die aktiwiteite is toegespits op taalaspekte soos woordsoorte, woordfamilies en woordgroepe. Die blokraaisels toets leerders se vaardigheid met onder meer leestekengebruik, formele en informele taalgebruik en meervoudsvorme." 'n Belofte word ook gemaak dat "wanneer gebruikers die aktiwiteite in hierdie boek baasgeraak het, sal hulle die nodige vaardighede hê om met gemak enige woordeboek met vrug te gebruik". Die fokus is op ontwikkeling van taal- en woordeboekvaardighede. Dit word ook vermeld dat leerders die aktiwiteite tuis of in die klas kan doen.

Kenmerkend van Pharos se boeke is dat dit "leerderboeke" genoem word en nie "werkboeke" soos die ander uitgewers s'n nie. Met die leerdergerigte titels word die fokus volledig op die (skool)leerder geplaas in ooreenstemming met die neiging van die gebruikersperspektief waarvolgens leksikografiese produkte met 'n spesifieke teikengebruiker in gedagte opgestel word.

Ten opsigte van die *HAT Aktiwiteite en Speletjies by die Afrikaanse Skoolwoordeboek en Afrikaanse Sakwoordeboek* (HATS) word verduidelik dat "die aktiwiteite en speletjies geskryf is om gebruikers met die *HAT Afrikaanse Skoolwoordeboek* vertrouwd te maak. Werkblaaie mag in die klas gebruik word om leerders speel-speel inligting uit woordeboeke te leer ontsluit". Opvoedkundige inligting word dus aangebied vanuit die perspektief om die leerervaring genotvol te maak.

Werkboeke is afgestem op doelwitte om teikengebruikers op te lei in vaardighede om inligting in woordeboeke suksesvol te ontsluit. Die fokus is op die verbetering van die taal- en woordeboekvaardighede van leerders. Woordeboekvaardighede verwys onder meer na kennis van die alfabetiese struktuur, kennis van woordeboekstrukture, byvoorbeeld die makro- en mikrostruktuur van woordeboeke, asook die verstaan van die funksies van 'n woordeboek, byvoorbeeld vertaling van woorde. Die verbetering van die genoemde vaardighede lei daartoe om leerders sogenaamd "woordeboekgeletterd" te maak, wat 'n beduidende rol kan speel in die skep van 'n woordeboekkultuur in skole en die samelewing. Die term *woordeboekgeletterd* verwys na die vermoë van 'n persoon wat oor kennis beskik om 'n woordeboek sinvol te gebruik en inligting daaruit suksesvol te ontsluit. So 'n persoon moet dus oor kennis van die alfabetiese struktuur, die woordeboekstrukture en -funksies beskik.

Beyer (2010: 60) verwys na die onderrigleersituasie waar woordeboekgebruik op die een of ander manier deur die onderwyser gestimuleer word. Een van die maniere (Beyer 2010: 60) kan wees die aanbod van onderrig in woordeboekgebruik deur die onderwyser en die praktiese toepassing daarvan met behulp van woordeboekwerkboeke.

Benaderings tot taalleeronderrig

Die opvoedkundige leksikografie het skoolwoordeboeke en aanleerderwoordeboeke in bestek. Daar word in hierdie artikel na werkboeke van skoolwoorde-

boeke, byvoorbeeld HATS, PLS en OTW en aanleerderwoordeboeke, byvoorbeeld PLA en WSG verwys. In die vorige afdeling is die doel van werkboeke beskryf en die gemeenskaplike doelwitte wat deur talle opstellers van werkboeke uiteengesit is, is die onderrig van taalvaardighede. Dit is dus belangrik vir werkboekopstellers om kennis te neem van die Department van Onderwys se benadering ten opsigte van taalleeronderig vir huistaal, addisionele taal en tweede addisionele taal.

Volgens die Kurrikulum en Assesseringsbeleidverklaring (KABV) vir Afrikaans as huistaal behoort die aspekte van taalstruktuur en/of kennis en begrip van die taal as interafhanklik van al die ander taalvaardighede beskou te word en moet dit hoofsaaklik kommunikatief en geïntegreerd in natuurlike en aktuele gebruiksvrbande onderrig word.

Taalleer behoort informeel in noue korrelasie met al die fasette van die onderrig van die huistaal te staan en leerders behoort 'n toepaslike kennis van en insig in die betrokke inhoud te hê en behoort in staat te wees om hierdie kennis en begrip effektief in 'n verskeidenheid situasies te gebruik.

Die KABV vir Afrikaans as huistaal en addisionele taal stel dit duidelik dat die benadering tot taalleeronderrig "teksgebaseerd, kommunikatief, geïntegreerd" asook "proses-georiënteerd" moet wees.

Die teksgebaseerde en kommunikatiewe benaderings is beide afhanklik van die voortdurende gebruik en produksie van tekste. 'n Teksgebaseerde benadering stel leerders in staat om vaardige, vrymoedige en kritiese lesers en skrywers (en ontwerpers) van tekste te word. Dit sluit die luister en kyk na, en die lees en ontleding van tekste in om te verstaan hoe dit saamgestel is en watter effek dit het. 'n Woordeboekartikel kan as 'n teks beskou word. Deur hierdie kritiese interaksie ontwikkel leerders die vermoë om tekste te evalueer. Die hoofbron van inhoud en konteks vir die kommunikatiewe, geïntegreerde leer en onderrig van taal is outentieke tekste. Die teksgebaseerde benadering sluit ook die produksie van verskillende tekste vir spesifieke doelstellings en teikengroepe in. 'n Begrip van die manier waarop tekste saamgestel word, rugsteun hierdie benadering.

'n Leerdergerigte onderrigbenadering

Die KABV gee ook bepaalde wenke oor hoe die onderrig benader behoort te word. Taal kom voor waar mense met mekaar verkeer en met mekaar wil kommunikeer. Betekenis staan daarom sentraal in enige kommunikasie-gebeure. Maniere waarop mense betekenis skep, asook die effektiewe kommunikasie van betekenis, behoort dus die fokus van taalonderrig te wees. Die bevordering van taalgebruiksvaardighede is gevolglik belangriker as blote kennis van taalstrukture (taalleer).

Taal bestaan nie los van die mens nie en kan dus nie as 'n afgebakende lys inhoud bestudeer word nie. Die interaktiewe en sosiale aard van taal beïnvloed die onderrigbenadering, onder meer ten opsigte van metodiek, evalue-

ring en handboekgebruik. Combrink et al. (1997: 6) beskryf dié benadering as sosiale konstruktivisme. Volgens die interaktiewe benadering skep mense taalkennis en taalprodukte in interaksie met ander mense, met ander woorde, kennis met 'n sosiale basis. Tipiese handeling in 'n leerdergerigte klaskamer is gesprekke, spontane redenering, gesamentlike probleemoplossing en refleksie. 'n Leerdergerigte benadering (Combrink et al. 1997: 6) beteken dat leerders nie net passiewe ontvangers van taal is nie, maar ook skeppende gebruikers van taal. Kennis van taal is belangrik, asook die genot- en betekenisvolle gebruik daarvan en die ontwikkeling van interaktiewe vaardighede.

Hoe vind dié onderrigbenaderings gestalte in werkboeke?

Leerdergerigte onderrigbenaderings vind op verskeie maniere plaas in Afrikaanse werkboeke. Aangesien woordeboekartikels as tekstipe beskou word, word dit in werkboeke verskaf en vrae word daarvoor aan die leerders gestel. Verskeie werkboeke maak van die teksgebaseerde benadering gebruik.

OTSW gebruik byvoorbeeld die artikel **snaaks** (ook verskaf in die werkboek) in teksgebaseerde aktiwiteite onder die hofie: "Hoe om inskrywings te verstaan." Sien voorbeeld 1 in dié verband.

1. Lees die hele inskrywing vir die woord **snaaks** en beantwoord dan die vrae.

(a) Skryf die eerste twee betekenisse van **snaaks** neer.

- 1 _____
- 2 _____

(b) Skryf die Engelse vertaling van **snaaks** in elk van die volgende sinne.

	Engels
Sy grappe is altyd baie snaaks.	1.
Hy was snaaks genoeg nie vies vir my nie.	2.

Voorbeeld 1

Twee vrae word gevra, naamlik die eerste twee polisemiese onderskeidings van die woord moet neergeskryf word, asook die Engelse vertaling moet verskaf word. Met die vraag kan 'n algemene fout, naamlik dat leerders die eerste polisemiese onderskeiding van 'n woord as korrekte antwoord kies, voorkom word. Die vraag dien as voorbereiding vir die volgende moeiliker vraag by die artikel van **maar**. Leerders moet verder as die eerste vertaling, naamlik **but**, lees en tot die gevolgtrekking kom dat **even though** ook 'n vertaling van die woord kan wees. Met verwysing na Bloom se taksonomie vir die evaluering van kognitiewe ontwikkeling (Combrink et al. 1997: 18, 19) word kennis en

begrip van die leerder gevra. 'n Hoërvlakvraag, naamlik toepassing, word gevra waar leerders die voegwoord **maar** moet gebruik om twee verskafte sinne te verbind. Hier handel dit oor sintaktiese orde, aangesien **maar** 'n newe-skikker is en onafhanklike woordorde teweegbring.

WSG bevat 'n hele aantal voorbeelde waar die teksgebaseerde benadering gebruik word. Vergelyk die aktiwiteite rakende woorde soos **stad** (mediostrukturele inligting), **polisie** (uitspraak-inligting), **boek** (grammatiese inligting), **tyd** (betekenisinligting), **bly** (semantiese betrekkinge). Die artikel **hak** (polisemiese onderskeidinge) kan ter voorbeeld gebruik word. Die artikel word verskaf en 'n opdrag daarvoor word gegee. Dit lui soos volg: "Sommige woorde het meer as een betekenis. Die betekenis word met syfers aangedui, byvoorbeeld 1 en 2. Kyk na die voorbeeld van die inskrywing in *Nuwe Woordeboek sonder grense*. Soek die woorde in *Nuwe Woordeboek sonder grense* op. Lees hul definisies en voorbeeldsinne. Besluit dan watter Engelse betekenis die beste by watter definisie pas. Skryf die nommer in die blokkie."

Aangesien WSGW op aanleerders gerig is, word daar telkens net een inligtingskategorie op 'n slag behandel. Dit pas by die gebruikers se vermoëns en vaslegging vind deur die inoefen van die konsep plaas. Dit is didakties suksesvol. Die teksgebaseerde benadering is goed om te gebruik, aangesien die woordeboekartikel in die werkboek herhaal word en die leerders word dus in kontak met die woordeboek gebring.

MacMillan maak uitstekend van die teksgebaseerde benadering gebruik. 'n Teks, bv. 'n artikel of storie word verskaf en relevante vrae word gevra wat tot woordeboekgebruik oor die betrokke teks lei. Sodoende kan die gebruiker die voordeel van woordeboekgebruik ervaar. Dit lei tot WST-woordeboekondersteunde taalonderrig.

Volgens die KABV behels die kommunikatiewe benadering dat leerders baie en ryke blootstelling aan die teikentaal moet kry. Om dit te bereik, moet leerders vele geleenthede gebied word om taal te gebruik en te oefen. WSG maak uitstekend van dié benadering gebruik deur 'n Afrikaanse liedjie se woorde (met musiek) te verskaf met die opdrag om dit in die klas te sing. In dié liedjie kom baie verkleinwoorde voor en leerders word gevra om die getal verkleinwoorde in die liedjie te soek en te identifiseer, asook om hul meervoude in die woordeboek na te slaan.

PLA maak ook van die kommunikatiewe benadering gebruik deur 'n teks te verskaf wat aanleerders uit Afrikaans in Engels kan vertaal, asook 'n Afrikaanse teks met spel- en leestekenfoute wat geredigeer kan word.

Die kommunikatiewe benadering is uiters geskik vir aanleerders van 'n taal, maar volgens die konstruktivistiese leerteorie waar die sosiale aard van taal beklemtoon word, kan dit vanweë die leerdergerigtheid daarvan ook baie suksesvol vir huistaalsprekers gebruik word. Leer word volgens die konstruktiviste as 'n sosiale en dus 'n taalgebaseerde aktiwiteit beskou (WKOD 2006). Konstruktivistiese leeromgewings ondersteun die bou van kennis deur sosiale bedinging en is afhanklik van die dialoog tussen leerder en opvoeder en die leerders self.

Die induktiewe metode as onderrigleeraktiwiteit

Beyer (2010: 61, 63) onderskei tussen woordeboekonderrig, woordeboekondersteunde taalonderrig en woordeboekgeïntegreerde taalonderrig. Hy wys daarop dat die induktiewe metode baie suksesvol tydens woordeboekgeïntegreerde taalonderrig, oftewel WIT, gebruik kan word. WIT (Beyer 2010: 61) verwys na die bereik van 'n lesdoelwit deur leerdergerigte onderrigleeraktiwiteite. Die gebruik van werkboeke tydens woordeboekgeïntegreerde taalonderrig kan 'n belangrike rol speel in 'n leerdergerigte benadering.

Die induktiewe metode (Meij et al. 1985: 6-7) bestaan daaruit dat die leerlinge aan voldoende voorbeelde van 'n bepaalde taalverskynsel bekend gestel word en deur die onderwyser se leidende vrae self daarin slaag om die betrokke reël, beginsel of verskynsel af te lei of te "ontdek". Taalleeronderrig is dan funksioneel, en nie formeel nie. Sodoende val die klem op taalgebruiksvaardigheid. Dit stel leerders in staat om self afleidings te maak en op dié manier ontdekkings beter te onthou. Daar is heelwat voordele verbonde aan die gebruik van die induktiewe metode vir taalleeronderrig.

Volgens Meij et al. (1985: 6-7) is leerders aktief betrokke by die onderrigproses; dit maak die les interessanter. Dit wat die leerders self ontdek, onthou hulle langer. Hul kennis is dus van 'n blywende aard. Hierdie metode prikkel leerders se denkvermoë en bevorder begrip. Die vlak van aanbieding word sterk beïnvloed deur die bydraes van leerders — dis dus nie te moeilik of te maklik nie. Die onderwyser kom gou agter wat leerders se probleme is en kan dit dan onmiddellik oplos.

Voorstelle vir onderrigleeraktiwiteite in werkboeke

In hierdie afdeling volg vyf voorbeelde van aktiwiteite volgens die induktiewe metode wat in werkboeke opgeneem kan word. Die onderwerp van die aktiwiteit word telkens verskaf en dan volg die aktiwiteit. Die onderwerpe vorm deel van die KABV vir Afrikaans huistaal vir graad vier. Dit sluit ook aan by Stark (1990) se kontrolelyns vir woordeboekvaardighede, soos in die volgende afdeling verwys.

Taalstrukture en -konvensies met die byvoeglike naamwoord: trappe van vergelyking as onderwerp. Lees die onderstaande sin:

Die rooi boek is **dikker** as die blou boek. Die vergrotende trap van die woord **dik** is in die sin gebruik. Soek die vergrotende en die oortreffende trap van die volgende woorde in die woordeboek op en skryf dit neer.

- (1) lui
- (2) blou
- (3) bitter

- (4) bleek
- (5) loom

Voltooi die volgende sin: Die vergrotende trap van 'n byvoeglike naamwoord word gevorm deur ____ agter die woord te voeg en die oortreffende trap van 'n byvoeglike naamwoord word gevorm deur ____ agter die woord te voeg.

Voorbeeld 1

Taalstrukture en -konvensies met die selfstandige naamwoord: meervoude as onderwerp.

Soortname het kenmerkend meervoude en dit word in woordeboeke aangedui. Slaan die volgende soortname in die woordeboek na en skryf die meervoud neer.

- (1) bal
- (2) jas
- (3) kam
- (4) nar
- (5) pet

Voltooi nou die volgende sin: As 'n woord 'n ____ klank in die middel het, verdubbel die ____ in die meervoud.

- (1) beer
- (2) haas
- (3) blaar
- (4) peer
- (5) skool

Voltooi nou die volgende sin: As 'n woord ____ klinkers in die middel het, verloor hy die tweede ____ in die meervoud.

Voorbeeld 2

Taalstrukture en -konvensies met die selfstandige naamwoord: verkleinwoorde as onderwerp.

Dit is kenmerkend van soortname om verkleinwoorde te hê en dit word in woordeboeke aangedui. Slaan die volgende soortname in die woordeboek na en skryf die verkleinwoord daarvan neer.

- (1) blaf
- (2) bek
- (3) dop
- (4) kas

Voltooi nou die volgende sin: om 'n verkleinwoord te vorm, sit ons 'n ____ aan die einde van 'n woord by. Indien die woord op __, __, _ en _ eindig en voorafgegaan word deur 'n kort, beklemtoonde vokaal, verdubbel die konsonant.

Slaan die volgende soortname in die woordeboek na en skryf die verkleinwoord daarvan neer.

- (1) hond
- (2) naald
- (3) pad
- (4) geut
- (5) woord

Voltooi nou die volgende sin: om 'n verkleinwoord te vorm, sit ons 'n ____ aan die einde van 'n woord by wat op 'n _ of _ eindig.

Voorbeeld 3

Taalstrukture en -konvensies met die bywoord as onderwerp.

Lees die volgende sinne deur en soek die woord in skuinsdruk in die woordeboek op. Watter woordsoort is die woord?

- (1) Die motors ry *vinnig* op die snelweg.
- (2) Die atleet oefen *hard* vir die byeenkoms.
- (3) Sy lees *gereeld* 'n speurverhaal.
- (4) Die doel het die doelpaal *rakelings* gemis.
- (5) Die atleet oefen *besonder* hard vir die byeenkoms.

Voltooi die volgende notas oor die bywoord: indien 'n bywoord ons meer vertel van ____ iets gedoen word, noem ons dit 'n bywoord van wyse. Indien 'n bywoord ons meer vertel van ____ iets plaasvind, noem ons dit 'n bywoord van plek. Indien 'n bywoord ons meer vertel van ____ iets plaasvind, noem ons dit 'n bywoord van tyd. Indien 'n bywoord ons meer vertel van 'n ander ____ of 'n ____, noem ons dit 'n bywoord van graad.

Voorbeeld 4

Woordsoorte kan nooit in isolasie getoets word nie, aangesien die rol wat die woord in die sin vervul, die funksie van die woord bepaal. Vergelyk die volgende voorbeelde:

Duik is my gunstelingtydverdryf. (selfstandige naamwoord)

Die dolfyn *duik* diep in die brander. (werkwoord)

Bogenoemde is voorbeelde van leerdergerigte aktiwiteite wat in werkboeke opgeneem kan word. Leerders word gelei om self die betrokke reël of verskynsel af te lei. Dit stel leerders in staat om self afleidings te maak en op dié wyse ontdekkings beter te onthou. Dit sluit aan by die konstruktivisme se idee dat leer die bou van kennis behels. Tydens die bou van kennis vind die volgende plaas: redenasie en kritiese denke, probleemoplossing, ontsluiting, verstaan en gebruik van inligting en sorgsame besinning oor ervaring. In bogenoemde gevalle vervul woordeboeke 'n kognitiewe funksie, aangesien data in verband met taalkunde deur die gebruiker nageslaan word.

Ontwikkeling van leerders se vaardighede

Stark (1990) stel 'n kontrolelys van woordeboekvaardighede voor na 'n evaluering van werkboeke:

- Stel vas watter leksikale item in die leesteks problematies is.
- Soek 'n leksikale item in die makrostruktuur:
 1. Bemeester die alfabetiese ordening van lemmas: samestellings, vaste uitdrukkings, kort vorme, frase- en voorsetselwerkwoorde.
 2. Soek 'n leksikale item in 'n inskrywing met 'n verskillende lemma: kruisverwysings, redes vir nie-nabye plasing.
 3. Soek 'n meerwoordige leksikale item: idiome en vaste uitdrukkings.
 4. Kies tussen homonieme, gebruik grammatiese of semantiese inligting.
 5. Kies tussen verskillende onderskeidings van 'n poliseem.
 6. Gebruik meer as een lys in die makrostruktuur van 'n woordeboek: voorwerk en aanhangsels rakende afkortings, eiename, lande en nasionaliteite, diername, verwantskapsterme, affikse, asook werkwoordverbuigingstabelle.
- Verstaan die beginsels van lemmatisering in woordeboeke: waarom sekere woorde lemmastatus het.
- Soek inligting in die mikrostruktuur
 1. Soek inligting oor die spelling of woorde: spelpatrone en verboë vorme, spelling van afkortings, akronieme, spelling van multileksikale items, hoofletters en punktuasie.
 2. Soek inligting oor gebruik: styl en register, kollokasies, gebruiksnote en variëteite.
 3. Soek inligting oor die grammatika van woorde: woordsoorte, werkwoordvorme en infleksie, meervoudsvorme, voorsetsels, asook selfstandige naamwoorde wat net in die enkelvoud of meervoud voorkom.

4. Soek inligting oor die betekenis van woorde: sinonieme, betekenis, kies van die regte definisie, kollokasies, korrekte lemma, polisemiese onderskeidinge, idiomatiese en vaste uitdrukkings.
5. Soek verwante woorde: plek van verwante woorde, sinonieme, paronieme, en antonieme.
6. Gebruik ekwivalente van 'n tweetalige woordeboek.
7. Soek inligting oor die herkoms van woorde.
8. Soek inligting oor die uitspraak van woorde: lettergreepverdeling, klem, uitspraak van homografe, alternatiewe uitspraak, asook homofone.
9. Kies die geskikte woordeboek op grond van die tipe leksikale item en die inligtingssoort wat verlang word.
10. Weet wat om van woordeboeke te verwag in die algemeen en van elke woordeboek in die besonder.

Bogenoemde lys van vaardighede kan uitgebrei word deur na Afrikaans as huistaal en addisionele taal se doelwitte in die KABV te verwys. Werkboekopstellers moet hulle vergewis van die nasionale kurrikulumverklarings se vereistes en benaderings vir taalleeronderrig en dit moet in werkboeke se inhoud gereflekteer word.

Met die gebruik van werkboeke kan leerders se woordeboekgeletterdheidsvlakke aansienlik verhoog word en opstellers kan van opvoedkundige beginsels kennis neem om die gehalte van werkboeke te verbeter. Die HATS maak van 'n belangrike tegniek in die didaktiese situasie gebruik, naamlik doelwitstelling. Elke aktiwiteit word van 'n uitkoms, d.w.s die doel van 'n aktiwiteit, voorsien, bv. "die aktiwiteite in hierdie afdeling sal jou help om die alfabetiese ordening van trefwoorde en samestellings in die woordeboek te verstaan".

Bloom se taksonomie vir die evaluering van kognitiewe denke (Combrink et al. 1997: 18, 19) is 'n nuttige klassifikasiesisteen vir die beskrywing van die denkvlakke van leerders. Hy identifiseer ses denkvlakke, naamlik kennis, begrip, toepassing, analise, sintese en evaluering. Dié kennisvlakke behoort aandag te geniet in werkboeke om te verseker dat alle denkvlakke en nie net die eerste twee, soos so dikwels voorkom, in werkboeke beslag vind nie. Dit kan verseker dat leerders aan die hoogste vlak van kritiese denke blootgestel word.

Slot

Dit is van die uiterste belang dat teikengebruikers en die doel van woordeboekwerkboeke gespesifiseer word ten einde hulp en leiding aan ouers en onderwysers te verskaf, aangesien daar groot onkundigheid oor woordeboekwerkboeke

bestaan. Bewusmaking van die didaktiese waarde van woordeboekwerkboeke is uiters noodsaaklik. Verdere navorsing oor effektiwiteit van werkboeke is nodig waar teoretiese leksikograwe, praktiese leksikograwe en kurrikulumkundiges leksikografiese navorsing in die klaskamer doen.

Bibliografie

- Beyer, H.L.** 2010. Kontekstuele bepalers by die beplanning van skoolwoordeboeke. *Lexikos* 20: 52-72.
- Bloom, B.S.** 1956. *Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain*. New York: David McKay.
- Chi, Man L.A.** 1998. Teaching Dictionary Skills in the Classroom. Fontenelle, Thierry, Philippe Hilgismann, Archibald Michiels, André Moulin en Siegfried Theissen (Reds.). 1998. *ACTES EURALEX '98 PROCEEDINGS, Communications soumises à EURALEX '98 (Huitième Congrès International de Lexicographie) à Liège, Belgique/Papers submitted to the Eighth EURALEX International Congress on Lexicography in Liège, Belgium*: 565-577. Leuven: Departemente Engels en Nederlands, Universiteit van Leuven.
- Combrink, L., N. Faasen, H. Geysers en A. Kloppers.** 1997. *Taalkonteks*. Kenwyn: Juta.
- Departement van Basiese Onderwys.** 2011. *Nasionale Kurrikulum- en Asseseringsbeleidverklaring Graad R-12*. Pretoria: Staatsdrukkery.
- Du Plessis, M.** 1998. *Tweetalige Aanleerderswoordeboek/Bilingual Learner's Dictionary Afrikaans-Engels/English-Afrikaans*. Kaapstad: Pharos Woordeboeke.
- Gouws, R.H.** 2001. Lexicographic Training: Approaches and Topics. Emejulu, J.D. (Red.). 2001. *Éléments de lexicographie gabonaise. Tome I*: 58-94. New York: Jimacs-Hillmann.
- Gouws, R.H.** 2006. Leerdergedrewe leksikografie. *Tydskrif vir Taalonderrig* 40(1): 72-84.
- Gouws, S. en M. Lähti.** 2005. *Nuwe Woordeboek sonder grense Werkboek Senior fase Graad 7-9*. Kaapstad: Maskew Miller Longman.
- Gouws, R.H. en D.J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.
- Gouws, R.H., M. Stark en L. Gouws.** 2004. *Nuwe Woordeboek sonder grense*. Kaapstad: Maskew Miller Longman.
- Gouws, R.H. en S. Tarp.** 2008. A New Step Forward for South African Learners' Lexicography: *The Oxford Afrikaans-Engels/English-Afrikaans Skoolwoordeboek/School Dictionary*. *Tydskrif vir Taalonderrig* 42(1): 65-79.
- Kromhout, J.** 2004. *Skoolwoordeboek Afrikaans-Engels/School Dictionary English-Afrikaans*. Kaapstad: Pharos Woordeboeke.
- Louw, P. (Red.).** 2007. *Oxford Afrikaans-Engels/English-Afrikaans Skoolwoordeboek/School Dictionary*. Kaapstad: Oxford University Press Southern Africa.
- Luther, J.** 2009. *HAT Afrikaanse Skoolwoordeboek*. Kaapstad: Maskew Miller Longman.
- Luther, J.** 2010. *HAT Aktiwiteite en Speletjies by die Afrikaanse Skoolwoordeboek en Afrikaanse Sakwoordeboek*. Kaapstad: Maskew Miller Longman.
- Meij, K., T. Kühn en R. Snyman.** 1985. *Vakdidaktiek: Afrikaans moedertaal in die sekondêre skool*. Pretoria: De Jager-HAUM.
- Paizee, D. en M.L. Kriel.** 2006. *Oxford Tweetalige Skoolwoordeboek Werkboek*. Kaapstad: Oxford University Press.

- Smith-Müller, W.** 2006. *Pharos Leerderboek by die Aanleerderswoordeboek*. Kaapstad: Pharos Woordeboeke.
- Smith-Müller, W.** 2007. *Pharos Leerderboek by die Skoolwoordeboek*. Kaapstad: Pharos Woordeboeke.
- Stark, M.P.** 1990. *Dictionary Workbooks. A Critical Evaluation of Dictionary Workbooks for the Foreign Language Learner*. Exeter Linguistic Studies 16. Exeter: University of Exeter Press.
- Tarp, S. en R.H. Gouws.** 2004. Wie leer wat uit Afrikaanse (aan)leerderwoordeboeke. *Tydskrif vir Geesteswetenskappe* 44(4): 276-298.
- Underhill, A.** 2002. *Macmillan English Dictionary for Advanced Learners Workbook*. Oxford: Macmillan Publishers.
- Van der Merwe, M.F.** 2013. Kriteria vir woordeboekwerkboeke in Afrikaans. Botha, W., P. Mavoungou en D. Nkomo. 2013. *Festschrift RUFUS H. GOUWS*: 280-294. Stellenbosch: SUN PRESS.
- WKOD Geletterdheid- en syferkundigheidstrategie 2006–2016*. 2006. Wes-Kaap Onderwysdepartement.

Differences in the Inclusion and Treatment of Terminology in OALD3, OALD4 and OALD8*

Marjeta Vrbinc (*marjeta.vrbinc@ff.uni-lj.si*), Faculty of Arts
and
Alenka Vrbinc (*alenka.vrbinc@ef.uni-lj.si*), Faculty of Economics,
University of Ljubljana, Ljubljana, Slovenia

Abstract: This article deals with the inclusion and treatment of scientific and technical vocabulary in the third, fourth and eighth editions of *Oxford Advanced Learner's Dictionary*. The comparison of these editions is based on a random sample of 50 pages from OALD8 from the lemma *foot* to *gimmick*. The same lemma range was also studied in OALD3 and OALD4. First, different ways of indicating terminology were identified: i.e., subject-field labels, definitions and short cuts. Then all the lemmata or their senses marked with a subject-field label were found and a list of all subject-field labels used in this lemma range was compiled to see similarities and differences between individual editions. The comparison showed that the number of subject-field labels in all three editions is almost identical, but the subject-field labels differ from edition to edition. The issue of overly specific labels (e.g., 'anatomy', 'phonetics') and labels that are too broad (e.g., 'science', 'technical') is addressed. The next part of the article is devoted to the changes in the treatment of LSP lexical items in these three editions of OALD, from missing labels to changes in labels and ways of indicating terminology by means of definitions and/or short cuts. The conclusion suggests improvements in the subject-field labels themselves, a more consistent way of including subject-field labels even when the definitions indicate the subject field and an improved use of short cuts when the reference is to a certain subject field.

Keywords: DEFINITIONS, GENERAL DICTIONARIES, LEMMA, MONOLINGUAL LEARNER'S DICTIONARIES, OALD3, OALD4, OALD8, SCIENTIFIC VOCABULARY, SENSE INDICATOR, SHORT CUTS, SUBJECT-FIELD LABELS, TECHNICAL VOCABULARY

Opsomming: Verskille in die insluiting en behandeling van terminologie in die OALD3, OALD4 and OALD8. Hierdie artikel gaan oor die insluiting en behandeling van wetenskaplike en tegniese woordeskat in die derde, vierde en agste uitgawes van *Oxford Advanced Learner's Dictionary*. Die vergelyking van hierdie uitgawes is gebaseer op 'n willekeurige

* This article is an extended and revised version of a paper "Comparative Inclusion and Treatment of Scientific and Technical Vocabulary in OALD3, OALD4 and OALD5", presented by Marjeta Vrbinc at the Eighteenth Biennial Meeting of the Dictionary Society of North America (DSNA), which took place at the McGill University, Montreal, Canada, 8–11 June 2011.

steekproef van 50 bladsye van OALD8 vanaf die lemma *foot* tot *gimmick*. Dieselfde lemmareeks is ook in OALD3 en OALD4 ondersoek. Eerstens is verskillende maniere van die aanduiding van terminologie vasgestel: d.i. onderwerpsveldetikette, definisies en kortpaaie. Daarna is al die lemmas of hul betekenis wat met 'n onderwerpsveldetiket gemerk is, opgespoor en 'n lys van al die onderwerpsveldetikette wat in hierdie lemmareeks gebruik is, opgestel sodat ooreenkomste en verskille tussen die verskeie uitgawes bepaal kan word. Die vergelyking het getoon dat die aantal onderwerpsveldetikette in al drie uitgawes amper eenders is, maar die onderwerpsveldetikette verskil van uitgawe tot uitgawe. Die kwessie van oordrewe spesifieke etikette (bv. 'anatomy', 'phonetics') en etikette wat te breed is (bv. 'science', 'technical') word behandel. Die volgende deel van die artikel word gewy aan die veranderinge in die behandeling van TSD- leksikale items in hierdie drie uitgawes van OALD, vanaf ontbrekende etikette tot veranderinge in etikette en maniere van aanduiding van terminologie deur middel van definisies en/of kortpaaie. Die slot stel verbeteringe in die onderwerpsveldetikette self voor, 'n konsekwenter manier om onderwerpsveldetikette in te sluit selfs wanneer die definisies die onderwerpsveld aandui en 'n verbeterde gebruik van kortpaaie wanneer die verwysing na 'n spesifieke onderwerpsveld is.

Sleutelwoorde: DEFINISIES, ALGEMENE WOORDEBOEKE, LEMMA, EENTALIGE AANLEERDERSWOORDEBOEK, OALD3, OALD4, OALD8, WETENSKAPLIKE WOORDESKAT, BETEKENISAANDUIDER, KORTPAAIE, ONDERWERPSVELDETIKETTE, TEGNIESE WOORDESKAT

1. Introduction

General dictionaries, be they mono- or bilingual, are primarily concerned with general vocabulary and are consulted more often by users than any other type of dictionary. The users of general dictionaries expect to find in them different pieces of information, such as definitions, translation equivalents, spelling, pronunciation, fixed word combinations, collocations, usage notes, grammatical information, information on word-division, etymology and register. They should be distinguished from LSP dictionaries, which include and treat the terminology of various specialist fields. As far as the treatment of lemmata is concerned, enormous differences can be observed between general and LSP dictionaries. General dictionaries usually include technical terms, particularly those that everyone can encounter in everyday life, but LSP dictionaries, as a rule, do not include words used in general language only (cf. Svensén 2009: 3).

Any dictionary user hopes to find a word he/she is looking up in his/her dictionary; on the other hand, it is impossible to include all the words everyone might want, which means that the compilers of a dictionary have to make decisions about what to include in a dictionary and/or exclude from it (cf. Atkins and Rundell 2008: 178). These decisions also concern the inclusion of scientific and technical vocabulary in general dictionaries. It should be stressed that in the last few decades, the proportion of entries in general dictionaries devoted to scientific and technical vocabulary has increased. This is in line with the increasingly important role of science and technology and consequently, with

new words that are coined on almost daily basis to name new inventions, concepts, devices or achievements. As Landau (2001: 32-35) points out, the larger general dictionaries are becoming a collection of LSP dictionaries merged with a general dictionary, for two reasons. The first reason is that the number of scientific and technical terms is increasing more rapidly than the number of general vocabulary items; the second reason originates in the prevailing cultural view in our society that science and technology are of the highest importance. The final decision about whether to include scientific and technical vocabulary depends mostly on the market, the user profile, and the cost of production, but it can be claimed that some dictionaries, especially more comprehensive ones, will include a considerable number of scientific and technical terms, whereas other dictionaries, especially pocket dictionaries, may exclude all or almost all of them (cf. Atkins and Rundell 2008: 182; Jackson 2002: 162).

The advent of computational lexicography and the use of corpora mean that dictionaries, including general-language dictionaries, can be updated more frequently and in a more representative way than is possible with manual methods. Apart from including new senses of already existing lemmata and new general vocabulary items, each update also contains scientific and technological neologisms, as well as more established terms. The unity between term and concept is an essential requirement of unambiguous communication (cf. Hartmann and James 1998: 138-39, Ahmad et al. 1995: 7). However, experts as well as laypeople use the same terms or are confronted by them. This is why terms are to be found in general-language dictionaries. What, however, are the criteria on the basis of which the terms are included in general-language dictionaries? According to Ahmad et al. (1995: 7), this depends on the status of the term in question, since we can make a distinction between 'field-internal' or 'field-internal/-external' terms. Field-internal terms are not part of the general language, since they are used in expert-to-expert communication. Terms falling into the category of field-internal/-external terms are encountered and sometimes used by laypersons as well as experts, thus being the best candidates for inclusion in general-language dictionaries. It can be claimed with a high degree of certainty that certain domains may not be covered in general-language dictionaries. It may also be expected that only subsets of terms from more accessible domains will be included and defined by their usage in communicative situations that are not exclusively field-internal (cf. Ahmad et al. 1995: 9). However, lexicographers should be aware that the main problem in selection is consistency.

In any general dictionary containing words that have special meanings in a technical field or science, field labels should be employed. In the course of planning a dictionary, a list of the domains should be drawn up whose vocabulary will be included in the dictionary. Field labels are applied to terms that are important in the field and in such widespread use that they have appeared in popular articles or in specialized magazines for the amateur. They are also employed when a word is used in two or more different disciplines with dif-

ferent meanings, or if it is used in one sense technically and in another popularly. It can be seen that in some dictionaries, field labels are used abundantly, whereas in others, the user hardly ever comes across one. In many cases, the information that a word, expression or a sense belongs to a field of science can be inferred from the definition (cf. Landau 2001: 226) or in recent editions of monolingual learner's dictionaries from short cuts (which are also called signposts, guidewords or items in a menu in various English monolingual learner's dictionaries).

The aim of this article is to compare three editions of OALD, i.e., OALD3, OALD4 and OALD8, to see how terminology is included and treated in each individual edition. First, we were interested in the number of lemmata labelled with a subject-field label in these three editions. Next, we wanted to investigate the number and types of subject-field labels to identify similarities and differences between OALD3, OALD4 and OALD8. Finally, we intended to study other ways of indicating technical and scientific vocabulary.

2. Methodology

In order to be able to analyse the treatment of terminology in English monolingual learner's dictionaries, it is essential to choose dictionaries that have been on the market for a longer period of time. Among the learner's dictionaries, only two dictionaries satisfy this criterion, i.e., *Oxford Advanced Learner's Dictionary* (first published in 1948) and *Longman Dictionary of Contemporary English* (first published in 1978), while their competitors are all too new to facilitate comparison. For the sake of our analysis, we chose the oldest among the learner's dictionaries: *Oxford Advanced Learner's Dictionary*. Our analysis focuses on the 3rd, the 4th and the 8th editions, because our aim was to observe the developments concerning inclusion and treatment of terminology from the relatively old OALD3 (published in 1974 and referred to as a dictionary belonging to the second generation of learner's dictionaries by Cowie (1999: 82, 97-105)) to OALD4 (published in 1989, referred to as a dictionary belonging to the third generation of learner's dictionaries by Cowie (1999: 144, 148-151)) and to the most recent edition: OALD8 (published in 2010). OALD4 was chosen because it was published approximately in the middle of the 36-year time span between the 3rd and the 8th editions. Another reason for choosing OALD4 is that the 1980s marked a watershed in learner lexicography, and as Cowie (1999: 144) points out, it was clear early in this decade that computers would play an increasingly important role in the compilation of dictionaries and that the next phase of dictionary development would be affected by the increasing professionalism of lexicography.

The analysis was carried out on a sample of 50 randomly chosen pages from OALD8, from the lemma *foot* to *gimmick* (pp. 602-653). The same lemma range was then also studied in OALD4 (44 pages, pp. 477-521) and OALD3 (29 pages, pp. 339-368). First, all the lemmata or their senses marked with a subject-

field label were identified, then a list of all subject-field labels used in this lemma range was compiled and compared. Last but not least, we examined the definitions, sense indicators and short cuts to see whether they are also used to indicate specific subject fields.

3. Ways of indicating terminology

English monolingual learner's dictionaries use several ways of indicating terminology: subject-field labels, definitions, sense indicators and short cuts. Since the subject-field labels are the most obvious elements indicating terminology, we will first examine how subject-field labels are used in OALD3, OALD4 and OALD8.

On the inside cover of OALD3, a list of subject-field labels and abbreviations can be found referred to as 'Specialist English registers' (i.e., specialist or technical fields). Terms appearing in the 'Specialist English registers', such as 'botany (bot)', 'nautical (naut)' and 'rugby', are self-explanatory and refer to as many as 58 different fields or sub-fields:

accounts, aerospace, algebra, anatomy, architecture, arithmetic, art, astronomy, ballet, biblical, biology, book-keeping, botany, business, chemistry, cinema, commerce, computers, cricket, ecclesiastical, engineering, electricity, farming, finance, football, gambling, geology, geometry, grammar, history, journalism, legal, linguistics, mathematics, mechanics, medical, meteorology, military, music, mythology, nautical, pathology, philosophy, phonetics, photography, physics, physiology, politics, psychology, racing, radio telegraphy, rugby, science, sport, tennis, theatre, trigonometry, zoology.

The list of subject-field labels in OALD3 is comprehensive and contains labels that fall beyond the scope of learner's dictionaries. Labels such as 'accounts', 'book-keeping', 'algebra', 'arithmetic', 'trigonometry', 'pathology', 'physiology', 'radio telegraphy', belong to LSP dictionaries rather than general dictionaries, let alone learner's dictionaries. They are far too specific and technical for the target user of a monolingual learner's dictionary to comprehend.

When going through the list of subject-field labels in OALD3, the label 'computers (comp)' captured our attention, since this dictionary was published in 1974, i.e., in the early days of computers. One might wonder which technical terms used in the field of computer technology would merit inclusion in a learner's dictionary, given the fact that no corpora and consequently no frequency counts were available at that time. With the help of a computer expert, we compiled a list of ICT terms that were known and used at the end of the 1960s and at the beginning of the 1970s, i.e., in the period of time when the lemma list for this edition of OALD was compiled. These terms include: micro-processor, processor, bus, ISA bus, I/O, drive, floppy drive, diskette, floppy

diskette, printer, laser printer, FTP, CD, VoIP, mainframe, EPROM, memory, monolithic main memory, ATM, file, disc and Ethernet. These terms are either not included at all or included but without the specialized sense referring to the ICT field. On the other hand, terms such as computer, computerize, programme, programmer, hardware and software are included in the wordlist of OALD3, but they are not labelled 'comp'. Interestingly, the lemma *computer* is defined as 'electronic device which stores information on discs or magnetic tape, analyses it and produces information as required from the data on the tapes, etc.', but the lemma *disc* is not treated in the sense of 'a device for storing information on a computer, with a magnetic surface that records information received in electronic form'. The lemma *programme* is another interesting example. OALD3 includes just the British English spelling as a lemma and gives *program* as a variant spelling in brackets without further explanation. Sense 3 of the lemma *programme* is defined as 'coded collection of information, data, etc fed into an electronic computer'. In the fourth edition of OALD as well as in all subsequent editions of OALD, two lemmata are included: *programme* and *program*. The latter has two senses: the American English spelling for *programme* and the specialized sense used in computing. Considering all this, the question arises why the subject-field label 'comp' is listed among the subject-field labels. It is not to be found in the dictionary, so it seems to be redundant. Is it there simply to impress the target user of that time, given the fact that in the early 1970s computer technology was still in its infancy at least as regards general use of computers?

In OALD4, the user's guide (called Detailed Guide) appearing as part of the back matter also includes a chapter on style and field (pp. 1572-75) but only a short passage (cf. 12.7 Technical fields, p. 1574) is dedicated to the treatment of terminology. Nowhere in the dictionary, however, can a user find a list of subject-field labels. The latter also holds true of OALD8, whose user's guide is limited to the graphic presentation of entries and a list of labels used in the dictionary (p. i), among which the only reference to subject-field labels is the label 'technical', though quite a few other subject-field labels appear in the body of the dictionary. The subject-field label 'technical' is explained as "language used by people who specialize in particular subject areas, for example *accretion*, *adipose*" (OALD8, p. i), which is a vague and far from informative explanation (see also 4. Subject-field labels in the studied segment in OALD3, OALD4 and OALD8 compared).

One of the improvements in OALD4 which definitely contributes to less confusion in understanding the labels is that labels are typographically distinguished from sense indicators in that italics are used to mark subject-field labels. The same applies to OALD8:

<p>genus /.../ <i>n</i> (<i>pl</i> genera /.../) 1 (science) division of animals or plants within a family /.../</p> <p>OALD3, p. 364</p>

genus /.../ *n* (*pl* genera /.../) **1** (*biology*) group of animals or plants within a family(4), often itself subdivided into several species(1). /.../
OALD4, p. 515

Besides subject-field labels, the definition itself often suggests that the word being defined belongs to a specific subject field:

geriatric /.../ *noun* **1** **geriatrics** [U] *the branch of medicine concerned with the diseases and care of old people* /.../
OALD8, p. 648

Another method frequently used by lexicographers to indicate terminology is the use of sense indicators, which are an entry component and are thought up by lexicographers to help a user choose the appropriate sense of the lemma:

front /.../ *n* /.../ **6** [C] (*of weather*) forward edge of an advancing mass of warm or cold air
OALD 4, pp. 496, 497

foreclose /.../ *v* [I, Ip, Tn] ~ (**on sb/sth**) (*of a bank, etc that has lent money for a mortgage*) take possession of the property of (sb), usu because repayments have not been made /.../
OALD4, p. 481

game point *noun* (*especially in TENNIS*) a point that, if won by a player, will win them the game
OALD8, p. 637

One feature relatively newly introduced into monolingual learner's dictionaries to aid users with the disambiguation of polysemous items is called short cuts (in OALD, they were introduced in the 6th edition). Short cuts give the core meanings of highly polysemous words. They help the users to make mental connections with the word in the context in which they encountered it. It should be pointed out that in many cases the context in which the user has met an unknown word will prompt the choice of short cut. Short cuts do not replace the full definition, but rather form a quick menu for the user's eye to run down. Consequently, users should usually be able to select the right sense paragraph to read fully without having to read all the details in several other paragraphs first.

freeze /.../ ► **COMPUTER 8** [I] when a computer screen **freezes**, you cannot move any of the images, etc. on it, because there is a problem with the system /.../

OALD8, p. 620

foot /.../ ► **IN POETRY 7** [sing.] (*technical*) a unit of rhythm in a line of poetry containing one stressed syllable and one or more syllables without stress. Each of the four divisions in the following line is a foot: *For 'men / may 'come / and 'men / may 'go. /.../*

OALD8, p. 603

As can be seen from the above dictionary entry for *foot*, two different ways of indicating the subject field can be combined in one and the same entry or one and the same sense. Sense 7 in *foot* combines a short cut ('in poetry') and a subject-field label ('technical'). Another possible combination is a subject-field label ('finance') and a sense indicator ('especially of a bank'), as illustrated by the example below:

foreclose /.../ *verb* 1 [I, T] ~ **(on sb/sth)** | ~ **sth** (*finance*) (especially of a bank) to take control of sb's property because they have not paid back money that they borrowed to buy it /.../

OALD8, p. 606

4. Subject-field labels in the studied segment in OALD3, OALD4 and OALD8 compared

Lemmata accompanied by subject-field labels are the most obvious terminological lemmata. In OALD3, 30 words with subject-field labels can be found in the range of the lemmata studied, while OALD4 includes 43 such lemmata and OALD8 as many as 126. If the number of lemmata in these three editions of OALD is compared, we can see that there is not an enormous difference between OALD3 and OALD4, but in OALD8 the number of lemmata marked by subject-field labels rises considerably. Interestingly, the number of subject-field labels does not change drastically from edition to edition. Table 1 presents subject-field labels found in the lemma span studied: OALD3 includes 18 subject-field labels, OALD4 only one more (i.e., 19), and the same applies to OALD8. Labels appearing in all three editions are shaded grey (6 labels), if the entire cell is shaded black, it marks the labels that can be found in both OALD3 and OALD4 (3 labels), whereas the entire cell shaded grey indicates the labels found in OALD4 and OALD8 (6 labels).

OALD3	OALD4	OALD8
1. anatomy	1. anatomy	1. anatomy
2. biology	2. Bible	2. architecture
3. botany	3. biology	3. astronomy
4. chemistry	4. botany	4. biology
5. cinema	5. chemistry	5. business
6. commerce	6. computing	6. chemistry
7. football	7. commerce	7. computing
8. grammar	8. finance	8. finance
9. legal*	9. grammar	9. geology
10. meteorology	10. law*	10. grammar
11. military	11. French law	11. law*
12. music	12. mathematics	12. linguistics
13. nautical	13. medical	13. mathematics
14. theatre	14. music	14. medical
15. science	15. religion	15. music
16. sport	16. physics	16. physics
17. swimming	17. psychology	17. phonetics
18. tennis	18. politics	18. psychology
	19. sport	19. technical

*The label 'legal' as used in OALD3 was replaced by 'law' in OALD4 and OALD8, but both labels refer to the same field of science, thus they can be regarded as the same label.

Table 1: Subject-field labels in OALD3, OALD4 and OALD8

Sports terminology deserves special attention, since it represents quite a lot of vocabulary in various editions of OALDs, especially in OALD8. Labelling of sport, however, presents a huge problem, since most of the labelling is carried out by indirect methods of labelling and not by subject-field labels proper. In fact 'sport' is the only subject-field label proper (found in OALD3 and OALD4, but not in OALD8), while all other references to different sports are only indirect ways of indicating the subject field. Only OALD3 uses two additional labels referring to sport, i.e., 'football' and 'tennis'. The various sports disciplines referred to in the OALDs and used either in the form of sense indicators or short cuts (only in OALD8) are as follows:

OALD3: football, swimming, tennis, golf

OALD4: cricket, snooker, football, tennis, golf, hockey

OALD8: American football, basketball, baseball, cricket, football, soccer, rugby, tennis

In many cases, the delimitation of two or even three seemingly related labels is

unclear. One can wonder about the utility of the subject-field labels 'anatomy' and 'medical' (in OALD4 and OALD8). Is anatomy not a subfield of medicine? Is it necessary to bother a general dictionary user who is not a specialist with such specific subject fields? The same can be said of labels such as 'linguistics', 'grammar' and 'phonetics' in OALD8 or 'commerce' and 'finance' in OALD4 and 'business' and 'finance' in OALD8. On the other hand, OALD8 uses the label 'biology', but not, for instance, 'botany', while OALD3 and OALD4 employ both 'botany' and 'biology' as labels. This means that some subject-field labels are used more generically to refer to the entire field of the science, whereas some subject-field labels refer to sub-fields themselves.

Another two labels that must be mentioned are the generic 'science' in OALD3 and 'technical' in OALD8. These two labels are very general and do not refer to any specific field or sub-field. In OALD3, 'science' is used to label words such as *genus*, which could be labelled with the more specific label 'biology'. In OALD8, for instance, 'technical' is used to label lemmata such as *fruit* ('a part of a plant or tree that is formed after the flowers have died and in which seeds develop'), which could more logically be labelled 'biology' (or 'botany' if this label existed; in OALD3 and OALD4, the label 'botany' is used) and *geld*, which could also be labelled 'biology' (or 'zoology' or even 'veterinary' if one of these existed).

Contrary to the general labels 'science' and 'technical', we can find the label 'French law' in OALD4, which is a very specific label, considering the fact that the label 'law' is also used in this edition of OALD. The lemma *force majeure* is labelled 'French law' in OALD4, while in OALD8, it is more reasonably labelled as 'from French, law'. The interpretation of both ways of labelling is completely different: 'French law' implies that the term is used only to refer to the French law and 'from French, law' means that the expression comes from French and is used in legal terminology.

5. Changes in the treatment of technical and scientific words in OALD3, OALD4 and OALD8

When the three editions of OALD are compared, certain changes in the treatment of technical and scientific words can be observed which are discussed below.

(1) A term that clearly belongs to technical or scientific vocabulary lacks a subject-field label in OALD3, but includes one in OALD4 and OALD8:

<p>formula /.../ 2 statement of a rule, fact, etc esp one in signs or numbers, as in chemistry, mathematics, etc, eg 'Water = H₂O'</p> <p>OALD3, p. 344</p>
--

formula /.../ 1 [C] (a) (*chemistry*) set of symbols showing the elements that a substance is made of /.../ (b) (*mathematics or physics*) expression of a rule or relationship in algebraic symbols /.../

OALD4, p. 485

formula /.../ 1 [C] (*mathematics*) a series of letters, numbers or symbols that represent a rule or law /.../ 2 [C] (*chemistry*) letters and symbols that show the parts of a chemical COMPOUND, etc /.../

OALD8, p. 611

It is evident from the definition in OALD3 that *formula* is a term used in chemistry, mathematics, etc., but in spite of that, the label has not been included. In OALD4, the unlabelled sense 2 from OALD3 is divided into two subsenses both labelled with very specific subject-field labels ('chemistry' and 'mathematics or physics'). OALD8, however, treats both subsenses from OALD4 as separate senses with 'mathematics' and 'chemistry' as the subject-field labels. The biggest step forward is the development from OALD3 to OALD4, where the term is treated with greater precision.

Other examples of lemmata which are unlabelled in OALD3 but include a label in OALD4 and OALD8 are: *futures* (labelled 'commerce' in OALD4 and 'finance' in OALD8), *garbage* (labelled 'computing' in OALD4, not included in this sense in OALD8), *gastric* and *gastritis* (labelled 'medical' in OALD4 and OALD8), *genital* (labelled 'medical' in OALD4 and without a label in OALD8), *gender* (labelled 'grammar' in OALD4 and OALD8), *genesis* (labelled 'Bible' in OALD4 and 'formal', which is a style label, in OALD8) and *gilt-edged* (labelled 'finance' in OALD4 in OALD8).

(2) In some cases, OALD3 includes a subject-field label, OALD4 drops it, but OALD8 includes it once again.

foreclose /.../ vt, vi (legal) [VP6A,2A] use the right (given by a mortgage) to take possession of property (when interest or capital has not been paid at the required time) /.../

OALD3, p. 342

foreclose /.../ v [I, Ip, Tn] ~ (on sb/sth) (of a bank, etc that has lent money for a mortgage) take possession of the property of (sb), usu because repayments have not been made /.../

OALD4, p. 481

foreclose /.../ *verb* **1** [I, T] ~ **(on sb/sth)** | ~ **sth** (*finance*) (especially of a bank) to take control of sb's property because they have not paid back money that they borrowed to buy it /.../
OALD8, p. 606

In OALD3, the subject-field label 'legal' is used, which is replaced by the label 'finance' in OALD8, whereas OALD4 lacks a label.

(3) A lemma is marked by a subject-field label in OALD3, but lacks a subject-field label in OALD4 and OALD8 (e.g., *foxhole*). Another possibility is that a subject-field label in OALD3 is replaced by a sense indicator in OALD4 or a short cut in OALD8:

front /.../ *n* /.../ **7** (*met*) boundary between masses of cold and warm air
OALD3, p. 352
front /.../ *n* /.../ **6** [C] (*of weather*) forward edge of an advancing mass of warm or cold air
OALD 4, pp. 496, 497
front /.../ *noun, adj., verb*
■ *noun* /.../
► **WEATHER 12** [C] the line where a mass of cold air meets a mass of warm air /.../
OALD8, p. 624, 625

The same situation applies to the lemma *force* in the sense of 'authority'. At first sight, OALD3 and OALD4 both define this sense in exactly the same way: '(legal) authority'. A closer examination, however, shows that 'legal' is the subject-field label in OALD3 and a sense indicator in OALD4 (this confusion is due to the lack of italics in OALD3). OALD8, however, uses neither the label nor the sense indicator but instead employs the short cut 'authority':

force¹ /.../ *n* /.../ **5** (*legal*) authority; power of binding(6): *put a law into ~, make it binding. When does the new law come into ~? The rule/regulation is no longer in force.*
OALD3, p. 341
force¹ /.../ *n* /.../ **8** [U] (*legal*) authority: *This decree has the force of law behind it. ...*
OALD4, p. 480

force /.../
■ **noun** /.../
▶ **AUTHORITY 5** [U] the authority of sth: *These guidelines do not have **the force of law**.* ◇ *The court ruled that these standards have force in British law.*
/.../
OALD8, p. 605

(4) A lemma or one of its senses is included in OALD3 but not in OALD4 and OALD8 (e.g., *foretop*, *foot* in the meaning of infantry). This is understandable because every dictionary must stick to certain principles of selection and must exclude many words and expressions because they are obsolete, rarely used, or too specialized for a general monolingual learner's dictionary.

(5) Adding new lemmata and/or new senses: owing to the development of science and technology, new terms are invented on a daily basis, or sometimes new senses are added to the existing ones, which is evident in every new edition. In OALD4 and OALD8, the lemma *function key*, labelled 'computing', is added, while it is not included in OALD3. In OALD8, the lemmata, such as *function word* (labelled 'grammar') or *fuzzy logic* (labelled 'computing'), appear in the wordlist; these lemmata are not included in OALD3 and OALD4. Some lemmata have developed new senses, which is evident from the treatment of the noun *footfall* in OALD8, where sense 2, 'the number of people that visit a particular shop/store, shopping centre, etc. over a period of time' (labelled 'business') is added.

(6) Inconsistent use of subject-field labels: in OALD3 and OALD4, *formic acid*, which is clearly a scientific term, lacks the subject-field label. *Acid*, on the other hand, is labelled in both editions with the label 'chemistry'. This is certainly an example of an inconsistent treatment of terms, which should be avoided by either labelling both terms 'chemistry' or suggesting the subject field within the definition. In OALD8, this inconsistency has been corrected by labelling both terms:

acid /.../ *noun, adjective*
■ (*chemistry*) a chemical, usually a liquid, that contains HYDROGEN and has a Ph of less than seven. The HYDROGEN can be replaced by a metal to form a salt. Acids are usually sour and can often burn holes in or damage things they touch. /.../
OALD8, p. 12
formic acid /.../ *noun* [U] (*chemistry*) an acid made from CARBON MONOXIDE and steam. It is also present in a liquid produced by some ANTS.
OALD8, p. 611

Another example of inconsistent labelling can be observed in the case of the lemma *gene*, which is labelled 'biology' in all three editions of OALD under investigation, and its derivatives and all the compounds containing 'gene', 'genetic' or 'genetically', which are mostly unlabelled. In Table 2 below, words marked by a subject-field label in any of the three editions of OALD are indicated by the subject field in brackets and in italics.

Derivatives
OALD3: gene (<i>biology</i>), genetic, geneticist, genetics
OALD4: gene (<i>biology</i>), genetic, genetically, geneticist, genetics
OALD8: gene (<i>biology</i>), genetic, genetically, geneticist, genetics
Compounds
OALD3: /
OALD4: genetic code, genetic engineering
OALD8: gene pool (<i>biology</i>), gene therapy (<i>medical</i>), genetically modified, genetic code, genetic engineering, genetic fingerprinting, genetic fingerprint

Table 2: Labelling of derivatives and compounds

It is evident from Table 2 that the field of genetics has undergone rapid development in the last four decades, which has clearly exerted great influence on the inclusion of terms connected with it in these three editions of OALD. In all unlabelled terms, it is only the definition that tells the user that the word or expression belongs to a specific subject field. Inconsistent labelling can also be observed in terms denoting names of diseases, plants, animals and measures. For example, *foot-and-mouth disease* or *flu/influenza* are not labelled 'biology' or 'medical' but are only defined as diseases, while *gastritis* is labelled 'medical' in OALD8, although it is also defined as an illness. It certainly makes perfect sense not to label *flu*, because the word belongs to LGP, *gastritis*, on the other hand, belongs to LSP. Another example is *foot* as a unit of length measurement, which is not labelled, whereas *gigabit*, a unit used in computer science, is. Certainly, *foot* in this sense has been used in English for a long time, whereas *gigabit* (or more precisely *bit*, from which *gigabit* is formed) originates from the 1940s. It can nevertheless be claimed that labelling some words belonging to a certain semantic field but not the others belonging to the same field can be quite confusing for a dictionary user.

6. Conclusion

A label as a special symbol or abbreviated term used to mark a word, expres-

sion or sense as being associated with a particular subject field is by far the most obvious sign that a word, expression or sense belongs to LSP. Thus, from a user's perspective, labelling terms by means of subject-field labels is the most user-friendly way of indicating terminology. But some remarks should also be made concerning the use of subject-field labels. The first problem is the use of closely related subject-field labels (e.g. 'linguistics' vs. 'grammar', 'phonetics'; 'business' vs. 'commerce', 'finance'). A general dictionary user cannot be expected to recognize the subtle differences between such subject-field labels. If such labels are used, one would expect an explanation of the distinction between them, but taking account of the type of dictionary and the target audience, it can be claimed with a high degree of certainty that this is an unnecessary complication. In monolingual learner's dictionaries, one would expect that the subject-field labels would refer to fields of science only and would disregard the sub-fields. On the other hand, OALD3 and OALD8 make use of one general subject-field label ('science', 'technical') that is only vaguely defined (if at all). It is recommendable to use this type of subject-field label only to mark entries common to several domains, i.e., as a higher-level domain marker. Otherwise, lemmata belonging to lower-level domains should be labelled using a more specific subject-field label that is also listed as a subject-field label in a given dictionary ('technical' vs. 'biology' or 'chemistry').

The next remark concerns the labelling of a lemma by means of a definition. A dictionary user may be puzzled that some lemmata are equipped with subject-field labels whereas in other lemmata, this function is taken over by a definition. This is a more indirect way of indicating technical and scientific vocabulary. It is thus questionable whether a general dictionary user is aware that he/she is dealing with an LSP lexical item. A considerable number of LSP lexical items lack a subject-field label but instead provide this piece of information within the definition. A more consistent policy in this respect should be expected, which means that a subject-field label should be provided even though it is clear from the definition that the word, expression or sense belongs to terminology.

The final remark refers to the use of short cuts as the third way of indicating LSP lexical items. Those short cuts that are lexically identical with the subject-field labels are not problematic from the point of view of the user who can easily recognize a lexical item as one belonging to LSP. Unfortunately, some short cuts are not so transparent and resemble sense indicators. It is therefore doubtful whether such short cuts are recognized by a dictionary user as indications of a subject field.

To sum up, subject-field labels should be listed in the front matter of a dictionary and, if necessary, explained. Apart from that, they should also be used more consistently throughout the dictionary, even in those cases where the subject field is indicated only by a definition or a short cut.

References

Dictionaries

- Cowie, Anthony P. (Ed.).** 1989⁴. *Oxford Advanced Learner's Dictionary of Current English*. Oxford: Oxford University Press. (OALD4)
- Hornby, A.S. (Ed.).** 1974³. *Oxford Advanced Learner's Dictionary of Current English*. Oxford: Oxford University Press. (OALD3)
- Turnbull, Joanna (Ed.).** 2010⁸. *Oxford Advanced Learner's Dictionary of Current English*. Oxford: Oxford University Press. (OALD8)

Other literature

- Ahmad, Khurshid, Willy Martin, Martin Hölter and Margaret Rogers.** 1995. *Specialist Terms in General Language Dictionaries*. Available at <<http://209.85.129.132/search?q=cache:felngsLWmIMJ:ftp://ftp.ee.surrey.ac.uk/pub/research/AI/pointer/diction.ps.gz+SPECIALIST+TERMS+IN+GENERAL+LANGUAGE&cd=1&hl=sl&ct=clnk&gl=si>>. [Accessed 3 January 2013.]
- Atkins, B.T. Sue and Michael Rundell.** 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.
- Cowie, Anthony P.** 1999. *English Dictionaries for Foreign Learners: A History*. Oxford: Oxford University Press/Clarendon Press.
- Hartmann, Reinhard R.K. and Gregory James.** 1998. *Dictionary of Lexicography*. London/New York: Routledge.
- Jackson, H.** 2002. *Lexicography: An Introduction*. London/New York: Routledge.
- Landau, Sidney I.** 2001². *Dictionaries: The Art and Craft of Lexicography*. New York/Cambridge: Cambridge University Press.
- Svensén, B.** 2009. *A Handbook of Lexicography: The Theory and Practice of Dictionary-making*. Cambridge: Cambridge University Press.

Moroccorp: tien miljoen woorden uit twee Marokkaans-Nederlandse chatkanalen

Tom Ruetten, *KU Leuven, Leuven, België en Humboldt-Universität zu Berlin, Berlijn, Duitsland (tom.ruetten@hu-berlin.de)*

en

Freek Van de Velde, *KU Leuven, Leuven, België en FWO-Vlaanderen, België (freek.vandevelde@arts.kuleuven.be)*

Samenvatting: In dit artikel stellen we een nieuw corpus voor van computer-gemedieerde communicatie in het Nederlands door Marokkaans-Nederlandse taalgebruikers, dat bestaat uit tien miljoen woorden chat-materiaal. We behandelen de achtergrond, de compilatiemethode en de interne structuur van het corpus, en we leggen het verband tussen ons eigen werk en eerdere pogingen om een corpus van Nederlandse chattaal te bouwen. We hebben ook een *Stable Lexical Marker analyse* uitgevoerd en een gevalstudie over een welbekende morfosyntactische eigenschap van het Marokkaans Nederlands om op die manier de representativiteit van het corpus te beoordelen.

Trefwoorden: NEDERLANDS, MAROKKAANS NEDERLANDS, CORPUS, CHAT, STABLE LEXICAL MARKER ANALYSIS, ETNISCH NEDERLANDS, REPRESENTATIVITEIT

Abstract: Moroccorp: Ten Million Words from two Moroccan Dutch Chat Channels. In this article we introduce a new corpus of computer-mediated communication in Dutch by Moroccan-Dutch language users, consisting of ten million words of chat material. We treat the background, the compilation method and the inner structure of the corpus, and we relate our efforts to previous attempts to build a corpus of Dutch chat language. We also conducted a *Stable Lexical Marker analysis* and a case study on a well-known morphosyntactic feature of Moroccan Dutch to assess in this manner the representativity of the corpus.

Keywords: DUTCH, MOROCCAN DUTCH, CORPUS, CHAT, STABLE LEXICAL MARKER ANALYSIS, ETHNIC DUTCH, REPRESENTATIVITY

1. Inleiding¹

De studie van 'taalcontact' mag zich al een tijdje verheugen in een groeiende interesse onder taalkundigen, zoals blijkt uit het bestaan van een eigen tijdschrift (*Journal of Language Contact*, Brill) en tal van inleidingen bij gereputeerde uitgeverij zoals die van Thomason (2001) bij Edinburgh University Press, die van Matras (2009) bij Cambridge University Press of die van Hickey (2010) bij

Blackwell. De toename is voor een deel toe te schrijven aan de beschikbaarheid van dataverzamelingen, zoals bijvoorbeeld transcripties van taalproductie van twee- of meertalige kinderen in de Childes database (MacWhinney 2000), of multilinguale corpora (zie Xiao 2008 voor een overzicht). De studie van taalcontact bevindt zich op het snijpunt van de sociolinguïstiek, taalverwerving, twee-/meertaligheid en de historische taalkunde, en er wordt vrij breed aangenomen dat taalcontact een uiterst belangrijke aanstoker of determinant is in taalverandering (zie Weinreich 1953; Thomason en Kaufman 1988; Harris en Campbell 1995; Croft 2000; Heine en Kuteva 2003, 2005; Drinka 2010).

Ook voor het Nederlands is er intensief onderzoek verricht naar het effect van taalcontact. De Lage Landen zijn al eeuwenlang een gebied waarin sprekers van verschillende talen met elkaar in contact komen. Meer zelfs: het uitzicht van het Nederlands is in grote mate een resultante van taalcontact (Buccini 1995, 2010), en dat geldt eigenlijk voor de Germaanse taalfamilie in het algemeen (Hawkins 1990: 60-61; Roberge 2010). Die impact van vreemde elementen is er in recente jaren niet minder op geworden. De demografische samenstelling in sommige Nederlandse en Vlaamse steden zoals Rotterdam en Genk laat diepe sporen na in het Nederlands dat daar gesproken wordt (Cornips en de Rooij 2003).

Ook vanuit de sociolinguïstiek is er veel aandacht voor variëteiten of 'lecten' die zich niet of niet alleen op een geografische dimensie laten definiëren, maar die in belangrijke mate gedefinieerd worden volgens de sociaal-culturele dimensie. Voor het Nederlands valt onder meer te denken aan 'poldernederlands', 'tussentaal', 'straattaal' en 'cités'. In die variëteiten zit vaak wel een geografische component: poldernederlands (Stroop 1998) wordt gesproken in Nederland, tussentaal (o.a. Geeraerts 2002) wordt gesproken in Vlaanderen en heeft een Brabantse invloed (pace Taeldeman 2008), cités (Ramaekers 1998) is beperkt tot Genk en ook straattaal (Appel 1999) is regiogebonden. Maar een goed begrip van deze variëteiten kan niet om de sociale dimensie heen: poldernederlands wordt of werd in eerste instantie geproken door jonge, hoogopgeleide vrouwen (Stroop 1998), en de sociale stratificatie van de tussentaal is uitvoerig onderzocht in Plevoets (2008). Voor cités en straattaal geldt de (groot)stedelijke context als bepalend, en een van de factoren die een grote impact hebben op die context is de aanwezigheid van sprekers die het Nederlands niet als moedertaal hebben. Op die manier ondergaat het Nederlands invloeden van buitenaf en de studie van deze hedendaagse lecten (cf. Ruetten et al. 2014) kan dan ook een bijdrage leveren aan de historische taalkunde, net zoals omgekeerd de historische taalkunde inzichten kan bijbrengen over de langetermijneffecten van zulke vormen van beïnvloeding.

De toenemende belangstelling voor taalcontact en lectale variatie en de groeiende technologische mogelijkheden binnen de Digital Humanities, hebben ertoe geleid dat er corpora aangelegd zijn van wat met een enigszins beladen term wel eens 'etnisch Nederlands' genoemd wordt. In dit artikel willen we het fonds beschikbaar onderzoeksmateriaal uitbreiden door een nieuw corpus voor

te stellen en publiek beschikbaar te maken voor onderzoek binnen de historische taalkunde, de taalverwerving, de sociolinguïstiek en de studie van taalcontact. Het gaat om een Nederlands chattaal-corpus van 10.000.000 woorden geschreven door bezoekers van chatkanalen die zich richten op Nederlanders met een Marokkaanse achtergrond, wier taalgebruik zich laat omschrijven als een etnolectische variëteit die in de literatuur vaak met de term 'Marokkaans Nederlands' of 'Moroccan flavored Dutch' aangeduid wordt (zie Nortier en Dorleijn 2008). Dit corpus hebben we de naam 'Moroccorp' gegeven.

Op de taalkundige en ideologische problemen bij het afzonderen van zo'n etnolect komen we terug in Sectie 2.3. We willen er hier wel al op wijzen dat we de term louter descriptief of 'fenomenologisch' opvatten, zonder meteen ook ontologische uitspraken te doen over de status van het etnolect.

Het aanleggen van een corpus van Marokkaans Nederlands is geen nieuw idee, maar het bestaande arsenaal aan ruwe onderzoeksdata is aan uitbreiding toe. Sommige van de 'etnische Nederlandse' corpora zijn particuliere initiatieven en zijn niet vrij beschikbaar (b.v. het corpus van Jaspers 2004). Andere kunnen wel vrij geconsulteerd worden. Zo is er de Dutch Bilingualism Database (DBD), die vrij beschikbaar is via de website van het Max Planck Instituut (MPI) in Nijmegen² en een sectie Marokkaans Nederlands bevat. De opnames in de DBD betreffen gesproken Nederlands, wat aan de ene kant fijnmazige analyse toelaat, ondermeer op het klankniveau, maar wat aan de andere kant praktische beperkingen heeft opgelegd aan de omvang van het verzamelde materiaal. De precieze omvang in aantal tokens van het Marokkaans-Nederlandse DBD-corpus is niet bekend, maar beslaat ongeveer 30 uur aan opnames. Het Moroccorp is in vergelijking veel groter, aangezien het chattaal van meerdere maanden omvat.

Het ontwikkelen van het Moroccorp is een inspanning om *Kommunikationsereignisse* (Zeige Te versch.) te verzamelen die de gevolgen van taalcontact tussen Nederlands en Marokkaans of Berbers weerspiegelen. Het Moroccorp doet dit door een substantiële verzameling aan te leggen van Marokkaans Nederlands — dat wil zeggen: taalproductie door Nederlandstaligen met een Marokkaanse achtergrond — waarin redactionele inmenging zo klein mogelijk gehouden is. Zoals in Sectie 2 in detail beschreven wordt, hebben we online materiaal vergaard uit chatberichten. Het idee om chattaal op internet te gebruiken als 'proxy' voor spontane en informele taalproductie met een laag redactioneel gehalte, is niet nieuw. In Androutsopoulos (2006) wordt een overzicht gegeven van onderzoek naar zogenoemde computer-gemedieerde communicatie, waaronder ook chattaal valt. Het idee om chattaal te verzamelen is voor het Nederlands eerder al gevolgd in het ConDiv-corpus (zie Grondelaers et al. 2000) en in het chatcorpus van Vandekerckhove (zie Vandekerckhove en Nobels 2010). We willen hierbij aanmerken dat een chatcorpus uiteraard slechts een partiële relevantie voor het taalkundige onderzoek heeft. Een belangrijke beperking volgt uit het feit dat enkel het taalgebruik van een kleine groep mensen in één zeer specifiek register gedocumenteerd wordt, waardoor de

gevonden resultaten niet zonder meer veralgemeend kunnen worden. Hierdoor is er vooral aandacht gegaan naar specifieke kenmerken van zogenoemde *internet-taal* (Crystal 2011), vanuit het perspectief van de conversatie-analyse (Herring 2010) of van taalverloedering (Tagliamonte en Denis 2008). Het zou anderzijds te ver gaan om chattaal zondermeer als onbruikbaar ter zijde te schuiven. Voor het Marokkaans Nederlands is echter de bruikbaarheid van internetdata onder de aandacht gebracht door Boumans (2002), zij het met een veel beperkter corpus (ca. 40.000 woorden).

2. Samenstelling van het corpus

Het samenstellen van een corpus waarin de taal van één of meerdere chatkanalen wordt verzameld vereist een aantal weloverwogen methodologische keuzes. We bespreken de genomen stappen kort in Sectie 2.1. In die sectie argumenteren we ook de stelling dat het door ons verzamelde taal materiaal representatief is voor het Nederlands van taalgebruikers met een Marokkaanse achtergrond in openbare chatkanalen. Daarna, in Sectie 2.2, geven we een korte kwantitatieve inkijk in het corpus door een overzicht te geven van het aantal woorden, het aantal gebruikers, etc. Tot slot is het noodzakelijk om in Sectie 2.3 enkele deontologische afwegingen te maken.

2.1 Methodologie

Technisch gezien is het niet zo moeilijk om een corpus te maken op basis van de conversaties in een chatkanaal. De meeste chatprogramma's voorzien de functionaliteit om een log bij te houden van het kanaal waarop men aangemeld is. Met behulp van die functionaliteit is het mogelijk om snel en eenvoudig relatief grote hoeveelheden chatconversaties te verzamelen. Voor het Moroccorp maakten we gebruik van het open-source programma *irssi*.³ Deze methode wordt algemeen gevolgd bij het maken van chatcorpora, zoals bijvoorbeeld in het *Dortmunder Chat Corpus*.

De chatconversaties in het Moroccorp zijn afkomstig uit twee chatkanalen. Het eerste chatkanaal heet *#maroc* en is bereikbaar via de server *irc.marocchat.net*. Het tweede chatkanaal heet *#maroc.nl* en is bereikbaar via de server *irc.scarynet.org*. Met behulp van de automatische logfunctionaliteit in het gebruikte chatprogramma konden we de conversaties in deze chatkanalen gemakkelijk opslaan. Het Moroccorp is gebaseerd op de logs van beide chatkanalen uit de zomer van 2012 (van mei 2012 tot en met september 2012). In mei en in juni werd het eerste chatkanaal gelogd, en van juli tot en met september werd het tweede chatkanaal gelogd. Onze logactiviteiten werden geregeld onderbroken uit technische overwegingen en om ervoor te zorgen dat onze passieve aanwezigheid op het chatkanaal niet zou opvallen — meer hierover in Sectie 2.3.

Beide chatkanalen hebben het Nederlands als voertaal, en richten zich

tegelijkertijd tot een publiek met Marokkaanse achtergrond. Hoewel de chatkanalen in principe ook toegankelijk zijn voor niet-Marokkaanse Nederlanders kunnen we er door de specificiteit van de chatkanalen van uitgaan dat het leeuwendeel van de chatters van Marokkaanse origine is. Voor die assumptie zijn er enkele kwalitatieve aanwijzingen. Zo verwijzen vele gebruikers in hun gebruikersnaam naar een Marokkaanse identiteit, bv. "vrouwuitmarokka", "ZeTLa-Maroc", "Femmedumaroc", "safouan_maroc", "maroc-meid". Daarnaast bevatten bijna 22.000 lijnen uit het chatcorpus (ongeveer 1% van alle lijnen in het corpus) een referentie naar Marokko, wat aantoont dat dit een belangrijk onderwerp is voor de chatters. Principieel sluit dit niet uit dat er hier gechat wordt over, zeg maar, Marokko als vakantieland, maar zelfs een eenvoudige vluchtige kijk in het Moroccorp doet anders vermoeden. Verdere aanwijzingen dat het hier gaat over het Nederlands van taalgebruikers met een Marokkaanse achtergrond halen we uit de lexicale analyse en de morfosyntactische analyse in Sectie 3. Uiteraard is het niet uit te sluiten dat er ook enkele niet-Marokkaanse taalgebruikers op het (openbare) chatkanaal actief zijn. Een aanwijzing hiervoor vinden we in Fragment 1, waarin chatter "TurksDraakje31" in zijn pseudoniem een expliciet niet-Marokkaanse achtergrond uitdrukt. Dat zien we niet als een probleem, aangezien een zekere hoeveelheid ruis in elk corpus onvermijdbaar is. Zo bevatten bijvoorbeeld ook klassieke krantencorpora vaak niet-journalistieke teksten: reclame, lezersbrieven, aankondigingen enzovoort. In feite is een belangrijk doel van de taalkundige analyse in Sectie 3 het aantonen dat deze hoeveelheid ruis relatief beperkt is.

Daarnaast blijken beide chatkanalen opgebouwd te zijn rond een overzichtelijke groep van kernleden. Dit zorgt ervoor dat de conversatie bijna nooit chaotisch wordt — dus zonder verschillende gesprekken die door elkaar gevoerd worden. Dat is namelijk vaak een probleem in openbare chatkanalen, waar soms heel veel mensen tegelijkertijd chatten (Herring 2010). Echter, de vrij gebalanceerde tussenvorm in de door ons verzamelde chatkanalen, met nooit al te veel chatters tegelijkertijd en een goed te volgen conversatie, lijkt vanuit linguïstisch oogpunt bijzonder attractief: er is voldoende animositeit, waardoor de snelheid van communiceren (zoals in een druk chatkanaal) zijn invloed kan hebben op het taalgebruik, maar tegelijkertijd is de groep chatters niet zo groot dat er geen coherent 'gesprek' gevoerd kan worden. Dat maakt het Moroccorp, los van zijn mogelijke etnolinguïstische waarde, ook interessant voor conversatie-analyse.

Voorts zullen we tegen het einde van het artikel durven te beweren dat het Moroccorp representatief is voor Nederlandse chattaal van taalgebruikers met een Marokkaanse achtergrond (zoals die gesproken werd in de zomer van 2012), omdat dit — zover ons bekend is — de enige twee chatkanalen zijn waar de deelnemers overwegend van Marokkaanse origine zijn en in het Nederlands chatten. Dit neemt uiteraard niet weg dat er een kans bestaat dat er toch nog andere, misschien private chatkanalen zijn waarop Marokkaanse Nederlanders converseren. Desalniettemin lijkt ons de kwantiteit van tien miljoen woorden

chattaal gedurende een beperkte periode een aanwijzing voor de grote populariteit van de door ons opgenomen chatkanalen. Het is echter wel zo dat we enkel toegang hebben tot de conversaties in het publieke gedeelte van het chatkanaal. Beide chatkanalen geven de gebruikers immers ook de mogelijkheid om private gesprekken te voeren in een ad-hoc sub-chatkanaal. Uiteraard zijn die private gesprekken niet vrij toegankelijk en ze zijn ook niet opgenomen in het Moroccorp. Daarom is het goed om toe te voegen dat het Moroccorp enkel chattaal bevat van taalgebruikers die zich ervan bewust zijn dat ze zich op een publiek forum begeven. Tot slot moeten we ook nog opmerken dat het bij chatcorpora doorgaans onmogelijk is om precieze socio-demografische gegevens van de chatters te achterhalen. Zo kunnen we niet met zekerheid weten wat het sociale profiel is dat het Moroccorp beschrijft, inclusief de etniciteit van de chatters. Daarom wijdt dit artikel hieronder veel plaats aan het aannemelijk maken van de Marokkaanse etniciteit van de meerderheid van de chatters. Ondanks deze beperkingen, zijn we er van overtuigd dat het Moroccorp een nuttige bron voor het taalonderzoek kan zijn.

2.2 Enkele cijfers

In deze paragraaf belichten we het corpus van een technische en kwantitatieve kant.

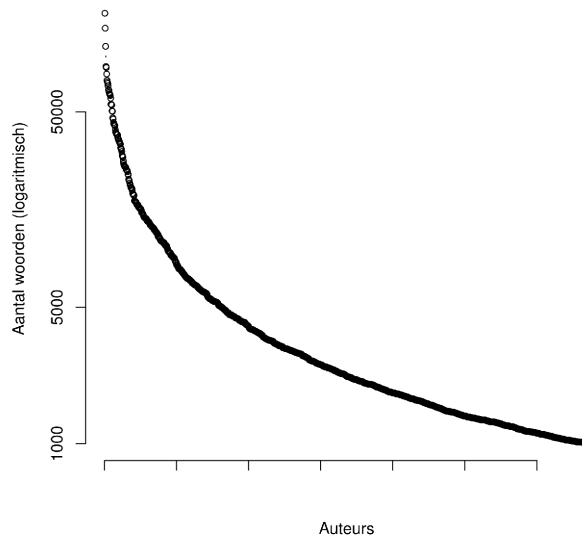
Tabel 1: Hoeveelheid woorden in het Moroccorp

chatkanaal	log	opgeschoond
marocchat.net:#maroc	5.542.949	2.739.330
scarynet.org:#maroc.nl	19.310.048	7.533.347
Totaal	24.852.997	10.272.677

Het volledige corpus zoals het werd gelogd in het chatprogramma telt bijna vijfentwintig miljoen woorden.⁴ Van het eerste chatkanaal hebben we vijf miljoen woorden in de logs, en van het tweede chatkanaal hebben we negentien miljoen woorden gelogd. Echter, de ruwe logs van het chatprogramma vereisen nog enkele opruimactiviteiten om de chatlogs in een bruikbaar corpus te veranderen. Ten eerste willen we boodschappen verwijderen die niet door mensen, maar door de computer gegenereerd zijn. Het is namelijk zo dat de software die nodig is om een chatkanaal uit te baten een heleboel automatische aankondigingen genereert. Daarnaast hebben de gebruikers van een chatkanaal de mogelijkheid om met behulp van zogenaamd *bots*⁵ de functionaliteit van het kanaal te vergroten. Deze teksten zijn formeel gemakkelijk te herkennen en kunnen dan ook eenvoudig en automatisch weggefilterd worden. Ten tweede bevat de ruwe chatlog informatie over de server waarop het chatkanaal zich

baseert, of over het chatprogramma waarmee een gebruiker zich heeft aangemeld, of de tijd waarop een berichtje de wereld werd ingestuurd. Deze meta-informatie wordt verwijderd zodat het corpus uiteindelijk regels bevat volgens het formaat *auteur {tab} boodschap*. Na deze opruimactiviteiten is het aantal woorden in het volledige corpus nog ruim tien miljoen (10.272.677). Met deze hoeveelheid woorden is het Moroccorp naar huidige standaarden niet zo groot, maar te vergelijken qua grootte met het Corpus Gesproken Nederlands (Nederlandse Taalunie 2004).

Figuur 1: Aantal woorden per auteur (enkel auteurs met meer dan 1000 woorden zijn hier afgebeeld)



Niet alle gebruikers dragen evenveel woorden bij aan het corpus, zoals we zien in Figuur 1. De grafiek uit Figuur 1 wordt iets tastbaarder als we de twintig *nicknames* (een soort van schuilnaam) die het grootste aantal woorden bij elkaar gechat hebben presenteren in Tabel 2. Het blijkt inderdaad dat er een beperkt aantal gebruikers zijn die een belangrijk deel van het chatcorpus vullen. De meest actieve gebruiker produceert maar liefst drie keer zoveel woorden als het nummer twintig van de productiefste chatters. Dat is in principe geen probleem voor de validiteit van het Moroccorp, maar wel voor de representativiteit: de gebruikers van het corpus kunnen niet zonder meer de bevindingen veralgemenen, en bij een kwantitatieve studie met inferentiële statistiek wordt het aangeraden om gebruik te maken van een *random effect* dat de specificiteit van iedere chatter ondervangt.

Tabel 2: De twintig meest productieve chatters

Nickname	Aantal woorden
Manal	160.258
koekje	134.301
Aketoef	108.481
lady_mamita	85.482
kipsate	84.435
groentebol	78.249
KYRA	72.585
nourdine	71.264
meisjepraatveel	70.069
awayagher	67.956
BijzondereMan	64.982
Zorro	64.620
Kanjer_	62.979
nabil_1984	61.912
Okegoed	61.422
Moslina21	60.688
TheMo	58.834
fessiaa	54.792
Fir3auwn	54.650
Dora	54.263

2.3 Deontologische afwegingen

Het compileren van een corpus van chatconversaties vereist dat de onderzoeker zich aanmeldt op het chatkanaal en zonder inmenging in de conversatie taal-materiaal verzamelt, teneinde de *observer's paradox* (Labov 1966) te ondervangen. Hoewel deze houding sociolinguïstisch methodologisch gebruikelijk is, kan men zich hierbij deontologische vragen stellen. In hoeverre is het ongemerkt 'opnemen' van spontaan taalgebruik zonder dat de taalgebruiker hiervan op de hoogte is aanvaardbaar? Daarnaast wordt het 'idlen' op een chatkanaal — aangemeld zijn zonder participatie — niet geapprecieerd.

Ter illustratie kunnen we een stukje chatconversatie weergeven van het moment waarop ontdekt wordt dat we al enkele dagen 'idlen'. Dit stukje is uiteraard ook opgenomen in het Moroccorp.

Fragment 1: Corpusmakers betraapt op chatloggen.

_adil_relax hij is de mol
...
_adil_relax undercover onderzoeker
...
_adil_relax kulnet.kuleuven.be
...
TurksDraakje31 denk dat ie zijn laptop vergete is dicht te klappe
TurksDraakje31 heb ik soms ook wel;))
...
_adil_relax 4 dagen vriend
_adil_relax bijna 5
TurksDraakje31 dalijk is ie dood ofso
_adil_relax zware chatlogger is dat
...
_adil_relax datamining
_adil_relax waar hebben we het over
_adil_relax ze zijn bezig met onderzoek
TurksDraakje31 weet ik veel
...
TurksDraakje31 zwaai dan ff
TurksDraakje31 misgien kom je in het nieuws:D
...
_adil_relax onderzoeker
_adil_relax undercover
...
_adil_relax research
...
_adil_relax ga een tageloud maken
TurksDraakje31 ja maar wrm kicke hem dan niet
...
_adil_relax "meest getype woorden lijst"
...
_adil_relax hahaha allemaal scheldwoorden

Uit dit fragment uit het eerste chatkanaal blijkt dat de chatters bekend zijn met het fenomeen chatloggen. Er ontstaat geen agitatie noch verontwaardiging, maar een zekere gelatenheid. De enige consequentie is dat onze aanwezigheid op het chatkanaal niet langer gewenst en toegelaten is. In Grondelaers et al. (2000) wordt daarom geadviseerd om op voorhand contact op te nemen met een verantwoordelijke van het chatkanaal. Wij hebben deze stap niet ondernomen omdat een chatkanaal per definitie openbaar is en het principieel niet nodig is om toestemming te vragen. Bovendien bestaat er duidelijkheid bij de chatters over de vrije toegankelijkheid van de chatkanalen. Het tweede chatkanaal plaatst zelfs een permanente boodschap bovenaan in het chatprogramma waarin wordt gewaarschuwd voor het ontbreken van privacy. Daarnaast hebben we geen enkele poging ondernomen om de chatconversaties te registreren die in een expliciet private context gevoerd werden. Aangezien de chatters hun eigen anonimiteit waarborgen door het gebruik van een pseudoniem, hebben we ook geen stappen ondernomen om het Moroccorp verder te anonimiseren.

Een andere deontologische overweging is dat etnolectisch corpusonderzoek een aantal risico's inhoudt, die gemakkelijk uit het oog verloren kunnen worden bij een (in het algemeen behartenswaardige) onbevungen aanpak. Dat heeft hiermee te maken dat het afgrenzen van etnolecten culturele implicaties heeft. Clyne (2000: 86) definieert etnolecten als "varieties of a language that mark speakers as members of ethnic groups who originally used another language or distinctive variety", en Androutsopoulos (2001: 2) hanteert de volgende definitie: "a variety of the majority language (or 'host language') which is used by and regarded as a vernacular for speakers of a particular ethnic descent and is marked by certain contact phenomena". Die definities zien er op het eerste gezicht vrij onschuldig uit, maar o.a. Jaspers (2008) wijst erop dat het afzonderen van een etnolect ideologische consequenties kan hebben door er al te vanzelfsprekend van uit te gaan dat taalgebruik in de eerste plaats aangestuurd wordt door etnie, en daardoor een onverantwoord homogeniserend, essentialistisch beeld op kan hangen van de sprekers, die daar als culturele minderheid hinder van kunnen ondervinden. Dat is natuurlijk niet onze bedoeling met het aanleggen van een corpus Marokkaans Nederlands, en we moeten bedacht zijn op dergelijke effecten. Toch kan geredelijk betwijfeld worden of we ons met het aanleggen van een corpus van een Marokkaans Nederlands etnolect schuldig maken aan cultureel essentialisme. Uit eerder onderzoek blijkt immers dat er zich in de taalfeiten wel een variëteit Marokkaans Nederlands laat onderscheiden (zie Nortier en Dorleijn 2008), en het lijkt ons interessant om collegae-onderzoekers de mogelijkheid te bieden die variëteit ook verder te onderzoeken in een wat groter tekstbestand (iets waar die collega's ook expliciet om vragen, zie Nortier en Dorleijn 2008: 140 en Hinskens 2011: 126). In principe biedt ons corpus ook de mogelijkheid om met taaldata in de hand de constructie van zo'n etnolect te falsifiëren. Als de eerder vastgestelde patronen in het taalgedrag van de sprekers van het etnolect niet stabiel zijn, of aanzienlijke variatie vertonen, of niet uniek zijn voor het etnolect, dan kan dat leiden tot een gemotiveerde revisie van de bestaande afbakening, maar niemand is gebaat bij een gebrek aan data. Ons corpus heeft alvast het voordeel dat het omvangrijker is dan wat er totnogtoe publiekelijk beschikbaar is. Met het gebruik van de term 'Marokkaans Nederlands' (cf. *infra*) en de naam van het corpus ('Moroccorp') hebben we geen ontologische pretenties, en we willen er in dit artikel verder ook uitdrukkelijk op wijzen dat het er ons niet om te doen is de variëteit in kwestie te bestempelen als 'onvolmaakt' of 'slecht' Nederlands — als die termen überhaupt al bruikbaar zouden zijn. Heel vaak zijn sprekers van een etnolect trouwens goed in staat de standaardtaal te spreken, maar zien ze daar in sommige contexten bewust van af (Hinskens 2011: 104). Trouwens, ook het standaardnederlands kan als een selectieve variëteit beschouwd worden, die geen usurperende claim kan leggen op het predicaat 'hét Nederlands'. Bovendien vestigen we nog even de nadruk op het feit dat wat standaardnederlands genoemd wordt, zelf de resultante is van verschillende periodes van diepgaand taalcontact in het verleden (zie Sectie 1, Inleiding), zodat je zou kunnen zeggen

dat elke variëteit van het Nederlands historisch gezien een etnolect is.

3. Hoe 'Marokkaans' is het Nederlands in het Morocccorp?

In deze sectie gaan we na hoe 'Marokkaans' het Nederlands is dat verzameld werd in het Morocccorp. Hiervoor zullen we twee taalkundige indicatoren gebruiken, namelijk een lexicale analyse en een morfologische analyse. Tegelijkertijd kunnen deze indicatoren aantonen hoe bruikbaar het Morocccorp is voor het taalkundige onderzoek.

Alvorens we tot de twee kleine gevalstudies kunnen overgaan moeten we bondig het Nederlands van taalgebruikers met een Marokkaanse achtergrond bespreken. Hoewel leden van de Marokkaanse gemeenschap in Nederland een verschillende moedertaalachtergrond hebben (Berber of Arabisch), en er uiteraard grote linguïstische verschillen zijn tussen eerste-, tweede- en ondertussen ook derde-generatie-Marokkanen, is het toch mogelijk een etnische variant van het Nederlands af te zonderen (zie El-Aissati et al. 2005). Die variant zal intern nog wel individuele verschillen vertonen, maar dat is bij de standaardtaal en bij sociolecten niet anders. Verschillende detailstudies en overzichtsartikelen hebben fonologische, lexicale en grammaticale kenmerken van dat Marokkaans Nederlands beschreven (El-Aissati et al. 2005; Nortier en Dorleijn 2008; Hinskens 2011). Een geschreven corpus zoals het Morocccorp, ook al ligt het zeer waarschijnlijk dichter bij de impromptu gesproken taal dan overvloedig geredigeerde journalistieke, wetenschappelijke of politieke teksten, is minder geschikt voor fonologisch onderzoek, maar op lexicaal en grammaticaal vlak is het corpus zeer wel bruikbaar. Hier moeten we uiteraard nog toevoegen dat het Morocccorp slechts een bepaald aspect van Marokkaans Nederlands belicht, namelijk het taalgebruik in de context van computer-gemedieerde communicatie.

Als eerste indicator van de Marokkaanse kleuring van het Nederlands in het Morocccorp voeren we een lexicale analyse uit. We proberen een gevoel te krijgen voor de inhoud en de kenmerken van de twee chatkanalen door op zoek te gaan naar woorden die typisch zijn voor de verzamelde chatkanalen. Een eenvoudige frequentielijst zoals in Tabel 3 is hiervoor niet bijzonder inzichtelijk, want we vinden enkel woorden terug die sowieso hoogfrequent zijn in het Nederlands. Dit versterkt uiteraard wel onze claim dat het Nederlands de voertaal is in de beide chatkanalen, en dat het taalgebruik in het Morocccorp een reële variëteit is van het Nederlands.

Om woorden te vinden die echt typisch zijn voor ons corpus maken we gebruik van de zogenaamde *Stable Lexical Marker analyse* uit Speelman et al. (2006, 2008) en De Hertog et al. (2013). Ruwweg vergelijkt deze techniek de woordfrequenties uit een doelcorpus met de woordfrequenties uit een referentiecorpus. Als de frequentie van een woord uit het doelcorpus statistisch gezien opvallend hoger is dan in het referentiecorpus, dan wordt dit woord een *Lexical Marker*. Als dat woord dan ook nog eens consequent frequenter is doorheen

willekeurige opdelingen van het doelcorpus, dan noemen we dit woord een *Stable Lexical Marker*. Als referentiecorpus nemen we hier de Nederlandse chatcomponent uit het ConDiv corpus (Grondelaers et al. 2000).⁶ De restrictie tot de Nederlandse chatcomponent (en dus de exclusie van de Vlaamse chatcomponent) is gemotiveerd door het bijzonder kleine aantal Vlamingen in de beide chatkanalen. De relatieve afwezigheid van Vlamingen hebben we vastgesteld met behulp van een telling van het aantal *ge*-vormen, een persoonlijk vornaamwoord voor de tweede persoon enkelvoud. Die *ge*-vorm is namelijk niet verspreid in Nederland. Het blijkt dat van de meerdere duizend chatters slechts ongeveer veertig chatters in het totaal ongeveer 100 *ge*-vormen produceren.

Tabel 3: Lijst van de twintig meest frequente tokens in het Moroccorp

Positie	Woord	Frequentie	Positie	Woord	Frequentie
1	je	290.435	11	van	79.986
2	ik	240.415	12	in	71.296
3	is	160.412	13	met	70.156
4	niet	128.978	14	ben	68.340
5	een	128.817	15	wat	64.696
6	de	118.019	16	jij	63.760
7	en	113.334	17	op	62.810
8	dat	101.814	18	maar	59.607
9	het	100.868	19	ze	56.332
10	die	84.419	20	heb	55.272

In Tabel 4 worden de resultaten van de *Stable Lexical Marker* analyse gepresenteerd. Technisch gesproken krijgt elk woord in een *Stable Lexical Marker* analyse een score toegekend, waarbij een sterk positieve score impliceert dat het woord opvallend frequent gebruikt wordt in het geteste corpus. Het zou echter niet inzichtelijk zijn om elk woord hier apart te representeren, en daarom hebben we de twintig hoogst scorende woorden hier in groepen gesorteerd.⁷ We vinden enkele woorden die duidelijk refereren aan Marokko en Marokkaans (groepen 1, 2 en 4), en ook een groep die verwijst naar de belangrijkste religie in Marokko (groep 6). Dat zijn lexicale domeinen die prominent zijn in het taalgebruik van Nederlanders met een Marokkaanse achtergrond, zoals El-Aissati et al. (2005: 174-175) opmerken — al kunnen die natuurlijk net zo goed voorkomen in het taalgebruik van andere Nederlanders wanneer ze het toevallig over deze onderwerpen hebben. Daarnaast zien we ook dat er enkele gebruikelijke Nederlandse chatafkortingen als markant worden aangeduid (groep 3), en enkele woorden die kunnen duiden op typische onder-

werpen in de chatkanalen (groep 5).

Tabel 4: Stable Lexical Marker analyse: groepering van de twintig meest typische woorden voor de Marokkaans-Nederlandse chatkanalen, in vergelijking met algemene Nederlandse chatkanalen

	Woorden	Uitleg
1	salaam, salam, wslm (wasalaam), ewa, beslama	Marokkaanse groetwoorden
2	marokkaanse, marokko, marokkanen	zelfreferenties
3	gwn (gewoon), wrm (waarom)	internet afkortingen
4	wollah (vloek), hmdl (hamdoullah)	Marokkaanse uitroepen
5	broeder, trouwen, dame, vader	topicwoorden
6	islam, moslim, ramadan, allah	religieuze termen

Deze lexicale analyse toont aan dat het Moroccorp een vorm van het Nederlands die aanleunt bij computer-gemedieerde communicatie (Tabel 4, groep 3), maar ook lexicale verwijzingen bevat die de veronderstelling dat de chatters een Marokkaanse achtergrond hebben bevestigt (Tabel 4, groep 1 en 4). Daarnaast wijzen ook de andere woordgroepen uit Tabel 4 op de in Marokko gangbare religie (groep 6). We beklemtonen dat deze lexicale analyse natuurlijk geen definitief uitsluitel geeft over de status van Marokkaans Nederlands als etnolect, maar wel onze overtuiging sterkt dat het Moroccorp het taalgebruik van Nederlanders met een Marokkaanse achtergrond beschrijft.

3.2 Morfologische analyse

Omdat tweede-generatie-Marokkanen het Nederlands meestal niet van huis uit hebben meegekregen, maar er doorgaans pas mee in aanraking zijn gekomen op de schoolbanken, vertoont hun taalgebruik hier en daar aspecten van wat met een technische term *fossilisatie* heet (Matras 2009: 75). Die fossilisatie neemt natuurlijk minder spectaculaire vormen aan dan die in het taalgebruik van eerste-generatie-sprekers, en het Nederlands op de chatkanalen is voor een ongetrainde blik vaak moeilijk te onderscheiden van dat van moedertaalsprekers, maar de omvang van het Moroccorp laat toe statistisch significante patronen aan te wijzen die onttrokken blijven aan het blote oog.

Een wel typisch te noemen verschijnsel in Marokkaans Nederlands, ook van die sprekers die vroeg met het Nederlands in aanraking gekomen zijn op school, betreft de adjectiefinflectie in het Nederlands. Het gaat om de inflectie in nominale constituenten van het type *een mooi verhaal*, waarin Marokkaans-Nederlandssprekenden de neiging vertonen 'onterecht' een flexie-e (sjwa) toe te

voegen. Als het Moroccorp dit patroon vertoont, zou hier opnieuw uit blijken dat het een waardevol instrument kan zijn voor de studie van het Nederlands door taalgebruikers met een Marokkaanse achtergrond.

Om te begrijpen wat hier precies aan de hand is, moet eerst iets gezegd worden over het eigenaardige systeem van adjectiefinflectie in het hedendaags Nederlands. Dat systeem kan in zijn meest eenvoudige vorm weergegeven worden als in (1), de talrijke kleine uitzonderingen en de regionale variatie niet te na gesproken in het Nederlands (zie daarvoor onder andere Blom 1994; Weerman 2003; Tummers et al. 2004, 2005; Plevoets et al. 2009).

- (1) (a) predicatief gebruik: adj- \emptyset (*dat boek is moeilijk*)
- (b) attributief gebruik: adj- \emptyset (*het moeilijk-e boek*)
- (b') behalve: [+sg -def +neutr] NPs: adj- \emptyset (*een moeilijk boek*)

Het systeem is syntactisch-semantic slecht gemotiveerd. Dat komt door de 'behalve'-clausule in (b'). Door die weg te halen ontstaat een relatief transparant, gemotiveerd systeem waarbij afwezigheid van inflectie predicatief gebruik markeert — daarmee aansluiting zoekend bij de adverbialia (zie Diepeveen en Van de Velde 2010) — terwijl de inflectionele sjwa attributief gebruik markeert. Door (b') te schrappen krijgt de inflectionele sjwa bij adjectieven met andere woorden een signifié. Het is precies deze vereenvoudiging die massaal toegepast wordt door (jonge) L2-leerders van het Nederlands (zie Ziemann et al. 2011), en met name ook Marokkaans-Nederlandssprekenden (zie Blom et al. 2008). Het resultaat is het systeem in (2) (zie ook Van de Velde en Weerman Te verschijnen 2014).

- (2) (a) predicatief gebruik: adj- \emptyset (*dat boek is moeilijk*)
- (b) attributief gebruik: adj- \emptyset (*het moeilijk-e boek, een moeilijk-e boek*)

De representativiteit van het Moroccorp kan nu afgelezen worden aan de mate waarin het dit soort sjibboletten bevat. Om nu te kijken of het corpus deze trend laat zien moeten we twee dingen doen. Allereerst moeten we op zoek gaan naar het patroon. Dat is moeilijker dan het lijkt: het aantal adjectieven en nomina dat in deze constructie kan voorkomen is groot, en zoeken in een niet-syntactisch geannoteerd corpus leidt dan bijna onvermijdelijk tot een hoop valse treffers, die manueel gefilterd moeten worden. Dit klemt te meer omdat onzijdige woorden veruit in de minderheid zijn in het Nederlands (Ziemann et al. 2011: 185), terwijl we daar net naar op zoek zijn. Uit praktische overwegingen is daarom gekozen een steekproef te doen bij vijf frequente onzijdige woorden, te weten: *verhaal*, *boek*, *land*, *onderscheid* en *verschil*.⁸ Concreet is er gezocht naar patronen van een onbepaald lidwoord (*een*) gevolgd door een willekeurig woord, gevolgd door een van deze vijf woorden. Dat levert, na het uitwieden van de irrelevante hits (b.v. *een harrypotter boek* (samenstelling), *een eigen land* (principeel onverbuigbaar adjectief) etc.) een verdeling op van geflecteerde versus ongeflecteerde vormen. De tweede stap is om deze verde-

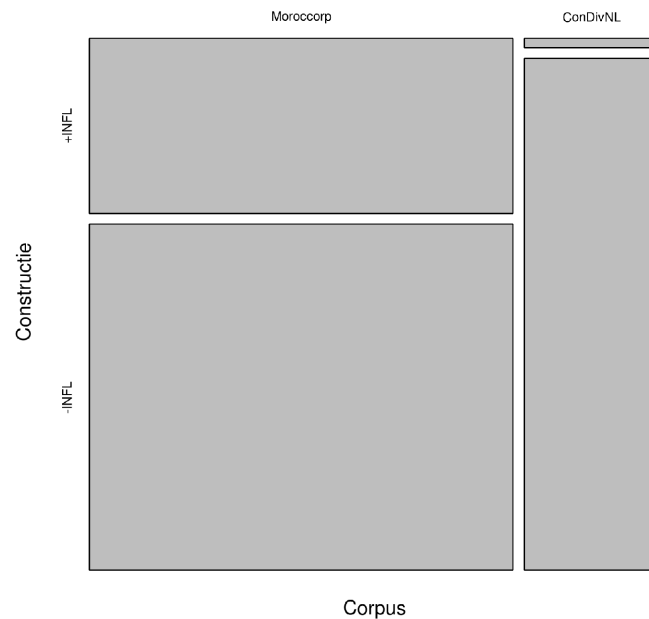
ling te vergelijken met een controlegroep. In principe is het immers best mogelijk dat het Moroccorp veel correct geflecteerde patronen bevat: het is beslist niet zo dat alle sprekers van het Marokkaans Nederlands altijd de 'onterecht' geflecteerde variant gebruiken, en principieel is het ook mogelijk dat niet-Marokkaanse Nederlandstaligen actief zijn op het chatkanaal. De vraag is daarom of de informanten in het Moroccorp de gemarkeerde vorm significant vaker gebruiken dan eentalig-Nederlandse chatters. Voor die controlegroep moeten we ons dus wenden tot een vergelijkbaar corpus. Daarvoor zijn we te rade gegaan bij het Nederlandse IRC-materiaal in het ConDiv-corpus — ongeveer 7 miljoen tokens in totaal (Grondelaers et al. 2000). Qua genre is dit een bijna perfecte match met het Moroccorp, en het lijkt aannemelijk dat er ook een grote gelijkenis is qua sociaal profiel. Een mogelijke storende factor, die we echter niet verder onderzoeken, zou kunnen worden gevonden in de toch wel tien jaar die ligt tussen het verzamelen van het ConDiv corpus en het Moroccorp. Uit de resultaten, weergegeven in Tabel 5, blijkt dat er een merkbaar verschil is tussen Marokkaans Nederlands en eentalig Nederlands (zie Figuur 2 voor een visualisatie in een mozaïekgrafiek). In het eentalige Nederlands komt de constructie eigenlijk niet voor, terwijl die in het Moroccorp in een derde van de gevallen gebruikt wordt. De associatie is statistisch significant.

Tabel 5: Verdeling van geflecteerde en ongeflecteerde adjectieven bij onzijdige nomina na een onbepaald lidwoord. (Chi-kwadraat, met Yates' continuity correction = 20,6927, $vg = 1$, p-waarde < 0,001)

	+INFL	-INFL
	<i>(een mooie verhaal)</i>	<i>(een mooi verhaal)</i>
bron:moroccorp	59	117
bron:condivnl	1	55

We merken bij deze kwantitatieve resultaten op dat de gemarkeerde +INFL vorm geen noodzakelijke en voldoende voorwaarde is om een bepaalde vorm van het Nederlands als 'Marokkaans' te bestempelen. Aan de ene kant komt de +INFL vorm ook (zeer beperkt) voor in het algemene corpus (ConDivNL), en aan de andere kant komt ook de ongemarkeerde -INFL vorm voor in het Moroccorp. Meer nog, de ongemarkeerde vorm blijkt duidelijk in de meerderheid. Dit toont aan dat zelfs een wijdverbreid stereotiep kenmerk van Marokkaans Nederlands op zichzelf wellicht niet voldoende is om deze vorm van Nederlands eenduidig af te zonderen als een etnolect. Een corpus van voldoende grote omvang, dat op een verantwoorde manier onderzocht wordt met behulp van statistische technieken, kan niettemin op een genuanceerde manier patronen blootleggen. En die patronen moeten verklaard worden. Het Moroccorp biedt die mogelijkheid.

Figuur 2: Mozaïekgrafiek van geflecteerde en ongeflecteerde adjectieven bij onzijdige nomina na een onbepaald lidwoord



4. Conclusie

Tot slot van dit artikel vatten we bondig de kenmerken van het Moroccorp samen. Het corpus heeft het taalgebruik opgenomen van een gemeenschap van (waarschijnlijk overwegend) Nederlanders met Marokkaanse afkomst. Het taalgebruik is beperkt tot conversaties in openbare chatkanalen waarin de voertaal het Nederlands is. Gedurende enkele maanden hebben we de conversaties in twee chatkanalen opgenomen totdat het corpus ongeveer tien miljoen woorden bevatte. De twee gevolgde chatkanalen zijn specifiek gericht op Marokkaanse Nederlanders, en hoogstwaarschijnlijk de enige van zulke aard die openbaar toegankelijk zijn. We durven daarom beweren dat het Moroccorp representatief is voor het Nederlands dat gebruikt wordt door Marokkaanse Nederlanders in openbare chatgroepen. Het Moroccorp bevat geen private berichten, en biedt daarom enkel inzicht in het taalgebruik van een groep mensen die zich bewust zijn van hun openbaarheid. Aangezien de chatters hun eigen anonimiteit waarborgen door het gebruik van een pseudoniem, hebben we geen stappen ondernomen om het Moroccorp verder te anonimiseren.

De omvang van het Moroccorp is ruim tien miljoen woorden, waarmee het een vergelijkbare grootte heeft als het Corpus Gesproken Nederlands. Hoewel het corpus voor een verzameling van geschreven taal relatief klein is, kun-

nen we erop wijzen dat het eenvoudig is om dit corpus verder uit te breiden. Van de opgenomen tien miljoen woorden werden een miljoen woorden geproduceerd door slechts dertien chatters. Dit wijst op een sterke Zipfianse curve (Figuur 1, waarin slechts een klein percentage van de taalgebruikers verantwoordelijk is voor de meerderheid van de opgenomen woorden. Aan de ene kant betekent dit dat de hoeveelheid taalgebruikers waarvoor we een voldoende grote hoeveelheid woorden hebben maar klein is, zodat de mogelijkheden tot veralgemening van het corpus niet overschat mogen worden. Aan de andere kant geeft dit ook aan dat de gebruikers van de chatkanalen een hechte gemeenschap vormen, waardoor het mogelijk wordt om ook fijnmazigere conversatie-analyse te doen. Ook voor andere veelgebruikte corpora (b.v. krantenmateriaal) geldt overigens dat ze relatief veel tekst van relatief weinig auteurs of sprekers exciperen.

Uit een kwalitatieve en kwantitatieve analyse van de chatconversaties blijkt dat de voertaal in de kanalen eenduidig Nederlands is, maar dat de uitingen van de gebruikers doorspekt zijn met Berberse of Arabische kenmerken (zie ook Boumans 2002). Op het niveau van het woordgebruik valt op dat typische Arabische en Berberse sjibboletten gebruikt worden in begroetingen en uitroepen. Het woordgebruik, zoals dat geanalyseerd werd in de *Stable Lexical Marker analyse*, verraaft ook dat de conversaties vaak over Marokko, Islam en religieuze tradities gaan. In dit artikel hebben we naast een lexicale analyse ook een stereotiep morfosyntactische kenmerk van zogenaamd 'Marokkaans Nederlands' besproken, namelijk de adjectiefflectie bij onzijdige nomina na een onbepaald lidwoord. Het blijkt dat de gemarkeerde vorm, die als kenmerkend voor Marokkaans Nederlands wordt beschouwd, statistisch significant meer voorkomt in het Moroccorp dan in een algemeen chatcorpus. Hoewel dit geen uitsluitel geeft over de al dan niet etnolectische status van Marokkaans Nederlands — iets waar dit artikel ook geen definitieve uitspraak over wil doen — sterken deze waarnemingen ons vertrouwen in de bruikbaarheid van het Moroccorp als een waardevolle bron voor onderzoek naar Nederlands van taalgebruikers met een Marokkaanse achtergrond. Het corpus is publiek beschikbaar voor onderzoeksdoeleinden. Wie het hebben wil, kan contact opnemen met de auteurs van dit artikel.

Eindnoten

1. De grafieken en de statistische analyses in dit artikel zijn uitgevoerd met het software pakket R. (R Core Team. 2012. R: A language and environment for statistical computing. Vienna. <http://www.R-project.org>).
2. <http://www.mpi.nl/resources/data>
3. <http://www.irssi.org>
4. De woordtelling in de logs werd uitgevoerd met het UNIX programma *wc*. Het precieze aantal woorden in de logs wijkt waarschijnlijk licht af van de gerapporteerde aantallen omdat de telling gebaseerd is op een bijzonder ruwe tokenisering. De woordtelling in de opgeschoonde versie is uiteraard wel accurater.

5. *Bots* zijn kleine programmaatjes die automatische boodschappen weergeven, zoals bijvoorbeeld quizvragen of, in het geval van deze chatkanalen, de gebedstijden per locatie.
6. De *Stable Lexical Marker analyse* werd hier uitgevoerd op woorden die minstens 30 keer voorkomen in elke opdeling van het doel- en referentiecorpus.
7. We hebben enkele spellingsvarianten, smileys en gebruikersnamen genegeerd.
8. Dit zijn niet de meest frequente onzijdige woorden. De reden daarvoor is dat woorden zoals *weer* en *haar* heel frequent zijn, maar ook homoniem zijn met functiewoorden, en daar hun frequentie aan te danken hebben. We hebben dan ook gekozen voor woorden die redelijk frequent zijn, maar eenduidiger. Volledigheidshalve geven we hier nog even de rang van de gekozen woorden in het corpus aan: *boek* (frequentie 778, rang 1724/333114); *land* (frequentie 1635, rang 943/333114); *onderscheid* (frequentie 96, rang 8202/333114); *verhaal* (frequentie 913, rang 1528/333114); *verschil* (frequentie 796, rang 1699/333114).

Literatuuropgave

- Androutsopoulos, J.** 2001. *From the Streets to the Screens and Back Again: On the Mediated Diffusion of Variation Patterns in Contemporary German*. Essen: LAUD.
- Androutsopoulos, J.** 2006. Introduction: Sociolinguistics and Computer-mediated Communication. *Journal of Sociolinguistics* 10(4): 419-438.
- Appel, R.** 1999. Straattaal. De mengtaal van jongeren in Amsterdam. *Toegepaste taalwetenschap in artikelen* 62(2): 39-56.
- Blom, A.** 1994. Het ondoorgroendelijk bijvoeglijk naamwoord. *Forum der Letteren* 35(2): 81-94.
- Blom, E., D. Polišenská en F. Weerman.** 2008. Articles, Adjectives and Age of Onset: The Acquisition of Dutch Grammatical Gender. *Second Language Research* 24: 297-331.
- Boumans, L.** 2002. Meertaligheid op de Marokkaanse elektronische prikborden. *Levende Talen* 3: 11-21.
- Buccini, A.** 1995. Ontstaan en vroegste ontwikkeling van het Nederlandse taallandschap. *Taal en Tongval. Themanummer* 8: 8-66.
- Buccini, A.** 2010. Between Pre-German and Pre-English: The Origin of Dutch. *Journal of Germanic Linguistics* 22(4): 301-314.
- Clyne, M.** 2000. Lingua Franca and Ethnolects in Europe and Beyond. *Sociolinguistica* 14: 83-89.
- Cornips, L. en V. de Rooij.** 2003. Kijk, Levi's is een goeie merk: maar toch hadden ze 'm gedist van je schoenen doen 'm niet. Stroop, J. (Red.). 2003. *Waar gaat het Nederlands naartoe? Panorama van een taal*: 131-142. Amsterdam: Bert Bakker.
- Croft, W.** 2000. *Explaining Language Change: An Evolutionary Approach*. Harlow: Longman.
- Crystal, D.** 2011. *Internet Linguistics*. London: Routledge.
- De Hertog, D., K. Heylen en D. Speelman.** 2013. Stable Lexical Marker Analysis: A Corpus-based Identification of Lexical Variation. Soares da Silva, A. (Red.). 2013. *Pluricentricity: Language Variation and Sociocognitive Dimensions*: 117-130. Braga: Aletheia.
- Diepeveen, J. en F. Van de Velde.** 2010. Adverbial Morphology: How Dutch and German Are Moving Away from English. *Journal of Germanic Linguistics* 22: 381-402.
- Drinka, B.** 2010. Language Contact. Luraghi, S. en V. Bubenik (Reds.). 2010. *Continuum Companion to Historical Linguistics*: 325-345. Londen: Continuum.
- El-Aissati, A., L. Boumans, L. Cornips, M. Dorleijn en J. Nortier.** 2005. Turks- en Marokkaans Nederlands. Van der Sijs, N. (Red.). 2005. *Wereldnederlands. Oude en jonge variëteiten van het*

- Nederlands*: 149-183. Den Haag: Sdu.
- Geeraerts, D.** 2002. Rationalisme en nationalisme in de Vlaamse taalpolitiek. De Caluwé, J., D. Geeraerts, S. Kroon, V. Mamadouh, R. Soetaert, L. Top en T. Vallen (Reds.). 2002. *Taalvariatie en taalbeleid. Bijdragen aan het taalbeleid in Nederland en Vlaanderen*: 87-104. Antwerpen: Garant.
- Grondelaers, S., K. Deygers, H. van Aken, V. van den Heede en D. Speelman.** 2000. Het ConDiv-corpus geschreven Nederlands. *Nederlandse Taalkunde* 5: 356-363.
- Harris, A. en L. Campbell.** 1995. *Historical Syntax in Cross-linguistic Perspective*. Cambridge: Cambridge University Press.
- Hawkins, J.A.** 1990. Germanic Languages. Comrie, B. (Red.). 1990. *The Major Languages of Western Europe*: 58-66. Londen: Routledge.
- Heine, B. en T. Kuteva.** 2003. On Contact-induced Grammaticalization. *Studies in Language* 27: 529-572.
- Heine, B. en T. Kuteva.** 2005. *Language Contact and Grammatical Change*. Cambridge: Cambridge University Press.
- Herring, S.C.** 2010. Who's Got the Floor in Computer-mediated Conversation? Edelsky's Gender Patterns Revisited. *Language@Internet* 7, article 8.
- Hickey, R.** 2010. *The Handbook of Language Contact*. Oxford: Blackwell.
- Hinskens, F.** 2011. Emerging Moroccan and Turkish Varieties of Dutch: Ethnolects or Ethnic Styles? Kern, F. and M. Selting (Reds.). 2011. *Ethnic Styles of Speaking in European Metropolitan Areas*: 101-129. Amsterdam/Philadelphia: John Benjamins.
- Jaspers, J.** 2004. *Tegenwerken, belachelijk doen: talige sabotage van Marokkaanse jongens op een Antwerpse middelbare school. Een sociolinguïstische etnografie*. Doctorale proefschrift. Antwerpen: Universiteit van Antwerpen.
- Jaspers, J.** 2008. Problematizing Ethnolects: Naming Linguistic Practices in an Antwerp Secondary School. *International Journal of Bilingualism* 12(1-2): 85-103.
- Labov, W.** 1966. The Social Stratification of (r) in New York City Department Stores. Labov, W. (Red.). 1966. *The Social Stratification of English in New York City*: 63-89. Washington, D.C.: Center for Applied Linguistics.
- MacWhinney, B.** 2000. *The CHILDES Project: Tools for Analyzing Talk. Volume 1: Transcription Format and Programs. Volume 2: The Database*. Mahwah, N.J.: Lawrence Erlbaum.
- Matras, Y.** 2009. *Language Contact*. Cambridge: Cambridge University Press.
- Nederlandse Taalunie.** 2004. *Corpus Gesproken Nederlands*. via TST-centrale: <http://tst-centrale.org/nl/producten/corpora/corpus-gesproken-nederlands/6-17>.
- Nortier, J. en M. Dorleijn.** 2008. A Moroccan Accent in Dutch: A Sociocultural Style Restricted to the Moroccan Community? *International Journal of Bilingualism* 12(1-2): 125-142.
- Plevoets, K.** 2008. Tussen spreek- en standaardtaal. Een corpusgebaseerd onderzoek naar de situationele, regionale en sociale verspreiding van enkele morfosyntactische verschijnselen uit het gesproken Belgisch-Nederlands. Doctorale proefschrift. Leuven: Katholieke Universiteit Leuven.
- Plevoets, K., D. Speelman en D. Geeraerts.** 2009. De verspreiding van de -e(n)-uitgang in attributieve positie. *Taal en Tongval* 22: 112-143.
- Ramaekers, W.** 1998. Mi, maak me geen eiers! Het Algemeen Cités. *Onze Taal* 67(4): 94-95.
- Roberge, P.** 2010. Contact and the History of Germanic Languages. Hickey, R. (Red.). 2010. *The Handbook of Language Contact*: 406-431. Oxford: Wiley-Blackwell.
- Ruetten, T., D. Speelman en D. Geeraerts.** 2014. Semantic Weighting Mechanisms in Scalable Lexi-

- cal Sociolectometry. Szmrecsanyi, B. and B. Waelchli (Eds.). 2014. *Aggregating Dialectology, Typology, and Register Analysis. Linguistic Variation in Text and Speech*: 178-198. Berlin: De Gruyter.
- Speelman, D., S. Grondelaers en D. Geeraerts.** 2006. A Profile-based Calculation of Region and Register Variation. Wilson, A., D. Archer and P. Rayson (Eds.). 2006. *Corpus Linguistics around the World*: 181-194. Amsterdam/New York: Rodopi.
- Speelman, D., S. Grondelaers en D. Geeraerts.** 2008. Variation in the Choice of Adjectives in the Two Main National Varieties of Dutch. Geeraerts, D., G. Kristiansen en Y. Peirsman (Eds.). 2008. *Cognitive Sociolinguistics: Language Variation, Cultural Models, Social Systems*: 205-233. Berlin/New York: Mouton de Gruyter.
- Stroop, J.** 1998. *Poldernederlands. Waardoor het ABN verdwijnt*. Amsterdam: Bert Bakker.
- Taeldeman, J.** 2008. Zich stabiliserende grammaticale kenmerken in Vlaamse tussentaal. *Taal en Tongval* 60(1): 26-50.
- Tagliamonte, S. en D. Denis.** 2008. Linguistic Ruin? LOL! Instant Messaging and Teen Language. *American Speech* 83(1): 3-34.
- Thomason, S.G.** 2001. *Language Contact: An Introduction*. Edinburgh: Edinburgh University Press.
- Thomason, S.G. en T. Kaufman.** 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.
- Tummers, J., D. Speelman en D. Geeraerts.** 2004. Quantifying Semantic Effects. The Impact of Lexical Collocations on the Inflectional Variation of Dutch Attributive Adjectives. Purnelle, G., C. Fairon en A. Dister (Eds.). 2004. *Le poids des mots*: 1079-1088. Louvain-la-Neuve: Presses Universitaires de Louvain.
- Tummers, J., D. Speelman en D. Geeraerts.** 2005. Inflectional Variation in Belgian and Netherlandic Dutch: A Usage-based Account of the Adjectival Inflection. Delbecq, N., J. van der Auwera en D. Geeraerts (Eds.). 2005. *Perspectives on Variation. Sociolinguistic, Historical, Comparative*: 93-110. Berlin: Mouton de Gruyter.
- Vandekerckhove, R. en J. Nobels.** 2010. Destandaardisatie en toe-eigening van schrijftaal. De chatcommunicatie van Vlaamse jongeren. Van der Wal, M. en E. Francken (Eds.). 2010. *Standaardtalen in beweging*: 173-191. Münster: Nodus Publikationen.
- Van de Velde, F. en F. Weerman.** Te verschijnen 2014. The Resilient Nature of Adjectival Inflection in Dutch. Sleeman P., F. Van de Velde en H. Perridon (Eds.). Te verschijnen 2014. *The Adjective in Germanic and Romance*. Amsterdam: John Benjamins.
- Weerman, F.** 2003. Een mooie verhaal; veranderingen in uitgangen. Stroop, J. (Ed.). 2003. *Waar gaat het Nederlands naartoe? Panorama van een taal*: 249-260. Amsterdam: Bert Bakker.
- Weinreich, U.** 1953. *Languages in Contact. Findings and Problems*. New York: Publications of the Linguistic Circle of New York.
- Xiao, Z.** 2008. Well-known and Influential Corpora. Lüdeling, A. en M. Kyto (Eds.). 2008. *Corpus Linguistics: An International Handbook*: 383-457. Berlin: Mouton de Gruyter.
- Zeige, L.E.** Te verschijnen. On Cognition and Communication in Usage-based Models of Language Change. Mengden, F. en E. Cousse (Eds.). Te verschijnen. *Usage-based Approaches to Language Change*. Amsterdam/Philadelphia: John Benjamins.
- Ziemann, H., F. Weerman en E. Ruigendijk.** 2011. Nederlands later geleerd: gebruik van lidwoorden en flexie van bijvoeglijke naamwoorden door duitstalige kinderen en volwassenen. *Internationale Neerlandistiek* 49(3): 183-207.

Lexicographie et enseignement de langues liées à l'immigration: note sur le contexte sociolinguistique de l'élaboration du dictionnaire finnois–lingala–français*

Kadima Batumona Adi Gilbert, *Centre de recherche pour l'instrumentalisation du cilubà (CRIC / Finlande), Centre de linguistique théorique et appliquée (CELTA / RD-Congo), Ghent University, Ghent, Belgium (adi_kadima@ymail.com)*

Les étrangers, y compris les migrants ont le droit de conserver leur langue maternelle, leur culture et leurs traditions.

Résumé: Le dictionnaire finnois–lingala–français dont il est question dans cette étude a été rédigé dans un contexte d'immigration. Il est l'illustration-même du soutien apporté aux migrants par leurs pays d'accueil en vertu du droit à la langue reconnu à tout être humain et au nom du droit international instituant les droits culturels des migrants. Notre contribution à cette thématique tout en se voulant linguistique, aborde quelques aspects de ces droits culturels des migrants en ce que leur reconnaissance a occasionné une activité lexicographique portant sur le lingala, langue dont la présence en Finlande ne se justifie que par l'immigration.

Mots-clés: LEXICOGRAPHIE, ENSEIGNEMENT DE LA LANGUE MATERNELLE, IMMIGRATION, DROITS DE L'HOMME, DROITS CULTURELS DES MIGRANTS, PLURILINGUISME, DROIT À LA LANGUE

Abstract: Lexicography and Language Teaching Related to Immigration: Note on the Sociolinguistic Context of the Development of the Finnish–Lingala–French Dictionary. The Finnish–Lingala–French Dictionary discussed in this article was compiled within the context of immigration. It illustrates the support for migrants in their host country under the right to language recognized for every human being and under international law establishing the cultural rights of migrants. Our contribution to this issue, intended to be linguistic, addresses some aspects of the cultural rights of migrants whereby their recognition has resulted in a lexicographical activity concerning Lingala whose presence in Finland can only be explained because of immigration.

* Texte remanié et mis à jour de la communication intitulée *Les problèmes posés par l'élaboration du dictionnaire finnois–lingala–français*, présentée au XVIIème Congrès des Romanistes Scandinaves, Université de Tampere, Finlande, 12–15 août 2008.

Keywords: LEXICOGRAPHY, LANGUAGE TEACHING, IMMIGRATION, HUMAN RIGHTS, CULTURAL RIGHTS OF MIGRANTS, MULTILINGUALISM, RIGHT TO LANGUAGE

1. Introduction

1.1 Objet de l'étude et objectif visé

L'intitulé du présent article nécessite quelque précision ou délimitation, car il laisse penser que l'objet de notre contribution consisterait en la description de la manière dont notre dictionnaire est utilisé en Finlande dans l'enseignement. Ce qui est une des possibilités, à première vue. Mais en nous basant sur la formulation de son sous-titre, son objet se laisse également saisir comme consistant en la description d'une part, du contexte sociolinguistique dans lequel ce dictionnaire finnois–lingala–français, destiné à être utilisé comme outil d'appui à l'apprentissage a vu le jour et d'autre part, à démontrer que l'immigration en tant que phénomène social, est à même de générer une activité lexicographique ou une dictionnaire.

En effet, ce dictionnaire qui a vu le jour dans ce particulier contexte d'immigration, nous sert d'exemple-type du soutien apporté aux migrants par les municipalités de leurs pays d'accueil et cela, au nom du droit international recommandant la reconnaissance ainsi que le respect de leurs droits culturels.

Mais notre contribution à cette dernière thématique demeure donc essentiellement linguistique.

1.2 Migration et langues liées à l'immigration (LLI)

La disposition de l'article 5.1.f de la Déclaration sur les droits de l'homme des personnes qui ne possèdent pas la nationalité du pays dans lequel elles vivent couvre juridiquement les migrants et résume à elle seule l'essentiel de leurs droits culturels. C'est au nom de ces droits culturels que dans beaucoup de pays européens les migrants se voient accordés certains droits spécifiques liés à leurs origines ainsi qu'à leurs langues et soutenus dans la pratique ainsi que dans l'enseignement de langues venant de leurs pays d'origine (ONU 1984).

D'ordinaire, les personnes qui émigrent à l'étranger prennent avec elles souvent quelques-uns de leurs effets personnels jugés importants, c'est-à-dire des bagages de première nécessité. Ces personnes soupçonnent rarement que tandis qu'elles émigrent, elles emportent également avec elles d'autres bagages plus importants: leurs langues maternelles, véhicules de leurs cultures et port d'attache à leurs origines traditionnelles!

Ces langues, étant donné qu'en Afrique, de nombreux pouvoirs publics ne les aménagent le plus souvent pas en vue de leur instrumentalisation, entretien ou équipement, doivent leur survie seulement grâce au fait qu'elles servent de moyen de communication au sein d'une même communauté linguistique ou entre les personnes issues de communautés linguistiques différentes. C'est par-

tant de cette situation de délaissement total que ces langues se retrouvent aujourd'hui parmi celles que les spécialistes de politique linguistique s'accordent à qualifier de *langues peu dotées*, ce qui veut dire qu'elles figurent parmi celles des moins outillées au monde.

Mais, une fois déballées sur le vieux continent, en vertu du principe de *respect aux droits culturels des migrants* ou celui du *droit à la langue maternelle*, ces *langues délaissées* se voient soudainement protégées et acquièrent des statuts qu'elles n'ont pas dans leurs milieux naturels, retiennent l'attention des pouvoirs publics des pays ayant accueilli leurs locuteurs et deviennent des matières potentielles d'enseignement pour les enfants des migrants. C'est dans ce contexte que diverses terminologies ont vu le jour en Europe et plus particulièrement dans les pays nordiques scandinaves pour désigner et différencier ces langues de *langues européennes: langues liées à l'immigration, langues d'origine, langues d'Europe, etc.*, terminologies que nous définirons plus loin.

1.3 Les LLI face à complexité linguistique de l'Europe

Nombreux parmi nous sont des personnes d'origine immigrée qui se souviennent de cette phrase: «L'Europe ne peut pas accueillir toutes les misères du monde», devenue depuis lors des plus célèbres et dont l'auteur n'est plus à présenter. Cette phrase revient le plus souvent dans les discours de certains leaders politiques dont nombreux se recrutent de l'extrême-droite et est utilisée aux fins de traduire l'hostilité ou la xénophobie que manifestent sur le territoire européen ces leaders face aux entrées effrénées, parfois massives et souvent illégales des populations étrangères.

Parallèlement à ces discours hostiles ou xénophobes, l'Europe ou mieux l'Union européenne, pour sa part, se montre humaniste, ouverte et prône une immigration qui respecte les droits des hommes et des États en même temps qu'elle s'interroge sur son identité, sur les possibilités d'en définir le contour sans pour autant pratiquer l'exclusion des populations qui viennent de tous les coins du monde pour y vivre¹. Ce qui va sans dire que l'Europe, en tant que terre d'immigration, est un espace où se parlent plusieurs langues, celles du terroir et celles liées à l'immigration, c'est-à-dire toutes celles que les migrants portent avec eux à leur installation sur le territoire européen. D'où la diversité linguistique de l'Europe qui compte dès lors non seulement ses propres langues (dites langues européennes) mais également toutes celles portées par les migrants (dites langues liées à l'immigration).

1.4 Quelques définitions utiles

— *Les langues européennes*

Dans les discours relatifs à la «diversité linguistique de l'Europe», les «langues européennes» sous-entendent uniquement les langues des natifs d'origine

européenne et comprennent les langues indo-européennes (entre autres les langues romanes), les langues finno-ougriennes et les langues sémitiques.

— *Les langues d'Europe*

Elles désignent aussi bien les langues européennes que les langues non-européennes (langues des minorités immigrées et réfugiées) en usage sur le territoire européen (comme c'est le cas du lingala, langue afro-congolaise et qui est la principale langue cible du dictionnaire dont il est question dans la présente contribution.

— *Les langues liées à l'immigration (LLI)*

L'emploi par Quentin (2002) de cette terminologie *langues liées à l'immigration* doit être perçu comme un relais du discours politique relatif à l'immigration. La réalité que couvre cette terminologie de «LLI» est diversement dénommée aussi bien dans la littérature scientifique que dans le vocabulaire politique européen relatif à l'immigration.

Aux côtés de cette terminologie de «LLI», nous en avons inventorié d'autres, à savoir: *langues de l'immigration* (Unesco 2003), *langues dites d'immigration* (Bouquin-Keller 2001) et *langues des enfants issus de l'immigration* (Huck 2012). Qu'il s'agisse de l'une ou de l'autre terminologie parmi celles que nous venons d'énumérer, nous nous trouvons dans tous les cas face aux dénominations non-consacrées et qui sont absentes du Cadre européen commun de référence aux langues (CECRL). La compréhension que nous faisons de ces terminologies est qu'elles se réfèrent toutes aux idiomes que les personnes immigrées ou toutes celles en instance d'immigration ont comme langues maternelles ou premières langues.

Toutefois, nous avons opté pour l'emploi de la terminologie *langues liées à l'immigration*, car elle traduit mieux cette jonction qu'il y a entre langue et immigration. C'est de cette jonction langue et immigration que découle la lexicographie de langues liées à l'immigration, notre centre d'intérêt dans le cadre de la présente description.

2. La lexicographie de langues liées à l'immigration (LLLI)

L'institution de cette terminologie de «langues liées à l'immigration», l'on s'en doute, ne peut rester sans conséquence pour d'autres branches du savoir humain qui sont liées à l'instrumentalisation ou mieux à l'équipement des langues. C'est dans ce contexte, par exemple, que la l'élaboration d'un dictionnaire comme celui portant sur le finnois, le lingala et le français, étant donnée les raisons justifiant la présence du lingala sur le territoire européen (entre autre en Finlande), convient d'être inscrite dans le cadre de ce que nous avons, par extrapolation, jugée bon de dénommer «la lexicographie de langues liées à l'immigration».

2.1 Quelques précisions terminologiques

Rappelons que selon le CNRTL ou le Centre national de ressources textuelles et littéraires (service spécialisé du CNRS), la lexicographie est la technique de confection des dictionnaires et l'étude scientifique et analytique des faits de lexique. C'est une discipline linguistique dont l'objet est la connaissance, l'étude des mots et des expressions d'une langue déterminée, et qui vise en particulier, à l'élaboration des dictionnaires (Larousse 2009).

La lexicographie de langues liées à l'immigration ou «LLLI», selon nous, pourrait s'entendre comme l'ensemble de pratiques dictionnaires qui visent l'équipement en outils lexicographiques des langues liées à l'immigration. Cette dénomination «LLLI» est de toute apparence nouvelle, mais lorsque nous examinons l'objet ainsi que les objectifs de la sous-discipline désignée, nous nous rendons vite à l'évidence que cette dernière a des rapports étroits avec la lexicographie d'apprentissage (LA) ou la lexicographie pédagogique (LP), dénominations des plus connues en sociolinguistique, en didactique ou en lexicographie.

2.2 Lien entre la LLLI et la LP ou la LA

Tout en tenant pour acquis les précisions terminologiques que nous venons de faire quant à la LLLI, examinons à présent le lien qui existerait entre elle (la LLLI) et ce que l'on nomme LA ou LP.

En effet, le GRELEP (Groupe de recherche en lexicographie pédagogique) décrit la lexicographie pédagogique ou lexicographie d'apprentissage comme une sous-discipline de la lexicographie qui se propose d'élaborer des dictionnaires (en papier ou électroniques) dont l'objectif est de répondre aux besoins des apprenants d'une langue, tant du point de vue réceptif (compréhension du message) que du point de vue productif (mise en discours d'un message).

Rappelons-le et ce, comme nous l'avons déjà mentionné ci-haut que notre «LLLI» se propose d'équiper une «LLI» en outils lexicographiques pour les besoins de ses locuteurs en leur conférant par le truchement de l'usage du dictionnaire la maîtrise des moyens d'expression ainsi que l'accroissement du savoir culturel. Comme nous pouvons bien nous en rendre compte, entre ce que nous désignons par la «LLLI» et ce que le GRELEP nomme la LP ou LA, la différence n'est que dénominationnelle, car les objets d'étude ainsi que les buts visés par les deux sous-disciplines lexicographiques demeurent identiques. Mais la particularité de la LLLI tient à la jonction entre langue et immigration dont elle est le produit ou le résultat. Bien plus, la LLLI est de par son objectif une LP ou LA tandis que toute LP ou LA n'est pas forcément une LLLI.

2.3 LLLI et LLI en Finlande: état de lieu

2.3.1 La lexicographie de langues liées à l'immigration

En Finlande, les dictionnaires sont élaborés et utilisés depuis plusieurs centaines d'années et interviennent dans toutes les formes d'interaction entre les locuteurs de langues différentes. Ils sont également utilisés comme matériels d'appui destinés aussi bien à la recherche qu'à l'enseignement ou l'apprentissage des langues maternelles et étrangères. C'est ce qu'il faut entendre par cette phrase en finnois de Romppanen (2001):

«Sanakirjoja on Suomessakin laadittu ja käytetty monen sadan vuoden ajan apuneuvoina koulutuksessa, tutkimuksessa ja kaikenlaisessa kanssakäymisessä erikielisten kesken».

Pour ce qui est du nombre des dictionnaires existants, on en dénombre près d'un millier. Il suffit de se rendre sur www.helmet.fi et d'interroger la base à l'aide des mots «sanakirja» (en finnois) ou «dictionary» (en anglais) pour se rendre vite à l'évidence que pour la seule région de la capitale Helsinki, il y a environ 1778 dictionnaires en circulation. Le tableau répertoriant pour l'année 2011 les langues étrangères enseignées en Finlande (cf. section 2.3.2.) mentionne aux côtés de chacune d'entre elles, les dictionnaires disponibles. L'absence de tout chiffre veut dire qu'il n'existe pas de dictionnaire pour la langue étrangère concernée. Ce que nous pouvons dire quant à l'état de lieu de LLLI en Finlande est que sur les 147 LLI parlées en Finlande, 53 langues ont été enseignées en 2011, 28 LLI parmi ces 53 suscitées disposent des outils lexicographiques de référence ou d'appui à l'apprentissage.

2.3.2 L'enseignement des LLI en Finlande

On peut lister plusieurs activités qui entretiennent des rapports étroits avec la lexicographie. Pour notre part, dans le cadre de la présente contribution, nous en avons identifié un, à savoir l'enseignement et/ou l'apprentissage de la langue maternelle, pour lequel les dictionnaires demeurent des ouvrages fondamentaux. En Finlande, cet enseignement n'est pas obligatoire, mais pour pouvoir l'organiser, deux conditions sont requises, à savoir: avoir *au minimum une classe de 4 apprenants et un enseignant ayant reçu, de l'Office finlandais en charge de l'enseignement, l'habilitation à dispenser cet enseignement.*

Le tableau ci-dessous fait état des cours de *langue maternelle, langue de la maison* ou *langue d'entretien* qui ont été organisés à travers la Finlande au cours de l'année 2011. Ces langues étant presque toutes étrangères à la Finlande (excepté le saami et le suédois), nous les considérons comme des LLI. Celles suivies d'un astérisque sont d'origine africaine et comme les chiffres l'indiquent, le lingala figure parmi les rares LLI d'origine africaine enseignées en Finlande. L'on notera en passant que le somalien figure parmi les grandes LLI

en Finlande, car il compte parmi celles qui ont un nombre plus élevé d'apprenants (2197 au total pour la seule année 2011) (cf. Opetushallitus s.d).

Langue enseignée	Apprenants	Dictionnaires disponibles	Langue enseignée	Apprenants	Dictionnaires disponibles
albanais	775		lingala*	20	1
allemand	192	40	lituanien	4	4
amharique*	14		népalais	44	
anglais	386	97	norvégien	27	4
arabe	1098	7	ourdou	77	
azéri	6		pashtou	22	
bengali	68		philippin	53	
birman	103		polonais	73	3
bosnien	127	1	portugais	33	4
chinois	334	3	roman	164	
dari	169		roumain	18	2
espagnol	262	17	russe	3770	56
estonien	740	16	sami	44	
farse	351	1	serbe	5	
français	164	37	somali*	2197	1
grec	63	1	suédois	17	65
hébreu	28	1	tamoul	14	
hindi	24		tchèque	7	
hollandais	18	4	tchéchène	25	
hongrois	85		thaï	285	1
islandais	11		tigrinya	20	
italien	104	12	turc	173	1
japonais	46	1	turkmène	7	
khmer	9		uighur	7	
kirundi*	13		ukrainien	8	1
kurde	454	4	vietnamien	476	1
letton	10				
	Nombre d'apprenants	13.224			
	Langues enseignées	53 dont 28 équipées de plus au moins un dictionnaire			
	Organisateurs	92			

Les LLI en Finlande étant pour la plupart des langues étrangères, leur enseignement ou apprentissage a lieu au cours de 10 premières années que dure l'enseignement fondamental. Pour ce qui en est des bénéficiaires, l'on notera que cet enseignement ou apprentissage est destiné aux enfants appartenant à l'une ou l'autre catégorie parmi celles ci-après: *enfants bilingues ayant pour langues maternelles autres langues que le finnois ou le suédois; enfants saamephones; enfants issus de familles d'origine tzigane; enfants finlandais ayant vécu à l'étranger et ceux ayant été scolarisés en langues étrangères tandis qu'ils séjournaient hors de la Finlande* (cf. Helsingin Kaupunki-Opetusvirasto s.d.).

2.4 La Finlande et la gestion de son plurilinguisme

Il est acquis en droit international qu'aucun État n'est obligé d'accueillir un non ressortissant sur son territoire. En revanche, lorsque ce non ressortissant, migrant ou étranger remplit les conditions de son admission sur le territoire, l'État concerné ne peut plus lui refuser la jouissance des droits garantis par les traités internationaux de droits de l'homme (cf. Agbetse s.d.). À titre d'exemple, parmi ces traités internationaux, nous avons mentionné en introduisant cette contribution, la *Déclaration sur les Droits de l'homme des personnes qui ne possèdent pas la nationalité du pays dans lequel elles vivent dont nous avons épinglé* l'article 5.1.f qui reconnaît aux personnes d'origine immigrée le droit de conserver leur langue maternelle, leur culture ainsi que leurs traditions.

Cette reconnaissance de jouissance des droits culturels des migrants, garantis par le droit international implique à son tour en Finlande la reconnaissance du droit à l'instruction pour lequel les municipalités se voient obligées de satisfaire aux exigences du programme d'enseignement en instaurant un soutien spécifique de langue d'origine aux enfants d'origine immigrée et qui n'ont ni le finnois ni le suédois pour langues maternelles.

En contexte de crise, garantir ce droit à la langue s'avère coûteux pour un pays d'accueil comme la Finlande qui se verrait obligée d'organiser par exemple les enseignements de 147 langues maternelles parlées sur son territoire, acheter le matériel didactique destiné à cet enseignement, financer l'élaboration des outils lexicographiques si faire se peut, payer les salaires aux enseignants de ces langues, etc. Ne disposant pas des moyens financiers suffisants, le gouvernement finlandais a planifié les choses comme bon lui semble en faisant la sourde oreille aux plaintes des mécontents: ainsi par exemple, tous les ressortissants congolais qui ont comme langues maternelles le ciluba, le kiswahili, le kikongo, etc. se voient obligés d'apprendre le lingala!

2.4.1 Statut des LLI en Finlande

Quel que soit le privilège dont les langues liées à l'immigration jouissent en vertu des droits spécifiques accordés ou reconnus à leurs locuteurs, du fait de leurs origines étrangères et ce, au nom du droit international ou simplement du droit à la langue, ces langues restent des langues étrangères.

Pour ce qui concerne aussi bien l'usage que l'apprentissage de ces langues, il importe de le souligner ainsi que le note l'Unesco que les États européens (surtout ceux du Sud et de l'Ouest) n'ont pas pris de dispositions particulières pour l'enseignement et la pratique des langues de l'immigration (entendez les LLI selon la terminologie que nous utilisons). La raison justifiant cette absence est que certains décideurs politiques européens dans ces derniers États considèrent que l'apprentissage ou l'enseignement des LLI constitue un des obstacles à l'intégration des populations immigrées (cf. Unesco 2003).

Par contre, dans les pays européens nordiques ou scandinaves et plus par-

ticulièrement en Finlande, il existe un programme officiel d'enseignement et d'apprentissage de toutes langues étrangères, localement dénommé «oman äidinkielen, kotikielen ja ylläpitokielen opetus» pouvant se traduire littéralement par *enseignement de sa propre langue maternelle, de la langue de la maison et de la langue d'entretien*. Le lingala étant non seulement une LLI, mais également une langue étrangère, son enseignement est assuré dans ce même cadre.

2.4.2 Les LLI dans la vie socioculturelle en Finlande

Les LLI ne sont pas seulement des matières d'enseignement ou des langues d'appui à l'enseignement en Finlande (les véhicules officiels sont le finnois et le suédois). Elles sont également des langues de travail, car elles sont présentes dans l'administration locale (hôpitaux, police, crèches, tribunaux, etc.) et sont utilisées par environ 22% de la population active.

En effet, tandis que le Centre finlandais des statistiques avait noté que pour l'année 2010, le nombre d'habitants en Finlande était de 5 375 276 personnes dont 248 135 étaient reconnues d'origine étrangère (cf. Väestökisterikeskus 2011: 2), l'Office finlandais au 31 décembre 2011, revoyait ce chiffre à la baisse et notait que la population immigrée résidant de manière permanente en Finlande était de 183 133 personnes (cf. Sisäasiainministeriö 2011: 4).

Pour ce qui est des langues parlées par ces populations d'origine étrangère, le même Centre finlandais des statistiques avance le chiffre de 147 langues dont 43 parmi ces dernières comptent plus au moins 500 locuteurs. (cf. Väestökisterikeskus 2011: 2). Il s'agit de:

afghan	chinois	lingala	russe
albanais	espagnol	néerlandais	serbo-croate
allemand	estonien	norvégien	somali
amharique	français	ourdou	swahili
anglais	grec	pakistanaï	tamoul
arabe	hindi	persan	télougou
bengali	italien	philippin	thaï
birman	japonais	polonais	turc
bosnien	kurde	portugais	ukrainien
bulgare	letton	roumain	vietnamien

Les diffuseurs institutionnels ainsi que les entreprises opérant dans domaine de la traduction et de l'interprétation dressent, pour leur part, la liste de ces 80 langues les plus demandées en Finlande (cf. Helsingin Tulkikeskus Oy):

akan	français	lituanien	roumain
albanais	géorgien	lu	russe

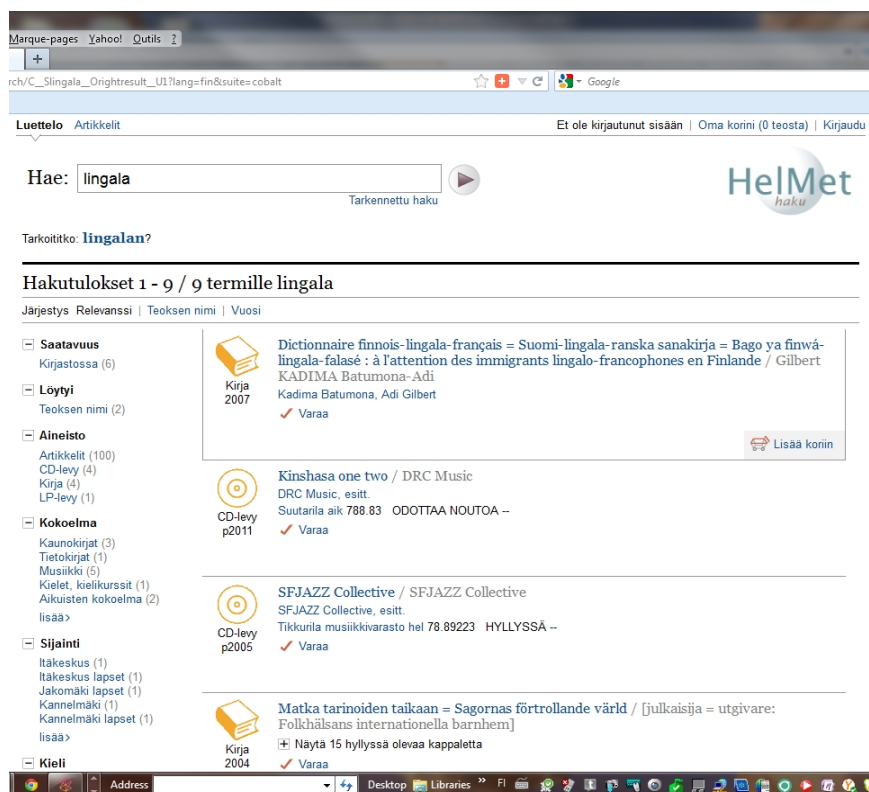
allemand	grecque	macédoine	serbe
amharique	haoussa	malayalam	serbo-croate
anglais	hébreu	manding	slovaque
arabe	hindi	marathi	somali
arménien	hongrois	néerlandais	suédois
azéri	igbo	norvégien	swahili
bengali	indonésien	nuer	tamoul
berbère	ingouche	oromo	tatar
birman	italien	ouïghour	tchèque
bosnien	japonais	ourdou	tchéchène
bulgare	jolla	ouzbek	thaï
chinois	kikongo	pachto	tigrinya
cinghalais	kikuyu	panjabi	turc
croate	kinyaruanda	persan	turkmène
dari	kirundi	peul	ukrainien
espagnol	krio	philippin	vietnamien
estonien	kurde	polonais	wolof
finnois	lingala	portugais	yoruba

2.4.3 Les outils lexicographiques disponibles

Il est quasi-universellement admis que les dictionnaires usuels, les grammaires et les atlas font partie des ouvrages de référence d'usage courant, dont la fréquence d'utilisation en classe est en général élevée et qui sont aussi considérés comme du matériel didactique de base (cf. MELS-Québec 2013).

En Finlande, le ministère de l'éducation partage cet avis et sur son site www.edu.fi, il a établi une classification des ouvrages de référence, semblable à celle en usage au Québec et dresse la liste des dictionnaires approuvés et qui peuvent être utilisés comme matériels d'appui dans l'apprentissage des LLI en usage en Finlande. Cette liste date de 2001 et n'est malheureusement pas actualisée. Le jour où elle le sera, elle reprendra certainement notre dictionnaire finnois-lingala-français, car approuvé par les spécialistes finlandais (cf. section 3.2) et il est utilisé depuis plus de six ans comme matériel didactique dans beaucoup d'établissements scolaires en Finlande, où sont organisés les cours du lingala comme langue étrangère et maternelle.

C'est à ce titre que la Bibliothèque municipale de Helsinki le répertorie dans sa banque des données parmi les quelques rares documents utilisés comme matériel didactique dans les établissements d'enseignement fondamental. Comme les lecteurs peuvent bien s'en rendre compte, c'est bien ledit dictionnaire qui vient en premier lieu à l'interrogation de la banque à l'aide du mot «lingala» comme critère de recherche.



3. Notre contribution: le dictionnaire finnois–lingala–français

L'implantation d'une langue liée à l'immigration peut susciter, dans certaines conditions, des recherches lexicographiques à même d'aboutir à la dictionnaire. C'est ce qui nous fait dire que l'immigration est un facteur de production dictionnaire. Mais qu'est-ce qu'un dictionnaire?

Murray (s.d.) écrit à propos du mot «dictionnaire» qu'il est un livre qui traite des mots isolés d'une langue afin de montrer leur orthographe, leur prononciation, leur dérivation et leur histoire ou au moins certains de ces faits. Pour la commodité du classement, les mots sont placés dans un ordre déterminé, alphabétique dans beaucoup de langues.

3.1 Corpus et méthodologie

De nos jours, une description de nature linguistique ou littéraire se conçoit mal sans un échantillon d'étude ou mieux sans corpus. C'est dans ce même ordre d'idées que Cabré (2000: 14) recommande que le «linguiste qui veut décrire le langage et les langues commence toujours avec quelques données sur les langues. La provenance de ces données est le discours oral ou écrit des sujets

parlants. Un linguiste concerné par la description des unités terminologiques doit les chercher dans les productions orales et écrites des spécialistes».

Ne pouvant pas déroger à cette exigence méthodologique, dans l'élaboration de notre dictionnaire finnois–lingala–français, nous avons pris soin de constituer un corpus dont les sources principales sont des outils didactiques, lexicographiques et terminologiques qui traitent du finnois, langue source de ce dictionnaire. Les plus importants des documents dépouillés ont été:

- (a) *Dictionnaire finnois–français–finnois* d'Hélène Lattunen et Kari Viljanen, 2ème édition de 2000;
- (b) *Suomi–englanti–suomi sanakirja* de Raija Hurme et al., 10ème édition de 2007; et
- (c) *Dictionnaire bangala–français–lingala* de Atibakwa Baboya Edema, 1ère édition de 1994.

Notre dictionnaire traduit celui de Lattunen et Vianen et c'est sur base de ce dernier que nous avons établi notre nomenclature. Nous avons recouru aux deux autres dictionnaires figurant sur la liste seulement lors de la recherche des équivalences.

Une remarque s'impose quant à l'absence parmi nos sources de référence de deux autres dictionnaires lingala des plus connus: il s'agit en effet du dictionnaire de Kawata Ashem Tem et du dictionnaire de René van Everbroeck auxquels nous n'avons pas pu accéder. Nous pensons nous corriger cette lacune avec la publication prochaine de l'édition à large diffusion en cours de préparation.

Pour ce qui est de la méthodologie nous avons eu recours à divers outils méthodologiques qui partent des observations passives mais soutenues aux dépouillement d'outils lexicographiques existant dans la langue source, le finnois.

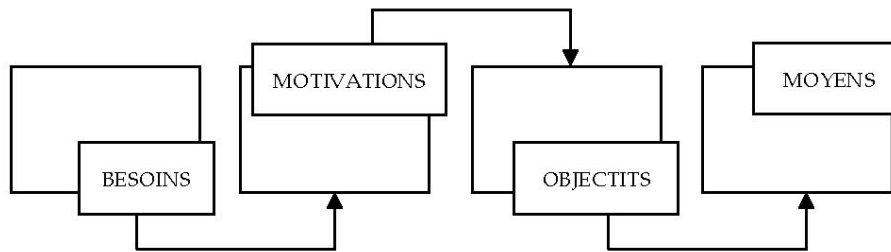
Ces observations qui avaient porté sur le paysage linguistique de la Finlande, nous ont conduit à poser un diagnostic, à savoir: l'inexistence d'un dictionnaire finnois–lingala. De ce diagnostic, nous avons, en tant que linguiste lexicographe, réalisé qu'il y avait réellement un «besoin» à satisfaire. Le mot «besoin» n'est pas à entendre ici seulement en termes de demande, de motivation ou d'attente, mais au sens lui conféré par les spécialistes de l'analyse de besoins au Québec, celui d'«écart mesurable qui existe entre une situation actuelle («ce qui est») et une situation désirable («ce qui devrait être») (Loubier 1994: 50).

Cette définition de la notion de besoin peut être ainsi représentée par ce diagramme:



C'est sur base de nos motivations et en fonction des besoins à satisfaire que nous avons pu identifier: (a) le manque ou l'inexistence d'outils didactiques et lexicographiques finnois–lingala, (b) apprendre le finnois, afin de nous intégrer en Finlande; et (c) élaborer un dictionnaire finnois–lingala–français à même d'aider aussi bien nous-mêmes que d'autres immigrants se trouvant dans la même situation que nous (quête d'intégration par la connaissance de la langue du pays d'accueil).

Pour atteindre ces objectifs, il fallait que nous puissions nous doter de ressources et moyens financiers conséquents. Nous avons procédé de la manière qu'illustre ce diagramme:



3.2 Outil informatique de confection

Par manque d'outil informatique approprié aux recherches dictionnairiques (les plus en vogue et les plus appropriés aux langues du domaine bantu étant les logiciels Shoebox/Toolbox² de la Société Internationale de Linguistique et TshwaneLex, logiciels payants), nous nous étions résolu de travailler avec nos moyens de bord en ayant recours à un outil informatique d'un maniement facile, à savoir le logiciel de traitement des textes Word de la firme Microsoft.

3.3 Le choix du lingala comme principale langue cible

Pour bien comprendre les raisons ayant motivé le choix du lingala comme principale langue cible de ce dictionnaire, il convient de rappeler de manière succincte sa situation sociolinguistique au Congo-Kinshasa. En effet, le lingala est avec le swahili, l'haoussa et le wolof, une des langues supranationales que l'Afrique connaît aujourd'hui. Il fait partie de quatre langues nationales de la République Démocratique du Congo, les trois autres étant le kikongo, le kiswahili et le cilubà.

Langue de la capitale politique de la RD-Congo qui est Kinshasa, le lingala est en passe de devenir un super véhiculaire sur l'ensemble du territoire national rd-congolais, étant donné que beaucoup de Congolais qui émigrent à l'étranger et qui ne sont pas issus de la zone linguistique lingalophone, le

comprennent et le parlent du fait qu'il est, entre autre, la langue de la musique congolaise moderne, la langue du théâtre national et la langue des affaires.

À l'étranger, c'est également le lingala qui, dans beaucoup de situations de communication, sert de langue d'intercommunication entre les membres de la diaspora congolaise issus d'aires linguistiques diverses. C'est ce qui justifie son choix comme principale langue cible de ce dictionnaire. Nous avons choisi le français comme langue cible seconde afin qu'il serve également les congolais issus de trois autres aires linguistiques. Ce choix découle de son statut de langue officielle en RD-Congo.



3.4 Fiche synoptique du dictionnaire

Intitulé du dictionnaire	Dictionnaire finnois–lingala–français
Auteur	Kadima Batumona Adi Gilbert
Statut	Docteur ès lettres; Agrégé de l'enseignement secondaire supérieur (option français)
Support du dictionnaire	papier
Pagination et format	323 pages A4
Classement	alphabétique
Public cible	(im)migrants lingalo-francophones établis en Finlande
Fonctions	référence, apprentissage, décodage, enrichissement lexical
Macrostructure	123.754 mots; 18.450 entrées
Microstructure	équivalents en langues ciblées (lingala et français); origine du mot si nécessaire; pas de marques d'usage, pas de prononciation
Illustrations	1 carte de 4 grandes aires linguistiques rd-congolaises

Ce dictionnaire a été élaboré aux fins d'aider de prime abord les congolais vivant en Finlande et servir à ces derniers d'outil d'appui dans l'apprentissage de la langue de leur pays d'accueil, le finnois. En tant qu'outil lexicographique, son rôle est de donner la maîtrise des moyens d'expression et d'accroître le savoir culturel de ses utilisateurs.

Avant sa validation comme dictionnaire pédagogique, il avait été soumis à l'approbation d'une commission chargée du contrôle des matériels pédagogiques de la *Fondation Diakonia*³. Cette commission comprenait un anthropologue, un professeur de langue finnoise (également francophone) et un spécialiste en sciences documentaires. L'approbation de notre dictionnaire par cette dernière commission tient entre autre au fait qu'il répond à ces exigences que le GRELEP pense comme celles, auxquelles doit devoir répondre tout diction-

naire d'apprentissage: — tirer parti des dictionnaires existants — être basé sur un corpus — se fonder sur l'expérience didactique de professeurs — être accessible par le biais de la langue maternelle de l'apprenant et enfin, — constituer un outil d'(auto-)apprentissage, c'est-à-dire de construction du savoir.

3.5 Extrait du dictionnaire

concis yö – butú - nuit yöeläin – nyama ya botámboli na butúbutú - animal nocturne yöelämä – ezalela ya butú - vie nocturne yökerho – ekalá - boîte de nuit, boîte yökyöpelii – motámbolampímpa, motámboli ya butú - couche-tard, noctambule yököttävä – ya bopési nkándá - écoeurant, dégoûtant yöleno – mobémbo ya mpépo ya butú - vol de nuit yöllinen – ya butú - nocturne	Zaire – Zaire - Zaïre zäirelainen – zairwá, zét, ya mbóka Zaire - zaïrois(e) zeniitti – zeniti - zénith zeppeliini – zepelé (< fr. zeppelin) - zeppelin Zimbabwe – Zimbábwe - Zimbabwe zimbabwelainen – zimbabweyé - zimbabwéen(ne) zirkoni – zirkó (< fr. zircon) - zircon zoomata – kozumé (< fr. zoomer) - zoomer zoomaus – zúm (< fr. zoom) - zoom zoomojektiivii – zúm - zoom
--	--

3.6 Caractéristiques techniques, macro- et microstructures

3.6.1 Caractéristiques techniques

Comme nos lecteurs peuvent bien s'en rendre compte en passant en revue le tableau synoptique ainsi que l'extrait de la microstructure ci-dessus, notre dictionnaire finnois-lingala-français est: *interlinguistique, trilingue* et *monodirectionnel*. En effet, il est interlinguistique trilingue parce qu'il porte sur trois issues de trois familles différentes: le finnois (famille finno-ougrienne), le lingala (langue africaine bantou) et le français (langue indo-européenne); quant à sa monodirectionnalité, elle réside en ce qu'il part de la langue source vers les langues cibles. C'est dire qu'à la différence d'autres dictionnaires multilingues (dont les bilingues surtout) qui sont bidirectionnels (se déployant de la source vers la cible et de la cible vers la source), et qui sont à la fois des dictionnaires de version et de thème, le nôtre est uniquement un dictionnaire de version étant donné qu'il se déploie dans un seul et unique sens.

Rey-Debove (2005) établit, pour sa part, la distinction suivante entre les dictionnaires monolingues et les multilingues:

Le dictionnaire multilingue réalise un trajet qui consiste en un décodage (traduction) et en un encodage (production): langue A → langue B → langue C. Ce n'est que le passage d'un code lexical à un autre et est généralement dépourvu de définition. Le dictionnaire monolingue, quant à lui, est destiné à un éclaircissement d'une langue A pour un lecteur A. Il révèle seulement le sens (= traduction) mais jamais la production (= encodage). Bien plus, il tente de décrire la langue par le lexique, en produisant des exemples de signes intégrés dans une syntaxe

avec leur morphosyntaxe. Ainsi donc, grâce à l'exemple on accède au sens du mot, à la grammaire, et à la sémiologie d'une société à un moment donné.

3.6.2 Macro- et microstructures

Du point de vue sa forme, notre dictionnaire finnois–lingala–français comme bon nombre de dictionnaires, comprend une macrostructure et une microstructure. La macrostructure est l'ensemble des adresses ou entrées tandis que la microstructure, elle, est l'ensemble des éléments informationnels accompagnant chacune des entrées d'un dictionnaire. Les détails relatifs à la macrostructure étant fournis dans la fiche synoptique ci-dessus, il ne nous reste qu'à apporter quelques éclaircissements quant à la manière dont nous avons organisé la microstructure du dictionnaire.

En effet, nous nous sommes référé à la tradition lexicographique qui est en vigueur en Finlande, où nous sommes établis et où le dictionnaire fut élaboré et destiné à y être utilisé. Dans ce pays et cela, en tenant compte de langues en présence, les microstructures d'un dictionnaire bilingue ou trilingue comme le nôtre s'articulent d'ordinaire autour des éléments ci-après: *prononciation, équivalents en langue(s) ciblée(s) et quelques marques d'usage (genre, nombre, etc.)*. Le lingala se lisant comme le finnois et ne faisant nullement pas de distinction quant au genre et au nombre, nous avons jugé bon de ne pas inscrire dans l'article ces marques d'usage ainsi que la prononciation. Ainsi donc, la microstructure de notre dictionnaire se limite à l'équivalent et dans certains cas s'étend à l'origine du mot s'il s'agit d'un emprunt. Dans l'extrait en question, les équivalents en lingala de ces trois entrées: *zoomata, zoomaus* et *zirkoni* sont accompagnés des marques relatives à leurs origines (ce sont des emprunts faits au français).

Voici à titre d'exemples, trois microstructures des dictionnaires bilingues en usage en Finlande et dont deux font partie de notre corpus. Pour des raisons d'économie, nous ne mentionnerons qu'une direction par dictionnaire.

— *Dictionnaire finnois–français–finnois de Kalmbach et Sundelin*
Finnois–français:

aaltoilu ondoisement <i>m</i> , ondulation <i>f</i>	~llä le matin, pendant la matinée, avant midi
aaltoliike mouvement <i>m</i> ondulateur; (<i>kuv</i>) va-et-vient <i>m</i> (<i>taipum</i>)	aamurusko aurore <i>f</i>
aaltomainen ondoyant, onduleux (<i>m</i>), onduleuse (<i>f</i>)	aamutak ki robe <i>f</i> de chambre, (<i>kevyt naisen aamutakki myös</i>)
aaltopahvi carton <i>m</i> ondulé	peignoir <i>m</i> ; -issa en robe de chambre, en peignoir
aaltopelti tôle <i>f</i> ondulée	aamutuimaan de bonne heure
aaltopituus (<i>kuv</i>) samalla aaltopituudella sur la même longueur d'onde	aamu-uninen lève-tard <i>mf</i> (<i>taipum</i>)
	aamuvirkku matinal (<i>mpl</i> <i>mati-</i>

— *Dictionnaire finnois–français–finnois de Lattunen et Viljanen*
Français–finnois:

Zaire <i>m</i> Zaire	zénith <i>m</i> huippu, zeniitti
zairois zairelainen	Z.E.P. (lyh. sanoista <i>zone d'éducation prioritaire</i>) (läh.) ongelmakoulualue
Zambie <i>f</i> Sambia	zéro nolla
zambien, zambienne sambialainen	zeste <i>m</i> 1 pala sirushedelmän kuorta
zantedescia <i>f</i> kalla	2 (kuv.) pieni määrä
zapper vaihtaa koko ajan TV-kanavaa; (kuv.) hyppiä asiasta toiseen	zézaïement <i>m</i> ässävika
zapping <i>m</i> kanavasurffailu	zézayer leskata
zébré juovikas	Z.I. (lyh. sanoista <i>zone industrielle</i>) teollisuusalue
zèbre <i>m</i> seepra <i>quel drôle de zèbre!</i> (ark., kuv.) mikä hassu tyyppi!	zieuter (sl.) katsoa
zebrina <i>f</i> (kasv.) juoru	zigouiller (sl.) tappaa

— *Dictionnaire finnois–allemand–finnois de Klemmt et al.*
Allemand–finnois:

zack <i>interj</i> nopeasti	zählbar <i>adj</i> joka voidaan laskea
Zack auf Zack sein (ihminen) olla valmiina/kunnossa, osata asiansa; (asiat) olla reilassa	zahlen <i>v</i> 1 maksaa <i>er hat zuviel dafür bezahlt</i> hän maksoi siitä liikaa <i>eine Rechnung zahlen</i> (ark) maksaa lasku
Zacke <i>die Zacke, die Zacken</i> piikki, (vuoren)huippu, (sahanterän) hammas, (kamman, haravan) pii(kki)	2 zahlen, bitte! saanko laskun?
zackig <i>adj</i> 1 rosoinen, rosoreunainen, pykäläinen 2 reipas	zählen <i>v</i> 1 laskea <i>die Gäste zählen</i> laskea vieraat, laskea paljonko vieraita on 2 sie zählt 50 Jahre hän on 50-vuotias 3 sie zählt ihn nicht zu ihren Freunden hän ei katso miestä ystäväkseen, hän ei katso miehen kuuluvan ystäväpiiriinsä 4 sie zählt zu den bedeutendsten Schriftstellerinnen dieses Jahrhunderts hän on yksi tämän vuosisadan merkittävimpiä (nais)kirjailijoita
zagen <i>v</i> empiä, aristella, haparoida (kuv)	5 kann ich auf dich zählen? voinko luottaa sinuun? voinko laskea sinun varaan?
zaghaf <i>adj</i> empivä, aristeleva, haparoiva	
Zaghafigkeit <i>die</i> epävarmuus, aristelu	
zäh <i>adj</i> 1 sitkeä, (neste) jähmeä	
2 (kuv) sinnikäs, sitkeä, sisukas	
3 hidas	
zähflüssig <i>adj</i> jähmeä	

3.6.3 Démarche lexicographique

Nous basant sur ces deux caractéristiques susmentionnées (interlinguistique trilingue et monodirectionnel), nous avons organisé nos adresses de manière onomasiologique, c'est-à-dire que nos adresses ou entrées sont présentées suivant l'ordre: vedette — équivalent ou synonyme; en cas de difficulté de trouver un synonyme, nous avons fait recours soit à l'emprunt, à un calque ou encore à une explication de la vedette en langue cible.

Comme dans tout outil lexicographique multilingue, la recherche d'un équivalent pose le plus souvent des problèmes, dans la mesure où il arrive qu'une notion dans la langue source n'ait pas de correspondant dans la (les)

langue(s) ciblée(s). Aux fins de faire face à cette difficulté, nous avons dû recourir aux techniques d'enrichissement qui sont d'usage courant aussi bien en lexicologie qu'en terminologie. Ces techniques sont résumées par Edema (2008: 72-73) en termes suivants:

Pour s'enrichir, toute langue dispose de trois bases: une base grammaticale qui, avec ses structures, aide à former de nouveaux mots, une base sémantique qui, avec des mots déjà existants, donne de nouveaux sens; et l'emprunt à d'autres langues, source également de nouveaux mots pour une langue. Les deux premières bases sont qualifiées de bases internes tandis que la dernière, à savoir l'emprunt, elle, est qualifiée de base externe.

De l'avis de Adamska-Salaciak (2011) qui réfère aux recherches menées par Zgusta (1971, 1987), les équivalences dans un dictionnaire bilingue ou multilingue peuvent être de différentes natures: cognitive, traductologique, explicative et fonctionnelle. À la question de savoir à quel type d'équivalences avons-nous eu recours de notre travail, nous répondrons qu'avant même que Adamska-Salaciak n'ait publié son étude, notre dictionnaire datant de 2007 et donc antérieur à cette étude, était déjà paru. Mais si nous essayons de scruter aujourd'hui les équivalences correspondant aux adresses en langue source, nous pouvons dire qu'elles sont de trois natures: cognitive, traductologique et explicative.

Exemples:

— *Équivalences cognitives*

Finnois	Lingala	Français
isä	tatá	père, papa
äiti	mamá	mère, maman
lapsi	mwána	enfant
koti	ndáko	maison
silmä	lísu	œil
koulu	etéyelo	école
aamu	ntóngó	matin
ilta	mpókwa	soir

— *Équivalences traductologiques*

Finnois	Lingala	Français
omakotitalo	ndáko ya moto mókó	villa
etunimi	nkómbó ya yambo	prénom
öljyvári	lángi ya mafúta	couleur à huile
yleistää	kosangisa esiká mókó	généraliser
virusoppi	boyébi ya maye matáli mikróbi	virologie
allekirjoittaa	kotiya sinyé	signer

— *Équivalences explicatives*

Finnois	Lingala	Français
divari	wenze ya kowet; wenze ya bwáká-nzóto	brocante
elokuu	sánzá ya mwambe	août
hedelmätarha	elanga ya malálá	verger
öljyváhinko	likámá eúti na bosopani ya pitoló	accident pétrolier
huomena	lóbi óy'ezalí koyá	demain

3.6.4 Grille de création lexicale

matrice interne	néologie de forme	construction	affixation	m i x a t i o n	préfixation
			composition		suffixation
		formation parasynthétique			
	juxtaposition				
	synapsie				
	néologie de sens	imitation et déformation			formation syntagmatique
changement de sens		onomatopée			
		métaphore			
néologie par emprunt		métonymie			
matrice externe			extension, restriction de sens		
					emprunt

Cette grille nous a servi d'outil pour la formation des nouvelles lexies lingala et ce, aux fins de combler des vides lexicaux. Ces quelques exemples qui suivent montrent les matrices lexicales auxquelles nous avons eu recours dans nos opérations de création lexicale. Les matrices les plus sollicitées sont: l'emprunt, la composition par juxtaposition, la composition synaptique, la formation syntagmatique, la dérivation parasynthétique, la préfixation et la suffixation.

3.6.5 Quelques exemples de création lexicale

— *L'emprunt*

Exemples:

Finnois	Lingala	Français	Origine
hissi	asanséli	ascendeur	français
juusto	fromázi	fromage	français
lumi	nézi	neige	français

sauna	saúna	sauna	finnois
kaukosäädin	telekomande	télécommande	français

— *La composition par juxtaposition*

C'est un mode de composition dans lequel les éléments lexicaux de forme pleine avec ou sans trait d'union sont réunis en mot composé (Boutin-Quesnel 1990: 24).

Exemples:

Finnois	Lingala			Français
	déterminé	déterminant	lexie composée	
budjetoida	kosála	bidié	kosála bidié	budgétiser
delegoida	kotínda	ntómá	kotínda ntómá	déléguer
dekoodata	kolongola	kóde	kolongola kóde	décoder
devalvoida	kokitisa	motúyá	kokitisa motúyá	dévaluer
elokuvata	kokanga	bilili	kokanga bilili	tourner un film

— *La composition synaptique*

C'est le mode de composition par lequel le déterminé est relié au déterminant par le truchement d'un mot de liaison couramment appelé connectif, joncteur ou déterminatif du point de vue fonctionnel.

Exemples:

Finnois	Lingala			Français
	déterminé	déterminatif	déterminant	
autoliikenne	botámboli	ya	mítuka	circulation automobile
autojono	molongó	ya	mítuka	queue de voitures
bussikaista	nzelé	ya	bísi	couloir réservé aux autobus
croissant	lípa	ya	krwasa	croissant
desinfiointiaine	nkísi	ya	kopétola	désinfectant

— *La formation syntagmatique*

Nous entendons par la formation syntagmatique un mode de composition qui conduit à la formation d'un syntagme lexical. Ce syntagme peut être construit sur le modèle de la composition synaptique, de la composition par juxtaposition ou une combinaison de ces deux formes de composition.

Exemples:

Finnois	Lingala	Français
Euroopan unioni	lisanga ya bambóka ya potó	Union européenne
isällisesti	na ndengé ya botatá	paternellement

desinfointiaine	nkísi ya koboma mikróbi	désinfectant
omakotitalo	ndáko ya moto mókó	villa
ompelimo	ndáko ya botongi bilambá	atelier de couture

— *La dérivation parasynthétique*

Il y a formation parasynthétique lorsqu'un substantif dérivé réunit en lui les deux modes d'affixation, à savoir la préfixation et la suffixation.

Exemples:

Finnois	Lingala					Français
	préfixe	radical	suffixe(s)	lexie dérivée	origine	
organisointi	bo-	-bong-	-is-i	bobongisi	kobongisa	organisation
painike	e-	-simb-	-el-i	esímbelí	kosímba	levier, poignée
pelokas	mo-	-báng-	-i	mobángi	kobánga	peureux
piirrin	n-	-lakis-	-a	ndakisa	kolakisa	style
ryhmitys	li-	-sang-	-a	lisangá	kosangisa	groupement

— *La préfixation*

Il y a donc préfixation lorsque l'affixation s'opère par l'adjonction des préfixes.

Exemples:

Finnois	Lingala					Français
	préfixe	radical	suffixe(s)	lexie dérivée	origine	
organisointi	bo-	-bong-	-is-i	bobongisi	kobongisa	organisation
painike	e-	-simb-	-el-i	esímbelí	kosímba	levier, poignée
pelokas	mo-	-báng-	-i	mobángi	kobánga	peureux
piirrin	n-	-lakis-	-a	ndakisa	kolakisa	style
ryhmitys	li-	-sang-	-a	lisangá	kosangisa	groupement

— *La suffixation*

On parle de la suffixation lorsque l'affixation s'opère par l'adjonction des suffixes.

Exemples:

Finnois	Lingala					Français
	préfixe	radical	suffixe(s)	lexie dérivée	origine	
organisointi	bo-	-bong-	-is-i	bobongisi	kobongisa	organisation
painike	e-	-simb-	-el-i	esímbelí	kosímba	levier, poignée
pelokas	mo-	-báng-	-i	mobángi	kobánga	peureux
piirrin	n-	-lakis-	-a	ndakisa	kolakisa	style
ryhmitys	li-	-sang-	-a	lisangá	kosangisa	groupement

3.7 Difficulté rencontrée

Le lingala, l'une des langues cibles de ce dictionnaire, comporte plusieurs variétés dialectales. En sus du problème lié au choix de la variété dialectale, nous avons eu aussi à faire face aux problèmes liés à l'orthographe. En effet, comme bon nombre des langues africaines «bantu», le lingala est une langue qui comporte des voyelles ouvertes et est une langue à tons.

Exemples:

- móto – tuli – feu
- moto – ihminen – être humain
- motó – pää – tête

Eu égard à la difficulté de nous procurer un logiciel approprié au niveau de la notation tant des tons que des voyelles ouvertes, nous avons estimé qu'il n'était pas du tout nécessaire de généraliser la notation des tons et des voyelles ouvertes, car, avons-nous estimé, étant donné que le dictionnaire traite des mots isolés d'une langue, tout lingalophone est à même de faire la distinction entre les homographes et les homophones selon le contexte.

4. Conclusion

Sous le titre de «lexicographie et enseignement de langues liées à l'immigration: note sur le contexte sociolinguistique de l'élaboration du dictionnaire finnois–lingala–français», nous nous étions proposé de retracer l'itinéraire méthodologique que nous avons suivi en vue la rédaction de notre outil lexicographique dont la présente contribution en constitue une description.

Dans les pages qui précèdent, nous nous sommes efforcé de montrer comment l'immigration en tant phénomène social a généré une recherche lexicographique qui s'est soldée par la dictionnaire dotant le lingala d'un ouvrage lexicographique trilingue.

Tandis que nous sommes en train de préparer la seconde édition dudit ouvrage lexicographique trilingue, nous avons estimé qu'il était nécessaire de recueillir des critiques quant à la méthodologie utilisée lors de la première édition. Nous sommes conscient que cette dernière est perfectible et qu'elle doit être enrichie de nouvelles entrées. Mais déjà, telle qu'elle se présente, si elle peut continuer à aider les immigrés congolais en phase d'apprentissage du finnois et ce, en vue de leur intégration en Finlande, nous trouverions en eux une raison de parachever dans le meilleur délai la seconde édition améliorée qui est en préparation.

Notes

1. Lire à ce propos les Propositions du Groupe des Intellectuels pour le Dialogue Interculturel Constitué à l'initiative de la Commission Européenne, dans «comment la multiplicité des

langues pourrait consolider l'Europe», Bruxelles 2008. Disponible en ligne sur: <http://www.europa.eu/languages/fr/document/106/5>.

2. Ce logiciel, dans sa conception automatique de dictionnaire, permet d'organiser en champs les données exportables et convertibles pour la suite en articles ordinaires de dictionnaire, sous forme de document Word.
3. La Fondation Diakonia, connue sous la dénomination de Diakonia Laitos est des plus présentes fondations qui sont visibles en Finlande dans le domaine de l'éducation-formation et qui se distinguent dans les oeuvres de charité. C'est elle qui avait financé la première édition du dictionnaire finnois-lingala-français.

Bibliographie

- Adamska-Salaciak, Arleta.** 2011. Between *designer drugs* and *afterburners*: A Lexicographic-Semantic Study of Equivalence. *Lexikos* 21: 1-22.
- Agbetse, Yao.** s.d. *Contribution à la réflexion sur la problématique de la réalisation effective des droits culturels dans le contexte de la migration internationale*. Disponible en ligne sur: <http://www.unifr.ch/iiedh/assets/files/DS/DS5-dc-travailleurs-migrants.pdf>.
- Boutin-Quesnel, Rachel et al.** 1990. *Vocabulaire systématique de la terminologie*. Disponible en ligne sur: http://www.oqlf.goav.qc.ca/ressources/bibliotheque/dictionnaires/voc_systematique_terminologie.pdf. Consulté le 12 novembre 2010.
- Bouquin-Keller, Hélène.** 2001. *Les langues dites «d'immigration» dans les bibliothèques municipales françaises*. Disponible en ligne sur: <http://bbf.enssib.fr/>. Consulté le 22 août 2013.
- Cabré, Maria Teresa.** 2000. Terminologie et linguistique: la théorie des portes. *Terminologies Nouvelles* 21: 10-15, juin. Terminologie et diversité culturelle.
- Centre national de ressources textuelles et lexicales.** *Lexicographie: définition*. Disponible en ligne sur: <http://www.cnrtl.fr/lexicographie/lexicographie>.
- Conseil de l'Europe: Unité des politiques linguistiques.** *Cadre européen commun de référence pour les langues: apprendre, enseigner, évaluer*. Disponible en ligne sur: <http://www.coe.int/lang-CECR>.
- Edema, Atibakwa Baboya.** 1994. *Dictionnaire bangala-français-lingala*. Saint-Maur: Sépia-ACCT.
- Edema, Atibakwa Baboya.** 2008. La construction du signifiant. Didi-Kidiri, Marcel (Ed.). 2008. *Le vocabulaire scientifique dans les langues africaines. Pour une approche culturelle de la terminologie*: 71-99. Paris: Karthala.
- Groupe de recherche en lexicographie pédagogique.** s.d. *Objectifs*. Disponible en ligne sur: <http://www.kuleuven.be/grelep/objectif/objectin.htm>.
- HCDH.** 1985. La Déclaration sur les droits de l'homme des personnes qui ne possèdent pas la nationalité du pays dans lequel elles vivent. Adoptée le 13 décembre 1985 par consensus et contenue dans la résolution 40/144.
- Helsingin Kaupunki-Opetusvirasto = Education Department of City of Helsinki.** s.d. *Oman äidinkielen, kotikielen ja ylläpitokielen. opetus — opas kouluille*. Disponible en ligne sur: <http://www.kotisivukone.com/aidinkielenopetus>.
- Helsingin Tulkikeskus Oy = Helsinki Interpretation Centre Ltd.** *Interpretation and Translations with I.A.* Disponible en ligne sur: <http://www.helsingintulkikeskus.fi/en/languages>.
- Huck, Dominique.** 2012. *Avant-propos: Les langues des enfants «issus de l'immigration»*. Disponible en ligne sur: <http://www.cahiersdugepe.fr>.

- Hurme, Raija et al.** 2007. *Suomi–englanti–suomi sanakirja*. 10ème édition. Helsinki: WSOY.
- Kalmbach, Jean-Michel et Seppo Sundelin.** 2008. *Suomi-ranska-suomi sanakirja*. 9ème édition. Helsinki: WSOY.
- Kawata Ashem Tem, François Thomas.** 2004. *Bagó-Dictionnaire lingála–falansé/française–lingála*. Paris-Saint-Pierre-lès-Nemours: Karthala.
- Klemmt, Rolf et al.** 1992. *Suomi-saksa-suomi sanakirja*. Jyväskylä: Gummerus.
- Lattunen, Héléne et Kari Viljanen.** 2000. *Dictionnaire finnois–français–finnois*. 2ème édition. Helsinki: Gummerus.
- Le Petit Larousse illustré*. 2009. Paris: Larousse.
- Loubier, Christiane.** 1994. *L'Aménagement linguistique au Québec: Enjeux et perspectives*. Coll. Langues et Sociétés. Québec: Office de la langue française.
- Ministère québécois de l'Éducation, du Loisir et du Sport.** 2012–2013. *Matériel didactique approuvé pour l'éducation préscolaire et l'enseignement primaire 2012–2013*. Disponible en ligne sur: http://www3.mels.gouv.qc.ca/bamd/Doc/Liste_primaire_fr_ref.pdf.
- Murray, James.** s.d. *Petite histoire des dictionnaires*. Disponible en ligne sur: <http://www.etudes-litteraires.com/histoire-dictionnaires.php>.
- ONU.** 1984. Dispositions internationales pour la protection des droits de l'homme des non-ressortissants (Publication des Nations Unies, N° de vente: F.80.XIV.2.).
- Opetushallitus.** s.d. *Omana äidinkielenä opetut kielet ja opetukseen osallistuneiden määrät vuonna 2011*. Disponible en ligne sur: http://www.oph.fi/download/142341_Omana_aidinkielenä_opetetut_kielet_ja_opetukseen_osallistuneiden_määrät_vuonna_2011-PDF.pdf.
- Quentin, Francine.** 2002. *L'héritage linguistique des immigrés en danger*. Archives-RFI (Langues). Disponible en ligne sur: http://www.rfi.fr/actufr/articles/026/article_13364.asp.
- Rey-Debove, Josette.** 2005. *Typologie des dictionnaires généraux monolingues de la langue actuelle*. Disponible en ligne sur: <http://amsacta.unibo.it/2301/i/Debove.pdf>.
- Romppanen, Birgitta** (2001), citée dans **Kotikielenseura.** 2002. *Uutta sanakirjatutkimusta*. Disponible en ligne sur: http://www.kotikielenseura.fi/virttaja/hakemistot/jutut/2002_300.pdf.
- Sisäasiainministeriö.** 2011. *Maahanmuuton vuosikatsaus*. Disponible en ligne sur: http://www.migri.fi/download/34214_maahanmuutto_2011_tilastokatsaus.pdf.
- Unesco.** 2003. *La promotion des langues des communautés issues de l'immigration en débat*. Disponible en ligne sur: <http://www.unesco.org/most/discuss.htm>.
- Väestökisterikeskus.** 2011. *Väestörakenne 2010*. Disponible en ligne sur: http://www.stat.fi/til/vaerak/2010/vaerak_2010_2011-03-18_fi.pdf.
- Zgusta, L.** 1971. *Manual of Lexicography*. La Haye: Mouton.
- Zgusta, L.** 1987. Translational Equivalence in a Bilingual Dictionary: Bahukosyam. *Dictionaries: Journal of the Dictionary Society of North America* 9: 1-47.

Making an Online Dictionary of New Zealand Sign Language*

Rachel Locker McKee (rachel.mckee@vuw.ac.nz)

and

David McKee (david.mckee@vuw.ac.nz)

Deaf Studies Research Unit, School of Linguistics and Applied Language Studies, Victoria University of Wellington, New Zealand

Abstract: The *Online Dictionary of New Zealand Sign Language* (ODNZSL),¹ launched in 2011, is an example of a contemporary sign language dictionary that leverages the 21st century advantages of a digital medium and an existing body of descriptive research on the language, including a small electronic corpus of New Zealand Sign Language. Innovations in recent online dictionaries of other signed languages informed development of this bilingual, bi-directional, multimedia dictionary. Video content and search capacities in an online medium are a huge advance in more directly representing a signed lexicon and enabling users to access content in versatile ways, yet do not resolve all of the theoretical challenges that face sign language dictionary makers. Considerations in the editing and production of the ODNZSL are discussed in this article, including issues of determining lexemes and word class in a polysynthetic language, deriving usage examples from a small corpus, and dealing with sociolinguistic variation in the selection and performance of content.

Keywords: SIGN LANGUAGE LEXICOGRAPHY, ONLINE DICTIONARIES, MULTIMEDIA DICTIONARIES, BILINGUAL DICTIONARIES, LEARNER DICTIONARIES, NEW ZEALAND SIGN LANGUAGE, VIDEO CONTENT, SIGN LANGUAGE CORPUS, POLYSYNTHETIC MORPHOLOGY, POLYSEMY, SOCIOLINGUISTIC VARIATION, SIGN LANGUAGE LINGUISTICS, USER PROFILE

Opsomming: Die maak van 'n aanlyn woordeboek van Nieu-Seelandse gebaretaal. Die *Online Dictionary of New Zealand Sign Language* (ODNZSL)¹ wat in 2011 bekend gestel is, is 'n voorbeeld van 'n hedendaagse gebaretaalwoordeboek wat gebruik maak van die 21ste-eeuse voordele van 'n digitale medium en 'n bestaande hoeveelheid beskrywende navorsing oor die taal, insluitende 'n klein elektroniese korpus van Nieu-Seelandse gebaretaal. Vernuwings in onlangse aanlyn woordeboeke van ander gebaretaale het bygedra tot die ontwikkeling van hierdie tweetalige tweerigting-multimediawoordeboek. Video-inhoud en soekhoedanighede in 'n aanlyn medium is 'n groot vooruitgang om 'n gebareleksikon meer direk aan te bied en gebruikers op veelsydige maniere toegang te verleen tot die inhoud, maar los nogtans nie al die teoretiese uitdagings op waarvoor gebaretaalwoordeboekmakers te staan kom nie. Oorwegings by die redigering en totstandbrenging van die ODNZSL word in hierdie artikel bespreek, insluitende kwessies soos die

* This article was presented as a keynote address at the Seventeenth Annual International Conference of the African Association for Lexicography (AFRILEX), which was hosted by the Department of African Languages, University of Pretoria, Pretoria, South Africa, 2-5 July 2012.

bepaling van lekseme en woordklas in 'n polisintetiese taal, die verkryging van gebruiksvoorbeelde uit 'n klein korpus en die hantering van sosiolinguistiese variasie in die keuse en werking van die inhoud.

Sleutelwoorde: GEBARELEKSIKOGRAFIE, AANLYN WOORDEBOEKE, MULTIMEDIA-WOORDEBOEKE, TWEETALIGE WOORDEBOEKE, AANLEERDERWOORDEBOEKE, NIEU-SEELANDSE GEBARETAAL, VIDEO-INHOUD, GEBARETAALKORPUS, POLISINTETIESE MORFOLOGIE, POLISEMIE, SOSIOLINGUISTIESE VARIASIE, GEBARETAALLINGUISTIEK, GEBRUIKERSPROFIEL

Introduction

Capturing the lexis of a signed language in a bilingual dictionary requires macro and micro decisions about issues of lemmatisation, ordering, variants, grammar, sense and usage. An even more fundamental challenge for sign language dictionary makers has traditionally been modality — attempting to describe a visual-gestural language in a static format. A decade ago, De Schryver (2003: 143) said this about electronic dictionaries:

The arrival of the modern computer set in motion a series of lexicographers' dreams without equal in the history of dictionary making. Achieving the wildest of those electronic-dictionary vistas has the potential to result in reference works beyond all recognition.

In the 21st century, the electronic medium has indeed transformed vistas for makers and users of signed language dictionaries. The digital, online platform affords a vastly better fit between the dynamic nature of a signed language and possible dictionary formats. This article describes how the electronic medium was leveraged in making an online dictionary of New Zealand Sign Language (McKee et al. 2011). The aims, design and production of the dictionary (hereafter abbreviated as ODNZSL) are described, and some perennial issues for sign language dictionary makers are illustrated, including providing usage examples, determining word class and citation forms, and dealing with polysemy and mouthing. Sociolinguistic considerations of lexical variation and performing the lexicon on video are also discussed.

Aims and impacts of sign language dictionaries

The making of sign language dictionaries has been important in defining Deaf people as distinct language communities. In many countries, the production of a national sign language dictionary has been the foundational piece of language documentation leading to further attestation of the linguistic and cultural status of Deaf communities. Involvement in language documentation

activities raises consciousness of linguistic identity within the primary language community, and the artifact of a dictionary can support recognition of language minority status (Lucas 2003, Padden and Humphries 2005, Hualand and Allen 2009). Documenting the lexicon serves to distinguish the identity of a national Deaf community from — as well as aligning them with — signing communities in other countries or regions. However, promoting the impression of a unified national sign language via dictionary content can be misleading, and potentially prescriptive, in contexts where sign language use is actually locally variable (Lucas 2003; Reagan 2010; Schmaling 2012). In the African context for instance, Schmaling (2012: 273) observes that informants to national dictionary projects have often not been representative of the country's Deaf population, having been exposed to foreign forms of sign language via schooling and contact with overseas sign language users in development projects; as such, dictionary products may obscure the localised linguistic diversity in the wider spectrum of Deaf people who live outside capital cities and who are not formally educated. The *Dictionary of Southern African Signs* (Penn et al. 1992–1994) is noted by Schmaling as a work that purposefully addressed representativeness by documenting variants from at least eleven regional varieties identified within the South African Deaf community.

Sign language dictionaries create an avenue for outsiders to knowledge about the language. Common motives for sign language dictionary projects are to support the use of sign language as a medium of instruction in deaf education, and to increase Deaf people's participation in wider society through language recognition and the training of interpreters. Typical aims are captured in Ashipala et al.'s (1994, as cited in Lucas 2003: 325) statement about a Namibian sign language dictionary project: "We want people to know that NSL is a real language like Oshiwambo or English. We want Deaf Namibians to be proud of their Deaf culture and sign language, and we want hearing people to respect our culture and our sign language." In similar vein, the introduction to the print *Dictionary of New Zealand Sign Language* (Kennedy et al. 1997: ix) expresses goals of legitimising NZSL as a 'real' language through research and documentation, improving communication between Deaf people and hearing people by providing a bilingual learning resource, and helping to improve Deaf people's access to education and other social services.

ODNZSL (2011) is a third generation dictionary, following from two previous dictionaries that achieved initial documentation and language recognition. NZSL was made an official language of New Zealand by statute in 2006 (NZSL Act 2006), only 21 years after the first dictionary was made (Levitt 1986), and nine years after a larger dictionary appeared (Kennedy et al. 1997). Although a combination of policy, education, advocacy and research initiatives have all contributed to status change for NZSL, dictionary publication has played a key role in consolidating its public profile as a bona fide language used by a New Zealand community.

Development of NZSL dictionaries

The first NZSL dictionary was produced in 1986 by Dan Levitt, an American interpreter contracted to train the first cohort of sign language interpreters in New Zealand. In the absence of documented NZSL resources, language classes relied upon rosters of Deaf visitors as live language models. During class sessions, their signing was video-recorded and subsequently described and compiled in a dictionary. Levitt's photographic dictionary features many older signers, making it a valuable record of earlier signs that have since been replaced by modern variants. In making the 1986 dictionary, community members expressed a preference for photographic illustrations; they could not easily visualise signs disembodied from their users. The 1200 entries are alphabetically ordered by English gloss, in a uni-directional English–NZSL format. Variant forms are included. The Levitt dictionary promulgated the name 'NZSL' in the discourse of Deaf and hearing communities, and was an important artifact in bringing the existence of the language into the public domain.

The next dictionary (Kennedy et al. 1997, hereafter DNZSL) was a collaborative effort between the national Deaf Association and Victoria University. The project was led by Professor Graeme Kennedy, an editor of the *Oxford Dictionary of New Zealand English*, who had also supervised a PhD thesis describing the grammar and lexicon of NZSL (Collins-Ahlgren 1989). The methodology of the DNZSL underlies the content and record structure of the current ODNZSL, and is therefore outlined briefly here; (a more detailed account is found in the Introduction of DNZSL).

The 1997 dictionary contains 4,500 entries and was completed in six years, comprising 20,000 hours of work by paid staff, and uncounted hours of voluntary work by Deaf community members (Kennedy, p. xi). Kennedy took a systematic approach to content selection by establishing a taxonomy of semantic domains common to most languages, as well as a list of words (in English and Māori) pertinent to New Zealand life and to Deaf culture. Using this 'concept net' of approximately 7,000 words as a guide, Deaf groups were recorded conversing freely about suggested topics, and responding to lexical elicitation prompts. From the hours of resulting video recordings, signs were identified, re-filmed as citation forms, and then viewed by representative Deaf groups in three regions who attested whether the sign was recognised and/or used in their region, whether another variant existed, and what the sign meant in context. Validated signs were phonetically analysed and transcribed using the Hamburg Notation System (HamNoSys; Prillwitz et al. 1989), then drawn with the addition of movement symbols, and described in terms of their formation and usage, to create records for dictionary entries.

HamNoSys is a sign transcription system for describing the phonological parameters of a sign that were originally identified by Stokoe (1960) as hand-shape, place of articulation, movement, and orientation of the palm and fingers. The set of HamNoSys symbols representing all permutations of these parame-

ters allows the phonetic features of any sign to be coded in a linear string, in the order of handshape, orientation, location (i.e. place of articulation), movement (as seen in figure 1). The DNZSL was one of few dictionaries at the time of publication to order its entries according to their handshape and location identity, rather than alphabetically by English translation (Zwitserslood 2010). DNZSL entries are grouped into 27 handshape categories, each of which includes up to eight related variants. A pictorial index of handshape categories, which also appear as section headers on dictionary pages, enables a user to identify the section in which a given sign should be found. Within each handshape section of the dictionary, signs are ordered by their place of articulation, starting from those made near the top of the head and moving progressively downwards in the signing space. Organisation by sign features is necessary to bi-directional access, allowing a user to search by the visual appearance of a sign if an English translation is not known. Privileging the formational features of signs in the dictionary layout also enhances users' receptive awareness about the form of signs, by having to attend to structural details that distinguish similar forms from each other. Nevertheless, this organisation is not familiar to new dictionary users, and can be challenging to use with precision. An English index at the back allows users to also search by English glosses. Figure 1 shows the elements and layout of an entry in the DNZSL.

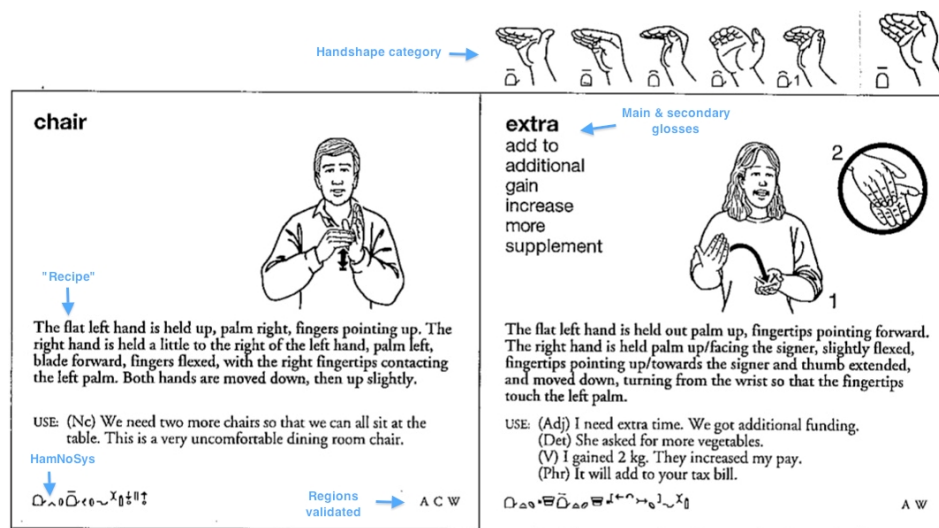


Figure 1: Entry structure in DNZSL (1997: 133)

"Online ease" factors

De Schryver (2003: 159) identified numerous advantages of electronic and online dictionaries described by lexicographers from 1980 onwards, and espe-

cially between 1990–2000. All of the 'online ease' factors that De Schryver lists are realised in the ODNZSL, and we comment on the relevance of these points (cited in italics) below.

"No space constraints, can handle huge quantities of data"

Migrating the existing NZSL dictionary database to an online interface allowed the addition of new data fields within records (such as sociolinguistic variation, semantic domain, usage examples) and enables the continuous addition of further entries. Schmaling (2012: 244) comments that while it is ideal for dictionaries to include all known variants, in practice this has often been unfeasible in sign language dictionaries because of limits on production resources and size. The Concise NZSL dictionary (Kennedy et al. 2002) containing the 2,000 most frequent signs was published in response to a need for a more portable volume than the original three-kilogram tome containing 4,500 signs on 800 pages. The online medium mitigates size and selection pressures, allowing for a more inclusive resource that is less constrained by the production, cost and usability parameters of a hard-copy text. According to Zwitterlood (2010: 454), most contemporary electronic sign language dictionaries have between 2,000–5,000 entries, and most print dictionaries have fewer than 2,000; so by current standards, both the print and online dictionaries of the NZSL are on the larger side, containing over 4,000 signs. Nevertheless, the ODNZSL represents only a portion of the established NZSL lexicon; its actual size is unknown, as is true for most signed languages (Johnston and Schembri 1999: 176). In light of the incomplete documentation of sign lexicons, online lexicons are vital to enabling ongoing description by future researchers.

"New types of information"

Moving to an online format allowed for re-design at all levels: the micro-structure of content within an entry, the macro-structure of dictionary organisation and user access to content, and the mega-structure of all elements in the dictionary (Hausmann and Wiegand 1989).

The micro-structure of entries was shaped by data fields established in the 1997 print dictionary; those that were transferred into the ODNZSL included a drawing, main gloss, secondary glosses, handshape category, grammatical and usage information. The ODNZSL entry components incorporated video clips of signs and usage examples (with English translation) and additional usage tags. The literal gloss of each example sentence not only shows the NZSL syntax, but allows users to click on any hyperlinked item in the sentence to go to the entry for that sign, providing cross-referencing.

The online format allowed for the addition of variation information (e.g., region, age, sign status) and hyperlinked tags explaining grammatical properties of signs. Since the website launch, Māori equivalents for all English main glosses have been added, to enable users to search for signs by a Māori word and to see lexical equivalence (or approximations thereof) between the three official languages of New Zealand.

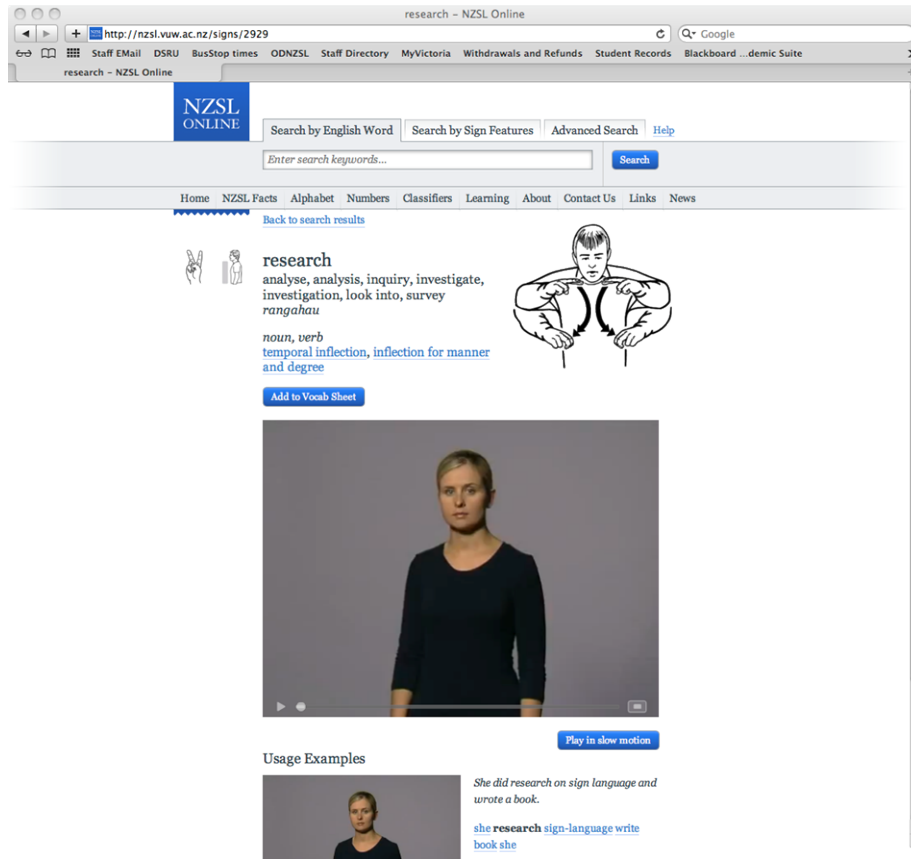


Figure 2: Entry display in ODNZSL (2011)

Relational searching enabled by the online medium considerably reduces the macro-structure issue of 'ordering'. Signs can be searched and displayed in ways that are more transparent than navigating a print volume organised according to phonological principles that are opaque to naïve users. Information about signs can be searched in various ways and combinations: via English word equivalents (which we also refer to in this article as main and secondary glosses), or by handshape and location features of a sign, or by 42 topic domains, or by tags for usage status (obscene, archaic, neologism, informal and rare). It was fortunate for the ODNZSL project that 1997 DNZSL editors made the (theoretically bold, at the time) decision to organise that dictionary according to sign form rather than alphabetically by English translations; hence, the immense task of phonetic analysis and coding of handshape and location features for each sign had been completed for all existing records. This HamNoSys coding was utilised as the basis for the 'search by sign features' in ODNZSL.

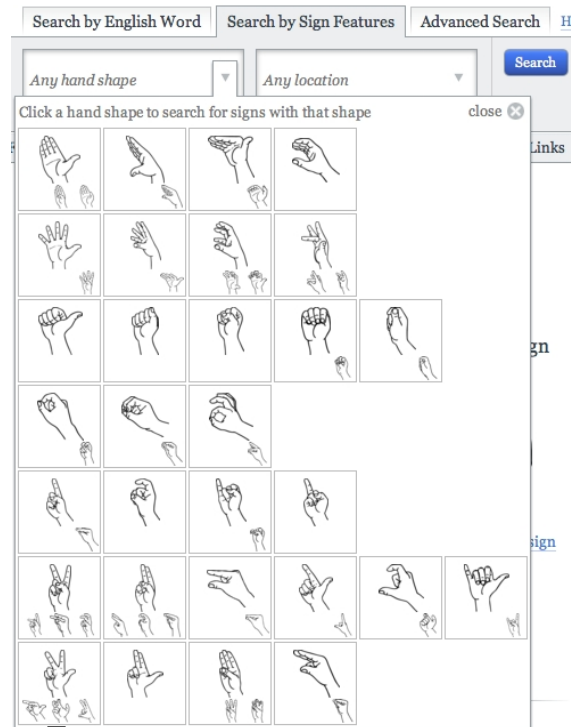


Figure 3: Handshape search menu

The dictionary needed to address both receptive and productive language needs of users, which requires being able to search by the visual form of a sign. Handshape and location features can be combined with topic tags in an advanced search, which is potentially helpful to a user who sees an unfamiliar sign used, and has an approximate knowledge of its form and domain, but not its meaning. Considering that users of the ODNZSL would be encountering a novel format for a dictionary, the design of 'search by sign' was informed by user consultation. As there is no equivalent of a conventional 'alphabet' for ordering signs, phonological features of handshape and location are used as the identifying parameters — which for users, is more analytically demanding than a 'search by word'. Handshape and location features need to be grouped and displayed in a logical manner in the 'search by sign' menus to facilitate user recognition of the most salient features. We examined recent online dictionaries of Danish Sign Language (Center for Tegnsprog 2008) and Finnish Sign Language (Kuurojen Liittory 2003), and their formats helped us to devise appropriate tables of NZSL handshapes and locations (see figures 3 and 4) that can be selected by clicking on pictures.

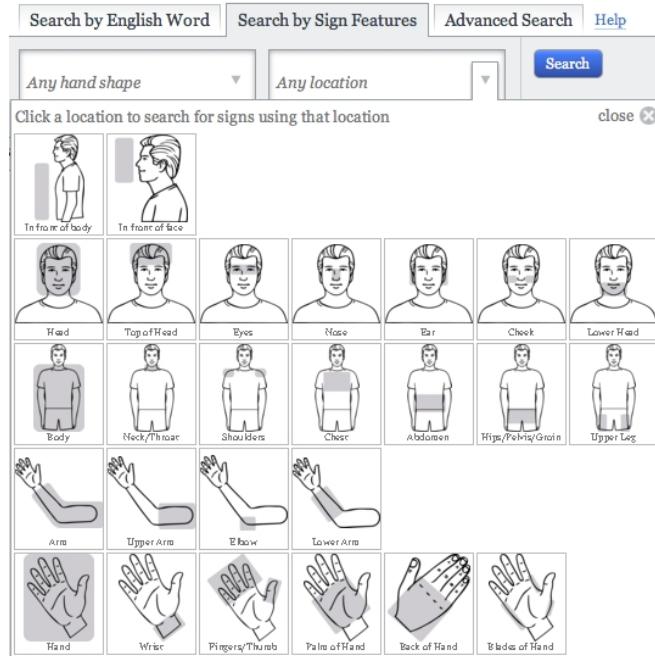


Figure 4: Location search menu

A handshape/location search yields a result list of forms that match the search parameters more or less closely; these are displayed by sign illustrations and hyperlinked glosses, which can be browsed directly from the results page, or refined further by returning to the sign feature menus. To determine the final format of these menus, adult NZSL students were given a pilot task to test how accurately they could identify salient handshapes and locations to find a set of target signs in the dictionary. Results indicated that users were able to make quite accurate choices, and pointed to cases of ambiguity which informed the final organisation of these menus (e.g., which handshapes to display at the top/category level, and which as sub-handshapes, and some that were not necessary distinctions for searching).

In its mega-structure, in addition to entry content, ODNZSL contains static information tabs that provide 'front' and 'back' matter about how to use the dictionary, history and grammatical aspects of NZSL, systems of numbers, finger-spelling, classifiers, attributions and acknowledgements, and links to NZSL-related organisations and learning resources. The front matter pages contain a video window that renders the information in NZSL translation, making the dictionary bilingual at the mega-structure level, and thus more accessible to Deaf NZSL users. A 'News' tab communicates relevant news about the dictionary to users, and offers a Facebook link, which has proven a useful means of dis-

seminating awareness of the dictionary. The homepage of the dictionary has a 'Show me a sign' display, which shows a sign illustration, gloss and entry hyperlink; these are randomly selected from within the database (excluding obscene signs after some early user feedback), and change at each log-in; this emulates a 'word of the day' feature in other online dictionaries.

"Video sequences, animation"

Signed languages are characterised by their visual-spatial production modality; transitions of movement and handshape are vital to distinguishing the form and meaning of signs. Analysing and representing these articulation features accurately in a dictionary is important both to the searchability of content, and to its value as a learning tool. The capacity for video content has been well utilised in recent online dictionaries of Flemish Sign Language (Van Herreweghe et al. 2004), Swedish Sign Language (Institutionen för Lingvistik 2009), Danish Sign Language (Center for Tegnsprog 2008), and Finnish Sign Language (Kuurojen Liittöry 2003). Although the line drawings in the print NZSL dictionaries (1997 and 2002) were of high precision, it is transformative to be able to represent signs and their contextualised use through video clips. Learners particularly appreciate the slow motion playback, which allows the articulation of signs to be closely observed. The two-dimensional view that video affords required consideration of camera angles when filming certain signs to maximise the perceptible contrast in hand position or movement. Access to video information increases language learners' independence in building vocabulary knowledge. Whereas learners previously needed to ask a teacher or native speaker, "What is the sign for (...)?", to get a demonstration, they can now ask the dictionary for a (repeatable) video rendition of the vocabulary item. Lexical knowledge is also enriched by incidental discoveries in the dictionary about usage, variants and translations that a teacher might not have offered up. Finally, video content furnishes a repository of data for other researchers to refer to in cross-linguistic studies.

"Links with other software (e.g., learning exercises)"

The user survey indicated that learner exercises would be 'nice to have' in an online dictionary resource. ODNZSL contains a link to an interactive practice website for learners also developed at Victoria University. A vocabulary 'save and print' function allows users to select, caption and print customised sheets of sign illustrations from entries, which is valuable for learning and teaching purposes. Further learning links will be added as resources allow.

"Up-to-date and dynamic repository"

Following migration of print dictionary content to the online database, all records were checked and many were updated with clarifications of form, meaning, usage and grammatical tags. The capacity to continuously and rapidly update content is a great advantage (the server integrates any revisions

into entry display overnight), particularly as NZSL is still developing and being described. Community ownership of the dictionary is enhanced by the capacity to show a rapid editorial response to user feedback about content which is invited via an email link in the website. (Facebook and YouTube options for user feedback are currently being explored.)

"Cheap, if not free"

For both practical and principled (social impact) reasons it was decided to make ODNZSL openly available without registration or subscription. Free access ensures that the primary language community of Deaf people has equitable access to it as a resource, and can readily refer others to it. Open access softens the borders of traditional (and historically private) domains of sign language transmission, such as deaf schools, homes, clubs and community events, and more recently, NZSL classes. Its availability on the web and the portability of the medium helps to normalise it in the linguistic landscape. Importantly, NZSL becomes more available as a resource to hearing families and educators of deaf children who are not socially connected to a Deaf community. Since the launch in June 2011, informal feedback reveals an unexpectedly wide variety of users, including workmates of Deaf individuals, service providers in customer contact positions, teachers of various subjects who incorporate NZSL as an additional modality into classroom teaching, 'baby sign' classes for parents of hearing infants, schools, and even a children's rugby coach who consults it for translating instructions to a deaf player. These adventitious uses of the dictionary in wider society indicate the extended reach of a free, online resource.

The dictionary carries a Creative Commons license,² which permits content to be shared and adapted, with attribution, for purposes such as developing learning resources. Under this license, an independent software developer has created a free application that allows key elements of the dictionary to run on mobile devices, and is now working on a flashcard application. Enabling such initiatives extends the educational potential of the dictionary and harnesses a wider pool of creative expertise than our research unit has at its disposal. As lamented by many publishers of electronic reference tools, however, the downside of being free is the absence of sales revenue; securing funding for maintenance and development of the dictionary promises to be more difficult than attracting initial project funding.

Steps in making ODNZSL

The ODNZSL project was the initiative of the Deaf Studies Research Unit at Victoria University, working in partnership with educational and community stakeholders. A grant of NZD\$750,000 for a three-year project was obtained from a government funding source.³ The editorial team comprised a managing editor, two consulting editors, a database manager, and two research assistants.

All personnel worked on the project on a part-time basis. None had formal lexicography training, but most had previous relevant experience with sign linguistics, dictionary and corpus projects. A technical editor (database programmer), illustrator, translators and web designers were also contracted.

Effective dictionary design proceeds from a realistic understanding of the purposes and skills of users (Atkins and Rundell 2008). The ODNZSL project team included NZSL teachers and interpreters who brought experiential knowledge of likely user needs; we knew that the dictionary should serve both receptive and productive uses, and expected that the most frequent use would be language learners seeking NZSL equivalents for English words. Although we had a good sense of the dictionary audience, user needs were investigated at the outset of the project by a written survey of people in the following categories: teachers and teacher aides of Deaf children; adult learners of NZSL (in tertiary and community education); NZSL teachers (mainly Deaf); Deaf community members; NZSL interpreters; deaf students in schools; mainstream school students and teachers; parents/families of deaf children/adults (mainly non-Deaf). 301 survey responses were received, informing decisions about functionality and content that would meet the expectations and needs of the majority of users.

The first step was to migrate the 1997 DNZSL data to an online database in which content could be edited and re-formatted. Fortunately the ODNZSL technical editor worked on both projects, which expedited this transition. The lexicographical tool designed by the technical editor was Freelex,⁴ an open source application that enables multi-user, online workflow, editing of data fields within records, quality assurance (QA) management, corpus search, and reports on user activity and record history.

Following migration, all record data was checked and updated as needed. New records were added, including further variants, neologisms and borrowings that had developed in NZSL in the 20 years since data was collected for the 1997 dictionary. In relation to a changing lexicon, it is relevant to note that during the past 25 years, provision of interpreting services and language recognition has lowered the barriers for NZSL users to participate in domains of higher education, white-collar occupations, the arts, media, and political advocacy. Deaf people's greater access to information and societal discourses has driven rapid growth in the NZSL lexicon, through coinages and borrowing from Australian, American and British Sign Languages.

All candidate signs to be added to the ODNZSL were validated by community groups in three regions. Although ODNZSL was intended as a general rather than a specialised dictionary, the two Deaf Education Centres were invited to submit signs relating to school curriculum subjects that they had already documented, for potential inclusion in the dictionary. In fact, many of the items submitted were very new coinages, nonce translations for English terms that had only contextually recognisable meaning, or coinages for concepts that could be translated by semantic extension of conventional signs

already in the dictionary (listed under an alternate English translation). As many of the suggested items apparently had marginal lexical status in NZSL, (some used only within a particular school or class), or already had conceptual equivalents in the established lexicon, approximately half of the suggested 'school signs' did not pass validation in a process of cross-school, and adult community consultation. Approximately 100 were added to the dictionary. For each new validated addition, a citation form was video-recorded for the purpose of making an illustration, and data fields for the record were populated.

During the revision of the whole database, some existing records from the previous dictionary failed QA for various reasons (e.g., duplicates or nearly identical phonetic variants) and are not displayed in the current ODNZSL. The total number of entries displayed in ODNZSL is 4070 with a further 2176 undisplayed records in the database. The total number of records in the database is 6,246.

The next main phase was to film citation forms and usage examples for each sign. The process of making usage examples from corpus data is discussed in more detail later in this article. A script for filming was created by typing example sentences in gloss form. Signs and usage examples (between one and four examples per entry) were filmed in five 'batches' of five-day blocks, over an 18-month period. Editing, uploading and QA of the clips, and further filming preparation was undertaken between batches of filming. Film personnel comprised a rotation of eight signing models (working in pairs for any given shoot), a cameraman, an assistant to log film shots, a director (the managing editor) to monitor accuracy of performance and consistency with dictionary illustrations, and a coach from the dictionary team to rehearse the signers. Over 11,000 clips were recorded, edited, quality assured, saved on a server in .mp4 format and linked as 'assets' to records in the database.

Each filmed usage example was translated into English, displayed beside the video window in the entry display. Below the translation in the entry is a literal syntactic gloss of the sentence in NZSL. To enable cross-reference to surrounding signs in the usage example, each sign gloss in the gloss sequence was tagged with the ID number of its corresponding record in the database, creating a hyperlink to the entry for that sign in the online dictionary. This feature was modeled on the Danish Sign Language Dictionary (Center for Tegnsprog 2008).

A web design company worked closely with the editing team to create a user interface for the website, and to integrate this with the database search engine. Ideas about features and functions of the website were drawn from review of recent online sign language dictionaries, in particular, the Danish and Finnish (Kuurojen Liittöry 2003) dictionaries. At the final phase, a range of users (first and second language signers) trialed a set of search tasks within the website, and provided feedback on problems encountered; this feedback was used to tweak functionality prior to launch.

Corpus-based evidence

De Schryver's final 'online ease' factor is "*Rapid access to lexicographical evidence in corpora*". The availability of electronic corpora has revolutionised the sourcing of lexicographical evidence for dictionaries of major spoken languages. The majority of sign language dictionaries, however, have lacked any usage information let alone contextualised examples based on attested use. Content selection is often based on native speaker introspection or consultation rather than upon systematic analysis of a large body of data (Johnston 2003; Wilcox 2003; Zwitserlood 2010). This is a potentially unrepresentative basis for a dictionary, as perceptions about language use often differ from actual usage (Atkins and Rundell 2008).

The chief constraint on corpus-informed dictionaries of signed languages is the practical difficulty of developing machine-searchable corpora for visual languages without a conventional written form.⁵ Rapid advances have been made recently in sign language corpus development, with the advent of software that enables searchable transcription and annotation of time-linked video files. ELAN (EUDICO Linguistic Annotator), originally developed for research on gesture, has become the standard tool for this, as it allows one to create, edit, visualise and search multi-tier annotations for video and audio data (Johnston and Schembri 2005; Leeson and Nolan 2008). A sophisticated multimedia relational database, 'iLex', has also been developed at the University of Hamburg for the purpose of sign corpus annotation and lexicography (Hanke and Storz 2008). Nevertheless, transcription of signed lexemes is done via a written gloss, and/or at sub-lexical level via a phonetic notation system — a hugely time intensive analytic process that poses issues such as the determination of lemmas, lexemes, and phrasal units in a signed language (Johnston and Schembri 1999), and the representation of non-manual, spatial and productive elements of the lexicon in use (Miller 2001; Brennan 2001). Sign language corpora are now being developed in numerous countries, although the scale remains relatively small in comparison to spoken languages, due to the labour intensive nature of transcription and the small number of people who have the analytic skills to do this work.⁶

Although it does not provide direct user access to a corpus, the ODNZSL is one of few signed language dictionaries to have utilised an electronic corpus as the main source of usage examples. The corpus is small and flawed in several respects, but its use is a step forward in providing lexicographical evidence for a sign language dictionary. Following the 1997 print dictionary of NZSL, the general editor, Graeme Kennedy, decided to create a first corpus of NZSL in order to analyse lexical frequency to guide compilation of a concise dictionary, and to inform teaching curricula (McKee and Kennedy 2006). The Wellington Corpus of NZSL (WCNZSL) consists of more than 50 hours of videotaped NZSL discourse by members of the Deaf community. 80 Deaf people

were videotaped taking part in 30 conversations and 43 monologues in discourse events that included conversations, narratives, speeches, meetings, discussions on more than 30 different topics, including school and education, sport, work, health, childbirth, and politics. Over one year, videotapes were transcribed into English glosses, creating a file consisting of 100,000 running signs, comprising 7,222 types or lemmas (McKee and Kennedy 2006).

An example of the transcribed corpus is shown in figure 5. By contemporary standards, this format is primitive in that the transcript is detached from the original video sources and lacks annotation; it is also imperfectly lemmatised, which demands completely consistent glossing of every type and token (see Johnston 2008). Non-manual syntactic markers and inflectional verb morphology expressed in movement and space are not captured (in other words, just 'bare' signs are represented). However, the text form of the corpus allowed the use of a concordance tool, WordSmith, to analyse lexical distribution (the findings of which are reported in McKee and Kennedy 2006). Analysis revealed that only 15 highly frequent signs (in the top 500) were missing in the original 1997 DNZSL, confirming that its coverage was quite representative. In the subsequent Concise DNZSL (2002) the most frequent 2,000 signs were supplemented with less frequent, but nevertheless, important signs, such as names of countries and cities, and vocabulary relevant to educational or health domains.

```
IX-2 KNOW IX-2 FAMOUS HORSE IX-loc AUSTRALIA IX-loc 1 MAN  
FROM HERE HORSE-TROTTING NAME fs-SHANUE fs-DYE PRAM  
KNOW IX-2 FAMOUS IX-3 MELBOURNE MELBOURNE CUP  
MELBOURNE CUP WIN FIRST SCL-1-horse-leads FIRST 3 YEAR PAST 3  
YEAR PAST IX-3 POS1 FRIEND IX-3 IX-1 LONG-TIME-AGO SMALL-  
CHILD IX-3 FATHER IX-3 fs-WAS POS1 FRIEND IX-3 IX-1 IX-3 NMS-  
nod IX-3 IX-3 BORN IX-3 IX-3 FATHER SISTER IX-3 FAMILY AREA  
OLD GOOD TOGETHER GOOD EACH-OTHER IX-3 IX-3 fs-SON IX-3
```

Figure 5: Extract from Wellington Corpus of NZSL

Between 2005–2008, a further 14,000 tokens from 81 conversational data clips, recorded in a sociolinguistic variation project, were added to the corpus. This data was transcribed using ELAN annotation software, a screenshot of which is shown in figure 6.

Text files from the two projects were combined to form a corpus linked to the ODNZSL database. Figure 7 shows the appearance of this corpus within the editing database.

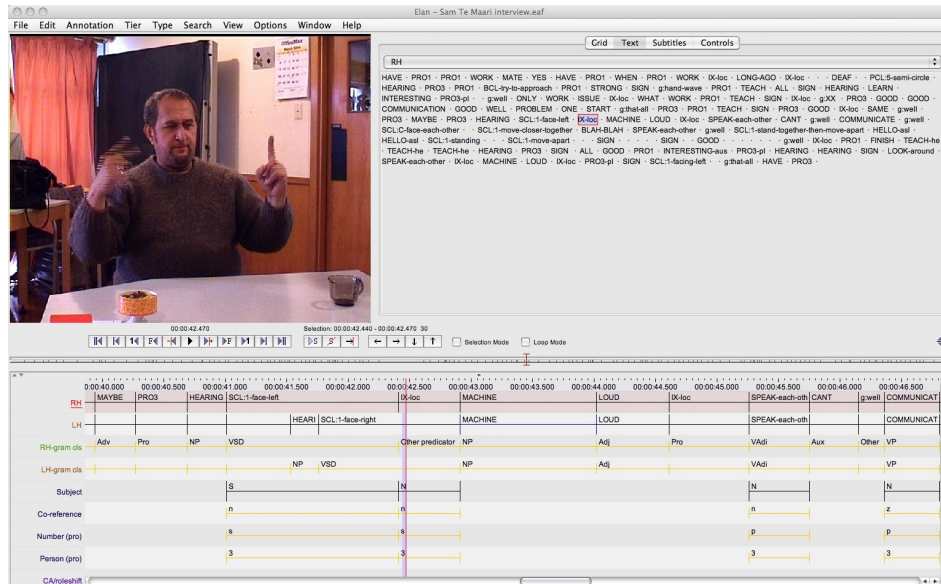


Figure 6: NZSL transcription and annotation in ELAN

As Atkins and Rundell (2008: 56) acknowledge, "Corpus creation is a pragmatic exercise, with compromise at all phases — design, data collection, and encoding." Although this corpus of 114,000 tokens is certainly compromised by its smallness (by spoken corpora standards) and imperfect sampling, it nevertheless represents a wide range of topics and speakers engaged in authentic discourse, and as such, provides valuable usage information about the lexicon. Corpus data informed the consideration of questions such as: Is the sign FIGHT used metaphorically as well as literally? Does the sign AFFAIR have a nominal or only a verbal sense? Which forms of HOT collocate with human vs. non-human subjects? In the editing process, the corpus was a valuable resource in making decisions about sense, word class, age and region-related usage (from metadata on signer identity), and provided the basis for example sentences, as explained in the next section.

The 1997 and 2002 NZSL print dictionaries include example sentences to illustrate sense and grammatical usage, but these are presented only in English, and so it is unclear whether they derive from authentic contexts of use in NZSL. Nor do they directly show a grammatical context in NZSL. It was therefore an advance for the ODNZSL to include filmed usage examples that are more authentic and informative with regard to both form and content. As the dictionary does not include definitions (meaning is indicated through English translation, sometimes supplemented by notes), usage examples are important for illustrating contextual features of both meaning and use.

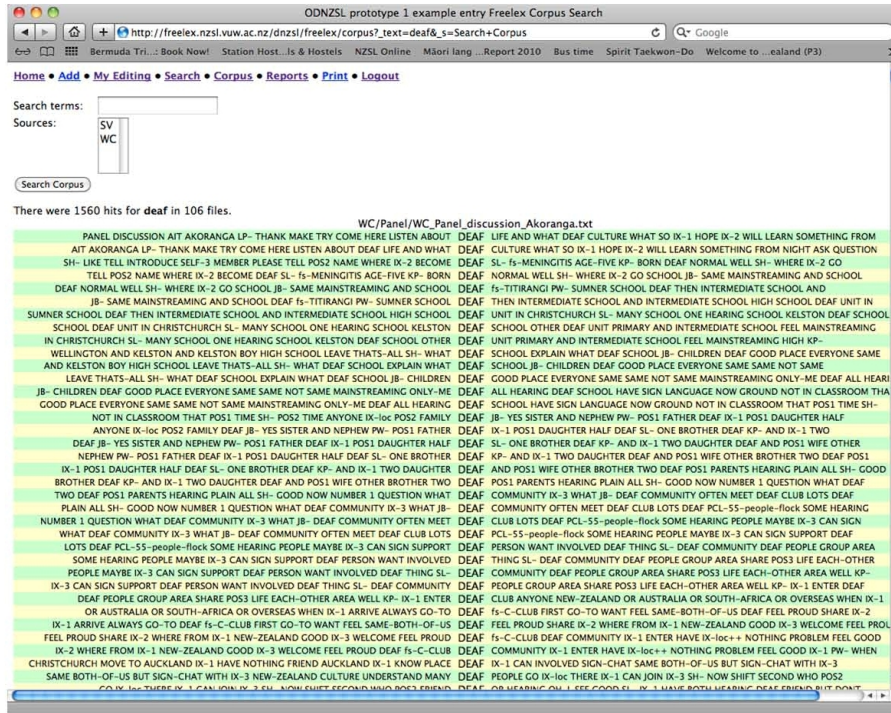


Figure 7: NZSL corpus in Freelex database (showing concordances of DEAF)

To create usage examples, the corpus was searched for tokens of each headword, and up to three utterances were selected as the potential basis for examples. Some utterances in the corpus could be used verbatim, but most required editing. Our experience concurred with Atkins and Rundell's (2008) advice that most examples in a learners' dictionary will need to be modified from raw data in order to be accessible and maximally informative to various users. We adopted Atkins and Rundell's guiding principles that usage examples should be: (i) *Natural & Typical*: the word is shown in its most usual context, syntax and collocation; not idiosyncratic usage; not mixing registers or varieties; (ii) *Informative*: the sentence gives informative context (helps understand the sense of the word). (iii) *Intelligible*: the sentence contains no words that are more difficult than the headword; clear structure; succinct.

Our procedure for creating examples from corpus citations included removing identifying references to person, place and event, and offensive matter that was not essential context for the target word. In light of intelligibility for learners, extraneous clauses or unusual contexts that did not clarify typical sense or grammatical use were removed. As a general rule, each headword illustrated in an example needed to appear in an uninflected form, whereas it might have been inflected in the corpus data. Examples also had to be decontextualised

enough to stand alone as an intelligible utterance, which is not always the character of discourse among interlocutors who share real-world experience. Citations that required implicit knowledge of the immediate or wider conversational context were discarded or re-worked to make comprehensible examples.

With learners in mind, we tried to ensure that most signs within examples were relatively high-frequency and appear in the dictionary, enabling consistent cross-referencing. The imperative for clarity, and the fact that each example would have to be performed to camera and later translated into English, also led us to avoid certain structures particular to signed languages, including the following: 'constructed action', a quotative device for enacting reported speech or action, which can be convoluted to translate 'classifier' (polycomponential) constructions that depict objects and actions local to the discourse context, which are potentially wordy to translate, and not possible to cross-reference to dictionary entries as they utilise productive rather than conventional lexicon (Johnston 2001).

Due to the nature of certain signs, we made some exceptions to the guidelines; for instance, it is impossible to illustrate a usage context for an imperative such as 'Stop it!', or an exclamation such as 'You-idiot', without embedding it in constructed dialogue. Certain signs typically collocate with other signs that are low frequency, colloquial, or offensive. In these cases, we favoured the principle of 'natural, typical' usage. In some cases, examples were modified in the filming studio by the editor and sign model, if it became apparent that a sentence was too complex, or lacked sense out of its original context.

The translation phase (post-filming and QA of video examples) revealed some complications that in hindsight could have been minimised; for example, we wished that we had been more strict in applying the rule articulated for the Danish dictionary (Kristoffersen 2010) of avoiding spatially-encoded anaphoric pronominal reference, or null person reference (pro-drop). These structures are typical in signed language grammar, but difficult to deal with in English translation, which allows little ambiguity of person reference. Also not ideal were examples in which the sense of the target sign in NZSL, in the context of the example, had to be translated as a different word form than the main gloss displayed (e.g. main gloss — 'equal', while translation had to be 'equality'). In such cases, we made use of a secondary gloss in translation.

The goal of deriving all usage examples closely from corpus data was not achievable, as the Danish Sign Language dictionary makers also reported (cf. Kristoffersen 2010). We met corpus-related limitations: almost 50% of headwords were not found in the corpus, due to its limited size and some inconsistency of type labels for signs in the corpus. Another problem was that variants or synonyms for a single meaning were not consistently identified by distinct glosses in the corpus, making it impossible to reliably identify different sign forms in the corpus. Conversely, some polysemous or homophonous sign forms were glossed inconsistently according to meaning in context, rather than by a uniform label (see Johnston 2008, 2010 re. ID glossing issues in signed corpora). Unfortunately,

these transcription flaws became more apparent years after the fact, when trying to utilise the corpus in a dictionary. In the end, the usage examples reflect "pragmatic compromise" (Atkins and Rundell 2008), being corpus informed, but modified in the interests of achieving accessibility for users, and performability and translatability. Approximately 30% of examples were devised from scratch by a native speaker in the editorial team.

Nonetheless, a noteworthy benefit of corpus-based examples in the ODNZSL is their reflection of the Deaf cultural frames that underpin the lexicon in use. Fillmore (2003: 284) explains frames as the "background set of beliefs, practices, institutions, or ready-made conceptualizations available to the speakers of the language as the necessary underpinnings of the way they speak". The way signs are used by Deaf people may invoke frames that are particular to their shared experiences and understandings. For instance, in ODNZSL the usage example for 'family' is glossed in NZSL as, MY FAMILY ALL HEARING ME ONE DEAF (translation: "My family are all hearing. I'm the only Deaf person"). This statement is common in Deaf personal narratives, invoking mutually understood implications of this isolated status with regard to family relationships, language acquisition, and social identity. The example also shows a frequent collocation of the signs 'one/only' and 'Deaf' in NZSL users' description of themselves in social contexts with hearing people. Similarly, the entry for the sign 'oral/speaking' illustrates a distinctly Deaf worldview; the example is glossed as, HE ORAL SPEAK GOOD CAN HE (translation: "He can speak well"). While the translation is understandable on face value, it cannot convey the implicit reference to a social categorisation of Deaf individuals into those who speak more intelligibly or more often, and might therefore have a stronger affiliation with the hearing world, in contrast to those who don't, and may identify (and be identified) more strongly as NZSL community members. Culturally rich examples such as these are a valuable element of ODNZSL content. Nevertheless, we agree with Fillmore's (2003: 269) argument that dictionaries could usefully go further in making explicit the background frames that underpin native use of lexical items, especially for non-native users wanting to understand why a word exists and why a speaker would choose to use it in a context — in effect, providing ethnographic reference. Augmenting dictionary entries with encyclopedic-type definitions, where relevant, is not unfeasible in an online format (given 'no space constraints'), and is an interesting prospect for future development.

Performance of usage examples and sociolinguistic identities

Issues of identity and social structure often surface in lexicographical work on minority languages (Axelrod et al. 2003). Producing video content for the ODNZSL highlighted sociolinguistic considerations about form and style of language. Video clips are embodied linguistic performance by individuals who have social identities and styles that are immediately recognisable to their own

language community, and which will be perceived by learners and other dictionary users as prototypical models of the target language. Decisions had to be made regarding who should demonstrate the language: what is the ideal, or acceptable, embodiment of gender, age, ethnicity, regional origin, native language status, body stature, hair colour/style, sexual orientation? Given the impact of these personal factors on signing style and user perception of normative language models, we aimed for authenticity and diversity of representation, accepting the fact that there is neither a neutral nor ideal speaker of a language. To cover as many bases as possible with regard to sociolinguistic identities and style, the eight sign models comprise a mix of younger/older, native/non-native, female/male, Māori/non-Māori/Pasifika, northern/southern region individuals. Metalinguistic awareness of phonetic and grammatical form was also necessary in the signing models, as was the ability to perform scripted language 'naturally' on cue, in front of a camera for hours at a time. These requirements, and the level of scrutiny that the dictionary models would attract, narrowed the pool of eligible candidates for the job.

Also considered in filming was whether lexical variants more associated with certain social groups should be performed by a signer with matched identity characteristics. Although content of usage examples was not authored by the sign models, its animation (Goffman 1981) by a particular person may be perceived by viewers as representing a connection with the visible identity of the signer. For practical reasons, we only attempted to match variant to signer characteristics in the case of signs used mainly by older or younger people, signs with Māori cultural reference, and certain gender sensitive signs (e.g., anatomical and sexuality related).

Descriptive challenges

Word class tagging

Determining grammatical category (part of speech) in signed lexicons deeply challenges lexicographers and linguists (Johnston and Schembri 1999; Wilcox 2003, Zwiterserlood 2010). Indeed, Schwager and Zeshan (2008: 514) describe the dearth of systematic analysis of word class in signed languages as "shockingly neglectful". There are, however, methodological and theoretical reasons for ambiguity on this topic. In a typological survey of part of speech across signed languages, Schwager and Zeshan (2008: 513) describe "a substantial amount of systematic ambiguity or vagueness in many sign languages. For instance, in Indo-Pakistani Sign Language (IPSL) many signs tend to have rather general meanings that are narrowed down by the context of the utterance, and since many grammatical categories can remain unmarked, a sequence of signs can be difficult to categorise structurally, although its meaning is entirely clear semantically." This description applies to NZSL, in which a single sign may express various semantic roles; this is partly to do with the

motivated nature of many signs. For example in the ODNZSL, the sign glossed as BUILD (figure 8) may also translate in context as 'builder, carpenter, construct, nail' — senses that are motivated by the action of using a hammer, and are distinguishable by semantic and grammatical context, and/or by mouthing of a corresponding English word. In agreement with Johnston and Schembri's (1999: 150) conclusion that separate entries for nominal and verbal uses of signs could not be justified in an Australian Sign Language dictionary on the basis of form, many entries in ODNZSL have multiple word class tags and a correspondingly varied set of English word class translations.

Word class information displayed in ODNZSL is based on semantic and morphosyntactic evidence from usage in the corpus data, editorial intuition about the primary character of a sign as verbal, nominal, adjectival, etc., and to some extent, its potential equivalence in English. Since these judgments are not based on thoroughly principled analysis according to morphosyntactic criteria (as proposed by Schwager and Zeshan 2008, for example), we consider word class information to be only indicative about usage. Another source of word class ambiguity in signed languages relates to the polycomponential nature of many signs, as we discuss below.

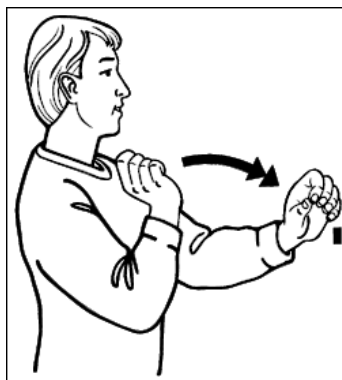


Figure 8: 'Build'

What qualifies as a lemma?

Defining signs, lemmas and lexemes for dictionary purposes is a problem previously examined by sign lexicographers (Brien and Turner 1994; Johnston and Schembri 1999; Brennan 2001). A key problem in defining the lexical status of signs is the polysynthetic capacity of signed languages to simultaneously combine meaningful elements in manual (hands), non-manual (face and body) and spatial dimensions to produce contextualised meanings. For example, figure 9 shows a sign conventionally glossed as 'reverse (a vehicle)'; this is actually a

polycomponential construction, or phrase, comprising several elements: a pro-form for 'vehicle' (right hand), a pro-form for 'ground' (left hand), a movement path of the vehicle's action (backwards), and manner (puffed cheek, indicating intensity of the agent's effort or focus). Although this construction occurs regularly and has a conventional meaning associated with it, in situated use, path and manner elements of the sign might vary to reflect details of a particular reversing event. Although this is a relatively straightforward example, such productive constructions do not easily fit morphological and lexical criteria for a fully lexicalised 'word' (Emmorey 2003; Schembri 2003).

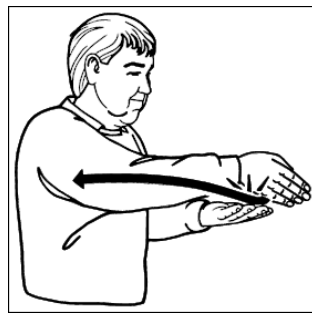


Figure 9: 'Reverse (a vehicle)' — a polycomponential construction

Johnston and Schembri (1999: 126) argue that many well-formed 'signs' observed in discourse, particularly polycomponential constructions, have reference local to a specific context but not established in wider use, and therefore have dubious status as entries in a dictionary. They state:

A lexeme in Auslan is ... a sign that has a clearly identifiable citation form which is regularly and strongly associated with a meaning which is (a) more specific than the sign's componential meaning potential, especially when cited out of context, and/or (b) quite unrelated to its componential meaning (i.e., lexemes may have arbitrary links between form and meaning).

On the other hand, many productive (polycomponential) and morphologically inflected sign forms acquire lexical status through frequency of use and thus warrant inclusion in a dictionary (Brennan 2001). As Johnston and Schembri also note in their discussion of lexical status of signs, demonstrating semantic equivalence between languages can be an important objective of a bilingual dictionary that supports the inclusion of productive and inflected forms that are important for second-language dictionary users to discover.

Lexicographers of polysynthetic languages regularly debate the merits of entering verbs and nouns by their stem or by their regularly inflected forms. In relation to the morphologically productive African language isiZulu, for example, Prinsloo (2011) evaluates the arguments for lemmatisation by word forms that include regular affixes, versus by stems. He concludes overall that the

word approach is more accessible to most users, as searching by stem requires considerable morphosyntactic knowledge to deconstruct the components of complex forms that are encountered in texts as 'words'. Similarly, Kosch (2013) suggests that the decision for any given dictionary should be guided by consideration of users' expectations and grammatical knowledge.

In our editorial team's consideration of whether commonly inflected verb forms should appear as entries, the native signers (who are also sign language teachers) tended to be more in favour of their inclusion than the non-native signers, who were more inclined to separate lexicon from grammar. However, in light of the value for second language learners of seeing frequent and semantically stable uses of productive verb morphology, we did include some inflected forms that are regularly associated with distinct English translations. Figures 10 and 11 illustrate such 'sets' of inflected entries, based on the verbs LOOK and WALK, respectively.


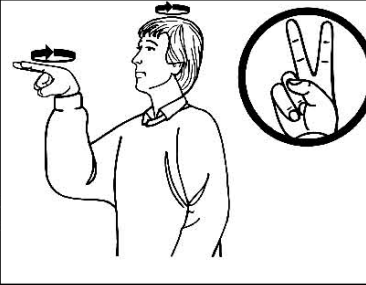
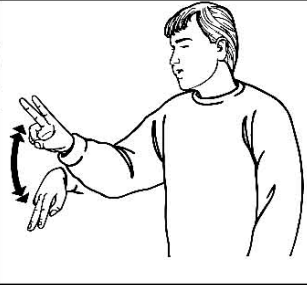

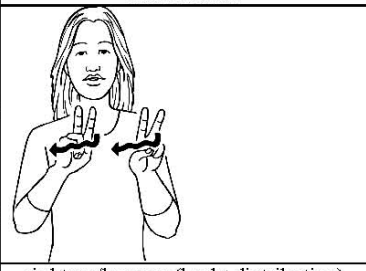
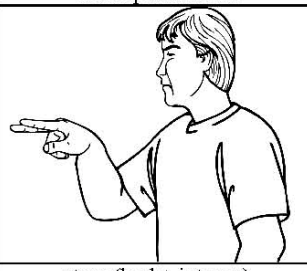
		
look at	look around	look up and down
		
search (look back & forth)	sightsee/browse (look+distributive)	stare (look+ intense)

Figure 10: Entries for conventionally inflected forms of 'Look'

During filming of headword signs, sign models often found it difficult to produce certain signs (usually verbs and adjectives) devoid of inflectional morphology (e.g., manner, degree, location), suggesting that while these features are, in theory, optional, in natural usage they are closely fused with the lexicon. Arbitrating on citation form for each item in a dictionary has something in common with the orthography decisions required in making dictionaries of

languages with an oral tradition (cf. Axelrod et al. 2003), and is a consultative process. As Axelrod et al. observe, native speaker perspective in this process is vital, but does not always guarantee agreement on judgments about 'accurate' or 'neutral' forms. We acknowledge the potential for ongoing editorial revision in this regard.

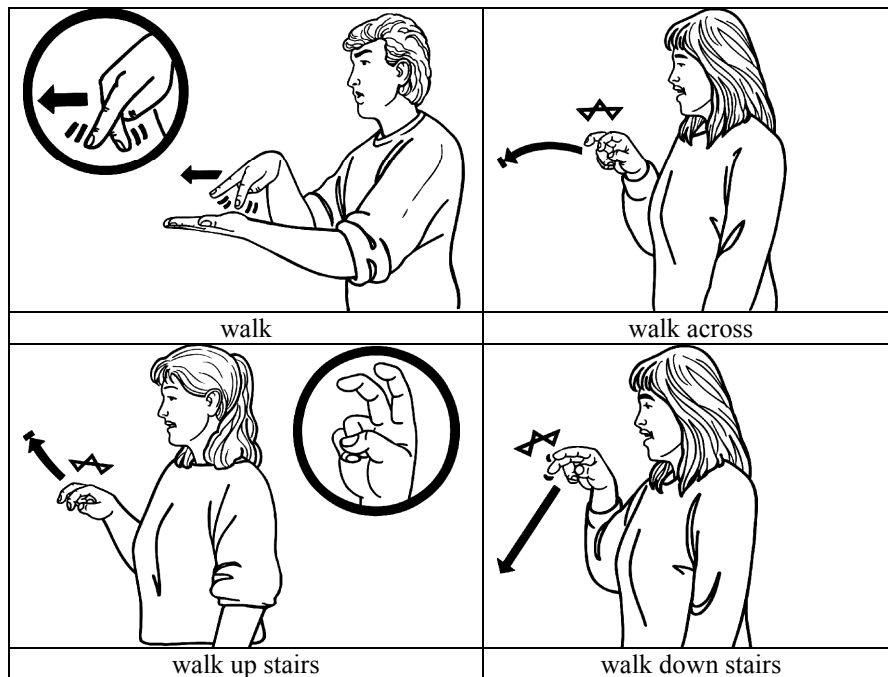


Figure 11: Entries for conventionally inflected forms of 'Walk'

Polysemy, mouthing and functional equivalence

NZSL has much polysemy and some homophony in its lexicon, which requires decisions about how to present multiple form–sense relationships. In addition to semantic context, an important means of disambiguating meanings of polysemous and homophonous signs in NZSL is the simultaneous mouthing of spoken English words, which stems from Deaf people's exposure to that lexicon in their everyday lives. Co-articulation of mouthed words with manual signs occurs in many signed languages, and the status of mouthing in relation to a sign lexicon is the subject of theoretical debate, which is beyond the scope of this article (e.g., Boyes-Braem and Sutton-Spence 2001, Ebbinghaus and Hessman 1996). Suffice to say that mouthing is significant in lexicographical decisions about polysemous signs, yet is not often discussed explicitly in relation to the practicalities of making of sign language dictionaries. An exception

is the online Danish Sign Language Dictionary project, which outlined their rationale (Kristoffersen and Boye Niemälä 2008) for deciding whether to treat word mouthings that are regularly associated with signs at the level of phonological form, or at the semantic level of contextualised meaning. Kristoffersen and Niemälä explain that if mouth patterns are treated as a formational element of a sign (i.e., as a stable, and contrastive 'phonological' feature), the result would be multiple dictionary entries for one manual form paired with alternate mouthings. For example, a manual sign with the prototypical meaning 'house' might have seven entries corresponding with seven dwelling-related Danish words that are regularly mouthed to extend the semantic range of this manual sign. On the other hand, if mouthing is treated at the level of meaning (i.e., not a conventional part of its form), the result is to make a single entry for a polysemous sign, with a listing of frequent senses that are expressed in combination with mouth patterns. Native speaker consultants on the Danish dictionary favoured treating mouthing at the level of meaning, as did the NZSL native speaker editors (for the mostpart); i.e., one sign that is typically produced with several semantically related mouthings has a single entry.⁷ An example in the ODNZSL is the entry for the sign BUILD (figure 8), which lists related senses that might be specified in discourse by mouthing and/or semantic context. As not all secondary glosses displayed in an entry are necessarily typical mouthings (but rather, alternate English translations), this is not an entirely satisfactory solution for informing users how signs are potentially combined with mouthing — which varies considerably in situated discourse.

A practical difficulty with representing polysemous items in ODNZSL is that video clips require a decision as to which, if any, mouthing the signer should show with the citation form of the headword. Our compromise was to film such citation forms with no mouthing (although this is not very typical in NZSL), and to show contextually appropriate mouthing in the usage examples. For instance, the entry LEARN/SCHOOL (secondary glosses — examination, lesson, schooling) (<http://nzsl.vuw.ac.nz/signs/5724>) displays a citation form with no mouthing, but two usage examples with natural mouthing of 'learn' and 'school' respectively.

Homophonous signs — a single manual form that has unrelated senses (and correspondingly unrelated word mouthings) — are treated as separate forms. For example, signs for LIVE (<http://nzsl.vuw.ac.nz/signs/565>) and TRUE (<http://nzsl.vuw.ac.nz/signs/440>) are homophonous, and are displayed as separate entries. This departs from the usual practice of organising a lexicon by contrasting forms, rather than meanings, thereby entering a headword once with all senses (related and unrelated) listed. Given the bilingual and pedagogical aim of this dictionary to present form–meaning correspondences for mainly English-speaking users, and considering search and display possibilities in the online medium (i.e., more 'space'), it seemed preferable to display distinct senses of homophones as individual entries. All analytic decisions about how signs should be arranged in a dictionary are, of course, theoretically

debatable; our decisions were influenced by native speaker intuition about the form–meaning correspondences, and by consideration of users' ability to navigate content without the need for extensive explanatory text, or over-cluttering the layout of individual entries. A resulting flaw in the ODNZSL macrostructure is that identical polysemous and homophonous forms which appear in separate entries are not cross-referenced to each other, which should be rectified in future.

Sociolinguistic variation: recording lexical variation and change

As unwritten languages of limited diffusion, signed languages tend to have considerable lexical and sub-lexical variation, which may be both socially patterned and idiosyncratic. Natural variation has often been intentionally or inadvertently disregarded in the making of sign language dictionaries, especially those produced by non-linguists for standardising purposes (Johnston 2003; Schmalting 2012). While lexical variation and innovation is not generally problematic for sign language users, it can be challenging for dictionary makers to capture, and sometimes for users (both native and non-native speakers) to accept in documented form. As Collins-Ahlgren (1994) commented about preparing an earlier dictionary of NZSL, whatever is contained between the covers will inevitably be regarded by second language users as the prescribed standard, and by the primary language community as an approved subset of the language — which may be contentious among members who have not contributed directly to its content or do not see their lexical preferences reflected. It is for such reasons of face validity of process and content that the preface to the DNZSL (Kennedy et al. 1997) lists, by name and region, the 90-plus Deaf informants who participated in elicitation and validation processes.

Sign language dictionaries have a role as a repository of evidence of language change (Wilcox 2003). This archival function was taken into account in making the ODNZSL and its predecessor; both document diachronic and synchronic variation in NZSL by including lexical and phonological variants. A search for the verb 'to die' for example, yields four variants, and three variants for the numeral 'nine'. Research on sociolinguistic variation (McKee and McKee 2011) shows that the lexicon of older signers is distinct from that of younger signers and that regional variation exists. From 1979, an Australian-based signing system was introduced into deaf education, which re-lexified a substantial proportion of NZSL from that time onwards. Variants recorded in the ODNZSL reflect both external and internal processes of change. External influences include vocabulary introduced from above through the education system, and from below via spontaneous borrowing by Deaf individuals travelling and living abroad, and via immigrants in the NZSL community. Internal change and innovation is prompted by new cultural referents and experiences, and in some cases by phonological and morphological processes. An example of morphological change is seen in early and modern variants of the concept

'open-minded' — originally a two-sign calque from the English phrase (shown in figure 12), which has evolved by relocating the sign OPEN to the location of MIND (forehead), thus incorporating two morphemes into a form that can be regarded as a compound lexeme (shown in figure 13).

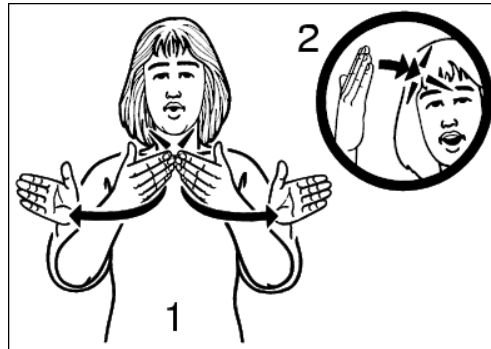


Figure 12: 'Open minded' — earlier form



Figure 13: 'Open minded' — contemporary form (morphologically reformed)

The boundaries of sign language lexicons are quite permeable compared to most spoken languages; individual borrowing and coining are commonly accepted strategies for filling lexical gaps. Editorial calls thus had to be made regarding when a foreign sign (usually from Auslan, ASL or BSL) qualifies as an established borrowing that should appear in a NZSL lexicon, and at what point neologisms are used widely and regularly enough to warrant an entry (e.g., recent coinages for 'Facebook', 'upload'). In practice, it was not possible to formulate rules to resolve these questions, which are a matter of varying perception. Our course of action was to include all signs that were recognised by the editorial team and community informants to be in use more widely than

one social grouping or region of country, and especially if they fill a lexical gap in NZSL (such as 'digital', or 'culture') rather than competing with an established sign. In the case of borrowings, language origin is displayed in the entry only when it is clear; however there are many untagged signs that have almost certainly been borrowed into NZSL by earlier generations, or by secondary routes (e.g., American and Irish signs that have entered NZSL via Auslan borrowings), making their origin now opaque. In the case of neologisms, users can provide feedback via the online feedback form and the Facebook page, which may inform future editorial revisions.

Conclusion

The ODNZSL illustrates the potential of the online medium to enable a more dynamic and authentic representation of a signed language lexicon and to offer users an interactive reference tool. A key element that enriches both descriptive and instructive value is the addition of video content, in addition to line drawings of signs, both of which can be downloaded for offline use.

The relational nature of online searching circumvents macrostructure issues of how to order entries and facilitates bi-directional searching, via the visual features of signs, or by word equivalents, topic domains or other tags. The 'search-by sign' function (using handshape and location features) promotes receptive use by learners who want to identify unknown vocabulary; this feature, and NZSL translation of front matter, also increases accessibility of the dictionary to Deaf NZSL users.

Editorial advantages of the online medium include the capacity to edit records easily, with overnight updating of the database, and to maintain a collaborative workflow in a multi-user team. New records can be created as new data comes to hand, and the microstructure of entries can potentially be adapted to display different fields from records in the database.

The electronic medium does not resolve longstanding lexicographical issues for sign language researchers, such as: deciding what qualifies for entry status (i.e., treatment of signs that are marginally lexicalised); determining morphologically unmarked citation forms; assigning word class; achieving lexico-pragmatic equivalence between the words of different languages (in this case NZSL, English and Māori); accounting for mouthing; and representing variation (in content selection, and in its embodied performance). The medium does, however, promote consultative and recursive editing processes, and allows for revision of content decisions in light of new evidence and ideas.

Dictionary content was not generated directly from an electronic corpus, but a transcribed corpus of NZSL informed selection, senses, and usage information. Although the format and scale of the corpus used to support the ODNZSL has been surpassed by recent advances in sign language corpus methods, the development of usage examples from a body of authentic language data is a significant advance in sign language dictionary making practice. Reference to corpus evidence in illustrating the use of signs not only

informs about typical linguistic contexts, but also offers glimpses into the cultural discourse of the NZSL community.

User feedback indicates that the availability of a user-focused online dictionary has widened societal access to knowledge about NZSL. It is hoped that this tool will support positive social change for the NZSL community, as well as contributing to improving the standard of contemporary online sign language dictionaries.

Notes

1. NZSL Online is found at: <http://nzsl.vuw.ac.nz/>.
2. See: <http://creativecommons.org/licenses/by-nc-sa/3.0/>.
3. The ODNZSL project was funded by the New Zealand Tertiary Education Commission's 'Encouraging and Supporting Innovation' Fund, 2008–2011.
4. Freelex is Open Source lexicography software by Dave Moskovitz. Download at: <http://www.matapuna.org/>.
5. Several writing systems have been created to capture the phonetic or lexical form of signs, but their use tends to be limited to linguists or educators who adopt them for technical or pedagogical purposes. They are not widely known by signers. See Zwitserlood (2010) for a recent summary of sign 'writing' systems.
6. An overview of recent developments in sign language corpus work is found in Crasborn et al. (2008). Also see the ongoing work of the Sign Linguistics Corpora Network hosted at Radboud University, Nijmegen: <http://www.ru.nl/slc/>.
7. Boyes-Braem (2001) reports the opposite decision by Deaf consultants for a lexical database of Swiss-German Sign Language, in which identical manual signs with contrasting mouthings are treated as separate entries. A different approach to polysemous signs with multiple mouthings is taken in the iLex project which is building a lexicographical database from a corpus of German Sign Language (DGS): the manual form is treated as the type (or lemma), and combinations of the manual form with different German word mouthings are treated as sub-types (lexemes) of the lemma. In this analysis, distributional properties of the manual sign can be analysed either independently, or in its various relationships with mouthing components in discourse (Reiner Konrad, 2012 Sign Corpus Linguistics Summer School, University of Hamburg, unpublished lecture notes.)

References

- Ashipala, S.P., D. Profelius, M.N. Haikali, N. Israel, F.T. Linus, H.H. Nilenge, T.F. Haiduwah and R.N. Hashiyannah. 1994. The Development of a Dictionary of Namibian Sign Language. Erting, C.J., R.C. Johnson, D.L. Smith and B.D. Snider (Eds.). 1994: 342-346.
- Atkins, B.T.S. and M. Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.
- Axelrod, M., J. Gómez de García and J. Lachler. 2003. The Role of Literacy and Collaboration in Documenting Native American Languages: A Report from the Jicarilla Apache Dictionary Project. *Sign Language Studies* 3(3): 296-321.

- Boyes-Braem, P.** 2001. A Multimedia Bilingual Database for the Lexicon of Swiss German Sign Language. *Sign Language and Linguistics* 4(1–2). *Special Issue: Sign Transcription and Database Storage of Sign Information*: 133-144.
- Boyes-Braem, P. and R. Sutton-Spence (Eds.)**. 2001. *The Hands are the Head of the Mouth: The Mouth as Articulator in Sign Languages*. Seedorf/Hamburg: Signum Verlag.
- Brennan, M.** 2001. Encoding and Capturing Productive Morphology. *Sign Language and Linguistics* 4(1–2). *Special Volume: Sign Transcription and Database Storage of Sign Information*: 47-62.
- Brien, D. and G. Turner.** 1994. Lemmas, Dilemmas and Lexicographical Anisomorphism: Presenting Meanings in the First BSL–English Dictionary. Ahlgren, I., B. Bergman and M. Brennan (Eds.). *Perspectives on Sign Language Usage: Papers from the Fifth International Symposium on Sign Language Research* 2: 391-407. Durham: The International Sign Linguistics Association (ISLA).
- Center for Tegnsprog.** 2008. *Ordbog over Dansk Tegnsprog*. (The Danish Sign Language Dictionary.) <http://www.viittomat.net/http://www.tegnsprog.dk/>.
- Collins-Ahlgren, M.** 1989. *Aspects of New Zealand Sign Language*. Unpublished Doctoral Thesis. Wellington, New Zealand: Victoria University of Wellington.
- Collins-Ahlgren, M.** 1994. Issues in Preparing and Presenting a Dictionary of New Zealand Signs. Erting, C.J., R.C. Johnson, D.L. Smith and B.D. Snider (Eds.). 1994: 335-341.
- Crasborn, O., T. Hanke, E.D. Thoutenhoofd, I. Zwitserlood and E. Efthimiou (Eds.)**. 2008. *Construction and Exploitation of Sign Language Corpora. Proceedings of the 3rd Workshop on the Representation and Processing of Sign Languages*. Paris: ELRA. http://www.lrec-conf.org/proceedings/lrec2008/workshops/W25_Proceedings.pdf.
- De Schryver, G.-M.** 2003. Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography* 16(2): 143-199.
- Ebbinghaus, H. and J. Hessman.** 1996. Signs and Words: Accounting for Spoken Elements in German Sign Language. Edmonson, W. and R. Wilbur (Eds.). 2008. *International Review of Sign Linguistics. Volume 1*: 23-56. Mahwah, NJ: Lawrence Erlbaum Associates.
- Emmorey, K. (Ed.)**. 2003. *Perspectives on Classifier Constructions in Sign Languages*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Erting, C.J., R.C. Johnson, D.L. Smith and B.D. Snider (Eds.)**. 1994. *The Deaf Way: Perspectives from the International Conference of Deaf Culture*. Washington D.C.: Gallaudet University Press.
- Fillmore, C.J.** 2003. Double-Decker Definitions: The Role of Frames in Meaning Explanations. *Sign Language Studies* 3(3): 263-295.
- Goffman, E.** 1981. *Forms of Talk*. Oxford: Basil Blackwell.
- Hanke, T. and J. Storz.** 2008. iLex — A Database Tool for Integrating Sign Language Corpus Linguistics and Sign Language Lexicography. Crasborn, O., T. Hanke, E.D. Thoutenhoofd, I. Zwitserlood and E. Efthimiou (Eds.). 2008: 64 -67.
- Hualand, H. and C. Allen.** 2009. *Deaf People and Human Rights*. Helsinki: World Federation of the Deaf.
- Hausmann, F.J. and 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. Berlin: Walter de Gruyter.

- Institutionen för Lingvistik.** 2009. *Svenskt teckenspråkslexikon.* (Swedish Sign Language Dictionary.) www.ling.su.se/teckensprakslexikon.
- Johnston, T.** 2001. The Lexical Database of Auslan (Australian Sign Language). *Sign Language and Linguistics* 4(1–2). *Special Issue: Sign Transcription and Database Storage of Sign Information:* 145–169.
- Johnston, T.** 2003. Language Standardization and Signed Language Dictionaries. *Sign Language Studies* 3(4): 431–468.
- Johnston, T.** 2008. Corpus Linguistics and Signed Languages: No Lemmata, No Corpus. Crasborn, O., T. Hanke, E.D. Thoutenhoofd, I. Zwitterlood and E. Efthimiou (Eds.). 2008: 82–87.
- Johnston, T.** 2010. From Archive to Corpus: Transcription and Annotation in the Creation of Signed Language Corpora. *International Journal of Corpus Linguistics* 15(1): 104–129.
- Johnston, T. and A. Schembri.** 1999. On Defining Lexeme in a Signed Language. *Sign Language & Linguistics* 2(2): 115–185.
- Johnston, T. and A. Schembri.** 2005. The Use of ELAN Annotation Software in the Auslan Archive/Corpus Project. Paper presented at the Ethnographic E-research Annotation Conference, University of Melbourne, Victoria, Australia, 15–16 February 2005.
- Kennedy, G., R. Arnold, S. Fahey and D. Moskovitz.** 1997. *A Dictionary of New Zealand Sign Language.* Auckland: Auckland University Press with Bridget Williams Books.
- Kennedy, G., D. McKee, R. Arnold, P. Dugdale, S. Fahey and D. Moskowicz (Eds.).** 2002. *A Concise Dictionary of New Zealand Sign Language.* Wellington: Bridget Williams Books.
- Kosch, I.M.** 2013. *Expectation Levels in Dictionary Consultation and Compilation.* Paper presented at the Seventeenth Annual International Conference of the African Association for Lexicography (AFRILEX), which was hosted by the Department of African Languages, University of Pretoria, Pretoria, South Africa, 2–5 July 2012.
- Kristoffersen, J.H.** 2010. *From Utterance to Example Sentence: A Lexicographic Approach to Exploitation of Corpus.* Poster presented at Sign Linguistics Corpora Network (SLCN) Workshop 4: Exploitation. Deaf Centre, Friedrichstraße 12, 10969, Berlin, 3–4 December 2010. Retrieved electronically 31 July 2012, from: http://www.ru.nl/slcn/workshops/4_exploitation/workshop-posters/.
- Kristoffersen, J.H. and J. Boye Niemälä.** 2008. How to Describe Mouth Patterns in the Danish Sign Language Dictionary. De Quadros, R.M. (Ed.). 2008. *Sign Languages: Spinning and Unraveling the Past, Present and Future.* TISLR9, *Forty Five Papers and Three Posters from the 9th. Theoretical Issues in Sign Language Research Conference, Florianopolis, Brazil, December 2006:* 230–238. Petrópolis/RJ. Brazil: Editora Arara Azul. <http://www.editora-arara-azul.com.br/EstudosSurdos.php>.
- Kuurojen Liittory.** 2003. *Suvi Suomalaisen viittomakielen verkkosanakirja.* (Finnish Sign Language Dictionary.) <http://suvi.viittomat.net>.
- Leeson, L. and B. Nolan.** 2008. *Digital Deployment of the Signs of Ireland Corpus in Elearning.* Unpublished conference paper, Language Resources and Evaluation Conference, Marrakesh, Morocco, June 2008.
- Levitt, D.** 1986. *Introduction to New Zealand Sign Language.* Auckland: National Foundation of the Deaf.
- Lucas, C.** 2003. The Role of Variation in Lexicography. *Sign Language Studies* 3(3): 322–340.

- McKee, D. and G. Kennedy.** 2006. The Distribution of Signs in New Zealand Sign Language. *Sign Language Studies* 6(4): 372-390.
- McKee, R. and D. McKee.** 2011. Old Signs, New Signs, Whose Signs? Sociolinguistic Variation in the New Zealand Sign Language Lexicon. *Sign Language Studies* 11(4): 455-527.
- McKee D., R. McKee, S. Pivac Alexander, L. Pivac and M. Vale (Eds.).** 2011. *Online Dictionary of New Zealand Sign Language*. Deaf Studies Research Unit, Victoria University of Wellington, New Zealand. <http://nzsl.vuw.ac.nz/nzsl>.
- Miller, C.** 2001. Some Reflections on the Need for a Common Sign Notation. *Sign Language and Linguistics* 4(1-2). *Special Issue: Sign Transcription and Database Storage of Sign Information*: 11-28.
- Padden, C. and T. Humphries.** 2005. *Inside Deaf Culture*. Cambridge, Mass.: Harvard University Press.
- Penn, C., D. Ogilvy-Foreman, D. Simmons and M. Anderson-Forbes.** 1992-1994. *Dictionary of Southern African Signs for Communicating with the Deaf*. 5 volumes. Pretoria: Human Sciences Research Council.
- Prillwitz, S., R. Leven, H. Zienert, T. Hanke and J. Henning.** 1989. *HamNoSys version 2.0. Hamburg Notation System for Sign Languages: An Introductory Guide*. Hamburg: Signum.
- Prinsloo, D.J.** 2011. A Critical Analysis of the Lemmatization of Nouns and Verbs in isiZulu. *Lexikos* 21: 169-193. Online at: <http://lexikos.journals.ac.za/pub/article/view/42>.
- Reagan, T.G.** 2010. *Language Policy and Planning for Sign Languages*. Washington DC: Gallaudet University.
- Schembri, A.** 2003. Rethinking 'Classifiers' in Signed Languages. Emmorey, Karen (Ed.). 2003. *Perspectives on Classifier Constructions in Sign Languages*: 3-34. Mahwah, NJ: Lawrence Erlbaum.
- Schmaling, C.** 2012. Dictionaries of African Sign Languages: An Overview. *Sign Language Studies* 12(2): 236-278.
- Schwager, W. and U. Zeshan.** 2008. Word Classes in Sign Languages: Criteria and Classifications. *Studies in Language* 32(3): 509-545.
- Stokoe, W.C.** 1960. Sign Language Structure: An Outline of the Visual Communication Systems of the American Deaf. *Studies in Linguistics: Occasional Papers No. 8*. Buffalo: University of Buffalo: Department of Anthropology and Linguistics.
- Van Herreweghe, M., S. Slembrouck and M. Vermeerbergen.** 2004. *Digitaal Vlaamse Gebarentaal-Nederlands/Nederlands-Vlaamse Gebarentaal woordenboek*. <http://gebaren.ugent.be>.
- Wilcox, S.** 2003. The Multimedia Dictionary of American Sign Language: Learning Lessons about Language, Technology and Business. *Sign Language Studies* 3(4): 379-392.
- Zwitserlood, I.** 2010. Sign Language Lexicography in the Early 21st Century and a Recently Published Dictionary of Sign Language of The Netherlands. *International Journal of Lexicography* 23(4): 443-476.

Introducing *Tlhalosi ya Medi ya Setswana*: The Design and Compilation of a Monolingual Setswana Dictionary

Thapelo J. Otlogetswe, *Faculty of Humanities, University of Botswana, Gaborone, Botswana (otlogets@mopipi.ub.bw)*

Abstract: This paper presents the design elements of a recently published monolingual Setswana dictionary, *Tlhalosi ya Medi ya Setswana* (Otlogetswe 2012), the fourth main Setswana monolingual dictionary to appear. The paper situates the dictionary within a recent, but growing Setswana monolingual dictionary tradition whose roots may be traced to 1976 when M.L.A. Kgasa's *Thanodi ya Setswana ya Dikole* was published. However, before this date, the development of Setswana lexicography lies wholly in a bilingual dictionary tradition which dates back to the 1800s. The paper also discusses the different features of *Tlhalosi ya Medi ya Setswana* demonstrating that it possesses some dictionary features which are not found in previous Setswana dictionaries such as frequency information, phonemic transcription and extensive cross-referencing.

Keywords: TLHALOSI, THANODI, DICTIONARY, SETSWANA, MONOLINGUAL

Opsomming: Bekendstelling van *Tlhalosi ya Medi ya Setswana*: Die ontwerp en samestelling van 'n eentalige Setswanawoordeboek. Hierdie artikel bied die ontwerpbeginsels aan van 'n onlangs gepubliseerde eentalige Setswanawoordeboek, *Tlhalosi ya Medi ya Setswana* (Otlogetswe 2012), die vierde belangrike Setswana eentalige woordeboek wat verskyn het. Die artikel plaas die woordeboek binne 'n onlangse, maar groeiende Setswana eentalige woordeboektradisie waarvan die wortels teruggevoer kan word tot 1976 toe M.L.A. Kgasa se *Thanodi ya Setswana ya Dikole* gepubliseer is. Voor hierdie datum egter lê die ontwikkeling van die Setswana-leksikografie geheel en al in 'n tweetalige woordeboektradisie wat terugdateer tot die 1800's. Die artikel bespreek ook die verskillende eienskappe van *Tlhalosi ya Medi ya Setswana* deur aan te toon dat dit 'n aantal woordeboekeienskappe besit wat nie in vorige Setswanawoordeboeke aangetref word nie soos frekwensie-inligting, fonemiese transkripsie en uitvoerige kruisverwysings.

Sleutelwoorde: TLHALOSI, THANODI, WOORDEBOEK, SETSWANA, EENTALIG

1. Introduction

The Setswana language has a fairly long lexicographic tradition characterised by low dictionary production. The tradition began with a bilingual Setswana-English dictionary by the London Missionary Society missionary, John Brown,

that was first published in 1875, enlarged and revised in 1895, and revised again in 1925 by John Tom Brown. For over 130 years this dictionary, which was updated by Matumo (1993) has been the only general Setswana–English dictionary. The Setswana monolingual dictionary tradition started in 1976 with *Thanodi ya Setswana ya dikole* (Kgasa 1976). And for about 20 years this was the only monolingual Setswana dictionary, although Setswana was taught in both Botswanan and South African schools (cf. Volz 2003). Much of this lexicographic tradition is dominated not by monolingual but by bilingual dictionaries. Jones (in Matumo 1993: vii) traces the origin of Setswana lexicography to John Brown's bilingual dictionary (1875) and to Robert Moffat's Setswana version of the Gospel of St Luke (1830), which has definitions of difficult words in its final back pages. Jones observes:

In 1830 Robert Moffat published a Setswana version of the gospel of St Luke, and at the back offered two pages of explanations of the more "difficult" words. Is it fanciful to regard this as the first small germ of a dictionary? ... but the first published dictionary of which the Botswana Book Centre has record is that of John Brown in 1875.

Alternatively, Cole (1955: xxviii) dates Setswana lexicographic research to later years to the plant names compilation of Miller (1936) and the list of kinship terms of Van Warmelo (1931). However, lexicographic research in Setswana may be traced to an earlier period, much earlier than Moffat's 1830 writings that Jones refers to, and certainly far earlier than Cole's botanical and kinship references. Research demonstrates that Lichtenstein in the second volume of *Travels in Southern Africa in the Years 1803, 1804, 1805, and 1806* had as an Appendix a list of about 270 Setswana words and phrases. The original book in German appeared in 1811. Therefore the earliest lexicographic activity, at least of a headword list with its English equivalents, can be traced to 1803–1806, in Lichtenstein's works. In 1815, Campbell in his *Travels in South Africa* gave a list of 80 "Bootchuana" words at the end of chapter 19 in the book. Salt (1814) in *A Voyage to Abyssinia* presents a list of 20 "Mutshuana" words and their English equivalents. Lexicographic work in Setswana, regardless of its size and detail, therefore existed before the work of Moffat, who only came to Southern Africa in 1816.

The first Setswana bilingual dictionary, *Lokwalo loa Mahúku a Secwana le Seeneles*, was compiled by Brown (1875) of the London Missionary Society. An enlarged and revised version appeared in 1895 and was reprinted in 1914 and 1921. In 1925 Brown produced the third edition of this dictionary with the aid of A.J. Wookey's research (Peters and Tabane 1982: xxiv). However, since the 1925 edition of Brown's dictionary to the mid-1970s, no Setswana dictionary was compiled. That represents a fifty year gap. It was not until 1976 that Kgasa published his 134-page monolingual dictionary, *Thanodi ya Setswana ya Dikole* (The Setswana Dictionary for Schools), whose main target users were primary school pupils. Kgasa's dictionary, which took ten years to compile (Rasebotsa

2008), is significant for it marks the start of the Setswana monolingual dictionary tradition. A smaller, but detailed, trilingual dictionary — Setswana, English and Afrikaans — was compiled by Snyman et al. (1990), *Dikišinare ya Setswana–English–Afrikaans Dictionary/Woordeboek*, whose target is secondary school and university users. This dictionary filled a gap that existed in the South African market where both English and Afrikaans were used widely by Setswana speakers.

The Compact Setswana Dictionary by Dent (1992) is an abridged 300-page dictionary "intended for those people who find more comprehensive dictionaries too cumbersome or too detailed for their needs" (Dent 1992: introduction). It has a large headword list, though it lacks detailed lexicographic metalinguistic matter.

Matumo (1993) revised Brown's 1925 dictionary into what is now the *Setswana–English–Setswana Dictionary*. Prinsloo (2004) revisits this dictionary suggesting how it may be revised by the design and use of a multidimensional Ruler and Block System for the measurement and balancing of alphabetical stretches in terms of time, average length of articles and number of pages per alphabetical category.

Cole (1995) compiled a dictionary of plants and animals called *Setswana — Animals and Plants (Setswana — Ditshedi le Ditlhare)*. While in the foreword L.W. Lanham notes that "the author ... eschews the label 'dictionary' for it, preferring to identify it as a 'lesser listing of vocabulary'" (Cole 1995: ix), it is a bilingual dictionary, Setswana to English and English to Setswana, with some of the entries included with their Latin names.

In 1998, in collaboration with Tsonope, Kgasa compiled the second monolingual dictionary *Thanodi ya Setswana* which until now remains the definitive Setswana monolingual dictionary used widely in Botswana schools and other domains both in Botswana and South Africa. It has been prescribed in Botswana junior and senior secondary schools for over fifteen years since its publication as a reference text for Setswana language and literature classes.

Créissels and Chebanne's *Dictionnaire Français–Setswana Thanodi Sefora–Setswana*, published in 2000, is the only French/Setswana bilingual dictionary. Its primary target group is students of French at secondary and university level. It is the first Setswana dictionary with Setswana phonemic transcriptions, though this part of the dictionary is limited. In 2007 Otlogetswe published a smaller bilingual dictionary, the *English–Setswana Dictionary*, for use as a reference work in Botswana primary schools.

In 2008 a fairly large Setswana dictionary of about 600 pages was compiled by Mareme with the aid of the Setswana National Lexicographic Unit (NLU) based in Mafikeng. Like Kgasa and Tsonope's dictionary, it is called *Thanodi ya Setswana*. This dictionary has not been popular in Botswana and is also not widely used in South Africa.

Cole and Moncho-Warren (2012) have published a Setswana–English, English–Setswana dictionary of nearly 1200 pages entitled *Macmillan Setswana*

and *English Illustrated Dictionary*. The dictionary will be of interest to language specialists, especially at the university, because, as the back cover states, the dictionary is "intended for use by scholars and teachers at tertiary level, and in libraries". G.B. Mareme, writing in the introduction of the dictionary, is probably right that "learners in the new South African education system, especially those studying life sciences, physics and chemistry, will find this dictionary invaluable".

The discussion above has outlined the Setswana dictionary landscape dating to the 1800s. It demonstrates that *Tlhalosi ya Medi ya Setswana* comes at a time when for the past 20 years there has been some exciting Setswana lexicographic projects and a growing Setswana monolingual dictionary tradition.

What follows is a description of the compilation and different features of *Tlhalosi ya Medi ya Setswana*.

2. Headword collection for the *Tlhalosi ya Medi ya Setswana*

The craft of lexicography demands not only the ability to collect data, but also the ability to make sense of it. The headword list collection for the compilation of *Tlhalosi ya Medi ya Setswana* was achieved through three strategies to capture the broad lexical representation of the Setswana language. The word list collection used a corpus, semantic domains and the consultation of other publications such as dictionaries, grammar books and language documentations.

2.1 Corpus contribution

Developments in corpus linguistics and corpus lexicography of the past 30 years, particularly in English lexicography, have resulted in unique techniques of lexical analysis for lexicography. The use of statistical analysis through the use of Corpus Querying Software (CQS) such as the *Oxford WordSmith Tools* (Scott 2004) and *SketchEngine* (Kilgarriff et al. 2004) has meant that it is now possible to analyse large amounts of text running into millions of words. *Tlhalosi ya Medi ya Setswana* was based on a 16 million token Setswana corpus (Otlogetswe 2008). The corpus is broad in its composition, with 90% of the text coming from written texts and 10% being transcribed speech. The corpus comprises novels, science texts, political texts, grammar books, poetry texts, religious texts, prose texts, spoken texts recorded from a variety of contexts such as funerals, Setswana classroom and family meetings. The corpus was analysed using the *Oxford WordSmith Tools* version 5 (Scott 2004). Frequency analyses were conducted to generate a dictionary headword list while word concordances were extracted to study how different words behave in context to aid sense discrimination. Some congrams were also run to identify word clusters of certain Setswana terms. The corpus was further used as a source of illustrative examples.

The contribution of a corpus to the dictionary making process has been discussed extensively in lexicographic literature by Béjoint (2000: 97), and Sinclair (1987, 1991) who demonstrate the crucial nature of a corpus and frequency lists for the dictionary compilation process. A corpus has increasingly become useful in providing definitive dictionary information for many modern dictionary compilations. However, claims and judgments on the definitiveness of the data abstracted from a corpus cannot be made independent of the claims and judgments on the quality and constitution of the corpus from which the data was abstracted (Otlogetswe 2008). Put differently: if the corpus data is skewed and badly structured, the data abstracted from the corpus will also reflect this — one cannot take from a corpus what has not been put into it. Focus must increasingly be directed to the structure and content of the corpus and not just what is abstracted from it (Biber 1993). The corpus should attempt to be as broad as possible in its coverage of the language varieties. Such a broad spectrum corpus would include novels, newspapers, magazines, spoken language, news, meeting transcriptions, health literature etc (Biber 1994). Since about 90% of the corpus used for this dictionary was written text and 10% was transcribed text from spoken language, semantic domains (Moe 2001) were used as a supportive measure to add words to the dictionary.

Many corpora in African languages are insufficient as sources of dictionary material since their text coverage is limited. Most corpora comprise published works which are a poor representation of the linguistic variation of a speech community. There are numerous oral domains in which African languages are used such as law, business, farming and health, which are largely excluded from a corpus of written texts. To fill this gap, other strategies of data collection were used.

2.2 Semantic domains

Apart from the corpora collection approach discussed above, Moe (2001) proposes a method of semantic domains to be used for the collection of words. He argues that the methodology is particularly attractive for minority languages, most of which have none or few written texts, or no corpora. His argument is that the methodology is 100 times faster than collecting words without a structure. He argues that 12,000 words have been collected in a few weeks through what is effectively a simple methodology but one which is able to produce a massive classified dictionary and thesaurus.

Moe analysed the domain classification of words as suggested by Murdock et al. (1987), Roget (1958 and 1985) and Louw and Nida (1989) and found them inadequate for eliciting vocabulary. What Moe (2001: 151) attempts to compile is "a universal list of semantic domains" that field lexicographers could use to prompt native speakers to think of words in their language. However, semantic domains have a greater relevance than mere elicitation of mother tongue speakers' words: "It could be used to collect words, it could serve to

classify a dictionary, and it could aid in semantic investigation" (Moe 2001: 152). Underlying this system is a mental approach to the lexicon: words are all linked together in the mind in a gigantic multi-dimensional web of relationships which cluster around a central nexus (Moe 2001: 4). The mental lexicon is not alphabetical but words congregate around key concepts which Moe calls semantic domains (Moe 2003: 216). His argument is therefore that related words should be collected at the same time. To guide field workers, Moe phrases domains as questions such as the following for the domain "sing":

What words refer to singing? sing, serenade, warble, yodel, burst into song
What words refer to singing without using words? hum, whistle

These series of questions are central to what Moe calls the Dictionary Development Process (DDP) which he used in Uganda in training lexicographers in collecting words for the Bantu language Lunyole. The DDP has 1,700 domains each with 8–10 questions which could elicit over 10 words per domain which means that the dictionary would have at least 17,000 entries.

This method was used to obtain specialised information which is difficult to find in a corpus such as the names of stars, types of cows, meat cuts, animal colours, terminology relating to traditional beer making processes etc.

2.3 Other literature

Setswana has a number of relevant literatures which were consulted to develop the headword list of the dictionary. Some are dictionaries such as Kgasa (1976), Kgasa and Tsonope (1995) and Mareme (2008). Others are grammar books such as Cole (1955) and Mogapi (1984). The headword list of *Tlhalosi ya Medi ya Setswana* was compared against that of Kgasa and Tsonope (1995). The grammatical analysis of Cole (1955) and Mogapi (1984) provided a useful resource for closed word class members such as demonstratives and pronouns. We also benefited from Cole (1995) from which a wealth of animal and plant names were culled.

3. Choosing *tlhalosi* over *thanodi*

Since this is the first Setswana dictionary called *tlhalosi*, it is important to provide a justification for the choice of the term *tlhalosi* over the more established Setswana term *thanodi* to refer to a Setswana monolingual dictionary. The first Setswana dictionary to carry the title *thanodi* was the monolingual dictionary Kgasa (1976). It was published about 100 years after the first bilingual Setswana–English dictionary (Brown 1875) appeared. Therefore, when Kgasa published the first monolingual dictionary, the term *dictionary* was very much synonymous with a bilingual dictionary, that is, a dictionary that offered translations from one language to another. This observation is important since

it may proffer an explanation why the term *thanodi* may have been chosen to mean "dictionary" in Setswana.

The word *thanodi* is derived from the Setswana verb *ranola* through the linguistic process of strengthening. It means "to unravel the signification of a hidden tongue or opaque semantics". However, in current usage, the word *ranola* means "translate". The term *thanodi* is therefore better suited to refer to bilingual or multilingual dictionaries such as the Matumo (1993) and Snyman et al. (1990) dictionaries, since they deal with translation and equivalents. The term *tlhalosi* on the other hand is derived from the Setswana verb *tlhalosa* which means "to explain". Since a monolingual dictionary explains words in a single language, the term *tlhalosi* is preferred as a term reserved for a monolingual dictionary. To use the term *thanodi* for *tlhalosi* would therefore be erroneous. That is why in part the current monolingual dictionary is termed *tlhalosi* and not *thanodi*, since it does not translate words from one language to another but it explains Setswana words using Setswana vocabulary. It has also been discovered that a number of Sotho-Tswana languages use the word *tlhalosi* to refer to either monolingual dictionaries or dictionaries in general.

For instance, the monolingual *Pukuntšutlhaloši ya Sesotho sa Leboa* (Mojela 2007) uses the word *tlhaloši* and not *thanodi* in its title. Other compilers use the term *tlhalosi* for dictionaries in general. An example is *Lenanentswe-hlhalosi: Seafrikaans–Seisemane–Sesotho sa Leboa* (Joubert and Mangokoane 1975). In the Lozi language the word for *dictionary* is *sitalusa-manzwi* (*setlhalosa mantswwe* in the Setswana orthography), "explainer of words". The word *manzwi* or *mantswwe* means "words".

Compared to *thanodi*, the word *tlhalosi* is also appropriate since it is derived from *tlhalosa* a verb which means "to explain". A word's explanation goes beyond its meaning. It includes its use in a sentence, its part of speech and noun class, its stylistic use, that is, whether it is formal or informal. It also includes a word's etymology, pronunciation and usage frequency. All these characteristics of a word are not translating but form part of explaining a word. This is why the word *tlhalosi* was given preference above *thanodi*.

4. Dictionary target users

The dictionary was designed with a broad spectrum of monolingual users in mind. In a community which has a vibrant and established dictionary culture, it is easier to divide target users into smaller specialised and focused groups. However, in a less vibrant dictionary environment like the one we are dealing with, it is important to consider a fairly broad spectrum of users for the dictionary, instead of restricting its scope of usage. This is significant because since 1976, *Tlhalosi ya Medi ya Setswana* is the forth Setswana monolingual dictionary to be compiled. It was therefore written with the following users in mind:

- (a) It was meant for use by Setswana language classes from secondary school level up to tertiary levels. Phonemic transcriptions with tonal markings were therefore included to cater for the needs of advanced language classes. Synonyms and etymological information were added to fulfil the needs of both secondary and tertiary level classes.
- (b) The dictionary is suitable for utilisation by Setswana writers and authors. It contains information on dialectal variation and stylistic markings which will aid writers in their work. These could be professional writers or students writing their essays at school or university.
- (c) The dictionary was compiled for Setswana speakers in general wherever they might live. Setswana has mother-tongue speakers in at least four countries: South Africa, Botswana, Namibia and Zimbabwe. The largest number of speakers is found in South Africa (over 4 million speakers, about 8% of the population) where Setswana is one of the eleven official languages. Zimbabwe has an estimated 29,000 Setswana speakers and Namibia has approximately 6,000. In Botswana, Setswana is used by circa 1.5 million speakers (70-90% of the population) as their mother tongue (Andersson and Janson 1997). Selolwane (2004: 4) observes that "the SeTswana language is the most dominant of all the language groups found in Botswana, with at least 70% of the population identifying it as a mother tongue and another 20% using it as a second language". In Botswana Setswana is the national language. It is studied as a compulsory subject by all Botswana students from primary up to senior secondary school. The dictionary is therefore written to instil pride in the Setswana language amongst Setswana speakers. This is important since Setswana is an important intangible heritage of its speakers.
- (d) The dictionary was written for Setswana language teachers and lecturers. It is designed to aid them in their instructional classes as well as in the development of the Setswana language.

5. *Tlhalosi ya Medi ya Setswana's* key features

The following are the microstructural features of *Tlhalosi ya Medi ya Setswana*.

5.1 Sense presentation

Each word has at least one core meaning, to which a number of polysemous senses may be attached. Core meanings represent typical, central uses of the word in question in current Setswana, based on corpus analysis and established research. The core meaning also represents the most literal sense that the word has in ordinary modern usage. This is not necessarily the same as the oldest meaning of the word, because word meaning change over time. In our analysis, the core meaning is based on frequency and an analysis of concordance lines. For instance, the entry for the word *bona* "see" is the following:

bôna¹ **** /bónà/ •ldr• 1^o go dirisa matlho go leba *Fa a bona koloji e tla e tabogile a fapoga tsela ka bofefo go iphemela*
2^o amogela sengwe *O bona madi a mantši ka go rekisa phane*
3^o go tihaloganyana sengwe *Jaanong ke a bona gore ga a nthate* 4^o go kopana le mongwe *Ba rile ba tlaa re bona teng*

Some related words have multiple parts of speech. The decision of which part of speech appears first is made according to the most common usage. In the dictionary, the core sense acts as a gateway to other related polysemous senses.

The word *kampa* "camp" can be used as noun as well as verb in sense 3.

kampa¹ /kámpá/ •ln. 9. n-, *10. din-• 1^o lefelo le go nniwang mo go lone ka nakwana bogolo jang le le kgakala le motse = BOTHIBÊLÊLÔ 2^o lefelo la bonno la masole, mapodisi kgotsa batlhokomedi ba dipologolo 3^o •ldr• go nna mo lefelong lengwe ka nakwana, bogolo jang le le mo sekgweng ⇐ SeE: camp

Other words have figurative meaning extension of the core sense. The core sense of the word *mokwatla* is "back part of the body" and the figurative meaning extension is "support of a system or organization".

5.2 Phonemic presentation

The dictionary is the first Setswana dictionary to have comprehensive phonemic transcriptions for each lexical entry. Since Setswana is a tonal language, each transcription has been given tonal markings. Phonemic transcriptions are important since they help users, language learners, linguists and teachers, to have a pronunciation reference point. The tonal markings in particular are important for disambiguating words which, though orthographically written the same, have different tonal distributions. These are words such as *mosimanyana* (mòsimànjàná) "small boy" and *mosimanyana* (mòsimànjàná) "small hole". Every headword in the dictionary is accompanied by phonemic transcriptions with the exception of multiword expressions such as idioms and proverbs. For instance:

gôpane o letsa kwadi •seane• go thata
Re lekile go thusa ka ditsela tsotlhe go agisanya bobedi jo, mme go supa-fala gore gopane o letsa kwadi fela bagaetsho

gôpane wa moikapari •leele• mohumanegi yo mogolo *Ke ne ke tlaa mo duela ka eng ke le gopane wa moikapari jaana?* = MOHUMANEGI, MODIDI YO O NTA E TLHÔGÔ E MOTÔPÔ

gôpanyane /χópánànì/ •ln. 1a. Ø, *2a. bo-• ngotlo ya lefoko "gopane"

gopê **** /χúpè/ •lbd• lebadi la maina a setlhopha 15 le le tlhobosang bole-ngteng jwa sengwe *Ke badile pego e yotlhe mme ga go na gope fa a tlhalosang gore monna yo o dirile molato teng*

gopêlêla /χúpélélà/ •ldr• go neela

mongwe magopelo <i>Re tlaa gopelela</i> <i>modisa le fa a ne a seyo fa re bua</i> <i>kgomo ya gagwe</i>	gore di wele mo go one, o khuru- metswe ka bojang le dikala <i>Batlho-</i> <i>komedu ba diphologolo tsa naga ba</i> <i>fulhetse go epilwe gopo mo gare ga</i> <i>sekgwa</i>
gopô /χòpò/ •ln. 9. n-, *10. din-• mosima o o epilweng go thaisa diphologolo	

5.3 Setswana dialectal and stylistic varieties

The dictionary is compiled to capture all the Setswana lexical varieties such as Sengwaketse, Sekgatla, Sengwato and others. For instance, the word for *tin* is lexicalised differently in a number of Setswana dialects. It is *mmolopita* in Sekgatla, *tsiri* in Sengwaketse and *sebagabiki* in Sengwato. In the dictionary, all these words are cross-referenced and indicated by a dialectal marker. The dictionary therefore captures high levels of synonymy. It has over 4,400 cross-references of synonyms from a variety of dialects. Synonyms appear at the end of head words and are preceded by the equal sign [=]. The headwords bear the following dialectal marks *AfBo.* (South African), *Bots.* (Botswana), *Ngwk.* (Sengwaketse), *Lete.* (Selete), *Kgat.* (Sekgatla), *Ngwt.* (Sengwato), *Rolo.* (Serolong), and *Kwen.* (Sekwena). It will be observed that some of the markers are broad, such as *AfBo.* and *Bots.*, while others, such as *Ngwk.* and *Kwen.*, are narrow. This is to capture both the dialectal variations associated with different Tswana ethnic groups while at the same time documenting the developing regional varieties of Botswana and South Africa used by the different ethnic groups found in these regions. As more research goes into Setswana dialectal studies, a more detailed characterisation of the Setswana lexical wealth will be revealed. Currently *Tlhalosi ya Medi ya Setswana* has 371 headwords with dialectal markings. The dictionary also has stylistic markings for *informal* entries and terms which may be considered *offensive* by the users and speakers of the language. Informal entries are marked by the replacement character [◆], while the offensive entries are marked by two exclamation marks [!!].

5.4 Frequency information

Tlhalosi ya Medi ya Setswana is the first Setswana dictionary to mark frequency. The most frequent 4000 headwords have been marked with the use of stars immediately following the headword. A frequency list was generated from a Setswana corpus of about 16 million tokens. The most frequent 1000 words were given four stars (****), words ranked between 1001 and 2000 got three stars (***) , words ranked between 2001 and 3000 got two stars (**), words ranked between 3001 and 4000 got one star (*), while words ranked 4001 and below got no star. Kilgarriff (1997: 135) notes: "A central fact about a word is how common it is. The more common it is, the more important it is to know it." Additionally, highly frequent words are more likely to be polysemous.

5.5 Parts of speech presentation

Like any general language dictionary, *Tlhalosi ya Medi ya Setswana* has an elaborate mark-up for parts of speech. These, amongst others, include *lethusi* (lths), *sekalediri* (skld), *lebopi* (lbp), *lebadi* (lbd), *lediri* (ldr), *lerui* (ler), *leetsisi* (leet), *legokedi* (lgkd), *lekopanyi* (lkpn), *lelatihelwa* (lltl), *leemedi* (lmd), *lesupi* (lsp), *lesoboki* (lsbk) and *letlhaodi* (ltlh). An elaborate mark-up in the dictionary is that of nouns since Setswana, like other Bantu languages, has a complex noun prefix system. Singular noun entries include mark-ups for both the singular noun class prefix and for the plural form of the noun. The structure of the singular entry therefore includes:

[In.] (*leina* "noun" abbreviation) [singular noun class]. [singular noun class prefix], [a star], [plural noun class of the headword]. [plural noun class prefix of the headword]

This translates to the following part of speech label:

•In. 9. n-, *10. din-• which means that a word is a noun *ln* of noun class 9 whose prefix is *n-* and that its plural belongs to noun class 10 whose noun class prefix is *din-*.

This may correspond to a real entry like *hōsetêlê* "hotel" below:

hōsetêlê /hōsítêlê/ •In. 9. n-, *10. din-• matlo a borobalo a a
agetsweng badiri kana bana ba sekolo *Go agiwa hosetele e
nngwe ya basetsana kwa sekolong sa rona* ≡ SeE: hostel

What makes this format attractive is that it marks not just the singular noun class of the headword and its class prefix, the plural noun class of the headword and its class prefix are also included in the mark-up. This gives users not just the grammatical information of the headword but also the formation of its plural form. As this dictionary is written primarily for monolingual speakers of the language, the above design was considered appropriate and sufficient. Nouns which have no plural form, carry only the mark-up for the singular form. For instance, the headword *bonana* "youthfulness" below:

bonana *** /bònánà/ •In. 14. bo-• mokgwa kana seemo sa motho
yo o mo dingwageng tse di mo magareng ga bongwana le
bogolo *Ga twe o akole bonana jwa gago fa o sa tshedile*

5.6 Etymology and borrowing

Setswana has for a long time been in contact with Afrikaans and English and therefore has borrowed extensively from these two languages. Other languages from which borrowings have taken place, include Zulu and Kalanga

(cf. Mathangwane 2008). Of the over 15,000 headwords in *Tlhalosi ya Medi ya Setswana*, 1,179 contain etymological information. Headwords with etymology information therefore constitute about 7.6% of the dictionary. The dictionary traces a word to the language of its origin as well as to its etymon (a word in the source language). Three strategies are used in the dictionary etymology mark-up. First, the source language as well as the etymon is given. For instance:

jê^mê /dʒémè/ •In. 9. n-, *10. din-• sejo se se borekereke se se botshe se se dirilweng ka maungo se se tshasiwang mo borothong *Ke rata botoro fela ka gonne jeme yone e na le sukiri e ntsi* ≡ SeE: jam

This entry *jê^mê* is a borrowing from English (SeE) *jam*.

The second strategy used is an elaborate descriptive one for words borrowed from other Setswana words. This is especially perceptible in the etymology of months. For instance:

Ferikgong /firiq^hòŋ/ •In. 1a. Ø, *2a. bo-• kgwedi ya ntlha ya ngwaga ≡ Kgwedi e e reilwe ka lephoi la mofiri kgotsa kofiri le le a bong le sela dikgonnyana, le aga sentlhaga, go tla le simolola go baya mae. Ka go nna jalo leina le le tswa mo go mofiri le dikgong

The entry *Ferikgong* "February" is traced to two Setswana words *kofiri* "dove" and *dikgong* "pieces of wood".

The final strategy is non-elaborate where the same word is traced to a Setswana word as shown in the examples below:

gôpane ** /χópáni/ •In. 1a. Ø, *2a. bo-• segagabi se setona se se tshwanang le kwena e nnye = KGWATHÊ ≡ gôpa²

mabarebare /màbàribàri/ •In. 6. ma-• dikgang tse di sa tlhoma-misegeng sentle *Ke utlwile mabarebare a gore o a nyala* = MAGATWE ≡ ba re

The entry *gopane* "iguana" is traced to the verb *gopa* "crawl on the belly", while the word *mabarebare* "rumours or hearsays" is traced to the expression *ba re* "they say".

5.7 Idiom and proverb presentation

The dictionary comprises 1,538 proverbs and idioms. This is about 10% of the dictionary's entries. The inclusion of these entries is important since it captures the wealth of the language's idiomatic expressions. Historically it has been argued that Setswana idioms (*maele*) begin with *go* (Mogapi 1984). This classifi-

cation of Setswana idioms is erroneous, since the *go* that is usually found at the beginning of many Setswana idioms is only the infinitive marker of Setswana verbs. Therefore in Setswana idioms it performs the similar function of the infinitive marker of a following verb. Defining Setswana idioms in this matter excludes other Setswana idioms which do not fit this syntactic characterisation such as:

- | | |
|----------------------------------|------------------------------|
| (a) mosamarea yo o molemô | (i) naga e ntse tshêtlha |
| (b) mosêkêla mpeng | (j) nakô kgolo ke eno |
| (c) mosi o tswa ka sekhurumêlô | (k) ntwá e e matlhô mahibidu |
| (d) motho a sa kgwe mathe | (l) ntwá ya tiitii pôtê |
| (e) motsetô o fapogile marago | (m) sebe sa phiri |
| (f) motsetsi o ja ka letsôgô | (n) motsetsi o ja ka letsôgô |
| (g) naga e jelê bore jwa thamaga | (o) tselapedi a tseye e fe |
| (h) naga e ntse kgologolo | |

In *Tlhalosi ya Medi ya Setswana* the *maele* label for Setswana idiomatic constructions is not restricted to those beginning with the infinitive marker *go* as in *go ja motho setshego* "to laugh at somebody". Setswana idioms have been characterised according to their figurative meaning and the fact that they fracture principles of semantic compositionality (Moon 1998). This is important, since the idiom label is not just syntactically determined, but most importantly it is semantically determined. This decision has a bearing on how the idioms are entered in a dictionary. In *Tlhalosi ya Medi ya Setswana* all idioms and proverbs are entered as separate headwords and not subsumed as subentries under a main entry. This is because, as they are semantically non-compositional, they are separate entities, semantically, syntactically and morphologically. For instance, *tsaya boipuso*, *tsaya dikgang*, *tsaya dikobô ka dikgoka* and others are not entered as subentries of *tsaya* but as independent headwords as shown below.

tsaya **** /tsájà/ •ldr• **1**• go ntsha sengwe fa se teng o se isa go sele
Tsaya ngwana o mo robotse mo ntlong **2**• go tsholetsá sengwe
3• go dirá gore sengwe se nne ka fa tlase ga taolo ya gago
Ditsouludi di ne tsa tsaya puso ka go tlhankgola tautona = gapa
4• go nyala mosadi *Mosimane yo a re o batla go tsaya ngwana wa moagisani wa rona*

tsaya boagedi •leele• go nna moagi wa lefatshe le lengwe

tsaya boipuso •leele• go simolola go ipusa ga lefatshe le sa busiwe ke lefatshe le lengwe *Ka 1966 ke fa lefatshe la Botswana le tsaya boipuso*

tsaya botshelô •leele• go fedisa botshelo kgotsa go bolaya *Bolwetse bo itshokile jwa bo jwa tsaya botshelo jwa ngwana*

tsaya dikgang •leele• **1**• go tlotla *Basadi ba fitlhile ba ema fa*

kgorong ya jarata ba tsaya dikgang ka lobaka 2° go ya mo lefelong motho a bo a gatisa ditiragalo tse di kgathisang mo lefelong leo, a bo a di isa kwa seromamoeng kgotsa kwa dikoranteng Tiro ya go tsaya dikgang e monate

tsaya dikobô ka dikgoka •leele• go tlhakanela dikobo le motho ntleng ga go rata ga gagwe *Mosimane yo o lebisitswe molato wa go tsaya dikobo ka dikgoka* = thubetsa, betêlêla

The decision to enter idioms and proverbs as separate headwords was taken because it is believed that such a strategy is linguistically motivated and especially user-friendly. The user is not left to guess under which headword the idiom or proverb is subsumed.

5.8 Middle section

The dictionary also comprises a rich middle section full of educationally and culturally important material. Amongst these are pages that explain punctuation marks, a detailed presentation of the etymology of Setswana months, kinship terminology, and the presentation of cattle colours, ear markings, cattle horn shapes and others in full colour.

6. Conclusion

This paper has presented Setswana's fourth monolingual dictionary *Tlhalosi ya Medi ya Setswana*. It is the first Setswana dictionary to carry the title *tlhalosi*. The dictionary is situated within a broad Setswana lexicographic tradition which dates to the 1800s. It has discussed some of its key features such as frequency mark-up and dialectal representation. In its compilation a 16 million word corpus, Moe's semantic classification strategy and the use of already published material in the Setswana language have been used. A Setswana monolingual dictionary tradition is still developing. However, there is much to be learnt from developments in lexicography elsewhere. Developments in English lexicography and developments in lexicography in the African languages such as Shona and Swahili can be instructive to the development of Setswana lexicography.

References

- Andersson, L.-G. and T. Janson. 1997. *Languages in Botswana; Language Ecology in Southern Africa*. Gaborone: Longman Botswana.
- Béjoint, H. 2000. *Modern Lexicography: An Introduction*. Oxford/New York: Oxford University Press.
- Biber, D. 1993. Using Register-diversified Corpora for General Language Studies. *Computational Linguistics* 19(2): 219-241.

- Biber, D.** 1994. An Analytical Framework for Register Studies. Biber, D. and E. Finegan (Eds.). 1994. *Sociolinguistic Perspectives on Register*: 31-56. Oxford: Oxford University Press.
- Brown, J.T.** 1875. *Lokwalo loa Mahúké a Secwana le Seeneles*. Johannesburg: Yates and Alexander, for the London Missionary Society.
- Brown, J.T.** 1925. *English Dictionary*. Johannesburg: Pula Press.
- Campbell, J.** 1815. *Travels in South Africa, Undertaken at the Request of the Mission Society*. London: Black, Parry and Co.
- Cole, D.T.** 1955⁹. *An Introduction to Tswana Grammar*. Cape Town: Longman.
- Cole, D.T.** 1995. *Setswana — Animals and Plants (Setswana — Ditschedi le Dithhare)*. Gaborone: The Botswana Society.
- Cole, D.T. and L. Moncho-Warren.** 2012. *Macmillan Setswana and English Illustrated Dictionary*. Johannesburg: Macmillan South Africa.
- Créissels, J. and A.M. Chebanne.** 2000. *Dictionnaire Française–Setswana, Thanodi Sefora–Setswana*. Mogoditshane: Tasalls Publishing.
- Dent, G.R.** 1992. *Compact Setswana Dictionary: English–Setswana, Setswana–English*. Pietermaritzburg: Shuter and Shooter.
- Joubert, P.J. and M.J. Mangokoane.** [1975–1978]. *Lenanentšú-tlhalosi: Seafrikaans–Seisemane–Sesotho sa Leboa*. Johannesburg: Suid-Afrikaanse Uitsaaikorporasie.
- Kgasa, M.L.A. (Ed.).** 1976. *Thanodi ya Setswana ya Dikole*. Cape Town: Longman Penguin Southern Africa.
- Kgasa, M.L.A. and J. Tsonope.** 1998. *Thanodi ya Setswana*. Gaborone: Longman.
- Kilgarriff, A.** 1997. Putting Frequencies in the Dictionary. *International Journal of Lexicography* 10(2): 135-155.
- Kilgarriff, A., P. Richly, 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. Available online at: <http://www.lexmasterclass.com/people/akcv.htm>.
- Lichtenstein, H.** 1815. *Travels in Southern Africa, in the Years 1803, 1804, 1805 and 1806. Volume II*. Translated from the original German by Anne Plumptre. London: Henry Colburn.
- Louw, J.P. and E.A. Nida.** 1989. *Greek–English Lexicon of the New Testament: Based on Semantic Domains. Volume 1*. New York: United Bible Societies/Kaapstad: Bybelgenootskap van Suid-Afrika.
- Mareme, G.B.** 2008. *Thanodi ya Setswana*. Pietermaritzburg: Nutrend Publishers.
- Mathangwane, J.T.** 2008. The Treatment of Loanwords in Setswana Dictionaries. Otlogetswe, T.J. and M.L.A. Kgasa. 2008. *A Pioneer Setswana Lexicographer*: 133-146. Cape Town: Centre for Advanced Studies of African Society (CASAS).
- Matumo, Z.I.** 1993. *Setswana–English–Setswana Dictionary*. Gaborone: Botswana Book Centre and Macmillan Botswana Publishers.
- Miller, O.B.** 1936. *Check-lists of the Forest Trees and Shrubs of the British Empire: No. 6. Bechuanaland Protectorate*. Oxford: Imperial Forestry Institute.
- Moe, R.** 2001. Lexicography and Mass Production. *Notes on Linguistics* 4(3): 150-156.
- Moe, R.** 2003. Compiling Dictionaries Using Semantic Domains. *Lexikos* 13: 215-223.
- Mogapi, K.** 1984. *Thutapuo ya Setswana*. Gaborone: Longman.

- Mojela, V.** 2007. *Pukuntšuthaloši ya Sesotho sa Leboa*. Pietermaritzburg: Nutrend Publishers.
- Murdock, G.P. et al.** 1987⁵. *Outline of Cultural Materials*. New Haven: Human Relations Area Files.
- Otlogetswe, T.J.** 2007. *English–Setswana Dictionary*. Gaborone: Pentagon.
- Otlogetswe, T.J.** 2008. *Corpus Design for Setswana Lexicography*. Unpublished Ph.D. Thesis. Pretoria: University of Pretoria.
- Otlogetswe, T.J.** 2012. *Tlhalosi ya Medi ya Setswana*. Gaborone: Medi Publishing.
- Peters, M.A. and M.M. Tabane.** 1982. *Bibliography of the Tswana Language/Bibiliokerafi ya fine ya Setswana*. Pretoria: State Library.
- Prinsloo, D.J.** 2004. Revising Matumo's *Setswana–English–Setswana Dictionary*. *Lexikos* 14: 158-172.
- Rasebotsa, N.** 2008. M.L.A. Kgasa; The Legacy of my Father. Otlogetswe, T.J. and M.L.A. Kgasa. 2008. *A Pioneer Setswana Lexicographer*: 26-39. Cape Town: Centre for Advanced Studies of African Society (CASAS).
- Roget, P.M.** 1958, 1985. *Roget's Thesaurus*. Harmondsworth: Penguin Books.
- Salt, H.** 1814. *A Voyage to Abyssinia and Travels into the Interior of that Country*. Philadelphia: M. Carey, and Wells and Lilly.
- Scott, M.** 2004–2006. *Oxford WordSmith Tools. Version 4.0*. Oxford: Oxford University Press.
- Selolwane, G.D.** 2004. Ethnic Structure, Inequality and Governance of the Public Sector in Botswana. *UNRISD Project on Ethnic Structure, Inequality and Governance of the Public Sector*.
- Sinclair, J.** 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sinclair, J.M. (Ed.)**. 1987. *Looking Up. An Account of the COBUILD Project in Lexical Computing and the Development of the Collins COBUILD English Language Dictionary*. London/Glasgow: Collins ELT.
- Snyman, J.W., J.S. Shole and J.C. le Roux.** 1990. *Dikišinare ya Setswana–English–Afrikaans Dictionary/Woordeboek*. Pretoria: Via Afrika.
- Van Warmelo, N.J.** 1931 *Kinship Terminology of the South African Bantu*. Ethnological Publications. Volume II. Pretoria: Government Printer.
- Volz, S.** 2003. European Missionaries and Tswana Identity in the 19th Century. *Pula: Botswana Journal of African Studies* 17(1): 3-19.

Le dictionnaire de chimie en lingála pour les élèves de Kinshasa

Bienvenu Sene Mongaba, *Département de Langues et
Cultures africaines, Université de Gand, Belgique*
(senemongaba@yahoo.fr)

Résumé: Pour réaliser un dictionnaire spécialisé dans une langue bien documentée, le lexicographe dépouille les documents disponibles, il en extrait des termes candidats et, éventuellement, leurs définitions et des exemples pertinents. Cette approche n'est pas adaptée pour des langues dont les documents correspondant à la spécialité sont rares ou absents. Tel est le cas du lingála, langue parlée dans la ville de Kinshasa, dans le domaine de la chimie. Outre le travail de dépouillement du corpus, des travaux de création terminologique doivent donc être réalisés, afin de permettre à ces dictionnaires de satisfaire les attentes des utilisateurs. Pour cela, le lexicographe adopte l'une des deux méthodologies suivantes.

La première méthodologie, fréquemment utilisée, procède par la traduction. Elle consiste à utiliser un corpus dans une langue source, en extraire des lemmes, des définitions et des exemples et, enfin, classifier ces données, toujours dans la langue source. Le document ainsi obtenu est ensuite traduit dans la langue cible.

La deuxième méthodologie consiste à définir des concepts et à produire les termes les désignant directement dans la langue source, sans pour autant toujours disposer d'un corpus correspondant dans cette langue. Cette approche nécessite donc que le lexicographe soit lui-même expert dans le domaine en question ou qu'il travaille avec un expert, qui va faire ce travail de production du savoir dans la langue cible. Ensuite, le lexicographe soumet ses résultats à la validation de locuteurs experts. Dans ce travail, nous avons adopté et nous décrivons cette deuxième approche.

En effet, cet article décrit la démarche que nous avons suivie pour la réalisation d'un dictionnaire de chimie bilingue lingála–français à l'intention des enseignants et des élèves de la troisième année secondaire des écoles de Kinshasa.

Mots-clés: TERMINOLOGIE, LEXICOGRAPHIE DES TERMES DE SPÉCIALITÉ, LINGÁLA, CONGO, KINSHASA, DICTIONNAIRE PÉDAGOGIQUE, DICTIONNAIRE DIDACTIQUE, DICTIONNAIRE BILINGUE

Abstract: Dictionary of Chemistry in Lingála for Kinshasa Students. In well-documented languages, the lexicographer aiming to make a specialized dictionary will process an available corpus to extract data (candidate terms, definitions and examples). This approach is not suitable for poorly documented languages, for which texts in any given specialized domain may be few and far between. This is the case with Lingála, spoken in the city of Kinshasa, in the field of chemistry. Besides processing the available corpora, the lexicographer needs to coin terms for concepts and to create definitions and examples in order to produce dictionaries meeting users' expectations. For this purpose, one of the two following methodologies can be adopted.

The first, and more commonly employed methodology adopts the translation approach. A corpus which is available in the source language is processed to extract lemmas, definitions and examples and eventually classify data in the source language. The document obtained in this way is then translated into the target language.

According to the second methodology, in the absence of well-documented specialized corpora in the target language, concepts are defined and coined directly in the target language. Examples are also produced directly in the target language. This approach calls for the lexicographer to be an expert in the given field or for him/her to work with an expert who can produce specialized knowledge in the target language. Then the lexicographer will submit his/her findings to other native speakers, who are also experts in that given field for validation. This is the methodology we have adopted and we are to describe in this paper.

In fact, this article describes the approach we have followed in order to make a Lingála–French bilingual dictionary of chemistry for students and teachers of the third year of secondary school in Kinshasa.

Keywords: TERMINOLOGY, LSP LEXICOGRAPHY, LINGÁLA, CONGO, KINSHASA, CHEMISTRY, PEDAGOGIC TOOL, PEDAGOGICAL DICTIONARY, DIDACTIC DICTIONARY, BILINGUAL DICTIONARY

Introduction

En République Démocratique du Congo, le français est utilisé comme langue d'enseignement et il jouit d'un privilège du point de vue représentationnel de la part de l'institution scolaire. Néanmoins, les élèves des écoles de la RD Congo, en particulier ceux des écoles de la ville de Kinshasa, maîtrisent de moins en moins le français, comme le montrent la plupart des études (Manduku Sasa 2004; Nyembwe et Koni 2004; Yawidi 2009; Nyembwe et Matabishi 2012; Kilosho 2013). Les enquêtes sociolinguistiques réalisées entre 2008 et 2012 (Sene Mongaba 2013b) ont révélé que, dans la ville de Kinshasa, capitale de la RD Congo, les enseignants sont souvent obligés de recourir au lingála pour réexpliquer les leçons pourtant données en français. En revanche, ils n'ont d'autre choix que d'employer un discours non spécialisé en lingála, puisque cette langue n'est pas enseignée à l'école et que les supports didactiques y font cruellement défaut. Les enseignants et les élèves utilisent donc la langue parlée dans la rue et dans le cercle familial. Ceci constitue un handicap dans le processus de transmission du savoir en milieu scolaire, qui exige un langage de spécialité adapté aux savoirs enseignés/appris. L'une des solutions consiste à confectionner des dictionnaires de spécialités qui pourront aider les enseignants et les élèves à construire un discours spécialisé cohérent.

Cet article décrit la démarche suivie pour la réalisation d'un dictionnaire de chimie bilingue lingála–français à l'intention des élèves et enseignants de la troisième année secondaire des écoles de Kinshasa.

Cet article comprend sept sections. Après cette introduction, la deuxième section traite du cadre théorique et méthodologique. La troisième section décrit

l'utilisateur-cible, tandis que la quatrième section expose les objectifs du dictionnaire. La cinquième section montre la démarche suivie pour sélectionner les lemmes et produire des définitions. La présentation de la structure du dictionnaire intervient à la sixième section qui précède la conclusion.

Le cadre théorique et méthodologique: L'extraction des données terminologiques

La terminologie, conçue comme une discipline scientifique prend en compte, comme le soutient Depecker (2009: 97), "les unités linguistiques dans les langues, les concepts par lesquels le réel est représenté dans la pensée, les objets et les représentations". La méthode adoptée dans un travail de terminologie dépend des objectifs poursuivis et des ressources disponibles. Chaque langue a une organisation bien définie à laquelle obéissent les unités linguistiques qui constituent les matériaux dont le terminologue a besoin pour organiser la terminologie dans un domaine bien précis. À côté de cette connaissance linguistique, les concepts, ainsi que les objets à nommer font partie d'un domaine, qui n'est pas la linguistique mais un domaine d'activités, par exemple la chimie, qui a aussi sa nomenclature et son système notionnel propre. La connaissance de cette organisation est également nécessaire dans un travail de terminologie. Enfin, les personnes, à qui est destinée cette terminologie, ont leurs représentations, aussi bien linguistiques que scientifiques ou technologiques, qu'il ne faut pas non plus ignorer.

Deux étapes préalables à la confection des dictionnaires sont importantes en lexicographie. Il s'agit de décrire les caractéristiques de l'utilisateur-cible et de déterminer les objectifs du dictionnaire que l'on veut réaliser. Ces deux étapes dictent les décisions que le lexicographe prendra, par la suite, pour la sélection des lemmes, des définitions, des exemples, du niveau de langue, ainsi que le choix des champs que va comporter chaque article (Prinsloo et al. 2000).

Aussi bien pour la terminologie que pour la lexicographie, l'extraction des termes/lemmes candidats marque le début du travail productif proprement dit. Dans plusieurs langues bien documentées, comme par exemple le français ou l'anglais, les terminologues et les lexicographes passent par le dépouillement des textes disponibles pour en extraire les termes candidats et éventuellement leurs définitions. Cette approche n'est pas adaptée à plusieurs langues africaines — tel est le cas du lingála — à cause de la faible quantité de documentation disponible. Cette faible production de textes ne permet pas un travail terminologique et lexicographique sur base de corpus écrit. En plus donc du travail de dépouillement du corpus, des travaux de création terminologique et de production de contenu doivent être réalisés, afin de permettre à ces dictionnaires de satisfaire les attentes des enseignants et des élèves. Pour cela, les lexicographes des langues africaines adoptent l'une des deux méthodologies suivantes.

La première méthodologie, fréquemment utilisée, consiste à utiliser un

corpus dans une langue source (comme l'anglais ou le français), en extraire des lemmes, des définitions et des exemples et pour enfin classifier ces données dans cette langue source. Le document classifié est ensuite traduit dans la langue cible (Matumele et al. 1995; Diki-Kidiri 2008: 113; Taljard et Nchabeleng 2011). Telle est l'approche que nous appelons traductionnelle.

Nous avons en revanche adopté la deuxième méthodologie, que nous qualifions d'approche terminologique. En effet, nous avons rédigé, en lingála, un manuel de chimie conforme au programme de la troisième année secondaire. Ainsi, nous avons défini les concepts et produit des exemples de contextualisation directement dans la langue cible. Ce travail de définition de concepts et de production d'exemples a ensuite permis la création de termes de spécialité qui se base sur les mécanismes de création de termes en lingála, notamment, *la dérivation*, *la composition* et *l'emprunt* (Sene Mongaba 2012). Cette approche a été possible grâce à notre parcours de chimiste d'une part et de linguiste d'autre part. Bien sûr, si le lexicographe n'est pas en même temps expert du domaine de création des termes, il va se prévaloir de la collaboration d'un expert, mais il sera malgré tout crucial, dans l'approche que nous préconisons, que l'expert produise effectivement du savoir dans la langue cible. Dans cette approche, la connaissance approfondie des mécanismes linguistiques de la langue cible permet la systématisation des méthodes de créations des termes, ainsi que le renforcement des capacités de cette langue.

Il y a lieu de faire une distinction entre cette approche "terminologique" et une approche de "lexicologie spécialisée". En effet, nous adoptons la définition de la lexicologie spécialisée comme étant "la branche de la lexicologie qui examine les unités lexicales des discours spécialisés en général au niveau formel et abstrait afin de dégager des méthodes de repérage, d'analyse et de description des unités lexicales qui lui sont spécifiques" (Valente 2002: 68-71). Nous n'analysons pas les unités lexicales de textes produits en lingála mais nous produisons, en lingála, un texte spécialisé. C'est à partir de ce texte spécialisé que nous extrayons la définition du concept et sur base de termes principaux contenus dans cette définition que nous déduisons une dénomination.

Cette approche n'est pas non plus seulement terminographique, dans la mesure où elle ne se limite pas à extraire et à ranger des termes et leurs définitions, mais, elle aborde aussi l'aspect théorique "en orientant l'utilisateur vers une solution terminologique acceptable linguistiquement et qui répondent pleinement à ses attentes de locuteurs" (Auger 2001: 220-221). La théorie adoptée dans cette approche est liée à la réalisation d'un dictionnaire bilingue à l'intention des élèves dont la langue première est une langue faiblement documentée alors que leur langue d'enseignement est une langue documentée.

Le lecteur gardera à l'esprit que ces deux approches — traductionnelle et terminologique — sont complémentaires, dans la mesure où l'approche traductionnelle permet une production quantitative de termes de spécialité dans un temps relativement court, alors que l'approche terminologique permet d'éprouver l'insertion des termes créés dans la syntaxe et la sémantique de la langue

cible et éventuellement y apporter des modifications. Enfin, le contexte culturel et socio-économique de la communauté linguistique cible y est aussi automatiquement intégré.

La caractérisation de l'utilisateur cible

Le dictionnaire de chimie visé dans cette démarche est destiné aux élèves et aux enseignants de la troisième année secondaire scientifique de la ville de Kinshasa.

Du point de vue du contenu, au Congo, le cours de chimie commence en troisième année secondaire, c'est-à-dire à la neuvième année d'école. Le Ministère de l'enseignement secondaire a publié un programme national de chimie que toutes les écoles sont tenues de respecter (Ministère EPSP 2009). À cet effet, plusieurs livres de chimie, écrits en français et conformes à ce programme, sont édités par des entreprises privées. D'une manière globale, le programme de chimie de troisième année secondaire, à raison de trois heures par semaine, prévoit les matières suivantes:

- les notions fondamentales (définition de chimie, définition de la matière, la transformation chimique et physique),
- l'étude de l'atome et de la molécule,
- la réaction chimique,
- la description de certains atomes, molécules ou mélanges tels que le chlore, l'oxygène, l'eau, l'air, le feu, etc.

Du point de vue sociolinguistique, il existe plusieurs variétés de lingála. Dans ce travail, nous nous intéresserons à la variété dite *lingála ya Kinshasa* qui correspond à la variété utilisée dans les villes de Kinshasa et Brazzaville. Cette variété a quatre principaux registres suivants:

- Le *lingála ya sóló*: il s'agit du registre élaboré. Il est utilisé dans les textes littéraires, les manuels de métiers, les revues de sensibilisation, ainsi que dans les versions récentes de la Bible et du Coran publiées après l'année 2000.
- Le *lingála facile*: il s'agit du lingála parlé dans ces deux villes. Il est désigné sous ce vocable à cause de son taux élevé de mots français incorporés dans le discours des locuteurs.
- Le *lingála ya bayanké* (indoubill): il s'agit du registre argotique.
- Le *langála*: il s'agit d'un registre encrypté du *lingála ya bayanké*.

Le lecteur peut se référer à l'article de Sene Mongaba (2012) en ce qui concerne les détails sur les variétés et les registres du lingála.

Nos enquêtes ont montré que, dans la ville de Kinshasa, le taux de mots français dans le lingála est d'environ 25% pour un locuteur faiblement instruit. Ce pourcentage peut s'élever jusqu'à 50% ou plus pour un locuteur instruit (Sene Mongaba 2013a: 15-135). Les élèves de la troisième année secondaire parlent couramment le *lingála facile* au taux moyen de 25%. Ils sont aussi relativement bien exposés aux textes écrits dans ce registre par l'internet (Facebook, Youtube, forum des sites congolais, ...), des SMS, de la radio et de la télévision. Il s'agit, souvent, de textes traitant les affaires de la vie courante (musique, théâtre, politique, religion, ...). Puisque, comme nous le rappelions plus haut, le lingála n'est pas enseigné à l'école, les élèves n'ont donc qu'un très faible accès aux textes élaborés (scolaires, littéraires, ...), aussi bien en *lingála facile* qu'en *lingála ya sóló*.

Il est utile de souligner à nouveau que le fait que les locuteurs du *lingála facile* utilisent un taux élevé de mots français ne signifie pas que ces locuteurs ont une bonne maîtrise de la langue française. En effet, ces mots sont utilisés, en grande partie, dans la syntaxe du lingála et souvent comme des emprunts et non comme de l'alternance codique.

Les élèves de la troisième année secondaire ont en moyenne 15 ans. Hormis les élèves des écoles élitistes et de quelques écoles des milieux aisés, les élèves des écoles secondaires de la ville de Kinshasa ont, en général, une très faible maîtrise du français (lire, écrire, parler, écouter, comprendre et reformuler). Telle est la raison pour laquelle les 58 professeurs de chimie interviewés dans le cadre d'une autre enquête sur la diglossie français-lingála (Sene Mongaba 2013b) sont convaincus qu'un dictionnaire bilingue lingála-français est nécessaire pour pouvoir réexpliquer les termes de chimie qui sont utilisés en français en classe.

Les élèves de la troisième année secondaire sont dans leur première année de chimie. Les définitions proposées dans le dictionnaire sont donc destinées à des novices en chimie, tout en gardant à l'esprit qu'il ne s'agit pas d'un document de vulgarisation. Ceci signifie que les définitions ainsi que les exemples devront être faits dans un discours de spécialité (chimie) sans toutefois rentrer dans des détails qui sortent du cadre de programme de chimie de la troisième année secondaire scientifique.

En ce qui concerne les enseignants, les métadonnées de l'enquête précitée auprès des 58 enseignants de chimie pour la validation des créations terminologique en lingála ont permis de caractériser le type d'enseignants de chimie que l'on peut rencontrer dans une classe de troisième année secondaire dans les écoles de Kinshasa. Les enseignants de la troisième année secondaire sont diversement qualifiés par rapport à l'enseignement de la chimie. Selon les règlements en RD Congo, les enseignants qualifiés sont ceux qui disposent d'un diplôme de trois ans (premier cycle universitaire) ou d'un diplôme de cinq ans (premier et deuxième cycle universitaire) en pédagogie option chimie-biologie. 30 enseignants (52%) remplissaient cette condition, 20 enseignants (34%) ont fait des études de chimie ou de biologie dans les filières des facultés des

sciences sans formation en pédagogie et 8 enseignants (14%) sont des personnes faiblement formées, ne disposant que de leur diplôme d'enseignement secondaire ou ayant fait, au plus, une ou deux années d'université.

Nous avons évalué la compétence linguistique des enseignants en *lingála facile* sur base de nombre d'années vécu à Kinshasa ainsi que sur leur capacité à répondre dans cette langue au cours des interviews. 25 enseignants sont nés à Kinshasa (43%), 23 vivent à Kinshasa depuis plus de 10 ans (40%), 6 vivent à Kinshasa entre 5 et 10 ans (10%) et enfin 4 habitent à Kinshasa depuis moins de 5 ans (7%). 56 enseignants ont répondu à nos questions en *lingála facile*, tandis que seulement 2 d'entre eux ont affirmé ne pas pouvoir le faire (3%). Cette enquête nous autorise à dire qu'en général ces enseignants ont les compétences linguistiques en tant que locuteurs de *lingála* mais ne disposent pas, de façon consciente, de la connaissance du fonctionnement des mécanismes de dérivation et de composition du *lingála*.

Les objectifs du dictionnaire

L'objectif poursuivi est celui de mettre à la disposition des élèves et des enseignants de la troisième année secondaire scientifique de Kinshasa un dictionnaire *lingála*-français qui servira de support pour la compréhension du cours de chimie. Deux situations possibles peuvent résumer les besoins des élèves et des enseignants.

Situation 1. Les élèves ont besoin de comprendre la signification d'une notion donnée au cours. Il s'agit donc de trouver l'explication en *lingála* d'un mot français. Le mot peut être un terme de chimie ou un mot de la langue générale. L'entrée de l'article sera donc en français mais devra comporter la définition et des exemples en *lingála*. Le terme équivalent en *lingála* favorise aussi la compréhension du concept et du terme en français. Comme démontré dans un travail précédent (Sene Mongaba 2012), le terme en *lingála* contribue également à motiver les élèves dans l'apprentissage de la leçon et à améliorer leur attitude face à leur propre langue.

Comme indiqué plus haut, l'enseignant est souvent amené à réexpliquer sa leçon en *lingála* lorsque les élèves ne comprennent pas le discours en français. La présentation du document sous forme de dictionnaire devra ainsi permettre à l'enseignant de faire une consultation ciblée, même en cours de leçon, pour trouver le terme équivalent ainsi que l'énoncé de la définition en *lingála*. Cette approche contribuera ainsi à réduire l'approximation du discours des enseignants. Le dictionnaire leur permettra d'éviter l'usage de mots génériques de la langue générale en lieu et place des termes de spécialité.

Situation 2. Dans son travail de reformulation didactique et linguistique: face à une nouvelle pratique qu'il a pu exprimer en *lingála*, l'élève ou l'enseignant peut être amené à vouloir transférer cette pratique en français. Dans ce cas, il cherchera un ou plusieurs mots dans le dictionnaire et trouvera son ou ses équivalents en français. En plus, grâce à la définition de ces mots en *lingála*, il

pourra aussi se rendre compte si ce terme lingála est adapté au concept qui désigne cette notion ou s'il n'est pas adapté. Dans le dernier cas, grâce aux définitions de différents mots dérivant du même radical, l'utilisateur pourra alors trouver le terme adéquat. Les différents renvois en lingála lui permettront, le cas échéant, de trouver le mot le plus adapté et ainsi l'apprenant pourra fixer durablement ces nouvelles notions apprises. Ceci est illustré par l'exemple sur le radical *-sang-* (mettre ensemble) que nous exposons plus loin dans ce travail.

Sur base des réalités sociolinguistiques décrites plus haut, nous pouvons affirmer que c'est le *lingála facile* qui est le plus accessible aux élèves. Par conséquent, les définitions et les exemples dans le dictionnaire seront énoncés dans ce registre *lingála facile lf* (les termes de spécialités, les nombres, les grandeurs sont en français), même si nous indiquerons entre parenthèses, des termes équivalents dans le registre élaboré *lingála ya solo ls*. La traduction en français, qui apparaît dans la colonne de droite, contribue à aider les utilisateurs dans leur travail de reformulation, étant donné qu'en classe ils sont obligés d'écrire en français.

Sélections des lemmes et des définitions

À l'heure actuelle, il n'existe qu'une liste de noms d'éléments chimiques (Mbikay 2001; Sene Mongaba 2009) et deux petits livres de chimie écrits en lingála: *Lexique de chimie français-lingála* (Mukinayi 2011a) et *Shimie, mambi ma ebandela* (Mukinayi 2011b). Les publications de Mukinayi constituent une introduction au cours de chimie générale destinée aux étudiants de la première année d'université. Elles ne couvrent pas le programme de chimie de la troisième année secondaire et en plus ils abordent des détails qui vont au-delà de ce qui est prévu pour la troisième année secondaire. Le lingála étant donc une langue très faiblement documentée en matière de chimie, nous avons donc dû, au même moment, tout en rédigeant un livre de chimie pour cette classe, créer aussi des termes de spécialité. Ce manuscrit du livre de chimie ainsi que les deux publications de Mukinayi et la liste des éléments chimiques de Mbikay et complété par Sene Mongaba nous ont servi de corpus. Les définitions en lingála contenues dans ce dictionnaire sont donc des définitions que nous avons produites tout au long de la rédaction de ce livre de chimie. Évidemment, nous avons consulté une bibliographie circonstanciée, constituée des livres de chimie écrits en français et destinés aux écoles congolaises (Mikalukalu 1984, 1986 et 1992, Lufimpadio 1998, Ndjungu Brahimu 2002), ainsi que d'autres livres des écoles secondaires d'autres pays comme la Belgique (Pirson et al. 2009), la Grande Bretagne (Ryan 2006), l'Italie (Tottola et al. 2008), l'Espagne (Del Barrio et al. 2008), la Malaisie (Yasin 2011), le Canada (Kotz et Treichel 2006) et les États-Unis d'Amérique (Moore 2004).

Sur le plan de la terminologie, nous avons d'abord cherché à expliquer, puis à définir le concept en lingála, sans nous soucier de comment cette notion

allait être désignée en lingála. Une fois la définition produite et le concept contextualisé à l'aide d'exemples, le terme de spécialité en lingála adapté à cette définition du concept découle de l'énoncé. Pour cela, notre connaissance des mécanismes de formation des mots en lingála (la dérivation, la composition, l'adaptation morphologique et phonologique des compositions savantes et l'emprunt) a été d'une grande utilité.

Enfin, nous avons travaillé en lingála en nous basant aussi sur la manière dont les différents locuteurs et auteurs des textes généraux contenus dans le corpus s'y sont pris pour définir les concepts dans cette langue de façon compréhensible. Les textes en français contenus dans les dictionnaires sont donc des traductions des textes lingála.

L'avantage de cette démarche est de constater directement si le terme créé s'insère aisément dans la syntaxe du lingála avec le moins d'ambiguïté possible.

Dans le cadre de ce travail, nous avons utilisé le logiciel Notepad++ pour la constitution du corpus. Le logiciel Unitex a été utilisé pour l'extraction des termes dans le corpus et pour évaluer leur fréquence et leur collocation. Le logiciel TswaneLex a servi pour la création du dictionnaire. Les logiciels Unitex et TswaneLex ont été localisés en lingála. Cela nous a permis de travailler dans un environnement informatique rédigé en lingála et obéissant à la structure linguistique et grammaticale du lingála. Cet aspect de notre travail ne rentre pas dans le cadre de cet article.

Présentation du dictionnaire et structures des articles

Le dictionnaire est composé de trois parties.

La première partie comporte les explications, en lingála, sur la façon d'utiliser le dictionnaire, les abréviations utilisées, un résumé sur la morphologie du lingála, le marquage des tons et la variété du lingála utilisée. Elle comporte aussi un guide de l'utilisateur sur la manière la plus efficace de rechercher un terme et les explications sur le concept désigné.

La deuxième et la troisième partie constituent le dictionnaire proprement dit. Chaque article apparaît dans un ordre alphabétique général des termes définis. Nous avons adopté l'ordre alphabétique par soucis d'accessibilité directe à l'information. Cette présentation est bénéfique à l'élève qui ne maîtrise pas encore la structure de son cours de chimie. Cette présentation est aussi favorable à l'enseignant qui peut s'y référer pendant qu'il enseigne. Le système de renvoi pallie le risque que l'utilisateur encourt en n'ayant pas une vue globale avec les autres articles en lien avec le concept défini.

La deuxième partie du dictionnaire concerne la direction français-lingála. Dans cette deuxième partie, il y a deux catégories d'articles: les articles du langage de spécialité et les articles du langage général.

Dans la première catégorie, il s'agit d'articles relatifs à des termes chimiques. Ces articles sont constitués des champs suivants:

- (1) Le premier champ est l'entrée elle-même. Le terme chimique est en français. Il correspond à l'une des notions de chimie prévues dans le programme de la troisième année.
- (2) Le champ *na Lingála* correspond au terme chimique équivalent en lingála.
- (3) Le champ *ezalí-níni* correspond à la définition chimique de la notion. Cette définition peut aussi être précédée d'une contextualisation dans le langage général, si cela est nécessaire. L'énoncé est fait en lingála avec des termes de spécialité en français.
- (4) Le champ *ndakisa* donne un ou plusieurs exemples de chimie, toujours en lingála, avec des termes de spécialité en français. Il arrive aussi que des exemples de la vie courante soient proposés dans le but de contextualiser le concept et de lui apporter un aspect concret.
- (5) Le dernier champ *tálá pé* correspond aux renvois permettant à l'utilisateur d'approfondir le sujet ou d'avoir une vue plus globale.

Tout l'article est traduit en français, dans le but d'aider l'élève ou l'enseignant à reformuler son discours en français. La page se présente donc sous forme d'un tableau à deux colonnes, où la colonne de gauche concerne le dictionnaire en lingála et la colonne de droite correspond à la traduction française du texte qui est à gauche.

<p>Mélange hétérogène: <i>sangísa ekesáná.</i> <i>Banzóto na shímí tó bacorps chimique, mibalé tó leká, etiámí esíká mókó, kasi, nzóto na nzóto ezalí komónana na ndéngé na yangó. Lisangá wáná ndé babéngaka mélange hétérogène tó sangísa ekesáná.</i></p> <p>Ndakisa: Sókí oté pilipíli, matungúlu pé pondú na eboka. Lisangá wáná ezalí <i>mélange hétérogène tó sangísa ekesáná</i>, pó, elóko na elóko ezalí komónana na ndéngé na yangó. Tángó otútí yangó, na súka ekokóma komónana káka néti elóko mókó: <i>pondú batútá</i>. Wáná ekómí <i>mélange homogène</i>.</p> <p>Tálá pé MÉLANGE, MÉLANGE HOMOGENE, SANGÍSA.</p>	<p>Mélange hétérogène: <i>sangísa ekesáná.</i> Deux ou plusieurs corps chimiques sont mis ensemble ou mélangés, mais chaque corps garde son apparence. Cet ensemble est appelé <i>mélange homogène</i> ou <i>sangísa ekesáná</i>.</p> <p>Exemple: Si l'on met du piment, de l'oignon et du <i>pondú</i> (feuille de manioc) dans un mortier, cet ensemble constitue un mélange hétérogène, puisque l'on voit séparément chaque ingrédient. Si vous pilez le tout, à la fin vous obtenez le <i>pondú pilé</i> (<i>pondú batútá</i>), qui est donc devenu un <i>mélange homogène</i>, puisque le tout ne se présente plus que comme un seul corps.</p> <p>Voir aussi MÉLANGE, MÉLANGE HOMOGENE, SANGÍSA</p>
--	--

<p>Mélange homogène: <i>sangísa esangáná.</i> Elingí koloba, banzóto ezalákí ya kokesana, esangání, ekómí komónana néti nzóto káka mókó.</p>	<p>Mélange homogène: <i>sangísa esangáná.</i> Cela signifie que des corps qui étaient distincts se mélangent et ne paraissent que comme un seul corps.</p>
---	---

Ndakisa: tálá na *mélange hétérogène*.
Tálá pé MÉLANGE, MÉLANGE HOMO-
GÈNE, SANGÍSA

Exemple: voir *mélange hétérogène*.
Voir aussi MÉLANGE, MÉLANGE HOMO-
GÈNE, SANGÍSA

Noyau: mukókólí.

Noyau ezalí motéma ya *atome*.
Ndakisa: Mukókólí ya manga ezalí
noyau ya manga.
Atome ekabwáná na biténi míbalé:
noyau pé ba*électron*. Na káti ya mukókólí
ya *atome* sóngóló, okokúta baprotón,
baneutron pé biténi misúsu ya mikélé-
kété. Na libándá ya mukókólí, okokúta
ba*électron* ezalí kotepatepa na bacouche
électronique to banzingánzingá ekabwá-
ná na milongó.

Ndakisa: *Noyau* ya *atome* ya H, idrojeni
(hydrogène), ezalí na *proton* mókó. Oyo
ya Pb, mbódi (plomb), ezalí na baprotón
82.

Tálá pé ATOME, NOMBRE DE MASSE,
NUMÉRO ATOMIQUE, ÉLECTRON, PROTON,
NEUTRON.

Noyau: mukókólí.

Le *noyau* est le cœur de l'*atome*.
Exemple: Le mukókólí de la mangue
est le *noyau* de la mangue.
L'*atome* est subdivisé en deux parties,
le *noyau* et les *électrons*. Au sein du
noyau, nous rencontrons des *protons*,
des *neutrons* et d'autres particules.
En dehors du *noyau* se trouvent des
électrons qui sont dans des *couches élec-*
troniques tout autour du *noyau*.

Exemple: Le *noyau* de l'*atome* H (hydro-
gène) a un *proton*. Celui de l'*atome* de
Pb (plomb) en a 82.

Voir aussi ATOME, NOMBRE DE MASSE,
NUMÉRO ATOMIQUE, ÉLECTRON, PRO-
TON, NEUTRON.

Soluté: emelami.

Ezalí *corps chimique* (nzóto na *shimí*) óyo
ya kantité ya moké emelamaka na
óyo ya kantité ya míngi. Óyo ya míngi
babéngaka yangó *solvant* (emela).
Lisangá ya *emela* ná *emelami* epésaka
solution (lisangana).

Ndakisa: Sókí otíe sukáli na káti ya
mái pó na kosála mái ya lóbo. Sukáli
ezalí *emelami* tó *soluté*. Máí ezalí *emela*
tó *solvant*.

Tála pé SOLVANT, SOLUTION, KOSA-
NGANA, KOSANGANISA, EMELA, LISA-
NGANA, KOMELAMA, KOMELAMISA

Soluté: emelami.

Un *corps chimique* généralement en
faible quantité qui est dissout dans
un autre *corps chimique* en grande quan-
tité. Ce dernier est appelé *solvant*. L'en-
semble forme la *solution*.

Exemple: Lorsqu'on dissout du sucre
dans l'eau pour faire de l'eau sucrée, le
sucre est le *soluté* tandis que l'eau est
du *solvant*.

Voir aussi SOLVANT, SOLUTION, KOSAN-
GANA, KOSANGANISA, EMELA, LISANGANA,
KOMELAMA, KOMELAMISA

Solvant: emela.

Na chimie: Ezalí *corps chimique* (nzóto
na *shimí*) óyo ezalí na kantité ya míngi
pé emelaka banzóto na *shimí* misúsu
ya bakantité ya miké (*bimelami* tó
basoluté) pé yangó nyónso ekosangana
pó na kosála *lisangana* tó *solution*.

Ndakisa: Sókí batíe múngwa ná sukáli
esíká mókó, ekozala ya kosangisa kasi

Solvant: emela.

En chimie: C'est le *corps chimique* en
grande quantité qui dissout d'autres
corps chimiques en petites quantités
(*solutés*). Le tout se mélange et forme
la *solution*.

Exemple: Si l'on met du sucre et du sel
ensemble, ils sont mélangés de façon

<p>esángáná té. Sókí babakísi mái, babalóli, okomóna esangáni. Máí emelí bangó. Máí ezalí <i>emela tó solvant</i>.</p> <p>Tálá pé SOLVANT, SOLUTÉ, EMELAMI, SOLUTION, KOSANGANISA, KOSANGANA, KOSANGISA</p>	<p>hétérogène (juste mis ensemble), mais ils ne se sont pas encore mélangés de façon homogène. Si on ajoute de l'eau, et que l'on agite, ils vont se mélanger de façon homogène. L'eau les a dissous. L'eau est le solvant.</p> <p>Voir aussi SOLVANT, SOLUTÉ, SOLUTION, KOSANGANISA, KOSANGANA, KOSANGISA</p>
--	---

Dans la deuxième catégorie d'articles, il s'agit de termes de langage général en français ayant une occurrence conséquente dans le discours de chimie et qui ont un équivalent lingála. Il s'agit en fait de termes répertoriés et définis dans la partie lingála–français. Les articles sont composés des champs suivants:

- (1) Le premier champ est l'entrée elle-même en français;
- (2) Le champ *na Lingála* correspondant à l'équivalent en lingála;
- (3) Le champ *tála pé* correspond aux éventuels renvois.

Dans cette partie, les définitions et les exemples ne figurent pas dans l'article, car ils sont déjà donnés dans l'article équivalent en lingála.

<p>Mélange: <i>sangísa</i>.</p> <p>Tálá pé MÉLANGE HOMOGÈNE, MÉLANGE HÉTÉROGÈNE, MÉLANGER, SE MÉLANGER</p>	<p>Mélange: <i>sangísa</i>.</p> <p>Voir aussi MÉLANGE HOMOGÈNE, MÉLANGE HÉTÉROGÈNE, MÉLANGER, SE MÉLANGER</p>
--	---

La troisième partie du dictionnaire est consacrée à la direction lingála–français. Le but de cette partie est de permettre à l'utilisateur de prendre conscience des mécanismes linguistiques qu'il utilise inconsciemment en tant que locuteur. Ceci lui permet d'améliorer la compréhension des énoncés définitoires des termes de spécialité figurant dans la deuxième partie. Dans cette troisième partie, il y a aussi deux sortes d'articles: les articles du langage de spécialité et les articles du langage général. Dans le premier cas, l'article est un terme lingála équivalent à un terme de chimie en français. Il s'agit en fait des termes répertoriés et définis dans la partie français–lingála. Dans le deuxième type, l'article peut être un verbe ou un substantif susceptible d'être utilisé lors de la reformulation des leçons de chimie en lingála. Les structures de ces deux types d'articles sont différentes.

Les articles correspondants aux termes de chimie en lingála sont composés de champs suivants:

- (1) Le premier champ est l'entrée elle-même en lingála;
- (2) Le champ *na français* correspondant à l'équivalent en français;
- (3) Le champ *tála pé* correspond aux éventuels renvois.

Bobóngolani na shimí: bo-bóng-ol-an-i/ <i>réaction chimique</i> Tálá pé KOBÓNGOLA, KOBÓNGOLANA, KOBÓNGWAMA, LIKOKANI NA SHIMÍ	Bobóngolani na shimí: <i>réaction chimique</i> Voir aussi KOBÓNGOLA, KOBÓNGOLANA, KOBÓNGWAMA, LIKOKANI NA SHIMÍ
Emela: e-mel-a/ <i>solvant.</i> Tálá pé SOLVANT, SOLUTÉ, EMELAMI, SOLUTION, KOSANGANISA, KOSANGANA, KOSANGISA	Emela: <i>solvant.</i> Voir aussi SOLVANT, SOLUTÉ, EMELAMI, SOLUTION, KOSANGANISA, KOSANGANA, KOSANGISA

Le dictionnaire contient aussi des termes de la langue générale (verbes ou substantifs) qui sont susceptibles d'être utilisés par les enseignants et les élèves dans la reformulation. Les articles relatifs à ces termes comportent des champs supplémentaires. L'énoncé est fait en lingála avec des termes de spécialité en français. Il s'agit des champs suivants:

- (1) Le premier champ est l'entrée elle-même (avec éventuellement sa forme plurielle);
- (2) Le deuxième champ correspond à la structure morphologique;
- (3) Le champ *na français*, correspondant à l'équivalent en français;
- (4) Le champ *ezalí-níni* correspond à la définition (éventuellement dérivationnelle) de l'entrée. Dans ce champ, deux sous-champs sont prévus: *na ndéngé ya chimie* correspond à la définition chimique du terme et *na ndéngé ya batu nyónso* qui correspond à une définition dans le langage général;
- (5) Le champ *ndakisa* est réservé aux exemples de chimie et/ou de contextualisation;
- (6) Le dernier champ *tálá pé* correspond aux renvois permettant à l'utilisateur d'approfondir le sujet ou d'avoir une vue plus globale.

Nous illustrons ce type d'article par quelques termes issus du radical *-sang-*. En chimie, plusieurs concepts relatifs à la notion de *mélange*, de *mise en commun* ou de *contact* interviennent dans la description des processus et des phénomènes. L'enseignant ou l'élève peut être amené à parler d'un de ces concepts. Dans le langage général, l'élève utilisera plus spontanément un verbe du radical *-sang-*, qui évoque l'idée de *mettre ensemble*. Les verbes *kosangisa* (mettre ensemble, mélanger) et *kosangana* (se mélanger) sont couramment utilisés dans le langage général. Si, par exemple, l'élève veut reformuler la notion de *homogénéiser un mélange*, il pourra éventuellement utiliser l'un de ces deux verbes. En vérifiant dans le dictionnaire, il verra que les définitions de ces deux verbes ne correspondent pas à la notion qu'il veut exprimer. Le renvoi lui permettra, par exemple, de consulter aussi le terme *kosanganisa*, qui est un terme moins usité par des locuteurs du lingála, mais qui est plus précis pour désigner le concept d'*homogénéisation*. Les renvois permettent à l'utilisateur de trouver le terme

correspondant au concept.

Kosangana: ko-sang-an-a/*se mélanger*

Na ndéngé ya shimí: Tángo basoluté míbalé tó ebelé etíamí esíká mókó pé ekómí komónana káka néti elóko mókó. Esálamaka tángó batié yangó na káti ya *solvant*. Basoluté esanganí ná *solvant* pó na kosála *solution* tó *lisangana*.

Na ndéngé ya bato nyónso: Tángo bilóko ezalí esíká mókó pé ekómí komónana káka néti elóko mókó. Bilóko yangó esangáni.

Ndakisa: Na mái ya lóbo, sukáli ná mái emónanaka káka néti elóko mókó. Tokoloba sukáli esangáni ná mái pó na kosála mái ya lóbo.

Tála pé MÉLANGE HOMOGÈNE, KOSANGANISA, KOSANGISA, KOBALOLA, KOBALUSA, KOKANGANA, KOKANGISA, KOBÓNGOLANA, KOBÓNGOLISA, KOBÓNGOLANISA, KOBÓNGWANANA, KOBÓNGWANISA.

Kosangana: *se mélanger*

En chimie: Quand deux ou plusieurs solutés sont mis ensemble dans un solvant et le tout ne paraît plus que comme un seul corps. Les solutés se mélangent au solvant pour former une solution.

De manière générale: Quand des choses qui sont ensemble ne paraissent plus que comme une seule chose. Elles se sont mélangées.

Exemple: Dans l'eau sucrée, le sucre et l'eau paraissent comme formant un seul corps. On dit que l'eau et le sucre se sont mélangés. L'eau et le sucre se mélangent pour former de l'eau sucrée.

Voir aussi MÉLANGE HOMOGÈNE, KOSANGANISA, KOSANGISA, KOBALOLA, KOBALUSA, KOKANGANA, KOKANGISA, KOBÓNGOLANA, KOBÓNGOLISA, KOBÓNGOLANISA, KOBÓNGWANANA, KOBÓNGWANISA.

Kosangisa: ko-sang-is-a/*mélanger*

Na ndéngé ya shimí: Tango otíé *bacorps chimique* (*banzoto na shimí*) míbalé tó ebelé esíká mókó. *Bacorps* yangó ekokí komónana néti elóko mókó (*mélange homogène*) tó pé elóko na elóko ekokí komónana na ndéngé na yangó (*mélange hétérogène*).

Na ndéngé ya bato nyónso: Kotía bilóko ékóma esíká mókó.

Ndakisa: Sókí batié sukáli na kópo. Babakísí míliki ya pusyér. Babalúsí yangó na lútu. Elingí koloba, basangísí yangó.

Tála pé SANGÍSA, MÉLANGE HÉTÉROGÈNE, MÉLANGE HOMOGÈNE, KOSANGANA, KOSANGANISA, KOBALOLA, KOBALUSA, KOKANGANA, KOKANGISA, KOBÓNGOLANA, KOBÓNGOLISA, KOBÓNGOLANISA, KOBÓNGWANANA, KOBÓNGWANISA.

Kosangisa: *mélanger*

En chimie: Quand on met deux ou plusieurs corps chimiques ensemble. Ces corps mélangés peuvent paraître comme un seul corps (*mélange homogène*) ou chaque corps peut apparaître distinctement (*mélange hétérogène*).

De manière générale: Mettre les choses ensemble.

Exemple: On met du sucre dans un gobelet et on y ajoute du lait en poudre. Ensuite, à l'aide d'une cuillère, on mélange.

Voir aussi SANGÍSA, MÉLANGE HÉTÉROGÈNE, MÉLANGE HOMOGÈNE, KOSANGANA, KOSANGANISA, KOBALOLA, KOBALUSA, KOKANGANA, KOKANGISA, KOBÓNGOLANA, KOBÓNGOLISA, KOBÓNGOLANISA, KOBÓNGWANANA, KOBÓNGWANISA.

Kosanganisa: ko-sang-an-is-a/homogénéiser
Kosanganisa tó kohomogénéiser ekesení ná *kosangana*. Ekesení pé ná *kosangisa*.

Na ndéngé ya shimí: Kosála ke bacorps chimique (banzóto na shimí), míbalé tó leká, ésangana yangó na yangó pé ékela *sangisa esangána* tó *mélange homogène*.

Na ndéngé ya bato nyónso: Kosála ke bilóko ésangana yangó na yangó tí ékoma néti elóko móko.

Ndakisa: 1. Sóki oté pilipíli ná múngwa na eboka. Wáná osangísí pilipíli ná múngwa (*kosangisa*). Sikóyo tángu otútí yangó pé esangáni ekómí káka potopóto móko (*kosangana*).

Yó moto osálí pilipíli ná múngwa ésangana. Wáná osangánísí yangó (*kosanganisa*).

2. Sóki osangísí mafúta na súde. Obalólí yangó tí ebimísí savón. Okomóna bisíká mosúsu esálí bambumambuma, bisíká mosúsu potopóto. Okosanganisa yangó tí bisíká nyónso ékóma potopóto ya ndéngé móko.

Tálá pé KOSANGANA, KOSANGISA, ESANGANISI, LISANGÁ, LISANGANI

Kosanganisa: *homogénéiser*

Homogénéiser ou *kosanganisa* diffère de *mélanger* (*kosangisa*) ou *se mélanger* (*kosangana*).

En chimie: Amener deux ou plusieurs corps chimiques à se mélanger pour former un mélange homogène.

En général: Faire que deux ou plusieurs objets se mélangent de façon à ne paraître que comme un seul objet.

Exemple: 1. Si tu mets du piment et du sel dans un mortier, tu les as mis ensemble. Tu as mélangé du piment et du sel. *Osangísí* du piment et du sel (*mélanger*).

Ensuite, quand tu piles, le tout se mélange et ne forment plus qu'une seule pâte. Le piment et du sel se sont mélangés (*se mélanger*).

C'est toi qui as fait qu'ils se mélangent, tu les as homogénéisés, *osangánísí yangó*.

2. Tu mélanges de l'huile avec de la soude caustique. Tu malaxes jusqu'à obtenir du savon. Tu verras que par endroits il y aura des morceaux durs et par endroits une pâte. Tu vas homogénéiser ton produit jusqu'à avoir une pâte homogène.

Voir aussi KOSANGANA, KOSANGISA, ESANGANISI, LISANGÁ, LISANGANI.

Conclusion

Cet article a décrit le travail de conception et de rédaction d'un véritable outil pédagogique à l'attention des élèves de la troisième secondaire scientifique. Il s'agit du dictionnaire de chimie lingála-français. Ce dictionnaire respecte deux préalables lexicographiques, qui consistent à caractériser l'utilisateur et à déterminer les objectifs du dictionnaire.

Sur le plan théorique et méthodologique, cette étude a abordé l'approche terminologique dans la confection d'un dictionnaire bilingue langue documentée — langue faiblement documentée. Cette approche consiste à produire du savoir directement dans la langue cible et à le classer ensuite en rapport avec les termes de spécialité extraits de la langue source (documentée). Cette approche

est différente de l'approche traductologique, qui consiste à extraire les termes, les définitions et les exemples du corpus disponible en langue source et à procéder ensuite à la traduction. Si l'approche terminologique a l'avantage de régler la question de l'usage du terme créé dans la syntaxe de la langue cible, elle nécessite néanmoins que le lexicographe soit aussi expert du domaine ou que le lexicographe travaille avec un expert, qui, lui, va effectivement produire un savoir dans la langue cible. L'approche terminologique est complémentaire à l'approche traductionnelle pour un travail lexicographique global. Ainsi, cette étude a contribué à montrer que la terminologie et la lexicographie répondent, mieux que la linguistique du type saussurien, aux demandes actuelles des locuteurs des langues africaines, afin que ces langues puissent être utilisées comme langues d'enseignement.

Enfin, certains gouvernements (France, Belgique, Afrique du Sud, etc.) disposent de services de terminologie pour la standardisation des propositions (Depecker 2009, Taljard et Nchabeleng 2011). Un organe institutionnel de ce type n'a pas encore été mis en place au Congo, ni dans plusieurs pays d'Afrique. Étant donné que les langues nationales congolaises ne sont pas de langues officielles, le gouvernement congolais n'a pas encore intégré cette dimension dans sa politique linguistique, qui, par ailleurs, n'est pas encore clairement établie et active. La création terminologique reste encore une activité des chercheurs indépendants. Tel est le cadre dans lequel se situent ces travaux de terminologie en lingála. Il serait souhaitable dans l'avenir que des équipes de recherche soient constituées dans ce sens dans les institutions de recherche universitaires du Congo comme par exemple le centre congolais de terminologie (CECOTEL) qui vient d'être ouvert à l'université de Kinshasa.

Références

- Auger, P.** 2001. Essai d'élaboration d'un modèle terminologique/terminographique variationniste. *TradTerm* 7: 183-224.
- Del Barrio, J.I., J. Puente, A. Caamaño et M. Augustench.** 2008. *Física y Química 1*. Madrid: Bachillerato.
- Depecker, L.** 2009. La terminologie: ouverture d'un champ disciplinaire. Willems, Martine (Éd.). 2009. *Pour l'amour des mots*: 91-106. Bruxelles: Facultés universitaires Saint-Louis.
- Diki-Kidiri, M.** 2008. La description de la méthode. Diki-Kidiri, M. et al. (Éds.). 2008. *Le vocabulaire scientifique dans les langues africaines: pour une approche culturelle de la terminologie (théorie, méthodologie et applications)*: 113-142. Paris: Karthala.
- Kilosh Kabale, Sim.** 2013. Comment gérer le multilinguisme et le plurilinguisme dans les universités des Grands lacs et de l'Afrique centrale. Cas de la RD Congo. *Synergies Afrique des Grands Lacs* 2: 33-43.
- Kotz, J.C. et P.M. Treichel Jr.** 2006. *Chimie générale*. Bruxelles: De Boeck.
- Lufimpadio, Nd.** 1998. *Travaux pratiques de Chimie*. Kinshasa: Afrique éditions.
- Manduku Sasa, H.** 2004. Langues et Communication en milieux étudiants de Kinshasa. Cas de l'Université de Kinshasa. *Penser la Francophonie*: 545-551.
- Matumele, M., P.A. Ayibite et P. Epanga.** 1995. *Terminologie grammaticale et pédagogique*. *Lexique*

- Français–Lingála Lingála–Français*. Nyembwe, N. (Éd.). 1995. Kinshasa: EUA/CELTA.
- Mbikay, D.P.** 2001. *Molongo mwa kemi*. Inédit.
- Mikalukalu, K.K.M.** 1984. *Lexique de chimie. Vol.2*. Kinshasa: CRP.
- Mikalukalu, K.K.M.** 1986. *Lexique de chimie. Vol.1*. 2^{èd.} Kinshasa: CRP.
- Mikalukalu, K.K.M.** 1992. *La chimie avec du matériel de fortune*. Kinshasa: CRP.
- Ministère EPSP RD Congo.** 2009. *Programme national de chimie*. Kinshasa: Édition électronique.
http://www.epsp.cd/component/option,com_joomdoc/Itemid,84/gid,52/task,cat_view/
- Moore, J.T.** 2004. *Chemistry Made Simple*. New York: Broadway Books.
- Mukinayi, J.M.** 2011a. *Lexique de chimie français–lingála*. Kinshasa: Science pour tous.
- Mukinayi, J.M.** 2011b. *Shimie, Mambi ma ebandela*. Kinshasa: Science pour tous.
- Ndjungu Brahimu, W.K.** 2002. *Cours de chimie 3ème année. Chimie élémentaire*. Kinshasa: CRP.
- Nyembwe, N. et T. Koni.** 2004. La francophonie en milieux estudiantins de Kinshasa: Pratique–Perception–Perspectives. *Penser la Francophonie*: 471-478.
- Nyembwe N. et S. Matabishi.** 2012. Le devenir du français en République Démocratique du Congo et question de la norme. Blumenthal, Peter et Stefan Pfänder (Éds.). 2012. *Actes du colloque: Convergences, divergences et la question de la norme en Afrique francophone. Le Français en Afrique* 27: 109-119.
- Pirson, P., H. Bordet, D. Castin, P. Snauwaert et R. Van Elsuwe.** 2009. *Chimie 3e/4e*. Sciences de base et sciences générales. Bruxelles: De Boeck.
- Prinsloo, D., A.R. Chuwa et E. Taljard.** The Lexicon of Africa. Webb, Vic et E. Kembo-Sure (Éds.). 2000. *African Voices*: 220-244. Cape town: Oxford University Press.
- Ryan, L.** 2006. *New Chemistry for You*. Revised Edition for all GCSE Examinations. Cheltenham: Nelson Thornes.
- Sene Mongaba, B.** 2009. *Tableau périodique de chimie–Etanda ebandelaka ya bibuki ya kemi*. Kinshasa/Wavre: Mabiki.
- Sene Mongaba, B.** 2012. The Use of Lingála in the Teaching of Chemistry in DR Congo: A Socio-terminological Approach. Marlo, Michael R. et al. 2012. *Selected Proceedings of the 42nd Annual Conference on African Linguistics*: 308-320. Somerville, MA: Cascadilla Proceedings Project.
<http://www.lingref.com/document#2779>.
- Sene Mongaba, B.** 2013a. *Lingála na matéya ya nzebi na bitéyelo ya Kinshasa. Búku II*. Thèse de doctorat. Gand: Université de Gand.
- Sene Mongaba, B.** 2013b. Analyse de la diglossie Français–Lingála dans les écoles de Kinshasa. Omer, D. et F. Tupin (Éds.). 2013. *Éducatons plurilingues: L'aire francophone entre héritages et innovations*: 151-166. Rennes: Presses Universitaires de Rennes.
- Taljard, E. et M.J. Nchabeleng.** 2011. Management and Internal Standardization of Chemistry Terminology: A Northern Sotho Case Study. *Lexikos* 21: 194-216.
- Tottola, F., A. Allegrezza et M. Righetti.** 2008. *Nuovo corsa di chimica*. Milan: Minerva Italica.
- Valente, R.S.** 2002. *La "Lexicologie explicative et combinatoire" dans le traitement des unités lexicales spécialisées*. Thèse de doctorat. Montréal: Université de Montréal.
- Yasin, Y.** 2011. *Chemistry SPM 4.5*. Petaling Jaya: Sastadi sdn.
- Yawidi Mayinzambi, J.P.** 2009. *Pourquoi mon enfant a échoué*. Kinshasa: Mabiki.

The Role of the Learner's Native Culture in EFL Dictionaries: An Experimental Study

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

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

Abstract: This article aims to demonstrate the hypothesis that the use of native culture (C1) in EFL learners' dictionary definitions and/or examples is useful in the comprehension of the looked-up words. This is done by means of a survey involving more than 100 lower-intermediate EFL Catalan students. The subjects were first presented with a pre-test in which they had to translate 30 English words. Then they were divided into two groups, each of whom had to take a different test. Test 1 contained the definitions of the 30 words taken from a dictionary aimed at a global audience, whereas in test 2 the definitions were taken from a culturally nativized dictionary, that is, a dictionary that included C1 elements. In the tests, the students were asked to translate again the 30 English headwords given in the pre-test. After comparing the results of the pre-test with those of the tests, the study concludes that students who use an EFL dictionary that includes C1 references have more than double the possibilities of understanding the meaning of a new looked-up word than those who do not. The results obtained confirm for the first time in the field of pedagogical lexicography the tenets of schema theory, which highlights the importance of background (e.g. cultural) knowledge to improve reading comprehension. The main implication of this finding for lexicography is that it is desirable that designers of EFL dictionaries deploy nativized versions, especially at lower levels, in order to facilitate comprehension of the foreign language.

Keywords: EFL DICTIONARIES, PEDAGOGICAL LEXICOGRAPHY, PEDAGOGICAL DEFINITIONS, FOREIGN LANGUAGE LEARNING, CATALAN LEARNERS OF ENGLISH, READING COMPREHENSION, NATIVE CULTURE, NATIVIZATION, CULTURAL REFERENCES, GLOCALIZATION, LOCALIZATION, SCHEMA THEORY

Opsomming: Die rol van die aanleerder se inheemse kultuur in EVT-woordeboeke: 'n Eksperimentele studie. Hierdie artikel het ten doel om die hipotese te bewys dat die gebruik van inheemse kultuur in EVT-aanleerderswoordeboekdefinisies en/of -voorbeelde nuttig is vir die begrip van die nageslane woorde. Dit word gedoen deur middel van 'n opname van meer as 100 laer-intermediêre EVT- Katalaanse studente. Aan proefpersone is eers 'n voortoets gegee waarin hulle 30 Engelse woorde moes vertaal. Daarna is hulle in twee groepe verdeel wat elkeen 'n verskillende toets moes aflê. Toets 1 het die definisies van 30 woorde bevat wat geneem is uit 'n woordeboek gerig op 'n globale gehoor, terwyl in toets 2 die definisies geneem is uit 'n kultureel gemotiveerde woordeboek, dit wil sê, 'n woordeboek wat C1-elemente insluit. In die toets is

die studente gevra om weer die 30 Engelse trefwoorde te vertaal wat in die voortoets gegee is. Nadat die resultate van die voortoets met dié van die toetse vergelyk is, toon die studie dat die studente wat 'n EVT-woordeboek gebruik wat C1-verwysings insluit, meer as dubbel die kans staan om die betekenis van die nuwe nageslane woord te begryp as dié wat dit nie doen nie. Die resultate wat verkry is, bevestig vir die eerste keer die beginsels van die skemateorie op die gebied van die opvoedkundige leksikografie, wat dui op die belangrikheid van agtergrond- (bv. kulturele) kennis om leesbegrip te verbeter. Die hoofimplikasie van hierdie bevinding vir die leksikografie is dat dit wenslik is vir beplanners van EVT-woordeboeke om genativiseerde weergawes te ontwikkel, veral op die laer vlakke, ten einde begrip van die vreemde taal te vergemaklik.

Sleutelwoorde: EVT-WOORDEBOEKE, OPVOEDKUNDIGE LEKSIKOGRAFIE, OPVOEDKUNDIGE DEFINISIES, AANLEER VAN 'N VREEMDE TAAL, KATALAANSE AANLEERDERS VAN ENGELS, LEESBEGRIP, INHEEMSE KULTUUR, NATIVISASIE, KULTURELE VERWYSINGS, GLOKALISASIE, LOKALISASIE, SKEMATEORIE

1. Introduction

Since the late 1980s there has been a major shift in the research field of reading comprehension. According to Ajideh (2003), there are two ways of understanding reading: the first assumes that meaning resides in the text itself, whereas the second views reading as an active process in which the reader constructs meaning by means of a series of mental operations. More precisely, the second view of reading comprehension posits that there is an interaction between the text and the reader's own background knowledge, and that it is this interaction that facilitates comprehension. The interaction between the reader's background knowledge and the text is a multifaceted process that may involve a wide array of cognitive processes, such as expectations, predictions, comparisons, prior knowledge, selection among possible meanings, inferencing, remembering, reasoning and problem solving (Nassaji 2002, Ajideh 2003). All these cognitive phenomena are key components of what is known as 'schema theory'. Schema theory began with Bartlett's classical study *Remembering: A Study in Experimental and Social Psychology* (1932), yet Bartlett's ideas did not have an important impact on cognitive psychology until the 1970s and 1980s. The basic tenet of schema theory posits that people's understanding and remembrance of events is influenced by their expectations or prior knowledge. The mental representation of these expectations or prior knowledge is called 'schema' (plural 'schemas' or 'schemata'). To put it in other words, schema theory is based on the assumption that prior knowledge or past experience helps us understand new knowledge.

It is easy to see, then, how researchers in the field of reading comprehension have taken as their basis the presuppositions of schema theory in order to carry out experimental studies that support the view that reading is a process in which readers use their previous background knowledge when interacting with new texts. It is through this interaction, they argue, that meaning is con-

strued. Many studies in reading comprehension have been carried out to support the schema theory: Nassaji (2002: 440) and Rokhsari (2012: 46) mention quite a few of them; for a summary of schema theory and its implications in EFL learning and teaching, see Cook (1997); and for studies that criticize and point out the limitations of schema theory, especially in the field of reading comprehension, see Carver (1992) and Nassaji (2002).

Apart from giving new insights into the process of reading comprehension, the schema theory has also had a practical impact on instruction, and more particularly on the following aspects: (1) research into the improvement of reading comprehension; and (2) the way EFL materials are designed. As for the first aspect, namely research into how to improve reading comprehension, there have been two basic fields of development. On the one hand, a considerable number of studies (for example Ajideh 2003, Karakas 2005, and Alemi and Ebadi 2010) have focused on the beneficial effects of pre-reading activities: such activities provide background knowledge to the main text, and thus facilitate the understanding of new knowledge. On the other hand, there have been several studies that have shown that providing cultural background before reading improves reading comprehension, whether it be with native students belonging to the same racial community (Garth-McCullough 2008), with native students belonging to different racial communities (Reynolds et al. 1981) or with EFL learners (Floyd and Carrell 1987). More specifically, there have been studies (for example, Chihara, Sakurai and Oller 1989, Alptekin 1993, Post and Rathet 1996, Jiang 2000, Coffey 2001, Fuhong 2004, Alptekin 2006, Al-Shumaimeri 2006, Chang 2007, Jalilifar and Assi 2008, and Rokhsari 2012) that have focused on the role that 'nativization' has on EFL reading. Nativization is a process by which culturally unfamiliar words in texts written in the foreign language are changed into more familiar native culture (C1) words. The studies that nativize texts have found that cultural familiarity (i.e. making use of the students' prior cultural knowledge, namely native culture or C1) clearly facilitates the students' understanding of EFL texts.

As for the second aspect mentioned above, that is to say, how schema theory has affected EFL materials design, nowadays it is relatively common to find pre-reading activities in most EFL textbooks. Apart from this, though, there have been very few implementations in materials design that take advantage of the readers' prior knowledge. Alptekin (1993: 140-141) has listed some situations in which the use of the learners' schematic knowledge may be beneficial when learning a foreign language: reading comprehension, writing and information retention. And Coffey (2001) has applied schema theory to the design of classroom activities. However, the schema theory has not been empirically studied in the case of understanding the definitions in EFL learners' dictionaries: this study aims to break new ground into this hitherto neglected aspect, for the studies carried out so far centre on reading comprehension of texts (or stories), whereas here we will focus on an idiosyncratic aspect of reading, namely reading dictionary definitions, which differs from reading texts or stories mainly in the length of the text and in the number of ideas encountered (for a

typology of texts and their differences, see Hatim and Mason 1990: 153-160).

The purpose of this study, then, is to explore the role that the learner's native culture (C1) plays in the comprehension of L2 (i.e. English) dictionary definitions and/or examples. In order to do so, this article compares the effectiveness of two kinds of dictionary: the *Easy English Dictionary with a Catalan-English Vocabulary (EED)* and the *Longman New Junior English Dictionary (LNJED)*. The *EED* is aimed at 10-14 year-old Catalan learners of English and it is devised for a particular cultural background end-user, namely native Catalan speakers whose primary culture (C1) is Catalan and Spanish (Catalan students can speak both Catalan and Spanish and are immersed in both Catalan and Spanish culture). On the other hand, the *LNJED* is aimed at 10-14 year-old learners of English as a foreign language and it is devised for a non-specific cultural background end-user, that is, it does not take into account the students' primary culture. Given the differences between these two types of dictionary, this article aims to answer the following research question: does the presence of C1 in EFL learners' dictionaries improve the comprehension of the words defined? And, if so, to what extent? To put it another way, this article aims to test whether a native cultural element (C1) that appears in the L2 definition or example of an EFL learners' dictionary entry can be determining in understanding the meaning of the defined word. Our research hypothesis is that the presence of C1 in EFL learners' dictionaries significantly improves the comprehension of the words defined, a fact that would seem to be supported, a priori, by the schema theory, and which we first hypothesized in 2008, stating that when reading an EFL dictionary definition, Catalan students might activate their previous linguistic and sociocultural knowledge of their native language and culture, as well as their knowledge of the world (Pujol, Masnou and Corrius 2008). In this article, we will leave aside the students' previous knowledge of the world and the role that their native language may play when trying to understand English. Instead, we will focus only on the role that the students' native culture (C1) may have when reading the foreign language. Our 2008 hypothesis was refined in 2010, when we (Corrius and Pujol 2010: 142) hypothesized that, among other strategies, localizing and glocalizing culture are useful mechanisms to improve comprehension when reading foreign language dictionary definitions. In other words, this article uses the schema theory to confirm Corrius and Pujol's (2010) hypothesis that dictionary users rely on their already established schematic cultural knowledge (C1) when trying to understand the meaning of new words.

2. Methodology

2.1 Subjects

The subjects who participated in the study were Catalan students of English as a foreign language. Their ages ranged from 11 to 13, and their level of English

was lower intermediate. The levels tested were 6th Primary Education (the last year of compulsory primary education, i.e. between 11 and 12 years old) and 1st ESO (the first year of compulsory secondary education, i.e. between 12 and 13 years old). A total of 114 students participated in the study (58 boys and 56 girls), but only 102 were assessed. This was due to the fact that 12 tests were discarded either because of the student's extremely poor level of English (most of them were foreign students who had just arrived in Catalonia and had serious difficulties in understanding English and Catalan) or because they did not follow the procedure correctly (it was detected that they read the third part of the questionnaire before the second part, so the results were declared invalid). Out of the 102 students assessed, 50% were female and 50% male.

The participants belonged to two different schools, one from Tona and one from Torelló (two Catalan villages near Barcelona, Spain). As will be seen in section 2.2, all the students took during the English class time (in May 2012) a questionnaire which consisted of three parts: (1) a section which asked for personal details (age, sex and course level); (2) a pre-test; and (3) a test (either test 1 or test 2). Out of the 102 subjects assessed, 60 6th Primary Education students answered the questionnaire: 20 from Tona plus 9 from Torelló took test 1, and 20 from Tona plus 11 from Torelló took test 2. On the other hand, 42 1st ESO students from Tona answered the questionnaire: 23 took test 1 and 19 took test 2. Thus, a total of 52 students took test 1 and a total of 50 students took test 2. These figures are summarized in table 1, where the data in italics refer to 1st ESO students and the data in bold type refer to 6th Primary students.

School	Level/(years old)	Test 1	Test 2	TOTAL
TONA	<i>1st ESO (12-13)</i>	27 (-4) = 23	26 (-7) = 19	42
TONA	6th Primary (11-12)	20	20	60
TORELLÓ	6th Primary (11-12)	10 (-1) = 9	11	
TOTAL		(23+29) 52	(19+31) 50	102

Table 1: Number of students who answered the questionnaire

The allocation of the students in each class into two subgroups (one taking test 1 and the other taking test 2) involved two different stages. First of all, we asked the teachers (in a previous interview in which we presented the experiment to them) that they divide each class into five levels according to the class marks that they had obtained during the course (from September 2011 to May 2012; the course ended in June). This, we told them, was easy to do: as in Catalonia the students' performance is evaluated on a scale from 0 to 10 (10 being the highest possible mark), the only function of the teachers was to place the students into five groups which corresponded to five grade scales: 0-2, 2.1-4, 4.1-6, 6.1-8 and 8.1-10. In this way, five subgroups were created, each of which

roughly represented 20% of the students. After this, the teachers were asked to randomly distribute the five subgroups into two groups, and to keep a note of their names so that at the time of the experiment they knew to which group each student belonged.

2.2 Materials

As stated in section 1, the aim of this study is to find out the role that C1 plays in understanding L2 (English) definitions in monolingual learners' dictionaries. In order to carry out the study, a questionnaire (reproduced in the Appendix) was used. The questionnaire had three parts, which the students had to answer consecutively and individually. The first part, which was in L1 (Catalan), was a general questionnaire on personal details (age, sex and course level). The second part, which had the instructions in Catalan, was a pre-test that consisted of a list of 30 English words which the students had to underline (if they thought they knew its meaning) and translate into their own language (Catalan). The third part of the questionnaire consisted of a test (test 1 or test 2). Both tests contained the same 30 words listed in the second part, but this time each entry was followed by an L2 definition and/or example. The students had to provide a translation for each of the defined words.

Test 1 contained the definitions from the *Longman New Junior English Dictionary (LNJED)* with the exception of the entries 'beech' and 'chestnut', whose definitions were taken from *The Oxford Illustrated Junior Dictionary (OIJD)*, and the entry 'tongue twister', whose definition was adapted from the *Easy English Dictionary with a Catalan-English Vocabulary*. These exceptions were due to the fact that the entries 'beech', 'chestnut' and 'tongue twister' did not appear in the *LNJED*. Test 2 contained the definitions from the *Easy English Dictionary with a Catalan-English Vocabulary (EED)*. The *LNJED* and the *OIJD* are aimed, as most EFL learners' dictionaries, at a global market. This means that the definitions and examples do not contain any reference to the learner's cultural background. The *EED*, on the other hand, is aimed at a local market. That is to say, the definitions and/or the examples chosen for test 2 contain a reference to the learner's cultural background.

In test 1, the entries (taken mostly from the *LNJED*) do not have any reference to Catalan or Spanish culture. In some cases (see 'draw', 'far', 'north', 'south' and 'take' in part three of the Appendix), there are references to British places such as London, Manchester or Oxford, which Catalan students (according to their teachers) cannot locate properly on a map. The rest of the examples provided are very general and aimed at a global market (see the entries 'against', 'apart', 'blind', 'branch', 'brand', 'coast', 'painting', 'picture', 'price', 'right' and 'used to' in part three of the Appendix). Thus, in either case, the examples do not mean much to Catalan students. It must be noted that there is a word in Test 1 ('beech') that was accompanied by an illustration (this is not reproduced in the Appendix, even though the indication "[+ illustration]" has been added after the definition.

As for test 2, it includes some of the entries from the *EED* which contain culturally-based material for Catalan learners of English. The entries chosen have Catalan and sometimes Spanish cultural references. Of the entries chosen, nine make reference to Catalan or Spanish places that students are familiar with (see the entries 'beech', 'coast', 'eastern', 'island', 'monk', 'north', 'sights', 'south' and 'used to' in part three of the Appendix), four make reference to famous Catalan or Spanish painters and musicians (see the entries 'cello', 'painter', 'painting' and 'picture'), three make reference to famous local football teams (see 'against', 'defeat' and 'draw'), three refer to distances between well-known cities (see 'apart', 'far' and 'take'), two refer to the European currency (see 'banknote' and 'price'), two refer to Catalan festivities (see 'chestnut' and 'firework'), one makes reference to the Spanish organisation for blind people (see 'blind'), one makes reference to the most important Catalan bank (see 'branch'), one refers to the most famous Spanish department store (see 'department store'), one makes reference to a Spanish fact (see 'right'), one makes reference to a Catalan linguistic specificity (see 'tongue twister'), one makes reference to famous airlines (see 'airline') and one refers to a very well-known toothpaste in Spain (see 'brand').

It is worth mentioning that in test 2 four of the words defined ('airline', 'right', 'take' and 'tongue twister') contain a glocal example, that is, an example which is both global and local. The *Oxford English Dictionary (OED)* defines 'glocalization' as: "The action, process, or fact of making something both global and local; *spec.* the adaptation of global influences or business strategies in accordance with local conditions." There are two key ideas in this definition: (1) the global and the local *coexist*; and (2) the global is *adapted* to the local. Coexistence and adaptation, then, are the two criteria for judging if glocalization occurs. These two ideas are confirmed by the first example of 'glocalization' recorded in the *OED*: "We've witnessed what you might have heard called 'glocalization': making a global product fit the local market." In other words: the local is adopted in global contexts, the result being a kind of hybrid that produces "simulacra of local authenticity" (*OED*, third example under 'glocalization'). Take, for instance, the entry 'tongue twister' as it appears in test 2:

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

The example in this entry is glocal in that it is aimed both at a global (international) audience and a local (Catalan) audience: it does not prioritize one culture at the expense of the other but presents both cultures on equal terms, which is something that global dictionaries do not do. This juxtaposition of C1 and C2 obviously has the advantages of localization, a phenomenon which hypothetically "leads to a potentially richer understanding of the word defined" (Corrius and Pujol 2010: 137). It is worth noting that in the example given for 'tongue twister' the global has not been deleted; rather, it coexists with the local, the result being that the whole example for the headword defined adapts

(by means of the equation local X + local Y = tongue twisters) the global cultural information (C2: "She sells sea shells on the sea shore") to the local cultural background of students (C1: "Setze jutges d'un jutjat mengen fetge d'un penjat"). This procedure (glocalization) is in line with Alptekin's view that in ELT a new "intercultural communicative competence" is needed so that instructional materials and activities "involve local and international contexts that are familiar and relevant to language learners' lives" (Alptekin 2002: 57 and 63).

2.3 Procedures

The administrators of the questionnaire were the authors of the study themselves, but they were aided by the students' EFL teacher. The students took the time they needed to complete the questionnaire, which was about 30 minutes. No training session was given to respondents, but clear instructions were given in Catalan so as to avoid misunderstandings.

The selection of the words in the pre-test and the two tests was made on the following basis. First, the words in test 2 were selected by a random browsing of different entries in the *EED*: the condition for being a word that could be chosen was that the definition and/or example of the headword should be nativized, that is, it should contain at least one C1 item. When these words had been selected, their equivalents in the *LNJED* (or, if this failed, the *OIJED* or the *EED*) were taken as the basis for the definitions and examples provided in test 1. A relatively small number of words (30) was chosen so that the students had time to answer the whole questionnaire.

When analyzing the data, it was checked whether each translation provided by the students in part two (pre-test) was correct or incorrect. The words with a proper translation were not taken into account in part three (tests 1 and 2). This methodology was followed because the conductors of the study wanted to find out exclusively the number of unknown words the students had understood by reading the definitions in part three. Therefore, each individual test had its own score and base.

It must be noted that some of the headwords (a total of 20%) in the pre-test were polysemous: they are the words 'branch', 'draw', 'picture', 'right', 'take' and 'used to'. First of all, it must be said that, among all the possible meanings for these words that appeared in the *EED*, the first (i.e. the most core) meaning was chosen for inclusion in tests 1 and 2. This drastically reduced the chance of eliciting translations in the pre-test that might potentially conflict with the meanings that appear in the tests. Nevertheless, there were a few occasions (5.3%) in which the students provided in the pre-test a correct translation for the headword which did not match the meaning provided in the tests. In these cases, it was decided, when doing the statistics, that the translations given in the pre-test and the test would not be taken into account. In other words: the polysemous words variable was controlled, the result being that there was no mathematical distortion that affected the validity of the results, as there was a correlation between the base and the score obtained by each student.

Another aspect regarding content validity affects the nature of the definitions and examples given in the tests. Because the definitions and examples of the *EED* were different from those found in the *LNJED* and the *OIID*, they are not exact parallels as far as grammar and lexical choice are concerned. The test designers were aware that these facts might affect the results, but preferred not to create artificial, ad hoc definitions for the items in test 1 for four reasons that led us to believe that the results obtained would not be invalid. First of all, even though the definitions in test 1 tend to be slightly longer, test 2 gives more examples after the definitions, so the differences in length are balanced out. Secondly, both test 1 and test 2 used a very basic vocabulary and uncomplicated grammatical structures. Thirdly, a previous interview with each of the teachers showed that, even though the definitions and examples in test 2 contained more words in the high frequency bands, there was roughly the same number of unknown words in each test. As for the grammar, the teachers told us that it would not be a hindrance insofar as all grammatical structures had already been studied. There is still a fourth reason why the testers did not create artificial definitions for items in test 1: it has been shown that in EFL learning, C1 background is more important for comprehension than linguistic complexity. Johnson (1981: 169), for example, states that "the cultural origin of the story had more effect on the comprehension of the ESL students than the level of syntactic and semantic complexity". And Carrell (1987) demonstrates that unfamiliar content causes more difficulty to the students than unfamiliar form, that is, in EFL reading comprehension, content schemata are more important than formal (i.e. linguistic and rhetorical) schemata, which implies that "reading familiar content even in an unfamiliar rhetorical form is relatively easier than reading unfamiliar content in a familiar rhetorical form" (Kang 2002: 41).

3. Results

After the students had completed the questionnaire, the data obtained were analyzed following the procedures outlined in section 2.3. The result of each test was a percentage of (previously unknown) comprehended definitions. Thus, the authors of the study made sure that they identified the successfulness of the different dictionary definitions in test 1 and in test 2.

A substantial statistical difference was found in the comprehension of new words in test 1 and test 2. As shown in table 2, in test 1 (which contained no Catalan or Spanish culturally-based references) students understood 18.9% of the words defined, whereas in test 2 (which contained Catalan and Spanish culturally-based references) students understood 46.2% of the words defined.

In 1st ESO 14.2% understood the words in test 1 and 37.1% understood the words in test 2. As far as 6th Primary Education is concerned, 21.3% understood the words in test 1 (14.9% in Tona and 27.7% in Torelló) and 50.8% understood the words in test 2 (50% in Tona and 51.5% in Torelló). These figures are summarized in table 2, where the data in italics refer to 1st ESO stu-

dents and the data in bold refer to 6th Primary students.

School	Level/(years old)	Test 1 (%)	Test 2 (%)
TONA	<i>1st ESO (12-13)</i>	14.2	37.1
TONA	6th Primary (11-12)	14.9	50
TORELLÓ	6th Primary (11-12)	27.7	51.5
TOTAL		<i>14.2 + 21.3 / 18.9</i>	<i>37.1 + 50.8 / 46.2</i>

Table 2: Percentage of headword comprehension in the whole survey

No significant statistical difference was found between boys and girls in both tests (test 1 and test 2). In test 1, the boys scored 18.5% in the comprehension of new words, while the girls scored 18.8%. In test 2, the boys scored 49.6% in the comprehension of new words, whereas the girls scored 41.7%.

4. Discussion

The results presented in the previous section show that the answer to our research question ("does the presence of C1 in EFL learners' dictionaries improve the comprehension of the words defined?") is positive. To what extent, though, does C1 improve comprehension of the words defined? The results show that students who use an EFL dictionary that includes C1 references have, on the whole, more than double (2.45 times) the possibilities of understanding the meaning of a new looked-up word than those who do not. The only exception to this are the Torelló 6th Primary subgroups, in which the correlation in headword comprehension is slightly less than double (1.8 times more). On the other hand, the subgroups that presented a higher variation were those of Tona 6th Primary level: in this case, the students who took test 2 more than triplicated (3.3 times more) the amount of word comprehension of test 1 students. The percentages in comprehension presented in table 2 are given in correlational form in table 3, where the data in italics refer to 1st ESO students and the data in bold refer to 6th Primary students.

School	Level/(years old)	Test 1	Test 2
TONA	<i>1st ESO (12-13)</i>	1	2.6
TONA	6th Primary (11-12)	1	3.3
TORELLÓ	6th Primary (11-12)	1	1.8
TOTAL		<i>1 + 1 / 1</i>	<i>2.6 + 2.3 / 2.45</i>

Table 3: Correlation of headword comprehension in the whole survey

The results presented in table 3 confirm our research hypothesis (Corrius and Pujol 2010: 142) that, among other strategies, localizing and glocalizing culture (the nativization processes found in test 2) are useful mechanisms to improve comprehension when reading foreign language dictionary definitions. The differences in headword comprehension between the students who took test 1 and those who took test 2 are both significant and meaningful. On the one hand, they are significant in that the probability that these differences occurred by chance alone is really low, if not null, owing to the methodology (see section 2 above) employed when carrying out the study, which controlled as many variables as was deemed necessary. On the other hand, the differences in headword comprehension in tests 1 and 2 are meaningful, because they are large: students who took test 2 outperformed those who took test 1 by a large margin (2.45 times more); similar comprehension results would have produced slight differences in the comprehension of the headwords in tests 1 and 2, but not more than the double.

The findings of our study, then, support the schema theory insofar as dictionary users rely on (i.e. activate) their schemata (i.e. their previous C1 knowledge) when confronted with new information (in this case, when trying to understand the meaning of new words). In other words: our results confirm the main tenet of schema theory that meaning is construed by activating previous (in our case, C1) knowledge. Apart from this, the side-by-side comparison of the results obtained in tests 1 and 2 is consistent with the findings in the literature that, given the appropriate cultural schema, "readers may be able to overcome their insufficiently developed linguistic knowledge" (Al-Shumaimeri 2006: 3). To put it another way, the findings in our study show that the absence of C1 elements in test 1 significantly hinders (and even prevents) the students' reading comprehension: the students who took test 1 had much lower scores than those who took test 2 (18.9% vs. 46.2% of words comprehended). From this it may be inferred that using C1 background knowledge in the classroom and when designing EFL dictionaries may be more effective than ignoring C1. In other words: C1 may be an aid to enhance learning, especially at lower levels in which the students' cultural and world knowledge highly exceeds their linguistic competence in the foreign language.

One final aspect worth pointing out is that, despite the high percentage of comprehension of previously unknown words (46.2% for students who took test 2), and despite the fact that C1 elements are present in every single item in test 2, a significant number of readers in test 2 (53.8%, as opposed to 81.1% in test 1) failed to understand the meaning of the word defined. These results in test 2 might be due to linguistic difficulties, to other cognitive processes not studied here (e.g. wrong inferencing or inability to process contextual information), or perhaps even to the fact that the readers' schemata do not match those of the dictionary writer. Of course, it may well be the case that readers cannot activate or invoke the relevant cultural schema necessary to understand the text they are reading. It must be acknowledged, though, that students reading texts

containing C1 encountered significantly fewer difficulties than those students who read texts in which the presence of C1 was null: C1 is an important factor (though not the only one) that contributes to reading comprehension.

5. Conclusions

By using the tenets of the schema theory, this article confirms Corrius and Pujol's (2010: 142) hypothesis that dictionary users rely on their already established cultural knowledge (C1) when trying to understand the meaning of new words. The examples found in test 1 (taken mainly from the *LNJED*) use British culture as the by-default standard when providing examples (see e.g. 'draw', 'far', 'south', 'take' and 'tongue twister' in test 1 in the Appendix), which reduces the chances that EFL students understand the meaning of the defined L2 headword. On the other hand, the examples in test 2 (taken from the *EED*) emphasize culturally-familiar (i.e. local) elements (in cases such as 'against', 'monk' and 'painter') and cross-cultural (i.e. glocal) elements (in cases such as 'airline', 'right', 'take' and 'tongue twister'), which increases the chances that students understand the meaning of the defined L2 headword.

It may be concluded, then, that C1 played a crucial role in understanding L2 definitions at lower-intermediate level (11–13 year-old students of EFL). As shown in table 2, in test 1 (which contains no Catalan or Spanish culturally-based references) students understand 18.9% of the words defined, whereas in test 2 (which contains Catalan and Spanish culturally-based references) students understand 46.2% of the words defined. These data show that the number of students in test 2 who understand the L2 headwords more than doubles the number of students who understand the same headwords in test 1.

To sum up, the hypothesis in this study, namely that the use of the students' own culture in the words defined helps them improve comprehension, has been tested and confirmed. As Jiang (2000: 332) summarizes, learners "swim confidently and rapidly when they are familiar with the water (i.e. within their native culture), but cautiously and slowly when it is unfamiliar to them (within a foreign culture)". It has been a long time since the schema theory hypothesized that students' previous knowledge could be helpful in understanding new information. However, most of the studies carried out so far are centred not on understanding dictionary definitions but on understanding passages or whole texts. This article shows for the first time in the literature that an EFL dictionary which uses cultural previous knowledge (C1) is a significantly more effective lexicographic tool than an EFL dictionary which is culturally neutral (see e.g. 'right' in test 1) or culturally biased towards C2 (see e.g. 'tongue twister' and 'apart' in test 1; the example in 'apart' contains the word 'mile' instead of 'kilometre', the first item being a unit of measurement not well-known by Catalan youngsters).

The results of our study of the effect of C1 on EFL readers' understanding of definitions has implications for curricular design. It seems desirable that, in

order to improve students' reading comprehension (and therefore EFL learning), lexicographers devise EFL nativized dictionaries (especially at lower levels) which take into account C1, and that teachers can choose between two types of dictionary: those that are culturally neutral or culturally biased towards C2 and those that incorporate C1 in their definitions or examples. Of course, in some cases, it may be desirable to incorporate C2 alongside C1 (see 'tongue twister') so that students perceive the similarities and differences between the two cultures and can learn about the foreign culture. Also, our study may have implications for classroom teaching. On the one hand, teachers may opt for minimizing their students' reading difficulties by incorporating C1 into the classroom; on the other, teachers may wish to administer pre-reading activities exploiting the knowledge of C1 as a basis for comprehending not only the foreign language, but also C2 (see 'tongue twister'). The same procedures might be applied to textbook design.

6. Limitations and further research

The present study is not without limitations. To begin with, its empirical scope is limited and therefore it would be unwise to claim universal validity for the results obtained. It must be acknowledged that, because of the sample characteristics, the results are only generalizable to Tona and Torelló students, or at best to schools with similar students. As Brown (1998: 203) puts it, "studies are *never absolute*. They must instead be viewed in relative terms — relative to probability, relative to other students, relative to theoretical frameworks." Therefore, a larger population would need to be tested in order to generalize the findings to other schools in Catalonia (and of course, to other schools outside the Catalan cultural context). Besides, it would be desirable to expand the number of items tested and/or try different items so as to provide more generalizable results. Moreover, further studies might control possible variables such as the students' stays abroad or the fact that they may have taken private lessons outside the school. Besides, even though the last part of section 2.3 gives reasons why we believe the language differences between test 1 and test 2 are not significant enough to invalidate the results, further tests could be carried out that would allow the researchers to control more tightly possible intervening variables such as the syntactical, lexical, and rhetorical differences between the two tests. For example, two identical tests in terms of grammar and vocabulary could be created, the only variable being the difference of the cultural elements present in each test (for studies focusing on text modification through nativization, see Rokhsari 2012: 46).

The limitations outlined above are a blueprint for further research. Other research studies that could be done in the future (using perhaps different dictionaries or materials from those employed here) are:

- (1) Replication of the study with students of different ages (the study was carried out only among 11–13 year-old students, so the results cannot be

generalized to students of different ages).

- (2) Replication of the study taking into account different levels of language proficiency: is it true, as Al-Shumaimeri (2006: 12) claims to have shown, that the lack of cultural knowledge affects "the low-ability students but not the high-ability students"?
- (3) The role of C1 could be investigated in relation to the other three skills: writing, listening and speaking.
- (4) Qualitative interviews or think-aloud protocols could be used to find out how students understand new knowledge:
 - (a) Can support be provided for the hypothesis that the higher reading comprehension when schemata are activated is due to the fact that, when students have to deal with familiar concepts or knowledge, a certain amount of cognitive space can be cleared up (Nassaji 2002)?
 - (b) Which cognitive processes exactly are involved when a schema is activated: inferencing, predictions, selection among possible meanings, reasoning, problem solving, remembering? For example, when reading the example in test 2 of the headword 'brand' ("Colgate' is a brand of toothpaste"), which schema is activated *in the context of the definition*: that of 'toothpaste' or that of 'brand'? And how do students arrive at the meaning of 'brand': by inferencing, reasoning, etc.?
 - (c) Are there any other mental processes (other than activating schemata) going on in the students' minds while trying to construe meaning when reading?

References

- Ajideh, P. 2003. Schema Theory-based Pre-reading Tasks: A Neglected Essential in the ESL Reading Class. *The Reading Matrix* 3(1): 1-14.
- Al-Shumaimeri, Y.A.N. 2006. The Effects of Content Familiarity and Language Ability on Reading Comprehension Performance of Low- and High-ability Saudi Tertiary Students Studying English as a Foreign Language. *Journal of King Saud University, Educational Sciences & Islamic Studies* 18(2): 1-19.
- Alemi, M. and S. Ebadi. 2010. The Effects of Pre-reading Activities on ESP Reading Comprehension. *Journal of Language Teaching and Research* 1(5): 569-577.
- Alptekin, C. 1993. Target-language Culture in EFL Materials. *ELT Journal* 47(2): 136-143.
- Alptekin, C. 2002. Towards Intercultural Communicative Competence in ELT. *ELT Journal* 56(1): 57-64.
- Alptekin, C. 2006. Cultural Familiarity in Inferential and Literal Comprehension in L2 Reading. *System* 34: 494-508.
- Bartlett, F.C. 1932. *Remembering: A Study in Experimental and Social Psychology*. Cambridge: Cambridge University Press.

- Brown, J.D.** 1998. *Understanding Research in Second Language Learning: A Teacher's Guide to Statistics and Research Design*. Cambridge: Cambridge University Press.
- Carrell, P.L.** 1987. Content and Formal Schemata in ESL Reading. *TESOL Quarterly* 21(3): 461-481.
- Carver, R.P.** 1992. Effect of Prediction Activities, Prior Knowledge, and Text Type upon Amount Comprehended: Using Rauding Theory to Critique Schema Theory Research. *Reading Research Quarterly* 27: 165-174.
- Chang, Y.-L.** 2007. The Cross-cultural Background Knowledge in the EFL Reading Comprehension. *Chia-nan Annual Bulletin* 33: 383-393.
- Chihara, T., T. Sakurai and J.W. Oller.** 1989. Background and Culture as Factors in EFL Reading Comprehension. *Language Testing* 6(2): 143-151.
- Coffey, S.** 2001. The L1 Culture in the L2 Classroom. *Modern English Teacher* 10(2): 54-58.
- Cook, G.** 1997. Key Concepts in ELT: Schemas. *ELT Journal* 51(1): 86.
- Corrius, M. and D. Pujol.** 2010. Linguistic and Cultural Strategies in ELT Dictionaries. *ELT Journal* 64(2): 135-142.
- Easy English Dictionary with a Catalan-English Vocabulary.* 2004. Vic: Eumo Editorial.
- Floyd, P. and P.L. Carrell.** 1987. Effects on ESL Reading of Teaching Cultural Content Schemata. *Language Learning* 37: 89-108.b
- Fuhong, T.** 2004. *Cultural Schema and Reading Comprehension*. [Tianjin]: Nankai University. <http://www.celea.org.cn/pastversion/lw/pdf/TanFuhong.pdf>.
- Garth-McCullough, R.** 2008. Untapped Cultural Support: The Influence of Culturally Bound Prior Knowledge on Comprehension Performance. *Reading Horizons* 49(1): 1-30.
- Hatim, B. and I. Mason.** 1990. *Discourse and the Translator*. Harlow: Longman.
- Jalilifar, A.R. and R. Assi.** 2008. The Role of Cultural Nativization in Comprehension of Short Stories in EFL Reading Contexts. *The International Journal of Language Society and Culture* 26: 62-79.
- Jiang, W.** 2000. The Relationship between Culture and Language. *ELT Journal* 54(4): 328-334.
- Johnson, P.** 1981. Effects on Reading Comprehension of Language Complexity and Cultural Background of a Text. *TESOL Quarterly* 15(2): 169-181.
- Kang, Y.** 2002. A Review of: Carrell, Patricia L. Content and Formal Schemata in ESL Reading. *LORE Journal* 11 March 2002: 39-42. http://rhetoric.sdsu.edu/lore/2_1/08_kang.pdf. Also available at <http://www.lorejournal.org/2002/03/review-content-and-formal-schemata-in-esl-reading-by-yunkyoung-kang>.
- Karakas, M.** 2005. The Effects of Pre-reading Activities on ELT Trainee Teachers' Comprehension of Short Stories. *Journal of Theory and Practice in Education* 1: 25-35.
- Longman New Junior English Dictionary.* 1993. Burnt Mill, Harlow: Longman.
- Nassaji, H.** 2002. Schema Theory and Knowledge-based Processes in Second Language Reading Comprehension: A Need for Alternative Perspectives. *Language Learning* 52(2): 439-481.
- The Oxford English Dictionary.* 2013. Oxford: Oxford University Press. <http://www.oed.com>.
- The Oxford Illustrated Junior Dictionary.* 1991. Oxford: Oxford University Press.
- Post, R. and I. Rathet.** 1996. On their Own Terms. Using Student Native Culture as Content in the EFL Classroom. *English Teaching Forum* 34(3): 12-17.
- Pujol, D., J. Masnou and M. Corrius.** 2008. Pedagogical Criteria for Effective Foreign Language Learning: A New Dictionary Model. Bernal, E. and J. DeCesaris (Eds.). 2008. *Proceedings of the XIII EURALEX International Congress, Barcelona, 15-19 July 2008*: 763-769. Barcelona: Universi-

tat Pompeu Fabra, Institut Universitari de Lingüística Aplicada.

Reynolds, R.E., M.A. Taylor, M.S. Steffensen, L.L. Shirey and R.C. Anderson. 1981. *Cultural Schemata and Reading Comprehension*. Champaign, Illinois: University of Illinois at Urbana-Champaign. https://www.ideals.illinois.edu/bitstream/handle/2142/17592/ctrstreadtechrepv01981i00201_opt.pdf?sequence=1.

Rokhsari, S. 2012. The Effect of Text Nativization and Schema-based Pre-reading Activities on Reading Comprehension of EFL Students. *Journal of Academic and Applied Studies* 2(5): 45-75.

Appendix: questionnaire

The appendix includes the questionnaire given to the students. Parts I and II were common for all students. Part III consists of two tests: students in group 1 took test 1 and students in group 2 took test 2. For ease of comparison, the definitions of each test are placed side by side. For the sake of clarity, a translation is provided in square brackets, after the original instructions in Catalan.

I. INFORMACIÓ PERSONAL [PERSONAL DETAILS]

Quants anys tens? [How old are you?]: _____

Ets noi o noia? [Are you a boy or a girl?]: _____

Quin curs fas? [Which course level are you taking?]: _____

II. CONEIXEMENT PREVI I TRADUCCIÓ DE PARAULES [PREVIOUS TEST: PREVIOUS KNOWLEDGE AND TRANSLATION OF WORDS]

Subratlla totes les paraules de les quals coneixes el significat i posa al costat el que volen dir en català. [Underline the words you know and write their meaning in Catalan next to the word.]

against
airline
apart
banknote
beech
blind
branch
brand
cello
chestnut
coast
defeat
department store
draw
eastern
far
firework
island
monk
north
painter
painting
picture
price
right
sights

south
take
tongue twister
used to

III. TEST [TEST]

Llegeix les definicions de més avall i escriu en català, al costat de cada paraula, el que creus que vol dir la paraula en **negreta**. [Read the definitions below and write in Catalan, next to each word, what you think the word in **bold** means.]

TEST 1	TEST 2
against not agreeing with someone or something: <i>I'm against killing animals for their fur.</i>	against <i>F.C. Barcelona is playing against Real Madrid.</i>
airline a company which carries people or goods by plane.	airline an aeroplane company: <i>Iberia and British Airways are international airlines.</i>
apart separately; away from another, or others: <i>The two villages are 6 miles apart.</i>	apart away from each other: <i>Barcelona and Vic are 60 km apart.</i>
banknote a piece of paper money.	banknote paper money: <i>a €10 banknote/a 10-euro banknote.</i>
beech a kind of tree [+ illustration].	beech a type of tree: <i>La Fageda d'en Jordà is a forest near Olot with lots of beeches.</i>
blind not able to see because you have something wrong with your eyes: <i>She was born blind.</i>	blind a blind person cannot see: <i>ONCE is an organisation for blind people.</i>
branch one part or one office of a business: <i>The bank has branches in all the big towns.</i>	branch an office in a big organization: <i>'La Caixa' has many branches in Catalonia.</i>
brand the name of a particular kind of goods made by one company: <i>What brand of soap do you like?</i>	brand the name of a product: <i>'Colgate' is a brand of toothpaste.</i>
cello a musical instrument like a large violin which you hold between your knees.	cello a musical instrument similar to a violin, but bigger: <i>Pau Casals played the cello.</i>
chestnut the shiny brown nut that grows on a chestnut tree.	chestnut a brown dry fruit that we usually cook and eat in October and November: <i>My mum always buys chestnuts on 1st November.</i>
coast the land next to the sea: <i>a town on the coast.</i>	coast the seaside: <i>The Costa Brava is a very famous coast in Catalonia.</i>
defeat to beat an opponent in a war, game, etc.	defeat to win: <i>F.C. Barcelona defeated Real Madrid: Barça 3 – Madrid 0.</i>
department store a type of shop that is divided into several parts, each of which sells a different kind of goods.	department store A very big shop that is divided into several floors: <i>El Corte Inglés is a famous department store.</i>

draw to end a game or match with an equal result so that nobody wins: <i>We drew with the London team. To draw a match.</i>	draw not to win or lose a match: <i>F.C. Barcelona drew with Real Madrid, 1–1.</i>
eastern in or of the east.	eastern from or in the east: <i>Catalonia is in the eastern part of Spain.</i>
far distant from a place: <i>How far is it to London? It isn't very far. It's too far to walk.</i>	far at a long distance from a place: <i>Miami is very far from Barcelona; Sabadell is very near Terrassa.</i>
firework a cardboard tube filled with special chemicals, which burns with a loud noise and makes bright lights in the air.	firework a small tube that explodes in the air and produces lights of beautiful colours: <i>We went to watch the fireworks on Saint John's Day.</i>
island a piece of land surrounded by water.	island a piece of land surrounded by water: <i>Minorca, Tenerife and Hawaii are islands.</i>
monk one of a group of men who live together and have given their lives to a religion.	monk a man that lives in a religious community: <i>Montserrat is a monastery where lots of monks live.</i>
north the direction that is on the left when you look towards the rising sun: <i>Manchester is in the north of England. The north part of the house doesn't get a lot of sun. Birds fly north in summer.</i>	north a cardinal point: <i>The Pyrenees are in the north of Catalonia.</i>
painter a person who paints pictures.	painter a person that paints: <i>Salvador Dalí was a famous Catalan painter.</i>
painting a painted picture: <i>a painting of a boat.</i>	painting a picture made with paint: <i>The art gallery was full of Picasso's paintings.</i>
picture something represented on paper as a drawing, painting, or a photograph: <i>She drew a picture of me.</i>	picture a drawing or a painting of a person, a place, etc.: <i>I like pictures by Dalí.</i>
price the money that you must pay to buy something: <i>The price of that house is too high.</i>	price the cost of something: <i>The price of this bicycle is €180.</i>
right towards the right side: <i>Turn right at the corner.</i>	right the opposite of 'left': <i>In Great Britain people drive on the left, but in Spain we drive on the right.</i>
sights the places that are interesting to visit in a city, country, etc.	sights interesting places that tourists visit: <i>The Sagrada Família is one of the most important sights in Barcelona.</i>
south the direction that is on the right when you look at the sun at the start of the morning: <i>London is in the south of England. The south side of the city, to travel south, living south of Oxford.</i>	south a cardinal point: <i>Andalusia is in the south of Spain.</i>
take to need a particular amount of time: <i>The journey to London takes three hours.</i>	take to need time: — <i>How long does it take from Barcelona to London? — It takes two hours by plane.</i>

<p>tongue twister a group of words difficult to pronounce when you say them quickly: <i>'She sells sea shells on the sea shore' is an English tongue twister.</i></p>	<p>tongue twister a group of words difficult to pronounce when you say them quickly: <i>'Setze jutges d'un jutjat mengen fetge d'un penjat' is a Catalan tongue twister; 'She sells sea shells on the sea shore' is an English tongue twister.</i></p>
<p>used to used with another verb to show that something was done often in the past, but is not done now: <i>He used to play football every Saturday when he was young. My father didn't use to smoke, but now he does.</i></p>	<p>used to 'used to' indicates that we did something regularly in the past: <i>When I lived in Platja d'Aro I used to go to the beach every day.</i></p>

From e-Lexicography to Electronic Lexicography. A Joint Review

Valeria Caruso, *Department of Literary, Linguistic and Comparative Studies, University of Naples 'L'Orientale', Naples, Italy*
(vcarus@unior.it/vcarus078@gmail.com)

Abstract: Two recently published books outline the main issues of the current debate on lexicography. The first, *e-Lexicography* edited by Fuertes-Olivera and Bergenholtz in 2011, presents the standpoints of the lexicographical function theory on the future developments of dictionaries, while, in some chapters, current innovative tools are described, tools which allow customizations according to the user's type of need. The second volume, *Electronic Lexicography* edited by Granger and Paquot in 2012, presents different opposing views on what the dictionaries of the future will look like, such as the linguistic-oriented stance of Hanks and that of Tarp concerning theoretical lexicography. The dictionary projects that are described within these pages offer an interesting basis of comparison with those developed by the representatives of the function theory.

Keywords: COMPUTER ASSISTED LANGUAGE LEARNING (CALL), CORPUS LINGUISTICS, CUSTOMIZATION, DATABASES, DICTIONARY SURVEY, DICTIONARY USE, EFFICACY, EFFICIENCY, ELECTRONIC LEXICOGRAPHY, INFORMATION SCIENCE, LANGUAGES FOR SPECIAL PURPOSES, LEXICOGRAPHICAL FUNCTION THEORY, LINGUISTIC THEORY, MONOFUNCTIONAL DICTIONARY, P-DICTIONARIES, PRACTICAL LEXICOGRAPHY, USABILITY TESTING

Opsomming: Vanaf e-leksikografie tot elektroniese leksikografie: 'n Gesamentlike beskouing. Twee onlangs gepubliseerde boeke skets die belangrikste kwessies in die huidige debat oor die leksikografie. Die eerste, *e-Lexicography* in 2011 geredigeer deur Fuertes-Olivera en Bergenholtz, gee die standpunte van die leksikografiese funksieteorie oor die toekomstige ontwikkelinge van woordeboeke, terwyl, in sommige hoofstukke, huidige innoverende gereedskap beskryf word, gereedskap wat pasmaking volgens die gebruiker se soort behoefte toelaat. Die tweede boek, *Electronic Lexicography*, in 2012 geredigeer deur Granger en Paquot, bied verskillende opponerende standpunte oor hoe die woordeboeke van die toekomst sal lyk, soos die linguisties georiënteerde benadering van Hanks en dié van Tarp betreffende die teoretiese leksikografie. Die woordeboekprojekte wat in hierdie bladsye beskryf word, bied 'n interessante vergelykingsbasis met dié ontwikkel deur die verteenwoordigers van die funksieteorie.

Sleutelwoorde: REKENAARGESTEUNDE TAALANLEER (RGTA), KORPUSLINGUISTIEK, PASMAKING, DATABASISSE, WOORDEBOEKOORSIG, WOORDEBOEKGEBRUIK, DOELTREFFENDHEID, DOELMATIGHEID, ELEKTRONIESE LEKSIKOGRAFIE, INLIGTINGSWETENSAP, TALE VIR SPESIALE DOELEINDES, LEKSIKOGRAFIESEFUNKSIETEOORIE, TAALTEORIE, MONOFUNKSIONELE WOORDEBOEK, P-WOORDEBOEKE, PRAKTIESE LEKSIKOGRAFIE, BRUIKBAAARHEIDSTOETSING

1. Different points of view in the current debate

The short span of time within which two quasi-homonymous volumes have been published, mirrors the intensity of the debate upon the issues currently at stake in electronic lexicography. The first book, edited by Pedro Fuertes-Olivera and Henning Bergenholtz, appeared in 2011 and has a shorter title but an explicative subtitle: *e-Lexicography. The Internet, Digital Initiatives and Lexicography*; the second, entitled *Electronic Lexicography*, was published in 2012 and edited by Sylviane Granger and Magali Paquot. Both are miscellaneous volumes, even sharing some authors and co-authors – namely Fuertes-Olivera, Leroyer, Lew, Tarp and Verlinde – but the editorial projects and intended aims are different, and the reading of both is highly recommended to obtain an exhaustive picture of the topic.

e-Lexicography, by Fuertes-Olivera and Bergenholtz, is a celebratory volume on the occasion of the conferment of a Doctorate degree in Lexicography to Henning Bergenholtz, a highly symbolic event which goes some way towards the legitimatization of Lexicography as a separate and independent field among the academic disciplines, one of the crucial claims of the 'Aarhus School' of which Bergenholtz is one of the main representatives. The volume is a collection of the papers presented at the conference organized for Bergenholtz's Doctorate, an event, says the co-editor, Fuertes-Olivera, in his preface, during which many "new and provocative ideas" were put forth. The remark is not trivial, since some of the "ground-breaking" (Leroyer 2011: 125) tenets of the *lexicographical function theory* proposed by the Danish lexicographers have been met with disbelief by some esteemed lexicographers (Atkins and Rundell 2008) and metalexigraphers (Béjoint 2010).

The story is well-known but nevertheless worth a mention here. The researchers of the Aarhus University have proposed a new vision of what Lexicography should be about, considering it as a theoretical discipline within the Information Sciences, devoted to the study of all kinds of reference needs and tools, not limited by the genera, e.g. dictionaries, nor by the medium, e.g. paper or the Internet. There are actually "artefacts which, all things considered, should also be counted as members of the vast family of reference works, such as almanacs, atlases, catalogues, directories, guides, handbooks, reference manuals and so on" (Leroyer 2011: 124). This assumption is a consequence of the emphasis placed on the nature of dictionaries, conceived as tools that people refer to in order to satisfy their needs. Therefore every dictionary must be compiled not only with its user in mind, but also considering the situation in which it will be accessed. At the intersection between the user and the situation, the lexicographer is able to identify the functions that his/her dictionary must fulfil, e.g. spell-checking for school children, translational equivalences for a scholar preparing his paper in an L2, and so forth. Therefore the task to be fulfilled by the lexicographer is an effective data selection and presentation, minimizing user effort whilst satisfying information needs.

Though the premises of the *lexicographical function theory* on the practical

nature of the dictionary and its usefulness for users are uncontroversial, the broadening of the field of lexicography and the conferring of a theoretical status to it was questioned by those who remark on the 'practical' nature of the lexicographer's job (Atkins and Rundell 2008) which, in their view, is devoted to the "systematic", "internally-consistent" description of language, "driven by what the language data is telling us", with "subjective judgments [...] kept to a minimum" (Rundell 2012b: 48). In this respect the contribution of linguists and linguistic theory is paramount, and the computational analysis of language data is the most valuable revolution brought about by the digital text era, since the lexicographer's main concern is not the "synthesis" of data (Atkins 1993), i.e. selection and presentation, but rather mining from the raw corpora material.

Setting aside the provocative slogans and disbeliefs, it seems fair to say that these two parties are actually looking at different parts of the same object: the Aarhus School is committed to the data holder and data presentation, relying on the resources of Information Technology; the 'practical lexicographers' are concerned with data *per se*, and credit computational linguistics with the needed insight for effective language descriptions. However, for the users' benefit, it must be hoped that the research from both sides can combine rather than exclude each other, as seems to be suggested also by the wide-ranging volume collected by Granger and Paquot, hosting contributions from practical lexicographers and lexicography theoreticians as well.

Before dealing with this challenging topic, reviewing in more detail the chapters devoted to the description of different dictionary projects, an outline of the two volumes will be provided, reporting on all the points of view that the editors of the books have collected, since they are thought to be the key topics in the current electronic lexicography debate.

Nevertheless, a selection was necessary, and the less general themes will be dealt with only briefly, since the debate is chiefly concerned with three crucial issues: what features should dictionaries have in the future, what is lexicography, and which methodologies should it use? The empirical research on dictionary use is indeed the other challenging topic, particularly for a theory devoted to the functions of dictionaries, "in which user needs are key" (Granger 2012: 7). Actually, even if the perfect customizations that Tarp has in mind (see below) dismiss the question of profiling the needs of the target user, empirical findings on what can be considered 'ergonomic', should be collected.

With these three main issues in mind, it is possible after some preliminary remarks to outline the contents found in the books. While in the volume by Fuertes-Olivera and Bergenholtz, the reader can find not only the newest available dictionary formats, but also the details of future ones (in the chapters by Bothma, Spohr and Tarp), Granger and Paquot are more concerned with the overall current issues, giving a more varied inventory of specific sub-topics (translator's needs, 'alternative e-dictionaries' and software management tools, among many others) and visions, such as the lexicological perspective on the development of dictionaries described by Hanks.

In the following sections, the overall contents of each volume will be given, while the chapters presenting innovative electronic tools will be discussed in two separate paragraphs, using the treatment of linguistic data within these tools as a basis of comparison for the different lexicographical practices.

2. Fuertes-Olivera and Bergenholtz's *e-Lexicography*

The book by Fuertes-Olivera and Bergenholtz presents fifteen contributions ranging from theoretical issues in lexicography to the illustration of innovative dictionary projects and the features that dictionaries will have in the future. Particularly significant is the inclusion of two chapters (4 and 5), that have Information Science as their topic, since one of the crucial claims of the Aarhus School is that lexicography should be an autonomous discipline within the field of Information Sciences. These chapters follow immediately after the introductory section, dedicated to the theoretical stances of the lexicographical function theory, and they precede the presentation of different e-dictionaries. The *Base lexicale du français*, a particularly promising e-tool for computer assisted language learning, is described in the closing section (chapter 13) along with a usability study (chapter 14) and a brief conclusive chapter (15), containing the main stances discussed at the congress held in Valladolid. Chapter 6 serves as a bridge between theory and practice, and presents "four lexicographic information tools"; while chapters 10 and 11 are concerned with language dictionaries: the first reviewing the online resources for the English language, the other presenting the issues posed by dictionary definitions to the function theory.

After the co-authored introduction by the editors, the book opens with a critical discussion of function theory within the stances of theoretical lexicography. The chapter by Gouws hails the radical shift in the lexicographical paradigm now having the users and the situations of use at its centre. The analysis of the different kinds of functions a dictionary may have with respect to the different contexts of consultation by each single user, allows the lexicographer to identify at least one function that is detached from any language problem. If someone accesses a reference tool in order to acquire new knowledge about something, he is faced with a cognitive problem that has nothing to do with communication, and thus language problems (Gouws 2011: 23). In this respect, lexicography proves to have its own theoretical space, and must not be credited as a linguistic discipline, but as a broader reference science, dealing with dictionaries among many other reference tools, e.g. tourist guides, informative web portals, instruction manuals and so on. After having sketched the theoretical background of the function theory, Gouws underlines that the new reference science demands a radical rethinking, and even "unlearning" of the old practices, dominated by a general "linguistic colonialism". Thus, within this framework, a new dictionary model is proposed, namely that of databanks from which different dictionaries can be extracted when accessing the resource

with specific query systems.

In the second chapter, Bergenholtz complains that the current lexicographical practice neglects to conform properly to the users' consultation needs. What is generally displayed by dictionaries is more suited for language experts than for real users. Access routes are only seldom implemented and there is unsatisfactory literature on dictionary use, since, first of all, it surveys unrepresentative population samples. For Bergenholtz, it is more promising to investigate the access paths of single subjects in task-based testing sessions. He reports on two studies, one of these proves that the fastest and most successful consultations were allowed by those dictionaries that require fewer search steps, something that is achievable only with a proper dictionary customization. The evaluative parameters used by Bergenholtz, such as the time and the goal achievement, are called 'efficiency' and 'efficacy' in Information Science; they are also used by Heid in a study presented in the same book (chapter 14, see below).

Bergenholtz goes on to illustrate the innovative features of the fastest dictionary in the survey, the dictionary called *Meaning of fixed expressions*, which is extracted from a database containing the necessary data to display three other 'monofunctional' dictionaries: one to assist with the use of fixed expressions, one to know more about them, and the other to find fixed expressions with a certain meaning. The 'monofunctional' dictionaries are the innovative tools compiled at Centlex of the Aarhus School: they are tailored to display only what is needed by the user in the access situation and they are extracted by a single database using one of the available search options.

This new dictionary model is indicated by Tarp, in the third chapter, as the only kind of reference work that can be reasonably considered as a product of e-lexicography, since it has been planned from the beginning as an adaptable instrument that users can access and customize to their specific consultation needs. Tarp (2011: 59) calls these dictionaries "T Ford", referring to the well-known anecdote of Henry Ford presenting his new car. Ford complained that if "he had consulted people before inventing" it, they would rather have asked for "faster horses". The claim by Tarp is for a radical shift in the lexicographical paradigm in order to make the best of the electronic medium and present-day technology. However, using these monofunctional dictionaries, a complete customization is not yet achievable. This could only become reality if the data display is completely dynamic, and adaptable to every single consultation need of all the possible users in all the possible consultation situations. If it existed, Tarp would call this dictionary a Rolls Royce.

However ambitious it may be, "modern information technologies can start addressing these issues", as Bothma (2011: 79) declares in the fourth chapter, which is dedicated to the analysis of the challenges of function theory from the point of view of the information scientist. Reviewing the tenets of the theory, Tarp (2011: 69) describes his vision of "the best dictionary in terms of needs satisfaction", namely "any dictionary [...] that allows *either* monofunctional

access *or* individualized access in the framework of its specific and foreseen functions". Therefore he illustrates the current technology that could make this wished-for tool become true: from search and navigation options to filters for user profiling and data selection. In particular, very promising for a fine-grained presentation of contents, are metadata, or markups, which are also used to describe the contents within a document. With such descriptors dictionary customization could improve tremendously, displaying only what is suited to the user; for example, it would be possible to show researchers highly technical entries, and laypeople short descriptions of the same topic extracted from the same databank.

With the aid of all the technologies described, Bothma is even able to figure out the details of Tarp's 'Rolls Royces'. These are tools that users can set according to their preferred profile, and change them on the basis of the situation of use, or according to the desired level of complexity. In this way, the system could automatically improve its adaptation to the user, with reference to his behaviour, and present him only with the necessary information. Moreover, the user would receive recommendations from the system, and could make public or private notes on it.

The next chapter (number 5) by Spohr is as enlightening as Bothma's, since it shows how the theory can be translated into reality: "in the e-lexicography environment, theory informs praxis and praxis informs theory – the one cannot exist or advance without the other" (Bothma 2011: 101).

Spohr presents the software architecture of a sophisticated linguistic database developed using the "so-called *Semantic Web* formalisms" instead of a simple relational database, as there were too many hierarchies and relationships to be encompassed. These pages merit special attention, since first of all the author brilliantly explains the details of software architecture in simple terms to computational novices. Secondly the topic itself is paramount, since it is concerned with the necessary treatment of data in order to achieve the most fine-grained customization of the dictionary contents.

The stored data are linguistic and, in order to elaborate a coherent "structured data model" for them, the component parts of this database have been identified in its "lexical entities" and in the mutual relationships that may exist between them.

The *Semantic Web* tools allow the lexicographer to represent hierarchical structures avoiding redundancy, thus it is possible to extract different information for the same item:

For example, one could say that statements like '*investigation* is a nominalization of *to investigate*' and '*investigative* is an adjectivization of *to investigate*' are more fine-grained ways of expressing that '*investigation* and *investigative* are derivations of *to investigate*'. (Spohr 2011: 107)

The first kind of information ('*investigation* is a nominalization of *to investigate*') could be useful for experts, while the second ('*investigation* and *investigative* are

derivations of *to investigate*) is less technical and could be displayed to lay-people. In the database, the relationship 'is a nominalization of' and 'is adjectivization of' are sub-relations of 'is a derivation of', and the system is able to infer the more general statement about the derivation of the terms automatically. This means much in terms of the customization options that will be available for the lexicographer, and these options can be implemented only by incorporating an "access and presentation model" into the system, namely a "layer" that

defines which of the entities (both classes, properties and instances) are relevant to which users in which situations. [...] In fact, this layer contains not only one but several such filters, each of which may let through or filter out different pieces of information, depending on the specific user and situation types (Spohr 2011: 114).

As a conclusion to this section, the chapter by Leroyer (number 6) deals specifically with the shift in paradigm advocated by the many points of view already presented. After a brief theoretical introduction, Leroyer exemplifies the outcomes of the function theory presenting four different "lexicographically designed information tools": a cancer dictionary (*Lexonco*) aimed at patients and their families customized with three different access modes; a guide for Danish purchasers of French real estate who speak some French and for those having no command of the language whatsoever; a mobile tourist guide with automated and user-driven options; a database that will be improved in order to become an aid instrument for scientific text production. This last project, called ARTES (*Aide à la Rédaction de TExtesScientifiques*) was developed in collaboration with the Université Paris Diderot and is the topic of the tenth chapter in the volume by Granger and Paquot (2012).

From chapter 7 to 10, different dictionary projects are presented that have been developed within the framework of function theory; they will be discussed in section 4 of this review, whereas chapter 8, by Fuertes-Olivera and Niño-Amo deals with some proposals to improve *El Diccionario Inglés-Español de Contabilidad*, which was compiled following the model of the *Accounting Dictionaries* project developed by Centlex of the Aarhus School (presented in chapter 7 by Nielsen and Almind). Fuertes-Olivera and Niño-Amo discuss how to support users more effectively when they are faced with cognitive problems and need to improve their knowledge on a new topic, such as accounting. Their idea is to use hyperlinks to external contents and to provide the dictionary with a "systematic introduction".

In chapter 11, Lew (2011) writes an overview of the current online lexicography for the English language. He addresses "general English dictionaries", "learner's dictionaries", "user-involved dictionaries", "diachronic (historical) dictionaries", "subject-field dictionaries", "dictionaries with restricted macrostructures", "dictionaries with restricted microstructures" and "onomasiological dictionaries". Moreover, he discusses general issues of online lexicography,

such as the so-called "dictionary aggregators", the "step-wise access" facility (Hulstijn and Atkins 1998), the customization options currently available in electronic dictionaries, and the multimedia facilities with which they are provided – graphics, audio and video files. The chapter ends with a report on the resources that have a "dictionary-like interface" (such as *ForBetterEnglish.com*, *JustTheWord* or the DANTE project) which is generated by sophisticated wrappers that conceal the data sources (databases or corpora) and are perceived as true dictionaries by lay web surfers.

Customization proves to be particularly promising also in the field of Computer Assisted Language Learning (CALL), as demonstrated in chapter 13 by Verlinde, who presents the new access paths for the *Base lexicale du français*. This web resource was initially created as a database for the French language but it has been progressively developed into a rich multilanguage assistant which is currently accessible with a device called *Interactive Language Toolbox*, presented in chapter 8 of Granger and Paquot (2012).

In Fuertes-Olivera and Bergenholtz (2011), Verlinde describes three applications of the *Base lexicale du français*, namely a task-oriented assistance with reading, translation and writing. The applications for these tasks require that a text is submitted to the system, thus all the options available are somewhat interactive, managed with pop-up boxes which contain translations or suggestions for more appropriate word choices. Moreover, the text can be analyzed by the system and thus complex lexical items, such as multiword expressions and collocations, can be automatically detected. Another valuable feature of the *Base lexicale du français* is that the underlying database is enriched with a corpus of academic words and one of learners' texts, thus the translation assistant can suggest the correct stylistic variant for academic terms, while the writing tool can identify the syntactic and lexical "problematic patterns" for learners, with special reference to Dutch native speakers whose specific errors have been added.

In the last contribution before the closing summary of the Valladolid Congress (chapter 15), Heid (chapter 14) reports on the results of a "set of experiments on usability testing of electronic online dictionaries" (Heid 2011: 288), which were carried out in laboratory sessions according to the usability tests protocols used in Information Sciences. Small groups of participants are required, since "good practice has shown that 12 to 15 lay testers of a homogeneous user group will be sufficient to identify the majority of usability problems" (Heid 2011: 293), while the analysis is made with reference to all the actions performed by the subjects in task-based test activities. Heid (2011: 298) describes how detailed the protocol is:

The sessions were carried out in a usability laboratory using the Morae Observer software for each task, and each subject, keystrokes, and mouse movements were recorded, to get a picture of the navigation behaviour of the subjects; furthermore, screen video and sound recording were used to capture think-aloud protocols.

During the experiment an observer also takes note of many aspects of the session development.

The aim of these investigations is to assess the "usability" of software products, namely, if they are effective when the users are given what they need, and if they are efficient when the effort to satisfy users' needs is adequate¹.

Even though these tests gave evidence of the very basic consideration that "simple search interfaces work better than more complicated ones" (Heid 2011: 300), it was found during the laboratory sessions that some access routes of the *BLF* are malfunctioning, while others work perfectly well.

Also promising are Heid's ideas about possible further developments with respect to mock-ups that could allow investigations upon single critical aspects rather than whole consultation processes.

3. Granger and Paquot's *Electronic Lexicography*

The first part of the volume by Granger and Paquot deals with "Lexicography at a Watershed". It opens with an overview by Rundell of the current editorial policies in dictionary-making. He explains that the shift from p- to e-dictionaries is still an ongoing process for which a specific "business model has not yet emerged". However, Rundell's point of view is far from nostalgic, since "dictionaries have at least found their ideal platform in the online medium" and the corpus revolution has brought forth incredible advances in language descriptions in terms of the "reliable generalizations" which are allowed about lexical patterns and meaning discrimination. Rundell sketches the main steps of dictionary-making regarding the new tools and technologies available that are definitively going to turn the lexicographer's work from the "origination of dictionary text to the validation of decisions made by the software" (Rundell 2012a: 28): from the software for lexical profiling, to workflow management tools and proformas that provide the outline of the entries.

The volume continues with a chapter (number 2) by Kilgarriff and Kosem, who present the current corpora technologies, with special reference to the Sketch Engine. Their contribution is conceived as a guide to the available resources and facilities provided by these tools. Particularly noteworthy for the development of corpus linguistics are the advantages provided by the 'word sketches' of the Sketch Engine, "one-page automatic, corpus-based summaries of a word's grammatical and collocational behavior" (Kilgarriff and Kosem 2012: 44). Since its first use for the Macmillan Dictionary, it has changed the lexicographer's view of large corpora, because "higher volumes of data helped to make the sketches an ever-more reliable reflection of real usage, but without adding to the lexicographer's workload" (Rundell 2012a: 22).

The overall importance of the corpora revolution on lexicography is stressed and analyzed by Hanks (chapter 4), who speculates about what the dictionaries of the future will look like: "contextualization and phraseology will

come to take centre stage. These dictionaries will be electronic products with hypertext structures and links" (Hanks 2012: 64). Whatever the case may be, the chapter presents Hanks's view of lexical meaning, which underlies all his considerations and criticisms, namely the fact that words have only

'meaning potential' rather than meaning as such, and that, at least in the case of verbs, words need to be put into context before any attempt is made to define their meaning. Different contextual patterns activate different components of a word's 'meaning potential'. (Hanks 2012: 68)

According to Hanks, these findings should also influence the lexicography of the future, whose descriptions will be dependent at least on the word classes. Nouns for example could still be described on the basis of their referents, but as for verbs or adjectives only "their normal phraseology" will count. This implies that lexicography, being based on the analysis of the real use of words, is "in a position to spread [...] radical new approaches to the theoretical understanding of meaning in language" (Hanks 2012: 76). Starting from these premises, the *Wiktionary*'s review by Hanks is sometimes tinted with humour:

Our sympathies may be with an anarcho-syndicalist approach to lexicography, but it is hard to imagine how a radical new approach to defining verbs or natural-kind terms and thus enhancing our understanding of the nature of language and meaning could be carried out systematically by large, uncoordinated groups of enthusiasts and volunteers, some with more expertise than others. (Hanks 2012: 82)

The next chapter (number 5) by Abel covers the management process of dictionary editing, and describes the main components of dictionary writing systems: the editing tool, providing the entry template, the database and "administrative tools", used for the whole project management, from the assignment of duties to the monitoring of the dictionary text. In the last paragraph, Abel reports on the latest uses of databases for the compilation of dictionaries. She mentions cases in which the database is the dictionary itself, such as the DANTE project (as it is also reported by Lew 2011, see above), and, vice versa, of multiple dictionaries extracted from the same database; the reference is of course to the "pluri-monofunctional tools" compiled by the Aarhus School, and in particular to the sophisticated databank system by Spohr (2012, see above).

Tarp follows Abel in chapter 6. He underlines the role that abstraction plays in reconsidering the nature of the needs that lexicography aims at satisfying: apart from those needs which are specific to every single dictionary, the general ones are simply 'information'. Therefore it is necessary to turn from the pure "art and craft" of lexicography to a general theory, in order to derive a comprehensive paradigm to design not only valuable present-day tools, but also those of the future. Abstraction has come into play in function theory also for profiling users' needs, since for a science it is necessary to elaborate valuable generalizations at least on its chief standpoints. Users were thus consid-

ered as 'types' for a while, and their dictionaries were the 'monofunctional' tools set for each of them, but Tarp (2012: 115) expresses his dissatisfaction with this simplistic generalization, since

no *type* of user has ever made a *type* of lexicographical consultation in order to access a *type* of data that may meet a *type* of information need occurring in a *type* of social situation. The only thing that has ever happened, and which happens every day, hour, and minute, is that an *individual* user with *individual information needs* occurring in an *individual situation* decides to make an *individual lexicographical consultation* in order to access the *concrete data* that may satisfy his or her individual needs.

The shift from the 'kind' of user to the real one implies a complete rethinking of dictionaries, from the 'monofunctional' to a more customizable tool, whose features can be changed manually by the user, or set automatically by the system detecting the user's behaviour. This is how function theory is going ahead now with its vision of future electronic reference tools.

Prinsloo (chapter 7), dealing with "Electronic Lexicography for Lesser-resourced Languages" and taking the South African context as its focus, complains that the situation is poor and, though the current technology could do a lot to manage the demanding morphology of the Bantu languages, there are still very limited resources for them.

The second part of the book, dedicated to "Innovative Dictionary Projects", starts with the presentation of the *Interactive Language Toolbox* (ILT) by Verlinde and Peeters (chapter 8), which is the new interface of the *Base lexicale du français*, designed on the basis of a thorough usability study which was also reported on by Heid in the volume by Fuertes-Olivera and Bergenholtz (2011, see above). The need for greater simplicity in order to satisfy users' needs has been converted into a single search box which allows 'incremental searches' (or "step-wise access", as Hulstijn and Atkins call it, see above) and has been improved in terms of user-friendly interface and layout.

The other dictionary projects, covering chapters from 9 to 12 and 14, are the object of specific analyses in the next section, while chapter 13, by Meyer and Gurevych, reviews three language versions of *Wiktionary*, the English, German, and Russian. The survey compares the collaborative Wiki platform to expert-built lexicons, testing terms and senses coverage, definitions, and indications about register and style. The findings of Meyer and Gurevych are very different from the considerations made by Hanks in the same volume (see above), since in these respects, the resource proved to be valuable; moreover, it offers a rich hyperlinking system, different access routes and illustrative graphics, and therefore it is to be expected that it can serve different types of users and needs.

The user's perspective is the topic of the third part "Electronic Dictionaries and their Users", which opens with a review by Dziemianko (chapter 15) on empirical studies about dictionary use. Unfortunately, research comparing

paper and electronic dictionaries proves to be disorienting not only for the different outcomes, but also with reference to the tasks administered, number of participants, and survey type – from eye-tracking to record sheets. In her conclusions, Dziemianko underlines that even for the few tendencies that emerge, such as positive vocabulary retention when there is more search effort, other investigations are needed. Her final claims are in line with those by Heid (2011, see above), stating the need for "prefabricated, purpose-built dictionaries" in order to control the parameters involved and allow more adequate comparisons.

In the following chapter (number 16), Lew (2012) evaluates the effectiveness of access to dictionary data, considering morphological (viz. inflected forms) and graphematic drawbacks: if the user does not know the right spelling or the base form of a word, how could electronic dictionaries help? Many options available in current dictionaries are illustrated and discussed. Moreover, Lew explains the advantages and disadvantages of multimedia files, urging caution on the use of video files for learning purposes.

Nesi (chapter 17), instead, reports on 'alternative e-dictionaries', or dictionaries that are not edited by the major publishing houses. Her review is focused on Eastern lexicography but the results are significant at every latitude since these resources are particularly popular among language learners but their quality is not guaranteed by experts.

In chapter 18, Bower writes a wish-list for the lexicographer, requiring specific information a translator may be interested in, such as general and specialized terms, phraseologisms, and multimedia files, but also notes on frequencies and warnings about inadequate word use.

Fuertes-Olivera, in chapter 19, reviews the features of free online business dictionaries and, comparing them to the inadequate printed bilingual resources, concludes that these dictionaries can support learners since they offer valuable extra-resources, such as videos, links to articles on specific topics, and updated data. Moreover, Fuertes-Olivera argues against the use of the available bilingual dictionaries not only because they lack coverage of the subject matter, but also because they offer the illusion of clear-cut cross-language correspondences.

The concluding chapter (number 20) by Müller-Spitzer, Kopleinig and Töpel reports on different surveys carried out applying the protocols of social sciences: many participants were involved, and accurate statistical analysis of data was performed. The findings of this three-year, "externally-funded" research has shown that online dictionaries are accessed using notebooks and desktop computers, rather than smartphones and small-screen devices, while the "tab view" modality of presenting dictionary information is the one that users prefer. It also proved the need for explicit instructions when innovative dictionary features are introduced, otherwise users do not realize their benefits.

Concluding their report, the authors address the objections made by Bergholtz and Johnsen (2005) on survey methodology, claiming that if the strict

parameters of social sciences protocols are applied, this research can be as valuable as the others.

4. Where have all the linguists gone?

The sections in both volumes devoted to innovative lexicographical projects offer a privileged standpoint of comparison. They will be analyzed here with special reference to the way they approach the treatment of linguistic data, because this topic has proved to be the crucial concern of the two opposing visions of "practical lexicography" and "function theory". Before presenting these electronic projects, preliminary remarks are necessary.

If compared to the volume by Fuertes-Olivera and Bergenholtz, it is immediately clear that among the pages of Granger and Paquot's book harsh criticisms have turned into more moderate positions, even though neither of the parties recognizes that there is room for both. Tarp paraphrases the metaphor of the "linguistic colonialism" (Gouws 2011: 22), affirming more fairly that it is simply wrong to believe that linguistics can "answer all challenges" of present-day lexicography (Tarp 2012: 118). On the other hand Rundell (2012a: 29), explaining why there will still be dictionaries in the future, quotes Tarp (2008: 40) himself and then concludes: "it is [...] clear that the dictionary is morphing from its current incarnation as autonomous 'product' to something more like a 'service', often embedded in other resources". It is not specified if these hybrid devices have anything to do with the "information tools" of function theory, namely "any tool, no matter what we call them, aiming to satisfy the needs users might have" (Fuertes-Olivera and Bergenholtz 2011: 3) in all possible situations.

Nevertheless, the speculation on the merging of the two stances is not a lazy intellectual exercise solicited by the broad-coverage of Granger and Paquot's book that urges the reader to put all the pieces of the current debate into a tidy picture. This game in speculation is indeed solicited by the lexicographical projects presented in the book, which in one way or another are indebted to the function theory, or at least make reference to it, but also deal with the linguistic issues of data selection and description.

Granger herself seems unable to perceive the contradiction. In her introduction to the book, she presents the topic of customization of data in the dictionaries' entries quoting different scholars who have addressed it, and after having started from a representative of the "Anglo-Saxon tradition" (Tarp 2012: 107), namely Sue Atkins, she ends with the Danish lexicographers Bergenholtz and Tarp, explaining that the function theory of lexicography "underlies several chapters in this volume" and that "users' needs have become a central issue not only for practical lexicography but also for lexicographic theory" (Granger 2012: 4).

Similarly, the dictionary projects presented in the book by Granger and Paquot deal with language descriptions but are also indebted in one way or

another to the function theory. In this respect, it is useful to consider the features of the DiCoInfo project (*Dictionnaire fondamental de l'informatique et de l'Internet*), directed by L'Homme at the Université de Montréal, and of LEAD (*Louvain English for Academic Purposes Dictionary*), directed by Granger at the Université catholique de Louvain, ARTES (*Aide à la Rédaction de Textes Scientifiques*), coordinated by Pecman at the Université Paris Diderot, the *Danish Sign Language Dictionary*, created by the Centre for Tegnsprog in Denmark, and the *Transpoetika Dictionary* managed by the Belgrade Centre for Digital Humanities.

The ergonomic customization of DiCoInfo (*Dictionnaire fondamental de l'informatique et de l'Internet*), a lexical database organized according to the principles and formalisms of Mel'čuk's lexical functions, has been carried out by the researchers of the Observatoire de linguistique Sense-Texte, L'Homme and Robichaud, in collaboration with Leroyer of the Aarhus School. The authors explain that

the adaptation of DiCoInfo to user-friendly representations has combined the resources of sophisticated linguistic encoding, a functional approach to specialized lexicography, and innovative computational programming for efficient data access and presentation. In the functional framework of lexicography (Tarp 2008, and this volume), dictionaries – and this can apply to specialized dictionaries – are considered as products designed for specific purposes, i.e. tools designed to meet information needs, and therefore solely defined according to the functions they are meant to fulfil. [...]

As far as the adaptation of DiCoInfo to user needs is concerned, two types of assistance were identified – assistance in connection with text production in L2, and assistance with translation from L2 to L1 – both belonging to the category of communicative user situations. (L'Homme, Robichaud and Leroyer 2012: 224-225)

In the closing statements, the authors enthusiastically declare that their sophisticated linguistic analysis and data structure have benefited a great deal from the quoted "functional principles" used to present them:

We attempted to show that dictionaries that contain rich linguistic information encoded with a formal system can still become user-friendly tools. (L'Homme, Robichaud and Leroyer 2012: 235)

Linguistic data were also carefully considered for compiling the *Louvain English for Academic Purposes Dictionary*, a "writing and learning-aid tool" (Paquot 2012: 171) for academic English, supporting non-native speakers. The writing and learning functions required a preliminary linguistic investigation in order to collect the necessary data, and the key lemmata for scientific discourse were identified on the basis of a specific study which extended the inventory proposed by the existing literature with high-frequency words – such as *namely*, *compare*, *aim* – expressing "rhetorical functions that are particularly prominent in academic discourse (e.g. give examples, express cause and effect, conclude, and express possibility and certainty)" (Paquot 2012: 171). Moreover, the learn-

ing function required a specific corpora analysis of text productions by non-native speakers, while the dictionary makes extensive use of the six different corpora (four of "expert writing" and two of learner writing) it contains, as explained by Paquot (2012: 184):

Examples of collocations and lexical bundles are automatically extracted from discipline-specific corpora. Corpora of learner writing are used to inform contrastive error notes targeting specific L1 learner populations. Customization is also implemented at the mesostructure level. According to users' profiles, lexical entries are linked to relevant concordance lines in discipline-specific corpora. As the corpus-query tool is fully integrated into the *LEAD*, users also have access to these specialized corpora to search words that are not in the dictionary.

Conversely, Kübler and Pecman complain that the lack of a consistent description of some language phenomena prevents the desired development of *ARTES (Aide à la Rédaction de Textes Scientifiques)*, a bilingual resource of scientific texts hosted by the Université Paris Diderot. The database aims to offer assistance for translators, professionals and linguists with the most fine-grained issues of specialized phraseology, such as "semantic prosody" and different types of collocations, i.e. "generic collocations" and domain-specific collocations, called by the authors "semantic preference". The semantic prosody is particularly challenging for translation, since it refers to the connotative aspect of meaning that words acquire because of their collocates, and there are few cross-linguistic correspondences in this respect, as Kübler and Pecman (2012: 203-204) explain:

The English verb *commit* [...] has a negative semantic prosody, as it co-occurs with nouns like *crime*, *murder*, *mistakes*, and *suicide*, all words having a negative connotation. The French equivalent *commettre* also presents a negative semantic prosody, as it also occurs with nouns having a negative connotation, such as *crime*, *délit*, *attentat terroriste*, *vol*, *erreur*, and *faute*. A more thorough analysis, however, shows differences between the French and English sets: the French noun *suicide* cannot be used with *commettre*, as an equivalent of *to commit suicide* (**commettre (un/o) suicide*), [...] and is best translated by the French synthetic verb *se suicider*. Not all typical collocates of the French *commettre* have a negative connotation: in the expression *commettre un roman/une oeuvre/une pièce de théâtre* [...], the nouns themselves do not have a negative connotation, but through the aura of meaning of the verb *commettre*, they take on a negative connotation, yielding a certain irony.

As for the other phraseologisms considered, Kübler and Pecman distinguish those collocates "which are common to a variety of" specialized domains ("generic collocations") and therefore can be used in all scientific texts – e.g. *my concern here is with, in much the same manner, by contrast* – and domain-specific phraseology, since "evidence that the semantic preference of a lexical item is not the same in different varieties of language is overwhelming". For example, "the verb *run* does not co-occur with the same semantic set in general science and computer science English" (Kübler and Pecman 2012: 203).

Though ARTES offers its users privileged onomasiological access to collocations and provides translational equivalences for them, the database architecture seems to be unable to account for the complex requirements of the semantic prosody, since the system encodes lexical units as wholes and each collocation is stored as a single piece, while it should be the other way around, as the authors implicitly admit: "we can instead decide to encode semantic prosodies and preferences as a specific type of relation between terms" (Kübler and Pecman 2012: 207).

More than a thorough linguistic theory, what appears to be missing is a proper data storage architecture, apparently caused by the ongoing nature of the project, which collects data from Master theses and "was designed to cater for teaching and learning needs", while now it seems to be turning into something else for other user groups. Therefore it is useful to refer to Leroyer (2011), since he is collaborating on the analysis of the lexicographical "functional aspects" of ARTES:

the goal is to design a central database prepared for the import and indexing of several thousand terminological records presently stored in other database systems, and to provide advanced search options as well as flexible editing options. (Leroyer 2011: 138)

Linguistic problems are instead crucial issues in sign language lexicography, and prevent the advances that the electronic medium seems to be offering, as Kristoffersen and Troelsgård (2012: 311) affirm in the presentation of the *Danish Sign Language Dictionary*:

It may be argued that a sufficiently strong base of linguistic research results to provide a dictionary with information beyond a basic description of the form and meaning of the sign vocabulary has not yet been developed.

Though focused on one specific sign language, this paper is also concerned with the general issues at stake for all sign languages, which share more features among them than spoken ones². Kristoffersen and Troelsgård highlight how videos can partially answer the "overriding challenge" of "how to render signs in the absence of a written language", offering monolingual definitions and dictionary entry contents. Instead icons can be used as access routes starting from handshapes and place of articulation (the basic components of signs together with orientation and movement, Kristoffersen and Troelsgård 2012: 302), but also additional parameters can be combined in the search facilities that an electronic dictionary allows. However, more basic linguistic issues prevent an adequate lexicographical description of the Danish Sign Language, such as the treatment of synonymic relationships or even the identification of the base form of a word and the lexical class it belongs to. Since great lexical variation is actually a common feature of sign languages and there is no accepted rule allowing one to identify synonyms among them, the lexicographer needs to decide by himself whether two forms are allophones or different

lexical units with the same meaning. In cases like these, a "consistent" analysis, based on explicit parameters, is needed (see Kristoffersen and Troelsgård 2012: 305). On the other hand, only more research on word classes and base forms of signs can bring forth "firm generalizations" to identify the correct lemma for a dictionary entry.

Kristoffersen and Troelsgård (2012: 302) consider the possible benefits of function theory for the great variety of their target users, and say that

it might be more practicable to consider function theory (Tarp 2008), and describe modern sign language dictionaries not by assigning them to traditional dictionary types, but rather by looking at potential user needs in different communicative situations.

However, for the moment, no user customization is available and the present dictionary is defined as "multifunctional", "comprehensive and user friendly" (Kristoffersen and Troelsgård 2012: 302).

Similarly, the *Transpoetika Dictionary*, a WordNet-based Serbian-English pedagogical vocabulary, will probably be adaptable in the future, thanks to a "layered interface that helps users answer specific questions related to their particular communicative or cognitive needs rather than a unified dictionary entry" (Tasovac 2012: 256). However, for the moment, the project seems to be promising only from a linguistic point of view, since, as the author underlines, the WordNet format offers many advantages for a less-resourced language such as Serbian. First of all, the so-called 'expand approach' allows a systematic description of the language using the cross-linguistic comparison with the English Wordnet, and when lack of correspondence occurs, language-specific synsets can be added. Specific improvements of the original format are instead required by the dictionary pedagogical function, such as grammar descriptions, which can be implemented by additional reference to an external MorphoSyntax database, and labels specifying register, regional forms, taboos and all the linguistic variations of which a learner must be aware.

This brief description of the lexicographical projects presented in Granger and Paquot's book shows how many of them are basically concerned with linguistic issues but also consider the possibility of user-friendly access routes to their data. On the contrary, the electronic dictionary tools presented in the book by Fuertes-Olivera and Bergenholtz (2011) take the linguistic data for granted, with the one exception of Spohr's contribution which, however, is focused on how these data are computationally organized. Consider, for example, Nielsen and Almind (2011) in the chapter discussing the route "From Data to Dictionary" (Fuertes-Olivera and Bergenholtz 2011, chapter 7).

Their standpoint is exemplified describing the *Accounting Dictionaries*, one of the leading projects of the Aarhus School, comprising five different monofunctional dictionaries³, whose declared "theoretical basis [...] is not text linguistic" but the "lexicographical functions" (Nielsen and Almind 2011: 141), or the kind of help a dictionary should provide to its users in order to satisfy their

needs.

The focus is therefore on the "technical options for accessing" data, not on the data *per se*, since these must be "presented in such a way that users can easily turn [them] into useful information" (Nielsen and Almind 2011: 155). User accessibility is thus a cognitive matter, since

printed and computerized dictionaries contain data and not information, but lexicographers must collect and present data that dictionary users can readily convert into information. The traditional linguistic and text linguistic approaches to lexicography have serious shortcomings, so dictionaries based on these approaches do not fully satisfy the needs for help and knowledge users have in specific types of situation. One way to address this problem is to re-assess the practical and theoretical foundations of online lexicography in light of the electronic options available to produce well-crafted reference tools. (Nielsen and Almind 2011: 166)

Nielsen and Almind describe the overall structure of these tools, which are made up of three different component parts: a database, a website, and a search engine. On the theoretical side, this structure is presented as a "triadic setup" (Nielsen and Almind 2011: 147) in which an "interface", or the website, replaces the static p-dictionary and displays all the contents that can be extracted from the database, using the search engine as a mediator between the database and the website. The conclusion is that, in the electronic environment, the dictionary is only an interface, and its "macrostructure has been replaced by a 'data presentation structure'" (Nielsen and Almind 2011: 148). The lexicographer's main task is to organize "how data elements relate to each other", and his/her main concern regards the type of database structure which will allow the most efficient data selection that, through the access routes of the search engine, is displayed to users. As for the data, in "lexicographically relevant databases [they] are of the type *text* and therefore rather uninteresting from a computational point of view" (Nielsen and Almind 2011: 142).

Computational issues are key, but are not the only ones by any means, since data architecture and management depend on the users' needs to be satisfied. This aim is achieved by "profiling" the intended users with the "diagnostic checklist", namely the listing of questions necessary to describe the specific needs that must be satisfied. For example: what is the users' native language?, what level of proficiency do they have in the foreign language? and "at what level do they master the special subject field" of the dictionary? (see Nielsen and Almind 2011: 150). The answers to these questions "will show the competences of the target group of the dictionary" and allow the selection of the necessary data that can effectively support users, particularly when they are likely to lack specific competences. In cases like these, the lexicographer can even assume a "proscriptive approach", and make recommendations to the user (Nielsen and Almind 2011: 165), suggesting, for example, the most appropriate spelling variant among the many used for the same word. The recommendation is made on the basis of a corpus data study, from which the lexicographer

can verify the contexts in which every variant occurs, and thus make the most valuable generalization on the proper spelling of the term.

At this point, corpus issues appear also within the function theory, which could be suspected to be a contradictory theme within the tenets of this conceptual framework, since it is concerned with linguistic problems and, as such, probably to be discredited. However, what Nielsen and Almind do in practice is to change the perspective from which the matter is seen. The "entire process of selecting data for the dictionaries" is ruled by the operative criteria of "relevance", namely "the quality of being directly connected with the subject field in question", which allows to "distinguish [...] data that directly support a lexicographical function" and therefore are "useful lexicographical data":

For example, collocations are selected because they are important when producing accounting texts, because they are important when translating accounting texts (and often difficult to translate between the languages concerned), and examples are selected because they specifically show how to write and translate accounting texts as well as provide data for knowledge building. (Nielsen and Almind 2011: 154)

As it appears from the description of this dictionary project, the linguistic data *per se* are not separate from the lexicographer's job, but they are only one aspect of the dictionary editorial process and in no way its starting point. Thus in the paragraph describing how the corpora of the *Accounting Dictionaries* were compiled ("Selecting Data for the Accounting Dictionaries Is a Multi-Stage Process"), the main concern regards the preliminary classifications needed in order to correctly "reflect the structure of the field of accounting" and build the proper corpora from which terms, collocations and all the data are extracted. In order to obtain this faithful representation, a careful analysis was carried out not only on the external limits of the subject field (what must not be included), but also on the internal structure of the subject considered and the corresponding terminological classification; while consulting experts of the field prevented "lacunae in the corpora and lack of factual knowledge" (Nielsen and Almind 2011: 153). In these accounting dictionaries, phraseological units – e.g. their extraction, management or presentation – are not shown as problematic issues, which is, however, the case with the DiCoInfo, the LEAD and the ARTES projects in the three aforementioned chapters in Granger and Paquot's volume. On the contrary, special linguistic attention is paid to homonymy and polysemy, considered to be "important aspects in connection with understanding the meaning of specialized terms". In this passage, the authors speak in strict metalinguistic terms when they address 'word class' as a "syntagmatic criterion":

The syntagmatic criterion 'word class' is used for treating terms as homonyms, so that the dictionaries clearly distinguish between homographs belonging to different word classes, for example the noun *expense* and the verb *expense*. (Nielsen and Almind 2011: 157)

They also go into the details of a purely linguistic analysis when dealing with polysemy:

Morphological criteria are generally used in cases of polysemy, so that homographs that can be both countable and uncountable are treated as being polysemous, such as the noun *authority* ('power to make contracts on behalf of another' (uncountable) and 'governmental agency' (countable), and words of the same word class that have different inflectional paradigms are treated as polysemous. (Nielsen and Almind 2011: 157)

Finally, assuming the standpoint of the "relevance" principle, it is possible to infer that the claim against linguistics is more a stance for a radical shift in the lexicographical paradigm, rather than a crusade against the role that linguistics can have in the correct treatment of language problems.

In this respect, the absence of linguists and linguistic theories during the compilation of the *Danish Music Dictionary*, which Bergenholtz and Bergenholtz underlines in chapter 9 of the same book (Fuertes-Olivera and Bergenholtz 2011), needs closer inspection. In this paper, Bergenholtz and Bergenholtz (2011: 189) declare:

Almost all of those who call themselves lexicographers, however, are of the view that lexicography is a linguistic discipline. This is remarkable. For the music dictionaries that are the subject of this contribution, there was no need for any cooperation with linguists or linguistic theories. Three types of experts participated in the planning and execution:

1. an expert in lexicography
2. an expert in music theory and the history of music
3. an expert in lexicographical databases

Obviously, there are dictionaries for which the co-operation of linguists is required, for example, in a general language text production dictionary [...]. For some dictionaries, but not for all, the co-operation of linguists is required.

The music dictionary is used by Bergenholtz and Bergenholtz to exemplify one of the possible applications of the function theory, namely the editing of 'mono-functional' dictionaries, which are those tailored for a specific task and therefore are expected to accomplish their users' needs better than the general 'poly-functional' dictionaries do. More precisely, the *Danish Music Dictionary* is made up of three tools, one for reception, and one for knowledge, while the other one is classified as a 'polyfunctional' dictionary extracted by wide-ranging search routes in the underlying database. This last one "may even serve as a production dictionary", but, since the amount of information provided in the database seems to be absolutely huge, it appears unlikely that this function could be effectively satisfied. The authors themselves prove to be surprised by some search results, such as the word "darbuka" after having typed the string "Berlioz".

However, resuming the initial statement about the fact that the dictionary

was compiled without linguists or linguistic theories, it seems that the implicit assumption is that they were not used simply because they were not needed, since there are no particular language challenges in this lexicographical project, whose functions are strictly receptive and cognitive. In cases like this, lexicography proves to exist on its own, without linguistics at its side. Moreover, the claims for freedom from linguistics also have political implications: "If lexicography were denied any form of being a science, linguistics would remain a discipline at our universities, but lexicography would not" (Bergenholtz and Bergenholtz 2011: 189), and the academic research in this field could be put into question. The metaphor introduced by Gouws of the "linguistic colonialism" in chapter 1 is also in line with these stances, and alludes to the broadening of the lexicographical horizons implied by the paradigm shift of the function theory. Dictionary compiling, as Gouws (2011: 27) states, is a task to be undertaken by a team, whose "members will typically come from different fields, depending on the type of dictionary to be compiled, and these different team members will acknowledge the diversity residing in and implied by the title of 'lexicographer'". In this vision, "linguists are not the only, or even the major, participants in the discussion".

The editorial project of a phrasal verb tool offers additional evidence of the choices and solutions adopted with respect to language data in the function theory. Andersen and Almind (chapter 10) created a database of English phrasal verbs from which three monofunctional dictionaries can be extracted to support Danish speakers with reception, translation, and production of English texts. There is no mention in this paper of any corpus from which the data were collected, while the authors claim to have referred to the British and American lexicographical practice of including in the lemma inventory not only combinations of verbs and adverbs, which is the strict "linguistic definition of phrasal verbs", but also combinations of verbs and a preposition (*to look into the matter*), or verbs with an adverb and a preposition (*to get on with it*). Andersen and Almind (2011: 217) admit that the data are not their concern, since they are the "traditional ones", "the innovative aspect of the project is that the database will be capable of generating three different dictionaries with three different functions". The database is described in the paper in one paragraph dealing with its structure and with the user interface, whereas other interesting outcomes of function theory on the linguistic data treatment can be found with respect to the decisions made on the "sense ordering, grammatical information and the style labels". Firstly, the lemma inventory was enriched with fixed expressions and free combinations that may be problematic for non-native speakers to understand, translate or use. The same criteria guided the ordering of the phrasal verb meanings within the displayed entries. For text reception and translation, it was assumed that the idiomatic meanings should come first, since these are "more likely to be consulted" (Andersen and Almind 2011: 223); instead, for text production the opposite holds: phrasal verbs with transparent compositional meaning are more commonly used by non-native speakers, who

would probably check their syntactic behaviour or recommendations about style and register. However the authors warn that the best solutions to address users' needs must be verified by specific empirical studies and analysis:

It must be emphasized that these observations are not based on empirical investigations, and any decisions about conscious sense ordering will have to rely on findings from investigations into these matters. (Andersen and Almind 2011: 223)

The topic of frequency is also discussed critically by Bergenholtz and Bergenholtz in chapter 9, with reference to the selection of specialist terms, for which "frequency is not an argument; subject relevance is" (Bergenholtz and Bergenholtz 2011: 193), since even the most rarely used technical terms are necessary in order to give a "systematic insight into" the field, regardless of whether they seldom appear in texts or not. The same is underlined by Nielsen and Almind (2011) for the selection of lemmata in the *Accounting Dictionaries*, in this case experts provided the necessary assistance to avoid omissions.

The challenging topic of tailoring definitions, partially dealt with by Bothma (2011: 90-91) who suggests using metadata markups, is addressed by Sánchez and Cantos in chapter 12. The authors apply the Lexical Constellation Model, a proposal for lexical semantics that they presented in two previous papers. Starting from these researches on corpus data, Sánchez and Cantos (2011: 262) elaborate their theory about lexical units which

result from the clustering of specific semantic features which are perceived as units by the speakers. These units are, however, not isolated entities; they may share part of their features with other lexical units, so that the units intervening in the same set of connections are not fully independent as regards their semantic properties. Such interconnectivity is the very foundation of a lexical constellation.

The "constellation model" seems to be inspired by the tenets of structuralist semantics (see Sánchez and Cantos 2011: 266), however its description would have been worth further explanation and more details, since the authors seem to emphasize the general problems of dictionary definitions and the shortcomings of present dictionaries more than illustrating their own model. Additionally, the example they provide to illustrate their theory is quite misleading, since they explain the word *lion* referring to the scientific taxonomy classification, and explaining that it is "categorized first as an 'animal', later 'a mammal', and then 'a cat'" (Sánchez and Cantos 2011: 262). To these general categories, physical features of the animal are added, e.g. its colour and size, and only in the end they underline what is crucial for the lexicon: "In any case, people may normally never mention that lions are animals or mammals, [...] while less common knowledge will be more decisive for identifying the world around us" (Sánchez and Cantos 2011: 263). However, even if the *lion* "may be taken as an illustrative example of the model [they] have in mind", when they "illustrate

[their] proposal at work", the methodology they use is by no means based upon a combination of taxonomies and pieces of cultural knowledge, but resorts to corpora analysis:

Mano, 'hand' in Spanish, is defined in a paper dictionary, GDUEsA, with 17 different meanings and more than 100 idioms and set phrases, in which *mano* appears as the head word. We looked for the meanings of *mano* in the corpus *Cumbre* and then we organized them in accordance with the LCM. (Sánchez and Cantos 2011: 266)

The aspect that seems to fit well into the function theory is the customization options that the semantic features implied by the theory allow, since it is possible to link a bundle of them to a particular user profile and thus to create definitions that can be increased and detailed with respect to the intended users.

5. No conclusion available yet

The careful analysis of the innovative dictionary presented in the books has underlined a difference in the agenda of the compilers, rather than a difference in their concerns. While all the projects of Centlex of the Aarhus School were conceived from the beginning in accordance with the parameters of the function theory, others started as lexicological archives and then evolved into something else: consider ARTES, but also the *Base lexicale du français* which has been improved step by step to reach smart personalization options. On the contrary, for those that still have major linguistic concerns, such as the *Danish Sign Language Dictionary*, the customization is only one of the many issues to be considered for future improvements.

However, as Spohr's challenging database project demonstrates, customization depends on data storage organization: the more detailed the analysis and archiving of these data, the better they will be displayed to the final user. When the data are linguistic, the general ontology and relational system must be of a linguistic kind, and linguistic theories will undoubtedly help a great deal.

This does not mean that the dictionary must display frequency labels and lists of collocations to school children, but that with one single database, provided with the finest language description, lexicographers would be able to compile one dictionary for linguists and one for primary school students at the same time. More precisely, these dictionaries will not be compiled, they will be displayed, since they will not exist before the user accesses them.

This dynamicity is one crucial point of function theory, since there was a radical shift in the 'user paradigm': the 'type' of user no longer exists, the future is only about real users (Tarp 2012, see above). The electronic tools that will be produced for them will allow the most tailored searches on the basis of manual settings by the users and automatic detection of the user behaviour by the system.

Within the framework of the function theory, one supposes that also empirical studies will play a key role, as claimed by Andersen and Almind (2011: 223), since not all the choices made by the lexicographer in order to satisfy a specific function are obvious, many of them remain arbitrary (e.g. sense ordering, specific entry contents and so on) and based on good common sense. Generally speaking, dictionary testing is one of the great concerns of researchers (Bergenholtz 2011: 30-53; Dziemianko 2012: 319-342), particularly with respect to the use of lexicographical resources by learners, and since there is a general discontent regarding the amount of research that has been carried out in this field, the proposal by Heid (2011: 287-304) of using 'usability testing' protocols seems to be at least new and worthy of further study.

To conclude, it must be underlined that the principles of function theory are taken into consideration in all but three chapters of Granger and Paquot's book, and they are also used in four innovative dictionary projects out of the six presented in the book, so one may suspect that function theory is the current academic fashion, or that Bergenholtz's adage is valuable: "nothing is more practical than a good theory".

All in all, the two volumes are about the future of lexicography and only time will tell what the outcomes of the current research in the field will be. Until that moment no conclusions on the topic can be written, it can only be observed, as Rundell says, that "we live in interesting times".

Notes

1. For the proper formulation of these concepts, see Heid 2011: 289-291.
2. This situation has been exemplified by Newport and Supalla (2000: 109) with the *dinner conversation paradox*: "A long dinner among Deaf users of different sign languages will, after a while, permit surprisingly complex interchanges."
3. These are the monofunctional dictionaries listed by Nielsen and Almind (2011: 154): *English Accounting Dictionary*, *Danish Accounting Dictionary*, *English–Danish Accounting Dictionary*, *Danish–English Accounting Dictionary*, and *English–Spanish Accounting Dictionary*.

References

- Abel, A. 2012. Dictionary Writing Systems and Beyond. Granger, S. and M. Paquet (Eds.). 2012: 83-106.
- Andersen, B. and R. Almind. 2011. The Technical Realization of Three Monofunctional Phrasal Verb Dictionaries. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011: 208-229.
- Atkins, B.T.S. 1993. Theoretical Lexicography and its Relation to Dictionary-making. *Dictionaries: Journal of The Dictionary Society of North America* 14: 4-43.
- Atkins, B.T.S. and M. Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.
- Béjoint, H. 2010. *The Lexicography of English*. Oxford: Oxford University Press.
- Bergenholtz, H. 2011. Access to and Presentation of Needs-Adapted Data in Monofunctional Inter-

- net Dictionaries. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011: 30-53.
- Bergenholtz, H. and I. Bergenholtz.** 2011. A Dictionary Is a Tool, a Good Dictionary Is a Monofunctional Tool. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011: 187-207.
- Bergenholtz, H. and M. Johnsen.** 2005. Log Files as a Tool for Improving Internet Dictionaries. *Hermes. Journal of Linguistics* 34: 117-141.
- 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.
- Dziemianko, A.** 2012. On the Use(fulness) of Paper and Electronic Dictionaries. Granger, S. and M. Paquot (Eds.). 2012: 319-342.
- 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.
- Gouws, R.H.** 2011. Learning, Unlearning and Innovation in the Planning of Electronic Dictionaries. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011: 17-29.
- Granger, S.** 2012. Introduction: Electronic Lexicography – from Challenge to Opportunity. Granger, S. and M. Paquot (Eds.). 2012: 1-11.
- Granger, S. and M. Paquot (Eds.).** 2012. *Electronic Lexicography*. Oxford: Oxford University Press.
- Hanks, P.** 2012. Corpus Evidence and Electronic Lexicography. Granger, S. and M. Paquot (Eds.). 2012: 57-82.
- Heid, U.** 2011. Electronic Dictionaries as Tools: Toward an Assessment of Usability. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011: 287-304.
- Hulstijn, J.H. and B.T.S. Atkins.** 1998. Empirical Research on Dictionary Use in Foreign-language Learning: Survey and Discussion. Atkins B.T.S. (Ed.). 1998. *Using Dictionaries. Studies of Dictionary Use by Language Learners and Translators*: 7-19. Tübingen: Max Niemeyer.
- Kilgarriff, A. and I. Kosem.** 2012. Corpus Tools for Lexicographers. Granger, S. and M. Paquot (Eds.). 2012: 31-55.
- Kristoffersen, J.H. and T. Troelsgård.** 2012. The Electronic Lexicographical Treatment of Sign Languages: The Danish Sign Language Dictionary. Granger, S. and M. Paquot (Eds.). 2012: 293-315.
- Kübler, N. and M. Pecman.** 2012. The ARTES Bilingual LSP Dictionary: From Collocation to Higher Order Phraseology. Granger, S. and M. Paquot (Eds.). 2012: 187-209.
- 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.** 2011. Online Dictionaries of English. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011: 230-250.
- Lew, R.** 2012. How Can we Make Electronic Dictionaries More Effective? Granger, S. and M. Paquot (Eds.). 2012: 343-361.
- L'Homme, M. Robichaud, B. and P. Leroyer.** 2012. Encoding Collocations in DiCoInfo: From Formal to User-friendly Representations. Granger, S. and M. Paquot (Eds.). 2012: 211-236.

- Newport, E.L. and T. Supalla.** 2000. Sign Language Research at the Millennium. Emmorey, K. and H. Lane (Eds.). 2000. *The Signs of Language Revisited: An Anthology to Honor Ursula Bellugi and Edward Klima*: 103-114. Mahwah, N.J.: Erlbaum.
- Nielsen, S. and R. Almind.** 2011. From Data to Dictionary. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011: 141-167.
- Paquot, M.** 2012. The LEAD Dictionary-cum-Writing Aid: An Integrated Dictionary and Corpus Tool. Granger, S. and M. Paquot (Eds.). 2012: 163-185.
- Rundell, M.** 2012a. The Road to Automated Lexicography: An Editor's Viewpoint. Granger, S. and M. Paquot (Eds.). 2012: 15-30.
- Rundell, M.** 2012b. "It Works in Practice but Will it Work in Theory?" The Uneasy Relationship between Lexicography and Matters Theoretical. Fjeld, R. Vatvedt and J.M. Torjusen (Eds.). 2012. *Proceedings of the 15th EURALEX International Congress, 7-11 August 2012*: 47-92. Oslo: University of Oslo, Department of Linguistics and Scandinavian Studies.
- Sánchez, A. and P. Cantos.** 2011. e-Dictionaries in the Information Age: The Lexical Constellation Model (LCM) and the Definitional Construct. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011: 251-274.
- Spohr, D.** 2011. A Multi-Layer Architecture for 'Pluri-Monofunctional' Dictionaries. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011: 103-120.
- 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. Theoretical Challenges in the Transition from Lexicographical *p-works* to *e-tools*. Granger, S. and M. Paquot (Eds.). 2012: 107-118.
- Tasovac, T.** 2012. Potentials and Challenges of WordNet-Based Pedagogical Lexicography: The Transpoetika Dictionary. Granger, S. and M. Paquot (Eds.). 2012: 237-258.
- Verlinde, S.** 2011. Modelling Interactive Reading, Translating and Writing Assistants. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011: 275-286.

An Analysis of the *Oxford Bilingual School Dictionary: Northern Sotho and English* (De Schryver 2007)

Inge Kosch, *Department of African Languages, University of South Africa,
Pretoria, South Africa (koschim@unisa.ac.za)*

Abstract: The *Oxford Bilingual School Dictionary: Northern Sotho and English* (De Schryver 2007) is a welcome addition to dictionaries that have been compiled for school use in particular. Its novelty and appeal lie in the fact that the lemmas and Northern Sotho mini-grammar are based on a corpus of general language usage and school textbooks. It reflects current language usage and concerns in the modern world that learners can easily relate to, making it a popular alternative to the more traditional dictionaries. This bidirectional, bilingual dictionary is equally useful to native speakers of Northern Sotho learning English and to English-speakers acquiring Northern Sotho. Though the number of lemmas is restricted to 5 000 in the lemma lists for Northern Sotho and English respectively, the compilers nevertheless succeeded in meeting the basic lexicographic and grammatical needs of the learner. This review article aims to take a critical look at various features of the dictionary.

Keywords: SESOTHO SA LEBOA, CORPUS, SCHOOL DICTIONARY, DICTIONARY CULTURE, LEMMATISATION, ACCESS ALPHABET, CROSS-REFERENCES, LEMMA, MINI-GRAMMAR, MODERNITY

Samevatting: Die *Oxford Bilingual School Dictionary: Northern Sotho and English* (De Schryver 2007) is 'n welkome toevoeging tot woordeboeke wat spesifiek vir skoolgebruik saamgestel is. Wat die woordeboek oorspronklik en aantreklik maak is die feit dat die lemmata en Noord-Sotho mini-grammatika op 'n korpus van algemene taal en skooltekste gebaseer is. Dit weerspieël hedendaagse taalgebruik en sake van belang in die moderne leefwêreld waarmee die leerder maklik kan assosieer en wat die woordeboek 'n gewilde keuse maak bo ander meer tradisionele woordeboeke. Hierdie tweerigting-, tweetalige woordeboek voorsien in die behoeftes van Noord-Sotho moedertaalsprekers wat Engels aanleer sowel as aan dié van Engelssprekendes wat Noord-Sotho leer. Ten spyte daarvan dat die lemmata beperk is tot 5 000 vir elk van die twee tale, het die samestellers nogtans daarin geslaag om aan die basiese leksikografiese en grammatiese behoeftes van die leerder te voldoen. Hierdie resensieartikel neem die kenmerke van die woordeboek krities in oënskou.

Sleutelwoorde: SESOTHO SA LEBOA, KORPUS, SKOOLWOORDEBOEK, WOORDEBOEKPRAKTYK, LEMMATISERING, TOEGANGSALFABET, KRUISVERWYSINGS, LEMMA, MINIGRAMMATIKA, NUWERWETSHEID

1. Introduction

The *Oxford Bilingual School Dictionary: Northern Sotho and English* (De Schryver 2007) has not without reason been awarded the esteemed SATI (South African Translators' Institute) prize in 2009. The user-friendly approach and layout in the two bilingual central lists together with a study section (which includes a corpus-based mini-grammar for Northern Sotho) and a reference section have set a benchmark for other African language school dictionaries to follow suit, such as the *Oxford Bilingual School Dictionary: Zulu and English* (De Schryver 2010). It plays an important role in filling a gap as a bidirectional dictionary for native speakers of Northern Sotho learning English and English-speakers learning Northern Sotho. This review article aims to take a critical look at various features of the dictionary based on the 3rd impression (2010) of the first (and thus far only) edition (2007). Errors that are noted should not be seen as criticism, but as contributions to the improvement of this work in a future revision cycle.

2. Fit for purpose

Gouws and Prinsloo (2005: 39) emphasize that "[t]he 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". In the case of school dictionaries it is particularly important to pitch dictionaries at the level of learners and not to produce scaled down versions of more comprehensive dictionaries as these "seldom meet the needs of the target users and often are more of a hindrance than a help" (Potgieter 2012: 262).

The dictionary under review is a perfect example of a reference work inspired and driven by the needs and dictionary consultation skill levels of school learners as a well-defined target user group (Grade 4–9). Having established the general level of dictionary education among the prospective target users in South African schools, the compilers designed this dictionary not only to serve as a source of information, but also as an educational tool in support of the development of a dictionary culture among learners. This paperback edition of 608 pages, measuring 210 x 150 mm, easily fits into any schoolbag. The binding appears to be robust enough to withstand rough and frequent handling by learners. Nothing spoils a user's look-up activities more than a dictionary that falls apart after just a few consultations.

Mindful that the dictionary is aimed at school learners, the compilers do not confront the user unnecessarily with linguistic technicalities or lexicographic terms. For example:

- The distinction between *verbs* and *verb stems* is not strictly maintained. The distinction would be important to a grammarian, but for the target

user of this dictionary it suffices to know that prefixes are attached to *verbs* (as explained in the study section, page S21), although strictly speaking they are attached to *verb stems*.

- Instead of phonetic transcriptions, the compilers have opted for what they call 'pronunciation fields' as a guide to correct pronunciation. This term is self-explanatory and thus easily understood by the learner. The pronunciation field slot is only filled in cases where mid-low vowels are involved (marked by means of a circumflex), which are pronounced differently to their unmarked mid-high counterparts.
- For a lexicographic term like 'lemma' a much more descriptive word like 'headword' is used throughout the dictionary.

On occasion the dictionary is referred to as *pukuntšu(ng) ya gago* or *your dictionary* (cf. S16 and S23 respectively in the study section). This personalises the relationship between the user and the dictionary and no doubt adds to the user-friendly approach adopted in the dictionary.

3. Outer texts

The title of the dictionary is formulated differently on the outside cover¹ and the first title page², which leaves one a little at a loss as to the correct way of referencing the dictionary. For the purpose of this discussion the title *Oxford Bilingual School Dictionary: Northern Sotho and English* was decided on, abbreviated as ONSD.³

There are three outer texts presented in both Northern Sotho and English.⁴ They are easily recognised by the grey marking along the edges of the pages. The front matter of 12 pages contains the table of contents, dictionary features and an introduction. The dictionary features (pp. iv-vii) can be grasped at one glance as they are presented diagrammatically. Unfortunately (and understandably so) the readability of this section had to be compromised by small print to fit in all the features. This might not be received well by a user who is in a hurry to have his/her look-up query answered and does not have time to first study the small print in the diagrams.


The middle matter text positioned between the Northern Sotho–English (NS–E) and English–Northern Sotho (E–NS) lemma lists comprises a study section of 28 pages (marked as S1–S28) including dictionary exercises in Northern Sotho and English for the inexperienced user. The activities are typical group activities, reminiscent of the outcomes based education (OBE) model. The activities encourage the students to collaborate, so that those in the group with (hopefully) some prior exposure to dictionary consultation would be able to assist the novices. Apart from useful templates on how to write e-mails, letters, etc., the study section also contains a corpus-based Northern Sotho mini-grammar in both Northern Sotho and English. The study section does not inter-

ferre with the numbering of the central list, i.e. it is wedged between the NS–E wordlist (pp. 1-254) and the E-NS wordlist (pp. 255-552). The back matter comprises a reference section of 16 pages (marked as R1-R16). It includes images and names of animals, fruits and vegetables, the human body and sport. Furthermore, the provinces and official languages of South Africa, terminology pertaining to the school curriculum, numbers, weights and measurements all constitute very useful information for school learners.

4. Corpus-based examples

A revolutionary feature of the dictionary and a first for an African language, is that it was "fully developed from a corpus, using the most modern, internationally accepted dictionary-development processes" (TshwaneDJe Blog).

The authenticity of example sentences was endorsed by a team of experts and mother-tongue speakers of English as well as Northern Sotho. The corpus from which lemmas were culled was based on general language texts and school textbooks. Receiving the SATI award for the dictionary on behalf of the team on 2 October 2009, Mamokgabo Mogodi explained how she decided on the example sentences: "For each word in the dictionary, I first studied hundreds of lines of text from the corpus. From this I saw how a word was really used and I could work out the different meanings it could have. For each meaning I then selected an authentic example sentence" (TshwaneDJe Blog). Despite this careful selection of example sentences, the meaning of a word represented by a lemma is not always deducible from the example sentence. The following serves as an example:

kadijela /kadijêla/*noun* 9/-  veteran ♦ Mna. Mokgobu ke yena **kadijela** ge go etla go tša politiki. *Mr Mokgobu is a veteran when it comes to politics.* (p. 81)

The word 'veteran' is not a focal point in the example sentence and the learner may not necessarily be able to derive from the sentence exactly what a veteran is. Potgieter (2012: 268) emphasises that sentences which clearly illustrate the use of a word "are of a much higher quality and are of much more use to the learners" than sentences in which the word represented by the lemma appears in a non-focal position. In defence of the compilers of the ONSD, however, one needs to take cognisance of the fact that example sentences had to be extracted from the existing corpus. The compilers' point of departure was that no example sentences were going to be fabricated and thus they had to make do with the best possible contexts.

Many words used in the example sentences are not included as lemmas. This, however, should not be viewed as a drawback but rather as a beneficial "extra" as the learner's active vocabulary is reinforced during look-up activities. As pointed out by Potgieter (2012: 264) it is important that example sentences

be selected carefully to "fit into the learners' world", no easy feat for the compilers of the ONSD, given the fact that they had to cater for the needs of two different language groups with divergent backgrounds.

In the E-NS section some English example sentences required a modification rather than a simple translation into Northern Sotho. Such modifications are well executed, a case in point being the lemma 'prefix' in the E-NS section (p. 452). In the English example sentence, 'prefix' is illustrated with reference to 'un-' in 'un-happy', but in the Northern Sotho example sentence the term is explained with reference to the prefix of a noun, such as 'di-' in 'di-kgoši' (chiefs), seeing that Northern Sotho does not form antonyms of adjectives by prefixation as English does in the case of 'un-happy'.

The example sentences and illustrative phrases are by no means stereotypical and depict an interesting array of topics in various fields, such as language, arts and culture, history, society, economy, science and politics. Through this the pedagogical function of the dictionary is well served and learners are afforded the opportunity to increase their general knowledge about a range of issues, apart from discovering the translational equivalent of their search word. The following may serve as illustrations:

Language: (NS-E)

p. 201 **segalo**: In addition to establishing that 'segalo' is *inter alia* translated as 'tone' in English, the user learns that Endemann was the first author to have realised the importance of tone in Sepedi.

Arts and culture: (E-NS)

p. 432 **opera**: The user not only learns that the English word 'opera' is also an acceptable term in Northern Sotho ('opera'), but that "La Bohème" is an opera composed by Puccini.

5. Modernity

One of the criteria by which the quality of a dictionary is judged according to MacMillan (in Landau 1989: 308) is modernity. "Ensuring that this dictionary reflects current usage and concerns" (De Schryver 2007: x) is listed as one of the key concerns of the ONSD. The lemma list and illustrative phrases thus include many current and most frequently used words of the modern world that have become part and parcel of the Northern Sotho vocabulary. Through this the dictionary is fulfilling its social role, unlike older dictionaries, most of which "were compiled in the context of the limited role the African languages played" (Chabata and Nkomo 2010: 73) and which have been found to be "limited in scope, perspective and function and hence less effective now that the languages are being assigned a greater social role" (ibid.).

The ONSD has not shied away from including loanwords as lemmas or

using them in illustrative sentences to reflect the speech of current Northern Sotho speakers. Some of these words include the following:

aphili (appeal), *profense* (province), *projeke* (project), *rekoto* (record), *pasetše* (passage), *rasiti* (receipt), *websaeteng* (on the website), *dipotfolio* (portfolios), *faki* (from Afrikaans 'vaatjie', i.e. barrel (of oil)).

It will be observed that many of these loanwords display consonant sequences that do not feature in indigenous words. However, in loanwords "foreign" combinations like *pr*, *kr*, *ft* and *bs* are acceptable and there is no need for them to be separated by vowels to achieve the typical open syllable structure of the African languages (Departmental Northern Sotho Language Board 1988: 22). Nevertheless, in some loanwords consonant sequences are regarded as "wrong", compare the lemma *praebete* where the user is referred to the "correctly spelt" lemma *poraebete* (p. 188).

Since a measure of flexibility can be expected when loanwords are formed, some words were found to appear with alternative spellings, for example *sesepe* and *sešepe* 'soap' (p. 67) (from Afrikaans 'seep'); *obarolo* 'overall' and its plural *dioborolo* (p. 436); *gempe* (p. 520) or *hempe* (p. 522) 'shirt' (from Afrikaans 'hemp').

In a number of cases loanwords have been used in the example sentences in the E-NS section, even though an indigenous word is available, for example

p. 353 English: He will fill the hole with **sand**.

N.Sotho: *O tla tlatša molete ka **santa*** (instead of *mohlaba*).

p. 489 English: Mpho and her **sister** look similar but they are not twins.

N.Sotho: *Mpho le **sesi** wa gagwe ba a swana, eupša ga se mafahla*.

(There are quite a number of possible terms for "sister" in Northern Sotho; however, the distinction between them is not of any relevance here and **sesi** functions as a convenient generic term in the translated sentence.)

6. Lemmatisation approach

Two main lexicographic traditions have evolved for Northern Sotho, namely the arrangement of entries according to stems on the one hand and words on the other. The stem approach was hailed by Ziervogel and Mokgokong (1975: Preface, p.87) as the only scientific one. In this approach words are lemmatised according to the first letter of the stem, while the prefix (if present) is disregarded, for example, the word *letolo* (knee) is lemmatised as *-tolo* under the letter T. The advantage of this approach is that lemmas which are inflectionally or derivationally related can all be accommodated under the same lemma, thus saving space and affording the user insight into how the words are related grammatically and semantically. The user will, for example, find *mmušo* 'gov-

ernment' and *pušano* 'democracy' under the lemma *-buša* 'reign' from which these words have been derived. The disadvantage of this method is that, unless the dictionary also lemmatises words like *mmušo* and *pušano* with a cross-reference to *-buša*, the less experienced user will be at a loss as to where to find *mmušo* and *pušano*. The stem approach is perceived as somewhat superior to word lemmatisation (Prinsloo 2009: 155), possibly because it serves the needs of the expert researcher. However, the stem approach is user-unfriendly and comes at a price for the novice in terms of accessibility and the time it takes him or her to find the required information.

The compilers of the ONSD have adopted the word-based approach for their macrostructural presentation. This statement needs to be qualified, however, since it does not apply to verbs: verbs are still entered as stems and not as words. Lemmatisation according to stems is the best option for verbs, according to Prinsloo (2009: 156).

The word-based approach is the most user-friendly approach for learners and optimizes their chance of finding required items. Even the general public with basic dictionary consultation skills will find the book accessible as no knowledge about the morphological structure of words is expected. They simply look up words in alphabetical sequence according to the first letter of the word, which is normally also the first letter of the prefix in the case of nouns. The needs of users who are unaware of the full forms of words and who consequently search for words under truncated spoken variants with zero prefixes, are also catered for. The truncated forms are listed with a cross-reference to their full forms, e.g.

p. 10 **bakeng** see SEBAKA

p. 67 **hlare** see SEHLARE

Lemmatisation according to the first letter of the word circumvents the problem of novices having to know the morphophonological rules of the language. Irregular nouns, for example, can be found in their expected alphabetical slots:

p. 141 **mmušo** 'government': Learners simply look up this word under 'M'. They do not have to know that this noun was formed from **mo-bušo* in order to be able to look up its meaning under 'B', the initial letter of the stem *-buša* 'rule, reign'.

An unfortunate effect of the word-approach is that learners are deprived of the opportunity to see the relationship between words which are listed at a distance from each other under different article stretches (cf. *mmušo* and *pušano*, under M and P respectively). A further disadvantage is that entries by words take up much more space than entries by stems and in the case of the ONSD the compilers had to limit themselves to 5000 of the most frequent lemmas in both the Northern Sotho and English section. A problem (germane to all printed dictionaries) is that the lemma list can impossibly include the complete

inventory of words in a language. The same restriction obviously does not apply to electronic versions of dictionaries. For further information about the nature, advantages and disadvantages of stem versus word lemmatisation, the reader is referred to Prinsloo (2009) who discusses these matters at length.

7. The nature of lemmas

Lemmas were selected on the basis of their frequency of occurrence in the corpus and limited to 5 000 items each for the two languages. An innovative feature is the star-grading applied to each lemma list, according to which the top 500 items are marked with three stars, the next 500 most frequent items with two stars and the third 500 most frequent items with one star. Not only words, but items larger than single words (compound words and phrases) as well as items smaller than words occur as lemmas (morphemes such as subject concords and object concords) — no doubt very useful to the beginner not yet able to distinguish between full words and parts of words. Apart from lemmas, there are also sub-lemmas, mostly confined to locative forms in the case of nouns (e.g. ▶ **nakong** under **nako** 'time') and relative or plural forms in the case of verbs (e.g. ▶ **bonago** and ▶ **bonang** under **bona** 'see').

8. Access alphabet

Whereas Ziervogel and Mokgokong (1975) found it practical to use a combination of the standard alphabet and other article stretches (e.g. bj, kg, tš, etc.), the ONSD follows a strict alphabetical ordering in its access alphabet, which is most convenient for users who do not have advanced dictionary skills to deal with these types of exceptions to the regular alphabet. Where diacritics are involved as in the case of *s* and *š*, the diacritic sign does not upset the alphabetic ordering, compare for example

p. 197 seaparo 'piece of clothing'
šeba 'eat on the side'
sebaka 'chance, opportunity'

Invariably, it may happen that two words are written exactly the same except for one diacritic sign. The typical principle applied here is that the unmarked form always precedes the marked form (cf. Gouws and Prinsloo 2005: 98), for example

p. 197 seba 'whisper'
šeba 'eat on the side'

The complete absence of the letters C, Q and Z in the Northern Sotho access alphabet simply conveys to users that the language does not contain words

starting with these letters. The combination of a capital and a small letter to mark the beginning of each article stretch, for example **Aa**, **Bb**, etc. is a most welcome eye-catching feature enabling learners to find their way quickly to the desired place in the macrostructure of the dictionary. Besides this, the alphabetic letters in grey shading at the edges of the pages provide additional guidance. Guide words at the top of each page, indicating the first and the last lemma on that particular page, are very useful.

Multiple word lemmas follow a strict alphabetical order. For this reason, a word like 'beard' precedes 'bear', because the latter is entered as a two word lemma 'bear fruit', in which the letter 'f' of the second word appears later in the alphabet than the final letter 'd' of 'beard'. The same principle applies to the lemmas 'theme' and 'them (in particular)' where the latter is regarded as appearing later in the alphabet compared to 'theme' because the word which follows 'them' starts with an 'i'. Because of the above system of alphabetisation, the entry 'we' is separated from 'we (in particular)' by 18 other unrelated lemmas (cf. p. 542/3).

9. Shaded information boxes

Shaded information boxes are devices that are commonly used as article-internal microstructural entries. Gouws and Prinsloo (2010) aptly summarise the nature and function of text boxes as "salient dictionary entries ... used to place more than the default focus on a specific data item". Text boxes provide example-specific guidance on matters such as pronunciation, restrictions on the range of application and contrast of related words.

Shaded boxes have been used judiciously in the ONSD and do not come across as intrusive. They highlight potential areas of difficulty, for example, in the NS-E section, the user is alerted to the fact that the subject concord is not translated when it appears together with its subject in a sentence. Text boxes also help to avoid common mistakes in pronunciation and stress. As far as vowels and diphthongs in English are concerned, the rhyming method is aptly used in the E-NS section. For example, 'bow' (p. 282) is explained as rhyming with 'no', in order to distinguish it from the word 'bow' (as in 'bow down'). Stressed syllables of certain items are marked in italics in shaded boxes to make the user aware of the role of stress in changing an item's syntactic category, e.g. '*perfect*' (adjective) versus '*perfect*' (verb) (p. 443). Stress has not been pointed out consistently, though — compare 'rebel' (p. 464) and 'record' (p. 465), for example, which can each function either as a noun or a verb depending on which syllable is stressed. In other instances again, shaded boxes are used to alert users to common errors in the choice of words, for example non-mother tongue speakers of English often confuse 'borrow' and 'lend' because the same term is used for both in Northern Sotho (*-adima*). The dictionary compilers, anticipating this confusion, have therefore done well to draw the user's atten-

tion to this difference in a shaded box under the lemmas 'borrow' as well as 'lend'.

Comment boxes would have been useful regarding discrepancies observed between the description of certain items as lemmas on the one hand and the way they feature in the example sentences of other lemmas on the other. The numerals for 'six' and 'seven' are a case in point: The structure of 'six' (p. 491) is given as '[DEM + SC +] selelago' or '[DEM + SC +] tshelago' (verbal relative construction). However, different constructions are observed in some of the example sentences, e.g. *dikgwedi tše tshela* (p. 309), in which the SC and relative suffix *-go* are absent. In other instances only the SC is absent, e.g. *dibeke tše selelago* (p. 494). The numeral for 'seven' also employs the verbal relative construction, namely '[DEM + SC +] šupago' (p. 484), but under the lemma 'shed' (p. 485) a possessive construction is used, i.e. *mengwaga ya šupa* 'seven years'. The learner would have benefited from additional information regarding the possible alternative structures.

10. Standard spelling

A practical feature of the dictionary is that words which tend to differ from the standard orthography and spelling rules for Northern Sotho have been included. Users likely to look up words under the wrong spelling will find a cross-reference to the entry with the correct spelling, for example

- p. 12 **bašimane**: correct spelling = **bašemane**
- p. 12 **baswa**: correct spelling = **bafsa**
- p. 15 **Bibele**: correct spelling = **Beibele**
- p. 27 **bontšhi**: correct spelling = **bontši**
- p. 32 **bowa**: correct spelling = **boa**
- p. 67 **hlalosa**: correct spelling = **hlaloša**
- p. 69 **hlokafala, hlokagala**: correct spelling = **hlokofala**
- p. 70 **hlokomelo**: correct spelling = **tlhokomelo**

11. Cross-references

The compilers were compelled to enter many nouns in their singular as well as plural forms as separate lemmas, because their inclusion was dictated by frequency of occurrence in the corpus. This "double listing" obviously takes up extra space, but the problem is addressed by means of cross-referencing and by only supplying an example sentence under one of the forms, usually the most frequently occurring lemma, for example

p. 106 **lebitla** * *noun* 5/6 (*pl.* **mabitla**) ☐ grave
◆ Banna ba tla tsogela go epa **lebitla** leo. *The men will get up early to dig that grave.*

p. 126 **mabitla** *pl. noun* 5/6 SEE sg. LEBITLA

Where a form acts as a plural form for two singular words with different meanings, the user is referred to the singular forms, for example

p. 38 **diphiri** *pl noun* 7/8, 9/10 See sgs. SEPHIRI, PHIRI

Some cases were observed where a cross-reference is made to one singular form only, although two singular forms exist. A case in point is the lemma 'dinoko' (p. 37), where the user is only referred to the singular form 'senoko' in class 7 with the meaning 'syllable' (p. 206), although 'dinoko' is also the plural form of a class 9 noun 'noko' with the meaning 'porcupine' (cf. p. 169). One can only surmise that the omission of the cross-reference of 'dinoko' to 'noko' was not considered by the compilers, because 'dinoko' referring to 'porcupines' did not occur in the 5 000 most frequent lemmas.

12. Suggested corrections

It is hoped that the errors and inconsistencies pointed out in the appendix will be useful should a revision of the dictionary be undertaken. The corrections have been grouped into four broad categories, namely inconsistencies in fonts and punctuation, typographical errors, translational inaccuracies and grammatical errors.

13. Conclusion

The compilers of this dictionary deserve great appreciation for their innovative approach. Just like the dictionary itself, the mini-grammar is also based on a corpus, constituting a first for African languages. What makes the dictionary particularly appealing as a school dictionary is that it is not rule-oriented. This means that all items can be accessed as lemmas and that there is no need for the user to first become acquainted with a set of derivational rules, which need to be applied to strip a word down to its stem, before it can be looked up (cf. Prinsloo 2009: 153-154). Prinsloo (2009: 157) points out that "learners generally lack sufficient knowledge of the morphology of verbs to isolate the verb stem". This dictionary is bound to make great inroads into schools towards popularising the acquisition and fostering of a deeper knowledge of Northern Sotho in the future. In the words of Landau (1989: 310) a dictionary is "an achievement of the intellect, and few works of the intellect are more useful to so many people over such a protracted period of time".

Acknowledgement

I wish to thank one of the anonymous reviewers for valuable suggestions and references.

Endnotes

1. Oxford Sesotho sa Leboa–Seisimane; English–Northern Sotho Pukuntšu ya Sekolo/School Dictionary.
2. Pukuntšu ya Polelopedi ya Sekolo: Sesotho sa Leboa le Seisimane/Oxford Bilingual School Dictionary: Northern Sotho and English.
3. This abbreviation was followed in emulation of Prinsloo (2009).
4. For an in-depth discussion of the outer texts, consult Chabata and Nkomo (2010).

References

- Chabata, E. and D. Nkomo.** 2010. The Utilisation of Outer Texts in the Practical Lexicography of African Languages. *Lexikos* 20: 73-91.
- De Schryver, G.-M. (Ed.).** 2007. *Oxford Bilingual School Dictionary: Sesotho sa Leboa–Seisimane, English–Northern Sotho*. First Edition. Cape Town: Oxford University Press Southern Africa.
- De Schryver, G.-M. (Ed.).** 2010. *Oxford Bilingual School Dictionary: Zulu and English*. First Edition. Cape Town: Oxford University Press Southern Africa.
- Departmental Northern Sotho Language Board.** 1988. *Northern Sotho Terminology and Orthography No.4*. Pretoria: Government Printer.
- 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.** 2010. Thinking out of the Box — Perspectives on the Use of Lexicographic Text Boxes. Available at: http://www.euralex.org/elx_proceedings/Euralex2010/034_Euralex_2010_2_GOUWS%20PRINSLOO_Thinking%20out%20of%20the%20box_perspectives%20on%20the%20use%20of%20lexicographic%20text%20boxes.pdf. [Accessed on 23 September 2013.]
- Landau, S.I.** 1989. *Dictionaries: The Art and Craft of Lexicography*. Cambridge: Cambridge University Press.
- Potgieter, L.** 2012. Example Sentences in Bilingual School Dictionaries. *Lexikos* 22: 261-271.
- McMillan, J.B.** 1949. Five College Dictionaries. *College English* 10(4): 214-221.
- Prinsloo, D.J.** 2009. Current Lexicography Practice in Bantu with Specific Reference to the 'Oxford Northern Sotho School Dictionary'. *International Journal of Lexicography* 22(2): 151-178.
- TshwaneDJe Blog.** <http://tshwanedje.blogspot.com/2009/10/tshwanedje-oxford-university-press.html>. [Accessed on 10 September 2013.]
- 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

1. Inconsistencies in fonts and punctuation

There is no consistent use of boldface for the description '+ pl. marker (ng)', cf. for example

bonang: '+ pl. marker (ng)' (p. 25) versus
dirang: '+ **pl. marker (ng)**' (p. 38)

- p. S15 Bold: moselakamanyi -go > **moselakamanyi** -go
- p. S15 Underline: *Banna ga ba age sekolo* > *Banna ga ba age sekolo*
- p. S15 Non-italic: ("*The students read English*") > ("The students read English")
- p. S15 Underline: ("They read English") > ("They read English")
- p. 331 **doctor**: delete question mark after '◇ group of doctors'.
- p. 443 **perform**: the font for 'Finders Keepers' in the Northern Sotho sentence should be non-italic, while the rest of the sentence (*ka phapošing ya borutelo*) should be italic.
- p. 447 **plot** (noun): the words 'ka botlalo' should appear in italics.
- p. R16 **Dikarabo, Mošongwana 5 e**: a question mark should be inserted, i.e. *Na o tseba go šomiša khomphutha?*

2. Typographical errors

- p. v dotlhalošo > ditlhalošo
- p. ix thlaloši > tlhaloši (two occurrences)
- p. 11 "... that the persons begin referred to ..." > "... that the persons **being** referred to ..." (two occurrences).
- p. 31 **botho**: one of the proper names is spelt in two different ways, i.e. **Thembi** and **Tembi**. The same discrepancy is observed in the example sentence on p. 396 under **kindness**.
- p. 47 **ešo**²: the possessive concord 'ta' should be 'tša':
- p. 70 **hlola**²: the spelling of 'ntsi' should be 'ntši'.
- p. 82 **kamogelo**: 'ga bomogwera' should be '**gabo** mogwera'. The same applies to the example sentence under **welcome** on p. 543.
- p. 88 **kgatla**: the space between the stabilizer *e-* and *hwa* should be removed, i.e. the sentence should read: '... go fihlela e **ehwa**' (not: go fihlela e e hwa).
- p. 109 **lehlaodi**: 'hlaolašago' should be 'hlaološago'.

- p. 113 **lema**¹: 'masemo' should be 'mašemo'.
- p. 130 **makgatheng**: 'wene' (you) should be 'wena'. See also the same example sentence on p. 415 under **midst**.
- p. 208 **serati**: there are two spellings for the same noun, i.e. 'mmapelo' and 'mapelo'.
- p. 226 **tholo**: 'bjale ka' should be 'bjalo ka'. See also the same example sentence under **kudu** on p. 396.
- p. 233 **tlholego**: 'hlaloseša' should be 'hlalošetša'.
- p. S14 Righthand column: 'botši' should be 'bontši'.
- p. S15 Lefthand column: 'kopantšwego' should be 'kopantšhwego'.
- p. 261 **AIDS**: the noun 'kalafi' should be 'kalafo' for 'cure'.
- p. 358 **fortune**: 'lehuno' should be 'lehumo'.
- p. 360 **frequently**: the sentence *Ke atiša ho swarwa ke hloga ka morao go nwa beine* should be *Ke atiša go swarwa ke hlogo ka morago ga go nwa beine*.
- p. 367 **go out**: *e* and *tšwa* have been separated over two lines, instead of being written as one orthographic word, i.e. *etšwa*.
- p. 420 **mug**¹: there should be a space between 'poured' and 'some'.
- p. 482 **secure**: 'it hope I won't slip' should read 'I hope it won't slip'.
- p. R13 Righthand column: legorola la 6 > legoro la 6.
- p. R14 *Meetse ka letangwaneng ka metara o 1 go ya fase*: 'o 1' should preferably be 'o tee'.

3. Translational inaccuracies or omissions

Unfortunately, where an example sentence in the NS–E lemma list contains an error, it invariably also surfaces in the E–NS section if the same example sentence has been used in the article treatment of the English lemma.

- p. 5 **anega**¹: the translation of the example sentence should be: Grandpa is telling one of the old-time tales, instead of Grandpa is telling you of the old-time tales.
- p. 28 **bopša**: the translation of the example sentence should read ... three or five people (instead of ... three or four people).
- p. 41 **ditlaleso**: the English sentence contains the additional words 'last night'.
- p. 41 **ditona**¹: 'molemirui yo' should be translated as 'this farmer'.
- p. 78 **ithekgile**: the phrase 'bjalo ka mehleng' (as always) is not reflected in the English translation.

- p. 78/79 **itokišeditše**: 'lehono' (today) is not reflected in the English example sentence.
- p. 82 **kanegelong**: 'ya rena' is translated as 'the' instead of 'our'.
- p. 96 **kopanya**: 'mafoko' is translated as 'words' instead of 'sentences'.
- p. 120 **leswiswi**: the word 'batswadi' is translated as 'women' instead of 'parents'. This error also occurs in the English entry under **darkness**, where the same example sentence has been used (p. 318).
- p. 129 **mahlahla**: 'ba lapa gabonolo' is translated as 'don't get tired easily' instead of 'get tired easily'. The same example sentence is used on p. 340 under **energy**, but there the English and Northern Sotho versions both correctly indicate the positive form of the statement.
- p. 146 & 344 **mohlala** and **example**: the word 'dipotšišo' has been translated as 'question' instead of 'questions'.
- p. 170 & 503 **nonwane** and **story**: the example sentences are not translated correctly in one respect, i.e. 'bana' (children) is given in the Northern Sotho version, whereas 'us' is given in the English version.
- p. 200 **sefako**: the English sentence contains the additional phrase 'during the night'.
- p. 203 **seke**²: the translation of **ga/sa/se (...)** **seke** is in the positive, i.e. '(must) become clear', whereas it should be in the negative.
- p. 216 **šupile**: the English example sentence should read: They realized that he was **referring** to them ... (instead of ... that he was **referred** to them ...). See also the same sentence on p. 466 under **refer**.
- p. 222 **temošo**: 'beke ye' (this week) is given in the English example sentence as 'next week'.
- p. S16 Righthand column: '...mahlaodi ao a dirišwago **gantši** mo Sesothong sa Leboa ...'
- The above is the opposite of what its English translated text conveys on p. S23 (lefthand column), i.e. '... adjectives that are **not frequently** used in Northern Sotho ...'
- p. S17 The last sentence before the heading **Segalo** is not reflected in the English version of the mini-grammar, namely 'Ga go na dithotharo tša dikantšufelo tše dingwe tšeo e lego gona polelong ya Sesotho sa Leboa'.
- p. S21 Under the heading **Suffixes, prefixes and negative morphemes used with verbs** there are two sentences in the English version which are not reflected in the Northern Sotho mini-grammar, namely 'Other single and combined verbal extensions are used less often. Your dictionary will always show you the presence of verbal extensions, and a verb with extensions is always cross-referred to the stem or main verb from

- which it is derived (on the condition that that verb stem is frequent enough)'.
p. 333 **dress**: 'Take off that dress ...' should be 'Take off that dress **of mine** ...' in a strict translation of the Northern Sotho example into English.
p. 336 **each**: Under the second bullet the sentence 'Each of you needs to help' has been incorrectly translated into Northern Sotho as 'mang le mang wa lena o hloka thušo', meaning 'Each of you is in need of help'.
p. 338 **elsewhere**: the words 'study' and 'studying' have incorrectly been translated in Northern Sotho as 'ruta', which means 'teach'. The equivalent in Northern Sotho should have been 'ithuta'.
p. 364 **gender**: the word 'refers' should be replaced by 'referring' as in '... don't use language **refers** to gender or race' > '... don't use language **referring** to gender or race'.
p. 441 **passionate**: the example sentence in Northern Sotho should read 'Ke **na** le kgahlego ya mmino...'.
p. 471 **reservation**: the plural has erroneously been given as 'resemblances', instead of 'reservations'.
p. 508 **supply**: 'The farm ...' in the English version of the example sentence should be 'The farmer ...'.
p. 497 **speaker**: the English sentence does not include the locative 'in the church', while *kua kerekeng* features in the Northern Sotho counterpart. On p. 199 under **seboledi** the same example sentence is given, except here 'in the church' is included in the English version.

4. Grammatical inaccuracies

Some of these could also simply be typographical errors.

- p. 18 **bobotse**²: In the last Northern Sotho sentence the demonstrative pronoun in the adjectival construction should be 'bjo' instead of 'bo': *O na le boitshwaro bo bobotse > O na le boitshwaro **bjo** bobotse.
p. 98 **kudukudu**: the adjectival construction in 'pula e ntši' should have made use of the demonstrative of class 9, i.e. 'pula **ye** ntši'.
p. 149 & p.411 **moko** and **marrow**: the demonstrative of the class 1 noun 'motho' is given as 'wo' instead of 'yo'.
p. 150 & p.508 **molaodi** and **supervisor**: the possessive concord for 'molaodi' is given as 'ba' instead of 'wa'.
p. 184 **phethegile**: the subject concord has been omitted in the example sentence, i.e. 'Mošomo phethegile ...' instead of 'Mošomo **o** phethegile ...'.

- p. 202 **sehloa**: the passive form 'kgethetšwego' should be in the active form, i.e. 'kgethetšego'. Cf. also the same example under **climax** (p. 300).
- p. 228 **thutlwa**: the demonstrative 'wo' in the adjectival construction 'wotelele' should be replaced with 'yo' as the antecedent is a class 1 noun 'morwa' (son).
- p. 229 **tiile**: the ending of the verb stem in the negative should be *-e*, instead of *-a*, thus 'ga di tekateke' and not 'ga di tekateka'.
- p. 231 **tlalelwa** and **tlaletšwe**: The analysis of both is given as *verb + passive (w)*. For the latter the description *perfect (ile)* should have been added.
- p. 233 **tlhaka**²: a possessive concord should have been inserted, i.e. 'tlhaka ya ntlo ya gagwe ...'.
- p. 257 **acid**: there is an incorrect combination of noun and subject concord in the example sentence, i.e. it should either have been 'Seenywa se na le tatso ...' or 'Dienywa di na le tatso ...'.
- p. 314 **credit**: 'four years' in the Northern Sotho sentence is given as 'mengwaga ye ba bane'. This should be 'mengwaga ye mene'.
- p. 342 **estate**: 'polase ya kgolo' should be 'polase ye kgolo'.
- p. 361 **full** (first diamond): the example sentence in Northern Sotho should read: '... o ile wa rotoga godimo ga dithaba' instead of '... e ile ye rotoga godimo ga dithaba'.
- p. 364 **gender**: '... don't use language refers to gender or race' should be '... don't use language referring to gender or race'.
- p. 383 **illustrate**: This book is illustrated with lots of beautiful drawings: 'lots' is not reflected in the Northern Sotho sentence.
- p. 410 **many**: 'kgoši' is used with the subject concord of class 1 in the example sentence. On p. 92, however, the lemma **kgoši** is indicated as belonging to class 9/10 only; despite this 'kgoši' features with a class 1 subject concord on p. 92 in association with the proper name David. A comment box indicating that '(di)kgoši' can be used with either class 1/2 or class 9/10 concords would have been useful.
- p. 422 **nail**¹ (verb): the pronoun 'He' in the example sentence should be reflected in the Northern Sotho sentence as the subject concord 'O' of the 3rd person instead of 'Ke' (1st person).
- p. 482 **secure**: the subject concord for 'bana' appears as 'be' instead of 'ba'.
- p. 497 **speak**: the adjective construction should have employed the demonstrative **ye** instead of **e**, i.e. 'kanegelo yeo e lego ye kopana'.
- p. 500 **stalk**: the verb 'a fišitše' should be in the passive form, i.e. 'a fišitšwe'.

Hanks, Patrick. *Lexical Analysis: Norms and Exploitations*. 2013, 462 pp. ISBN 978-0-262-01857-9. Cambridge, Mass./London, England: The MIT Press.

Patrick Hanks is well known as a lexicographer and as the author of several remarkable articles on phraseology, collocations and co-occurrences and on the description of meaning. He was the chief editor, or one of the editors, of several dictionaries, some of which are highly original, particularly in their treatment of polysemy. Many people were hoping that he would eventually develop his views in a book, and this had been 'announced as "forthcoming" for many years'. 'Some people ... had given up hope that it would ever appear' (xv), but now, at last, after a period of preparation of sixteen years (215), what began as a 'disjointed collection of short essays and other fragments' has become 'a coherent text' (xv) of almost 500 pages.

The book is divided into thirteen chapters, five of which (chapters 3, 4, 5, 6 and 8) are based on previously published papers. They were all rewritten, but there is inevitably a certain amount of repetition that those readers who are already familiar with Hanks's work may find unnecessary but that will be useful to all those who are not. The book can be seen as a sort of hypertext, in which each point is first introduced in the general context of the theory and then more thoroughly developed in the following chapters. The present review reflects some of these repetitions.

Each chapter begins by a short (sometimes very short) abstract (except chapter 1 and chapter 13) and ends with a summary of the main points (except chapter 13). Oddly enough, there is no introduction. The Acknowledgments has the usual listing of the names of all those who helped Hanks develop his theory or write his text (Gilles-Maurice de Schryver played a decisive role in the encouragement of the author and also, apparently, in the shaping of the book), but it is mainly the story of how Hanks's ideas developed. He was influenced by John Sinclair and James Pustejovsky, and, to a lesser extent, by M.A.K. Halliday, Yorick Wilks, Charles Fillmore, Anna Wierzbicka and no doubt other linguists, but he is above all a lexicographer. He happened to be active when corpora became available for dictionary-making, when John Sinclair published his most important work and when prototype theory appeared in linguistics, and he was one of the first who believed that all three were important for lexicography and for semantics. His book is 'rooted in practical experience of monolingual lexicography' (xiv). His first job, from 1965 to 1970, was the edition of the Hamlyn *Encyclopedic World Dictionary* (1971), a dictionary that has been forgotten but was one of the sources of the *Collins English Dictionary* (1979), that Hanks also edited. He then worked with John Sinclair on the preparation of the *COBUILD Dictionary of the English Language* (1987), and that is when he began working with a corpus, 7 million words at first, then 18, and a lot more since then. In 1990 he moved to Oxford University Press, where he was chief editor of Current English Dictionaries. At the end of the decade he

produced the *New Oxford Dictionary of English* (1998; with a second edition called *Oxford Dictionary of English*), to this day the general dictionary of English in which the presentation of word meanings was most influenced by his ideas. In 2000 he worked with James Pustejovsky for a software company in Cambridge, Mass., and since the early 2000s he has been teaching at various universities: Brandeis, Berlin, Prague, Brno, Wolverhampton and Bristol. This, presumably, gave him more time for research and theory than he had ever enjoyed as a lexicographer, and the experience that he accumulated over the years eventually became the theory presented in this book (see also De Schryver 2010).

'Words and Meanings: the Need for a New Approach' (23 pages) is the first and certainly the most important chapter: it sums up the essence of Hanks's theory of norms and exploitations (TNE), 'a lexically-based, corpus-driven, bottom-up' (17) theory of language that matured after 'a lifetime of editing and writing dictionary definitions — attempting to account for the meaning of words, wrestling with the problem of word meaning' (7). Basically, the theory is simple: (i) words have normal, i.e. usual, typical, common, conventional, usually frequent, usages and meanings, to be distinguished from marginal, unusual, uncommon, atypical, non-conventional usages and meanings; (ii) any normal usage may vary in lexis, semantics and syntax while remaining basically the same; and (iii) the meaning of a word is determined by its environment, i.e. its valencies, its phraseologies, its collocations and more generally its context. The latter is not very original: many other linguists have said the same, in various ways, from Saussure to Firth to Sinclair, and it is also what Wittgenstein is remembered for in linguistics. Except that the linguist working with a large corpus has access to more contexts in less time, and can therefore draw conclusions that are more robust than those that were accessible before. And the first two points above are even more interesting: they are Hanks's conclusions after having worked with corpora that became larger and larger over the last three decades, and those conclusions were simply out of reach for linguists who did not have access to a corpus. Most preceding linguists did not have one, and it is not certain that they would have used it if they had, Hanks says (and he will say it again in chapter 12). Those who had one have tended to use it as a 'fish pond', in which to angle fish that fit their theory, while fish that did not were thrown back into the pond. Hanks, on the contrary, wants 'to find out what sort of fish are in the pond' (7): TNE is corpus-driven and bottom-up.

Most linguists have relied on their own intuitions, like Fillmore or Apresjan, 'a recipe for self-fulfilling prophecies' (20), Hanks says, or on the acceptability judgements of native speakers, like Mel'čuk or Wierzbicka (or Cruse, but Cruse is never mentioned, not even in the References), but that did not allow them to distinguish the normal from the more or less abnormal. Acceptability judgements distinguish between what is possible and what is impossible, but that is not the main distinction, Hanks says: what matters most is the distinc-

tion between normal and 'less normal' utterances. Worse, judgements on one's own linguistic usage or on the usage of others are unreliable: they tend to focus on what is cognitively salient, i.e. striking, unusual, and to ignore what is socially salient, i.e. frequent, common (21); all lexicographers know that. Linguistics has neglected the more common usages and given too much attention to rare, unusual, marginal usages, for several reasons: because there are many, as the study of large corpora has shown (18), because they are cognitively salient, and because they make it possible to explore the limits of language, what the code accepts and what it doesn't, an implicit objective of much of modern linguistics. 'The linguist's tolerance of abnormality is unusually great' (20), Sinclair once wrote, and the consequences have been dramatic, Hanks says: it 'left linguistics drowning in a welter of imagined possibilities' (4).

What does Hanks find in his corpus? He finds evidence for the three points above: that every content word has a frequent usage, or a few frequent usages, that can therefore be considered normal, and less frequent, less normal usages; that every normal usage shows more or less important variations in lexis, semantics and syntax, and can be played with to produce special effects — so that the language users have a double competence, a competence to use words according to the norm and a competence to play with the norm, to exploit it; and that every meaning corresponds to a specific context, that can be more or less frozen. Often, though not always, one usage is clearly dominant. An example (given in chapter 13): in the British National Corpus, *spoil* is used in the sense, and in the pattern, of 'spoil an event that should be enjoyable' 60% of the time, the other usages being much less frequent ('spoil a view' 18%, 'spoil a child' 11%, 'food spoils' 3%, 'be spoiled for choice' 3%, 'spoil a paper ballot' 1%) (427).

TNE is a theory, in the sense that it aims at describing how language works, with lexis at its centre, but it begins as a practice, a method for the study of corpora in order to extract meanings based on textual evidence, not on intuition or acceptability judgements, to determine which are dominant, and to describe them. How does Hanks proceed? He starts from a word (a notion discussed in chapter 2; see below) and notes all the contexts in which the word is used, and in what form (though the influence of the variations of form on the use of a lexical item is one of the points that could have been developed further in the book). Those contexts can be reduced to patterns, with all their important elements, other content words and function words, arranged in a syntactic form — TNE is not only lexically based, corpus driven and bottom up; it is also pattern based. The analyst then proceeds to isolate those patterns that illustrate the normal uses of the word, for which there are a substantial number of similar, if not exactly identical, examples in the corpus. At this early stage, all the other uses are left on the back burner: they are either performance errors, in which case they will be ignored, or exploitations of the norm, i.e. rare uses based on normal patterns but differing from them in a way that is not governed by any discernible rule (see chapter 8); those will be examined later. They are

potentially interesting, because they are the inventive, the bold, the playful, deliberate or inadvertent uses of journalists, of fiction writers, of poets and other actors of linguistic creativity, most of which will be forgotten but some of which will survive and become new norms, but they are not the first priority. When the normal patterns have been listed, the lexical analyst tries to determine the meaning of the word in each pattern, surely the most difficult stage of the process, in which intuition and previous knowledge of the language and of the world play a more or less important role. The results are in the form of a list of all the normal patterns of the chosen word, with a meaning corresponding to each pattern, or vice versa, arranged according to their 'degree of normality'. This is radically different from what can be found in lexical bases such as, for example, WordNet, and in traditional dictionaries, where meanings are attributed to isolated words (19), where the meanings of polysemous words are listed chronologically, or according to some 'logic', and where all meanings are given equal weight, regardless of their normality.

Of course, the operation is not as simple as it looks, and the following chapters examine some of the difficulties encountered by the lexical analyst. One of the first is the definition of 'normal' discourse: the usages that are retained for analysis must be authentic, but authenticity is not enough; they must also be natural (not stilted like many of the examples invented by linguists), usual, typical, common, conventional, frequent enough, and socially salient, i.e. mastered passively and actively by 'all' the users of the language community. Deciding what is entirely normal and what is less so may be difficult, because the boundaries are fuzzy. The main criterion is frequency: if a use is very frequent it must be considered normal. But a lower frequency does not designate a marginal use. For example, the Oxford English Corpus of 1.5 billion words has only six examples of *tell* in the sense 'The strain was beginning to tell', but Hanks decides that it must be considered a norm, because the meaning is distinct enough (16) — a case in which meaning, not frequency, is the basis of the decision.

The patterns that correspond to meanings reveal lexical sets, i.e. groups of words that can be used — and are used — in a particular position of a particular pattern with a word in a particular meaning. For example, the direct objects of *fire* in one of its meanings (*gun, rifle, pistol, revolver, machine gun*, etc.) are a lexical set. In another meaning, the objects of *fire* are human beings, more precisely employees, another lexical set. The words of a lexical set are united by a semantic type, and semantic types can be placed in an ontology: for example, the direct objects of *fire*¹ are firearms, firearms are artefacts, etc. (13). This will be developed in chapter 5.

Chapter 2, 'What is a Word?' (40 pages), explores the concept of 'word' together with the neighbouring concepts of 'type', 'token', 'lemma', 'lexeme', 'phrase' or 'multiword expression', and 'lexical entry'. The question is what constitutes a unit for a theory in which the lexicon plays the central role. The discussion does not end with a clear conclusion recommending terms or con-

demning others (29), but in the rest of the book Hanks uses mostly *word* and *lexical item*, though the latter is 'used in different ways by different writers' (29); *lexical unit* is used on page 389 to describe Fillmore's work but is not defined and is absent from the index. The chapter ends with a discussion of neology and of proper names that could have been shorter. Neology concerns mostly nouns, less often adjectives, rarely verbs and virtually never function words (42). The lexical types that are created are mostly terms (i.e. lexical items with a precise meaning that has been stipulated by the specialists of the domain) and multiword expressions, two categories that are not well covered in traditional dictionaries. The pages on proper names (33 *sq.*), though not absolutely necessary in the general architecture of the book, are interesting because proper names are rarely mentioned in books on semantics. Hanks is a specialist: readers may remember that he edited a dictionary of first names and a dictionary of surnames, and that the *Hamlyn* and the *Collins* that he edited both had proper names, contrary to the mainstream tradition of general dictionaries in the twentieth century in England. Proper names are numerous in many sorts of texts, and will be more and more numerous in monitor corpora: 'In some large lexical databases, aiming at full coverage of a language, over 70% of the lexical entries already are proper names, and this percentage continues to increase' (35). They cannot be neglected, because they carry information, real-world knowledge that is part of the competence in a language, and future lexicographers, Hanks argues, will have to include more of them and describe them more fully. This is not pursued in the rest of the book.

Chapter 3, 'Do word meanings exist' (19 pages), is a new version of a paper published in *Computers and the Humanities* in 2000 that has been widely regarded as an important contribution to the literature. Hanks argues that words do not have meanings in isolation, that they only have meaning potentials, and that these potentials are activated when the words are used. Meanings, as a consequence, are best seen as events rather than as entities: they take place in contexts of space and time. The meaning potential of a word is made up of semantic components — the 'semes' of other linguistic schools, but Hanks does not use the word. For many words, one semantic component is salient, i.e. it is activated most of the time, if not all the time, but for others 'different combinations ... are activated in different contexts' (83). Readers may think of the semantic portrait of *game* by Wittgenstein, but there are many other examples. The identification of these components and the exploration of the ways in which they combine in different contexts (82) are among the major tasks of lexicographers, and they have not done very well so far, again because they have tended to define words in isolation.

Hanks then returns to the relation between meaning and context. One question is how far the lexical analyst should cast his net when trying to determine the meaning of a word on the basis of its contexts of use. 'In the overwhelming majority of cases, a correct meaning can be assigned to a keyword on the basis of clues in its immediate environment' (81), but in other cases

it is necessary to consider a wider context, particularly for nouns (see also chapter 5). This is out of the reach of the corpus analyst working with a KWIC list but it can be retrieved more or less easily in all modern corpora. Normally, consideration of the context and of phraseology points to the right meaning, but the operation is not 'a magic bullet': it solves many problems of ambiguity, but not all (82). For example, the two meanings of *check* ('inspect' and 'cause to slow down or stop') are used in the same syntactic pattern, and in some instances of use they even co-exist (76), i.e. the same context allows the two meanings, not one or the other but the two together. Polysemy, Hanks concludes, is much more complex than its presentation as a flat list of distinct meanings in most dictionaries suggests.

Chapter 4, 'Prototypes and norms' (28 pages), examines phraseology, i.e. the more or less frozen patterns that help define a meaning. Patterns have been neglected by dictionaries in general, particularly by English dictionaries — German, Czech and Modern Greek have done better. Lexicographers have 'tended to focus on conventions of meaning and to neglect conventions of phraseology' (104), Hanks writes, perhaps because phraseology is more difficult to identify and describe. Yet patterns are important, as chapter 1 has made clear, because they are keys for the identification of meaning. They are what the language user has to analyze and interpret to give each word its right meaning and understand the message. The lexicon is 'a store of shared beliefs or meaning potentials, each of which is associated with one or more phraseological norms' (87). Patterns are what the lexical analyst sees in a corpus, and frequent patterns suggest common usages. 'Any sizable corpus will usually show a very large number of very similar uses of each word — similar not only in terms of syntactic construction but also in terms of preferred collocations', Hanks writes again (91). A norm is identified by grouping similar corpus lines together around a phraseological prototype; note the word *similar*, meaning that the phraseologies that are grouped together are not necessarily identical; they are formally different manifestations of the same basic pattern, and 'judgment is required to decide what counts as "similar"' (92). Once this has been done, the frequency of each pattern can be calculated, and that will help in the identification of normal usages: '... unusual uses are put on one side for later analysis ..., and then the number of corpus lines in each group is counted to discover the comparative frequency of each pattern in the sample' (92). As we have seen in chapter 1, some rare usages must also be 'recognized as patterns, despite their rarity, because they have distinctive meanings' (92; an example is given in chapter 7; see below). When a pattern has been identified as a norm, with its variations (alternations and exploitations), 'the next step is to associate it with a meaning' (95). Patterns normally serve to determine meanings, and, in some rare cases, meanings serve to identify patterns; the apparent contradiction is probably unavoidable.

The number of patterns associated with each word is highly variable. For 'most words, just one or two patterns are salient (socially salient, i.e. frequent),

while the other patterns are less common' (95), as we have seen for *spoil*. The verb *climb* has two basic meanings (and is presented accordingly in the *Oxford Dictionary of English*), 'clamber' and 'ascend', and a number of other usages that are much less common. Each basic meaning is associated with groups of normal subjects, normal objects and normal adjuncts, lexical sets that correspond to semantic types but play a syntagmatic role: 'Different lexical sets in different syntactic roles can alter the meaning of the target word' (105). These sets are fuzzy: for example, some words may belong to one set with one verb but not with another verb with a very close meaning. Lexical sets are 'prototypical in character' (105), with words that are more central than others, because they are found in the lexical sets of all, or almost all, related words.

Chapter 5, 'Contextual dependency and lexical sets' (31 pages), continues the description of the operation in which the analyst collects patterns, defines lexical sets for each syntactic slot and then determines the meaning of the central word for each pattern: the 'identification of normal complementation patterns by the corpus analyst, not only in terms of valencies but also in terms of lexical sets, is an essential step in determining a word's meaning' (113). The 'semantics of each word in a language is determined by the totality of its complementation patterns', Hanks writes (113), echoing Firth's pronouncement: 'You shall know a word by the company it keeps' (Firth 1968: 179). For a verb, the essential elements to consider are subjects, objects and adverbials, and for a noun they are the words that are typically found in its wider context, even those that have no syntagmatic relationship to it. Corpus evidence shows 'what patterns of usage are normal, central, and typical and ... which patterns are the most frequent'. They are those that should be retained for description: the corpus analyst is 'concerned with the regular and the normal, not with the boundaries of linguistic possibility' (115), Hanks writes again. But, as we have seen, the identification of regular patterns is only the first step in the identification of meaning: 'corpora provide direct evidence for patterns of usage, but only indirect evidence for meanings' (116). For many words, one pattern is highly dominant; *spoil*, again, or the verb *urge*: 61% of its uses are in the pattern of 'a person urging another person to do something', while 'a person urging a steed or another person onward or upward (or in some other direction)' accounts for only 3.5% (117). Dictionaries never say what meanings are dominant, and in what proportions; all meanings are given equal weight.

Hanks then returns to the construction of lexical sets. They must be assembled with great care, he says: if they are too narrow or if they are too broad they are not much use in the identification of meaning. For example, one can urge a horse on, or a stallion, or a camel (or, I suppose, an ox, an elephant, etc.). Or a car, the corpus says, in a few examples that are clearly an exploitation of the basic pattern. But even if one such example describes a Ford Sierra being urged on over difficult terrain by its driver it would be unwise to include *Sierra* in the lexical set of the nouns that can be objects of the verb *urge*, because then any name could be a member as well (118). Similarly, it would be point-

less to include *John* or *Sylvia* in the lexical set of the objects of the verb *fire* in the sense of 'fire an employee'.

Some semantic sets are unified not only by their semantics, the meanings of their words, but also by their 'semantic prosody', a concept introduced by Sinclair (1991) and Louw (1993) to designate the positive or negative connotation of some words. An example used by Sinclair is the verb *set in*: it has a marked preference for subjects such as *rot, decay, malaise, despair, decadence, impoverishment, infection, prejudice*, etc., all with a negative semantic prosody, and therefore *set in* can also be said to have a negative semantic prosody (124) — note the word *preference*, meaning that other subjects are possible, but marginal. For nouns, the analyst must identify 'statistically significant collocates in the environment of the target word', and as we saw earlier such collocates 'do not necessarily have to be in a structured relationship' (134) with the word. Incidentally, a large corpus reveals things that could not be noted before. For example, Hanks notes that in his corpus the word *spider* co-occurs significantly with the word *bath*, but this seems to be restricted to English; other languages do not show the same co-occurrence (135). Why? Is it because there are fewer spiders in other communities, or fewer baths, or because people are less afraid of spiders, or because foreign spiders do not like baths, or because a spider in a bath is so common in other countries that nobody mentions it, or for another reason? Such correlations in the wider contexts are potentially interesting for language teachers.

The identification of the relevant elements in each relevant pattern is not an easy operation: 'they have to be teased out, often painstakingly and slowly' (141). Hanks concludes that much remains to be done: 'procedures have to be developed for distinguishing relevant features from mere noise. Appropriate levels of generalization have to be chosen at every step, for every pattern of every word' (141).

Chapter 6, 'Norms change over time' (27 pages), starts from a well-known fact: word meanings change with time, so that we sometimes find it difficult to interpret a word used in an old text. In TNE these changes are part of the evolution of norms. To study them, the lexical analyst needs a historical corpus with 'examples of "everyday" texts ... as well as great works of literature' (145). That is not easy, because 'the work of great writers of the past tend to have survived, whereas the mundane, everyday use of a language ... has left fewer traces behind' (158). Interesting examples are the words *enthusiasm* and *condescension* as used by Jane Austen. Hanks has discovered that *enthusiasm* was slowly changing from negative to positive semantic prosody in the eighteenth century when Austen was writing, while *condescension* was moving in the opposite direction, from positive to negative. With TNE the analyst can try to reconstruct the system at a given period of a given dialect, with its primary norms, its secondary norms, its alternations and its exploitations, and to understand how the whole system evolved (see chapter 10). In passing, Hanks discusses a view of literature that has ancient origins but became popular for a

while in universities in the late twentieth century, according to which as soon as it has been written the text acquires a life of its own and belongs exclusively to the reader, not to the author. There is some truth in this, Hanks says, but in its extreme version it is a 'first step down a road leading to the absurdity of Humpty Dumpty's position' (a word means whatever I want it to mean), 'a solipsistic universe ... in which other minds, other worlds, learning, and scholarship — the whole social consensus of meaning in language — all count for nothing. In such a universe, language itself becomes meaningless' (156). How refreshing, for those of us who have seen the devastation produced in the minds of young students by some quasi-fanatical advocates of the more extreme versions of that approach.

Chapter 7, 'Three types of alternation' (38 pages), details the variations that can take place in normal phraseologies. Alternations are of three types: a pattern can vary in lexis (*clutching / grasping at straws*), semantic-type (*treating injured people / injuries / injured legs*, etc.) and syntax (*He broke the window / The window broke*) and yet remain the same pattern. Chapter 8, 'Exploitations' (40 pages), is about the distortions of a norm that 'create new meanings ad hoc and ... say old things in new ways' (211), that Hanks chooses to call 'exploitations'. The difference may have been unclear when Hanks started discussing the two phenomena, but it clarifies when he writes: 'alternations are regular, second-choice elements within an overall pattern, exploitations are typically dynamic, creative, or graphic choices within the boundaries of possible language use' (212). Exploitations are typically infrequent, i.e. not socially salient, but created to produce a strong effect, i.e. cognitively salient (214), so that the 'distinction between alternation and exploitation is in many cases one of frequency, coupled with semantic or rhetorical effect' (216). Indeed, frequency is not always enough, as we have seen. In one example (given in chapter 10), Hanks decides that 'The industry is scratching its head' is an exploitation of a normal pattern of the verb *scratch*, not an alternation, because the two main words (*industry* and *head*) show 'incoherence' (291). Of course, here again, as often in linguistics, there is no sharp division between the two phenomena, alternation and exploitation, because normality is a cline: 'Some uses of words are more normal than others' (214).

The chapter continues with a review of the main tropes that can be used to exploit a norm: metonymy, synecdoque, zeugma, oxymoron, understatement, euphemism, etc. Here Hanks gets a bit carried away by his enthusiasm (in the modern sense), as when he discusses cases of hyperbole, irony or sarcasm, whose role in the exploitation of a norm is not obvious (236). The use of tropes in the creative use of language brings Hanks back to the question of ambiguity that he has discussed in many of his publications, because it is one of the main points where he differs from other linguists: 'Ambiguities are plentiful in the literature of linguistics, because they are based on invented examples isolated from any real context of utterance, but corpus evidence shows that genuinely baffling ambiguity arises in remarkably few cases' (243). In most cases, the

context leaves no doubt as to which meaning is being used.

The study of exploitations continues in chapter 9, 'Intertextuality: Literature and the exploitation of norms' (32 pages), with an exploration of the role of noteworthy language users, famous authors among them, but also journalists, teachers, lawyers, etc., in the evolution of word meaning through their exploitations of norms, and how their creations often become norms in their turn. Most of the chapter is about Shakespeare and the Bible, of course, and there is a very good — though more important for lexicography than for lexical analysis — passage about the lexical innovations of James Joyce in *Finnegan's Wake* (275 sq.). Should they be included in dictionaries? What are the best criteria for inclusion *vs* exclusion? How can they be defined if there is only one occurrence, or even if they are used by only one author? The *Oxford English Dictionary* has a few of these *hapax legomena*, but it is difficult to see what criteria were used to select them and reject the others.

Chapter 10, 'Word and pattern meaning: A complex linguistic gestalt' (21 pages), contains 'portraits' of words of varying complexity for the lexical analyst, *scratch*, *throw* and a few others. A norm can give rise to an exploitation that can become a secondary norm and/or be used as a basis for another exploitation, and so on, creating a web of, in some cases, extreme complexity. This raises the question of how the language users store such complexities, and how they can pick the right interpretation when the word is used (298).

Chapter 11, 'Meaning, philosophy of language, and anthropology' (42 pages) and Chapter 12, 'The role of the lexicon in linguistic theory' (62 pages, by far the longest chapter), could have been placed earlier in the book, but undoubtedly they are the two chapters that students (and teachers) of linguistics will need to have read carefully and keep for future reference. In chapter 11, Hanks notes that 'many of the developments that have most fundamentally affected our understanding of meaning in language took place in philosophy of language and anthropology rather than in linguistics' (306), but then proceeds to describe the work of linguists such as Ogden and Richards, Bar-Hillel, Wierzbicka as well as that of philosophers and anthropologists, Aristotle, Wilkins, Leibniz, Wittgenstein, Grice, Austin, Rosch, Putnam, Kripke, and others. Now that large corpora are available, Hanks writes, containing plentiful evidence of language as it is normally used, 'an immense task of sifting lies ahead, to determine which linguistic hypotheses can be maintained satisfactorily and which must be modified or abandoned' (307). One problem with what philosophers have written about language and meaning is that they have failed to distinguish between 'meaning in naturally occurring language and the stipulated scientific meaning of a rigorously defined concept' (311; Hanks does not say how the definition of a scientific term can be rigorous if the words used in its definition have fuzzy meanings, but that is a minor point in the book). For most philosophers, and for many linguists, 'the vagueness and fuzziness of meaning of ordinary words ... was an imperfection' (313), and it was not until the 1960s that they 'began to realize that far from being a minor imperfection, [the] fuzzi-

ness of word meaning is a central design feature of natural language, contributing a flexibility that allows existing words to be applied to new situations and to be exploited in new and interesting ways' (336). Many linguists, and most philosophers of language have tried to describe an idealized language that did not suffer from the imperfections of real discourse. They 'spent enormous amounts of time and effort speculating about possible but implausible sentences' (307), failing to distinguish between possible usage and normal usage. Their work is 'strewn with examples of self-fulfilling theoretical prophecies, in which bizarre examples are first invented, then judged to be acceptable (according to the researcher's intuitions), and then presented as evidence for conclusions about some aspect of the nature of language or linguistic rules. However, bizarre examples are conducive to bizarre theories' (307).

In Chapter 12, Hanks continues the presentation of the work of his predecessors, only linguists this time, situating the theory of norms and exploitations in relation to other theories of language, past and present (347). Having noted that the importance of the lexicon has been underestimated by a vast majority of linguists until the late twentieth century, he reviews the work of Humboldt, Saussure, Trier (and other advocates of semantic field theory), Apresjan, Mel'čuk, Chomsky, Jackendoff, Meyer, Bresnan, Pustejovsky, Langacker (and other cognitivists), Fillmore, Firth, Halliday, Sinclair, Hoey, Stubbs, etc. He stresses, again, that many of them did not use, or refused to use, a corpus and therefore based their theories on invented examples that led them to debatable conclusions. One can feel his pleasure when he discusses, for example, Chomsky, acknowledging his importance but at the same time arguing that his theories are far removed from the reality of discourse: wrong objective (how to construct well-formed sentences), wrong method (introspection), constant hesitations and re-adjustments over the last fifty years, and eventually results that have not kept their promises in the clarification of how language works.

Hanks's review of the literature in those two chapters is good reading and will prove extremely useful. His culture is immense, as befits the editor of the monumental *Lexicology: Critical Concepts in Linguistics*, published in 2008, and his views are challenging because they are those of a linguist with strong convictions against which the work of his predecessors can be measured. Sometimes, however, one is left wondering how exactly the work being described relates to TNE, or even if it does at all. Could Hanks, again, have been carried away by his enthusiasm? How, for example, can Ogden and Richards' semiotic triangle be useful in the lexical analysis of a corpus? (329); How can Wittgenstein's well-known discussion of the meaning of *game* be useful to TNE? (325); How does Wierzbicka's treatment of *game* 'stand up in the light of corpus evidence'? (326); What use are Austin's notions of locution, illocution and perlocution for Hanks's work? (334), etc. No doubt Hanks could answer these questions, but I wish he had been more explicit in the book. Other (admittedly minor, in the context) questions are not answered: Why does Wilkins's work 'seem very odd to modern readers'? (314); Why has Leibniz's immense influ-

ence on thinking about language been 'often based on misunderstanding of his work'? (316); How exactly was Firth influenced by Malinowski? (335); etc.

Chapter 13, 'The broader picture' (21 pages), sums up the main points of TNE: Making 'predictions about probable usage is much more useful than speculating about the boundaries of possibility' (415); TNE is about 'rules for using words, rather than [about] rules for constructing sentences' (416); 'The normal meanings of a word can be extrapolated from statistical study of a large number of its uses' (410); 'The difficulty lies in achieving just the right level of generalization' (411); Most of the time, though not always, 'different patterns activate different meanings' (an exception is *hazard*, as in 'hazard one's life' and 'hazard a guess') (413). The final pages examine the fields in which TNE could be useful: the semantic web, natural language-processing, artificial intelligence, language learning, language teaching, and of course computational lexicography. Hanks mentions his own *Pattern Dictionary of English Verbs*, saying that 'at the time of writing (August 2010), approximately 12.5% of the *PDEV* is complete, after five years of work. At the current rate of progress if there is not a substantial injection of funds enabling the project to recruit a professional lexicographic staff, it will not be completed until 2045, when the author will be 105 years old' (427). Good luck, Patrick.

It is difficult not to be convinced when Hanks says that words have normal usages and meanings, that normal usages vary in lexis, semantics and syntax, that the meaning of a word is determined by its environment and that words in isolation only have meaning potentials. What could the unconvinced say? They could note that the method raises a chicken-or-egg question. In TNE, the meaning of a word is determined by its contexts, but, as we have seen in one or two instances, there are cases where Hanks has to admit that it is the presence of a word that makes it possible to interpret the context. A (marginal?) example: in 'Doctors treating Michael Gibson', the 'sense of the named entity ("Medical Patient") is activated in this context by the verb *treat*' (178). More precisely, the chicken-or-egg question is between semantics and syntax. Does meaning determine syntax, or does syntax determine meaning? Hanks argues that meaning is too vague and variable to be the basis of the operation but his analyses sometimes work the other way round.

One could also say that much of what Hanks says is not really new. This is partly true, but partly only. Hanks shows chapter after chapter that the analysis of word meaning in a large corpus has already uncovered many aspects of language use that had never been discussed, and that more can be expected in the future. One could even say that Hanks's theory solves only those problems of identification and definition of word meaning that could already be solved by more traditional methods, if not by pure and simple intuition. But, again, that is not really fair: many of Hanks's word studies end with conclusions that differ from what is on offer in traditional dictionaries. It is true that even after a careful study of all contexts, the linguist may be left with a residue of cases that resist analysis. Hanks himself readily admits that TNE does not always work

smoothly. He knows that in some cases the analysis of the corpus must be aided by 'introspection and a minimal amount of editorial art' (148). But this may be because the problem has no solution: 'in many cases, there is no single "correct" answer to the lumpers/splitters debate' (180); or it may be because the method still needs to be improved. Much remains to be done, Hanks writes, to find effective procedures for selecting relevant information in the context, to explain the 'rules' of semantic-type alternations and of exploitations, to discover the rules of ellipsis (198 *sq.*), etc.

One of the central characters in Hanks's book is the dictionary. Most of his mentions are to point at the weaknesses of existing general dictionaries: they fail to distinguish normal and less normal usages, they fail to adequately describe the contexts of use, they are weak on phraseologies, etc. But for him the objective of the general dictionary is clear: it is first of all to record common, frequent, normal usages. Most modern lexicographers will agree, not only in England: in the age of the electronic corpus, that is what they want to do, and that is what a dictionary is for, they say. Perhaps, but the users need most help on the marginal usages that are found in the work of past and present fiction writers, poets or journalists and in the discourse of specialists of a science or a technique. In fact, one wonders, apart from foreign students and the specialists of lexical databases and their various applications, who needs the more common entries in a dictionary. What does Hanks think? Can the same dictionary attain the two objectives, or do we need two types of general dictionary? Not to mention a third objective, telling the users not how most people speak and write but how people should speak and write, or how the best users, the *élite*, speak and write. Should this traditional type be abandoned? These questions are not central for Hanks in this book, but I still wish he had said more on them.

Lexical Analysis is remarkably well produced, as indeed one would expect from the MIT Press. I found only eight (small) errors in its almost 500 pages: 188, 226, 246, 274 (*Duchamps* should be *Duchamp*), 362 ('grincer les dents' should be 'grincer des dents', at least in my dialect), 371, 404, 423. Also, on page 40, the reader is referred to examples '(6)–(8)' that I could not find, probably a relic of the preceding version of chapter 2.

Lexical Analysis is a big book, it is not cheap but it is exceptionally good value. It is theoretical and practical, it is limpid and extremely subtle, it is highly personal and can also be used as a manual for students. It is not always easy to read, but the main ideas are repeated in different chapters and the message is clear. It is 'intended to serve as part of the foundation for future empirical research in many language-related disciplines: corpus linguistics, cognitive linguistics, computational linguistics, historical linguistics, philosophy of language, and language teaching, among others. It offers a new foundation for a variety of practical tasks such as computation of meaning in ordinary language, dictionary making, and textbook writing' (22). *Lexical Analysis* is indeed an important book, and Hanks is right to be ambitious: his book is a must for spe-

cialists of several disciplines. TNE is not only convincing; it seems obvious, and many people will regret that Hanks took such a long time before developing it and offering it in book form. Who could now envisage exploring and describing word meaning in the traditional way? *Lexical Analysis* makes us all want to try TNE, and no doubt many have already tried. Every semanticist, every lexicologist, every lexicographer will want to know how far the theory applies to his/her discipline and to his/her own language. It will be interesting to see how TNE influences semantics and how it changes the dictionaries and lexical databases of English and other languages in the near future.

References

- De Schryver, G.-M. (Ed.)**. 2010. *A Way with Words: Recent Advances in Lexical Theory and Analysis. A Festschrift for Patrick Hanks*. Kampala, Uganda: Menha Publishers.
- Firth, J.R.** 1968. A Synopsis of Linguistic Theory, 1930–1955. Palmer, F.R. (Ed.). 1968. *Selected Papers of J.R. Firth, 1952–9*: 168–205. London: Longman.
- Hanks, P.** 2000. Do Word Meanings Exist? *Computers and the Humanities* 34: 205–215. Also in Fontenelle, T. (Ed.). 2008. *Practical Lexicography: A Reader*. Oxford Linguistics. Oxford/New York: Oxford University Press.
- Hanks, P. (Ed.)**. 2008. *Lexicology: Critical Concepts in Linguistics*. 6 volumes. London/New York: Routledge.
- Louw, B.** 1993. Irony in the Text or Insincerity in the Writer? — The Diagnostic Potential of Semantic Prosodies. Baker, M., G. Francis and E. Tognini-Bonelli (Eds.). 1993. *Text and Technology: In Honour of John Sinclair*. Amsterdam: John Benjamins.
- Sinclair, J.M.** 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Henri Béjoint
Centre de Recherche en Terminologie et Traduction
Université de Lyon
Lyon
France
Henri.Bejoint@univ-lyon2.fr

Publikasieaankondigings / Publication Announcements

Willem Botha, Paul Mavoungou and Dion Nkomo (Editors). *Festschrift Rufus H. Gouws*. 2012, XIV + 299 pp. ISBN 978-1-920338-95-4 (Paperback). Stellenbosch: SUN PReSS. Price: R250.

Gilles-Maurice de Schryver (Editor-in-chief). *Pukuntšū ya Polelopedi ya Sekolo. Sesotho sa Leboa le Seisimane. E gatišitšwe ke Oxford / Oxford Bilingual School Dictionary. Northern Sotho and English*. 2007, xii + 552 + R1-16 pp. ISBN 978 0 19 576555 7. Cape Town: Oxford University Press South Africa. Price: R129.95. (Review article in this issue.)

Pedro A. Fuertes-Olivera and Henning Bergenholtz (Editors). *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. 2011, XIV + 341 pp. ISBN 978-1-4411-2806-5 (Hardback). London/New York: Continuum. Price: £76.50. (Review article in this issue.)

Sylviane Granger and Magali Paquot (Editors). *Electronic Lexicography*. 2012, XIV + 517 pp. ISBN 978-0-19-965486-4 (Hardback). Oxford: Oxford University Press. Price: £75.00. (Review article in this issue.)

Patrick Hanks. *Lexical Analysis: Norms and Exploitations*. 2013, XVI + 462 pp. ISBN 978-0-262-01857-9 (Hardcover). Cambridge, MA: The MIT Press. Price: \$60/£41.95. (Review in this issue.)

Mariusz Kamiński. *A History of the Chambers Dictionary*. 2013, XII + 297 pp. ISBN 978-3-11-031250-8 (Hardback). Lexicographica. Series Maior 143. Berlin/Boston: Walter de Gruyter. Price: €89.95.

Le Dictionnaire des Mwanas avec Papito et Jolica. J'apprends le Fang. 2013, 62 pp. (Hardback). Gabon: Éditions Unik Africa.

Le Dictionnaire des Mwanas avec Papito et Jolica. J'apprends l'Inzèbi. 2013, 62 pp. (Hardback). Gabon: Éditions Unik Africa.

Le Dictionnaire des Mwanas avec Papito et Jolica. J'apprends le Lembaama. 2013, 62 pp. (Hardback). Gabon: Éditions Unik Africa.

Le Dictionnaire des Mwanas avec Papito et Jolica. J'apprends l'Omyènè. 2013, 62 pp. (Hardback). Gabon: Éditions Unik Africa.

Le Dictionnaire des Mwanas avec Papito et Jolica. J'apprends le Yipunu. 2013, 62 pp. (Hardback). Gabon: Éditions Unik Africa.

Timothée Mukash Kalel. *Dictionnaire Kanyòk-Français.* 2012, XV + 768 pp. ISBN 978 99951-26-61-3 (Hardback). Kinshasa: Centre de Recherches Pédagogiques.

H.S. Ndinga-Koumba-Binza and S.E. Bosch (Editors). *Language Science and Language Technology in Africa: A Festschrift for Justus C. Roux.* 2012, XX + 362 pp. ISBN 978-1-920338-79-4 (Paperback). Stellenbosch: SUN PReSS. Price: R260.

Oskar Reichmann. *Historische Lexikographie. Ideen, Verwirklichungen, Reflexionen an Beispielen des Deutschen, Niederländischen und Englischen.* 2012, XII + 588 pp. ISBN 978-3-11-028255-9 (Hardback). *Studia Linguistica Germanica* 111. Berlin/Boston: Walter de Gruyter. Price: €139.95.

VOORSKRIFTE AAN SKRYWERS

(Tree asseblief met ons in verbinding (lexikos@sun.ac.za) vir 'n uitvoeriger weergawe van hierdie instruksies of besoek ons webblad: <http://www.wat.co.za>)

A. REDAKSIONELE BELEID

1. Aard en inhoud van artikels

Artikels kan handel oor die suiwer leksikografie of oor implikasies wat aanverwante terreine, bv. linguïstiek, algemene taalwetenskap, rekenaarwetenskap en bestuurskunde vir die leksikografie het.

Bydraes kan onder enigeen van die volgende rubrieke geklassifiseer word:

(1) **Artikels:** Grondige oorspronklike wetenskaplike navorsing wat gedoen en die resultate wat verkry is, of bestaande navorsingsresultate en ander feite wat op 'n oorspronklike wyse oorsigtelik, interpreterend, vergelykend of krities evalueerend aangebied word.

(2) **Resensieartikels:** Navorsingsartikels wat in die vorm van 'n kritiese resensie van een of meer gepubliseerde wetenskaplike bronne aangebied word.

Bydraes in kategorieë (1) en (2) word aan streng anonieme keuring deur onafhanklike akademiese vakgenote onderwerp ten einde die internasionale navorsingsgehalte daarvan te verseker.

(3) **Resensies:** 'n Ontleding en kritiese evaluering van gepubliseerde wetenskaplike bronne en produkte, soos boeke en rekenaarprogramme.

(4) **Projekte:** Besprekings van leksikografiese projekte.

(5) **Leksikonotas:** Enige artikel wat praktykgerigte inligting, voorstelle, probleme, vrae, kommentaar en oplossings betreffende die leksikografie bevat.

(6) **Leksikovaria:** Enigeen van 'n groot verskeidenheid artikels, aankondigings en nuusvystellings van leksikografiese verenigings wat veral vir die praktiserende leksikograaf van waarde sal wees.

(7) **Verslae:** Verslae van konferensies en werksessies.

Bydraes in kategorieë (3)-(7) moet almal aan die eise van akademiese geskrifte voldoen en word met die oog hierop deur die redaksie gekeur.

2. Wetenskaplike standaard en keuringsprosedure

Lexikos is deur die Departement van Onderwys van die Suid-Afrikaanse Regering as 'n gesubsidieerde d.w.s. inkomstegenererende navorsingstydskrif goedgekeur.

Artikels sal op grond van die volgende aspekte beoordeel word: taal en styl; saaklikheid en verstaanbaarheid; probleemstelling, beredenering en gevolgtrekking; verwysing na die belangrikste en jongste literatuur; wesenlike bydrae tot die spesifieke vakgebied.

3. Taal van bydraes

Afrikaans, Duits, Engels, Frans of Nederlands.

4. Kopiereg

Nóg die Buro van die WAT nóg die African Association for Lexicography (AFRILEX) aanvaar enige aanspreeklikheid vir eise wat uit meewerkende skrywers se gebruik van materiaal uit ander bronne mag spruit.

Outeursreg op alle materiaal wat in *Lexikos* gepu-

bliseer is, berus by die Direksie van die Woordeboek van die Afrikaanse Taal. Dit staan skrywers egter vry om hulle materiaal elders te gebruik mits *Lexikos* (AFRILEX-reeks) erken word as die oorspronklike publikasiebron.

5. Oorspronklikheid

Slegs oorspronklike werk sal vir opname oorweeg word. Skrywers dra die volle verantwoordelikheid vir die oorspronklikheid en feitelike inhoud van hulle publikasies.

6. Gratis oordrukke en eksemplare

Skrywers ontvang vyf gratis oordrukke van elke artikel of resensieartikel van hulle wat gepubliseer is asook een gratis eksemplaar van die uitgawe waarin sodanige artikel(s) verskyn het. Skrywers van suiwer evalueerende resensies en van bydraes tot die rubrieke Leksikonotas, Leksikovaria, Projekte en Verslae ontvang vyf gratis oordrukke van hulle bydraes. In laasgenoemde vier kategorieë kan die redaksie egter, afhangend van die aard en omvang van die bydraes, besluit om ook 'n eksemplaar van die betrokke uitgawe aan 'n skrywer toe te ken.

7. Uitnodiging en redaksionele adres

Alle belangstellende skrywers is welkom om bydraes vir opname in *Lexikos* te lewer en aan die volgende adres te stuur:

Die Redakteur: LEXIKOS

Buro van die WAT

Posbus 245

7599 STELLENBOSCH

Republiek van Suid-Afrika

B. VOORBEREIDING VAN MANUSKRIP

Die manuskrip van artikels moet aan die volgende redaksionele vereistes voldoen:

1. Lengte en formaat van artikels

Bydraes moet verkieslik nie 20 getikte A4-bladsye met teks in dubbelspasiëring en ruim kantlyn (ongeveer 2,5 cm) oorskry nie. Manuskrip moet verkieslik in elektroniese formaat as ASCII-teks, as volledig geformateerde Microsoft Word (DOS of Windows) lêers of as WordPerfect (DOS of Windows) lêers op rekenaar-skyf (360 KB tot 1.44 MB) voorgelê word. 'n Rekenaardrukstuk van die artikel moet die skyf vergesel. Elke artikel moet voorsien wees van 'n Engelse opsomming van tussen 150 en 250 woorde, sowel as tussen 10 en 30 Engelse sleutelwoorde.

2. Grafika

Een stel duidelike oorspronklike illustrasies, tabelle, grafieke, diagramme, of kwaliteitsafdrukke daarvan, moet voorgelê word. Die plasing van grafika binne die teks moet duidelik aangedui word.

3. Bibliografiese gegewens en verwysings binne die teks

Kyk na onlangse nommers van *Lexikos* vir meer inligting.

INSTRUCTIONS TO AUTHORS

(For a more detailed version of these instructions, please contact us (lexikos@sun.ac.za) or refer to our website: <http://www.wat.co.za>)

A. EDITORIAL POLICY

1. Type and content of articles

Articles may deal with pure lexicography or with the implications that related fields such as linguistics, general linguistics, computer science and management have for lexicography.

Contributions may be classified in any one of the following categories:

- (1) **Articles:** Fundamentally original scientific research that has been done and the results that have been obtained, or reflecting existing research results and other facts in an original, synoptic, interpretative, comparative or critically evaluative manner.
- (2) **Review articles:** Research articles presented in the form of a critical review of one or more published scientific sources.

Contributions in categories (1) and (2) are subjected to strict anonymous evaluation by independent academic peers in order to ensure the international research quality thereof.

- (3) **Reviews:** An analysis and critical evaluation of published scientific sources and products, such as books and computer software.
- (4) **Projects:** Discussions of lexicographical projects.
- (5) **Lexiconotes:** Any article containing practice-oriented information, suggestions, problems, questions, commentary and solutions regarding lexicography.
- (6) **Lexicovaria:** Any of a large variety of articles, announcements and press releases by lexicographic societies which are of particular value to the practising lexicographer.
- (7) **Reports:** Reports on conferences and workshops.

Contributions in categories (3)-(7) must all meet the requirements of academic writing and are evaluated by the editors with this in mind.

2. Academic standard and evaluation procedure

The Department of Education of the South African Government has approved *Lexikos* as a subsidized, i.e. income-generating research journal.

Articles will be evaluated on the following aspects: language and style; conciseness and comprehensibility; problem formulation, reasoning and conclusion; references to the most important and most recent literature; substantial contribution to the specific discipline.

3. Language of contributions

Afrikaans, Dutch, English, French or German.

4. Copyright

Neither the Bureau of the WAT nor the African Association for Lexicography (AFRILEX) accepts any responsibility for claims which may arise from contributing authors' use of material from other sources.

Copyright of all material published in *Lexikos* will be vested in the Board of Directors of the Woordeboek van die Afrikaanse Taal. Authors are free however to use their material elsewhere provided that *Lexikos* (AFRILEX Series) is acknowledged as the original publication source.

5. Originality

Only original contributions will be considered for publication. Authors bear full responsibility for the originality and factual content of their contributions.

6. Free offprints and copies

Authors will receive five free offprints of each of their articles or review articles published, as well as one complimentary copy of the issue containing such article(s). Authors of purely evaluative reviews and of contributions to the categories Lexiconotes, Lexicovaria, Projects, and Reports receive five free offprints of their contributions. In the case of the latter four categories, the editors may, however, depending on the nature and scope of the contributions, decide to grant the author a copy of the issue concerned.

7. Invitation and editorial address

All interested authors are invited to submit contributions for publication in *Lexikos* to:

The Editor: LEXIKOS
Bureau of the WAT
P.O. Box 245
7599 STELLENBOSCH
Republic of South Africa

B. PREPARATION OF MANUSCRIPTS

Manuscripts of articles must meet the following editorial requirements:

1. Length and format

Contributions should not exceed more than 20 typewritten A4 pages with double spacing and ample margins (about 2,5 cms). Manuscript should preferably be in electronic form on a (360 KB to 1.44 MB) floppy disk as either ASCII text, fully-formatted Microsoft Word (DOS or Windows) or WordPerfect (DOS or Windows) files. A computer printout of the article should accompany the disk. Each article must be accompanied by an English abstract of 150 to 250 words, and between 10 and 30 English keywords.

2. Graphics

One set of clear original drawings, tables, graphs, diagrams or quality prints thereof must be submitted. The locations of graphics must be clearly indicated in the text.

3. Bibliographical details and references in the text

Examine recent issues of *Lexikos* for details.

HINWEISE UND RICHTLINIEN FÜR AUTOREN

(Nehmen Sie bitte uns Kontakt auf (lexikos@sun.ac.za) für eine ausführlichere Wiedergabe dieser Hinweise oder besuchen Sie unsere Webseite: <http://www.wat.co.za>)

A. REDAKTIONELLE ZIELSETZUNGEN

1. Art und Inhalt der Artikel

Es können Artikel aufgenommen werden, die sich mit Themen der Lexikographie befassen oder mit Zusammenhängen, die zwischen der Lexikographie und benachbarten Fachgebieten wie z.B. Linguistik, allgemeiner Sprachwissenschaft, Lexikologie, Computerwissenschaft und Management bestehen.

Die Beiträge sollten einer der folgenden Kategorien entsprechen:

(1) **Artikel**, die grundlegend über neue Forschungsansätze und deren Ergebnisse berichten, oder die bestehende Forschungsergebnisse und andere Informationen selbständig, interpretativ, vergleichend oder kritisch bewertend wiedergeben.

(2) **Rezensionsartikel**, die in der Form eines Forschungsartikels eine oder mehrere veröffentlichten wissenschaftlichen Quellen kritisch rezensieren.

Beiträge in Kategorien (1) und (2) werden streng anonym von unabhängigen wissenschaftlichen Experten begutachtet, um ein internationales fachliches Niveau in *Lexikos* zu gewährleisten.

(3) **Rezensionen**, die veröffentlichte wissenschaftliche Quellen und Produkte, wie z.B. Bücher und Software, analysieren und kritisch bewerten.

(4) **Lexikographische Projekte**, die vorgestellt werden.

(5) **Notizen zum Lexikon**, die praxisbezogene Informationen, Vorschläge, Probleme, Fragen, Kommentare und Lösungen hinsichtlich der Lexikographie enthalten.

(6) **Lexikovaria**, die unterschiedliche Beiträge, Ankündigungen und Pressemitteilungen lexikographischer Vereinigungen, die dem praktischen Lexikographen wichtig sein können, einschließen.

(7) **Berichte** über Konferenzen und Workshops.

Beiträge in Kategorien (3)-(7) müssen im akademischen Stil abgefaßt werden. Sie werden von der Redaktion unter diesem Gesichtspunkt beurteilt.

2. Wissenschaftliche Standards und das Beurteilungsverfahren

Das Erziehungsministerium der südafrikanischen Regierung hat *Lexikos* als eine subventionierte, d.h. einkommenerzeugende Forschungszeitschrift anerkannt.

Artikel werden auf Grund der folgenden Gesichtspunkte bewertet: Sprache und Stil; Sachlichkeit und Verständlichkeit; Problembeschreibung, Argumentation und Schlußfolgerung; Hinweise auf die neueste und wichtigste Literatur; wesentlicher Beitrag zum besonderen Fachgebiet.

3. Sprache der Beiträge

Afrikaans, Deutsch, Englisch, Französisch oder Niederländisch.

4. Das Urheberrecht

Weder das Büro des WAT noch die African Association for Lexicography (AFRILEX) übernehmen Verantwortung für Ansprüche, die daraus entstehen könnten, daß Autoren Material aus anderen Quellen benutzt haben.

Das Urheberrecht aller in *Lexikos* publizierten Artikel wird dem Direktorium unseres Büros übertragen. Es steht Autoren jedoch frei, ihren Beitrag anderweitig zu verwenden, vorausgesetzt, *Lexikos* (AFRILEX-Serie) wird als Originalquelle genannt.

5. Originalität

Nur Originalbeiträge werden begutachtet. Autoren tragen die volle Verantwortung für die Originalität und den sachlichen Inhalt ihrer Beiträge.

6. Sonderdrucke und Freixemplare

Autoren erhalten fünf Sonderdrucke ihrer veröffentlichten Artikel oder Rezensionsartikel gratis sowie ein Freixemplar der betreffenden Ausgabe. Rezensenten und Autoren von Beiträgen zu den Kategorien Notizen zum Lexikon, Lexikovaria, Projekte und Berichte erhalten fünf Sonderdrucke ihrer Beiträge gratis. Die Redaktion kann sich jedoch, abhängig von der Art und dem Umfang der Beiträge der letztgenannten vier Kategorien, vorbehalten, dem Autor ein Freixemplar der Ausgabe zu überlassen.

7. Einladung und redaktionelle Adresse

Alle Autoren, die interessiert sind, Beiträge für *Lexikos* zu liefern, sind herzlich willkommen. Sie werden gebeten, ihre Artikel an die folgende Adresse zu schicken:

Der Redakteur: LEXIKOS
Buro van die WAT
Postfach 245
7599 STELLENBOSCH
Republik Südafrika

B. VORBEREITUNG DES MANUSKRIPTS

Ein Artikelmanuskript muß den folgenden redaktionellen Anforderungen entsprechen:

1. Umfang und Format

Beiträge sollen nicht länger als 20 getippte A4-Seiten in zweizeiligem Abstand und mit Randabständen von ca. 2,5 cm sein. Das Manuskript sollte möglichst als elektronischer Text auf einer (360 KB bis 1.44 MB) Diskette vorgelegt werden, entweder im ASCII-Format, oder in formatiertem Microsoft Word (DOS oder Windows) bzw. WordPerfect (DOS oder Windows). Ein Ausdruck des vollständig formatierten Artikels soll mit der Diskette eingereicht werden. Jedem Artikel ist eine Zusammenfassung im Umfang von 150-250 Wörtern beizufügen. Ferner sollen etwa 10-30 inhaltskennzeichnende Stichwörter zu jedem Artikel angegeben werden.

2. Abbildungen

Ein reproduktionsfähiger Satz der originalen Abbildungen, Illustrationen, Tabellen, Graphiken und Diagramme oder Qualitätsabdrucke muß vorgelegt werden. Der Text selber sollte klare Hinweise auf die Position der Abbildungen enthalten.

3. Bibliographische Einzelheiten und Hinweise im Text

Zu Einzelheiten des bibliographischen Systems sind neuere Ausgaben von *Lexikos* einzusehen.

INSTRUCTIONS AUX AUTEURS

(Pour une version plus détaillée de ces instructions, contacter le Bureau du WAT (lexikos@sun.ac.za)
ou consulter notre website: <http://www.wat.co.za>)

A. POLITIQUE ÉDITORIALE

1. Caractéristiques et contenu des articles

Les articles seront consacrés à la lexicographie pure, ou aux rapports entre la lexicographie et les disciplines voisines telles que la linguistique, la linguistique générale, l'informatique et le management.

Les contributions pourront appartenir à l'une des catégories suivantes:

- (1) **Articles:** Recherches scientifiques originales, avec leurs résultats; ou présentations originales, synthétiques, interprétatives, comparatives, évaluatives et critiques des résultats de recherches en cours;
- (2) **Articles bilans:** Articles de recherche présentés sous forme de bilan critique de travaux scientifiques déjà publiés.

Les contributions appartenant aux catégories (1) et (2) seront soumises de manière anonyme à des experts spécialistes indépendants afin d'en assurer la qualité scientifique au niveau international.

- (3) **Recensions:** Analyses et évaluations critiques de travaux de recherche et de productions scientifiques, telles que livres ou logiciels;
- (4) **Projets:** Présentations de projets lexicographiques;
- (5) **'Lexiconotes':** Textes contenant des informations pratiques, ou des suggestions, des problèmes, des questions, des commentaires et des solutions concernant des activités lexicographiques;
- (6) **'Lexicovaria':** Articles, annonces, communiqués de presse émanant de centres de lexicographie et qui revêtent un intérêt particulier pour les lexicographes;
- (7) **Rapports:** Rapports sur des colloques et ateliers.

Les contributions dans les catégories (3) à (7) devront répondre aux exigences de qualité des publications scientifiques et seront évaluées dans cette optique.

2. Critères et procédures d'évaluation

La revue *Lexikos* est reconnue et subventionnée par le Ministère de l'Éducation du gouvernement Sud-Africain comme revue devant générer des revenus.

Les articles seront évalués selon les critères suivants: langue et style, concision et clarté, formulation de la problématique, raisonnement et conclusion, référence aux travaux les plus importants et les plus récents, contribution substantielle à la discipline.

3. Langue des contributions

Afrikaans, allemand, anglais, français ou néerlandais.

4. Copyright

Le Bureau du WAT ou l'*African Association for Lexicography* (AFRILEX) décline toute responsabilité en cas de réclamations motivées par l'utilisation d'autres sources par les auteurs.

Les droits d'auteurs des documents publiés dans *Lexikos* appartiennent au Conseil d'administration du *Woordeboek van die Afrikaanse Taal* (WAT). Cependant,

les auteurs sont libres d'utiliser leurs textes dans d'autres publications, à condition d'indiquer *Lexikos* (collection AFRILEX) comme source de la publication originale.

5. Originalité

Seules les contributions originales seront acceptées pour la publication. Les auteurs conservent l'entière responsabilité de l'originalité et du contenu factuel de leur texte.

6. Tirés-à-part et exemplaires gratuits

Les auteurs recevront gratuitement cinq (5) tirés-à-part de chaque article ou article de bilan, ainsi qu'un exemplaire gratuit de la publication contenant cet article.

Les auteurs des recensions et des publications dans les catégories 'Lexiconotes', 'Lexicovaria', Projets et Rapports recevront cinq (5) tirés-à-part de leur texte. Dans ces quatre dernières catégories, le responsable de la revue pourra néanmoins décider, en fonction de la nature et des dimensions des publications, d'accorder un exemplaire gratuit de la revue à leurs auteurs.

7. Adresse de la revue

Les auteurs intéressés sont invités à soumettre leurs propositions à:

L'éditeur: LEXIKOS
Bureau du WAT
Boîte postale 245
7599 STELLENBOSCH
République d'Afrique du Sud

B. PRÉSENTATION DES MANUSCRITS

Les manuscrits se conformeront aux exigences suivantes:

1. Longueur et format

Les contributions ne devraient pas excéder 20 pages dactylographiées, de format A4, avec double espacement et marges suffisantes (environ 2,5 cm), si possible en format électronique (disquette 360KB à 1,44MB), sous forme de fichier ASCII, complètement formaté sous Microsoft Word ou sous WordPerfect (DOS ou Windows). La disquette sera accompagnée d'un tirage papier. Chaque article sera pourvu d'un résumé en anglais de 150 à 200 mots, et de 10 à 30 mots-clés.

2. Tableaux et graphiques

Les dessins, tableaux, graphiques et diagrammes seront envoyés, soit sous leur forme originale soit sous forme d'une copie de bonne qualité. Leur place dans le texte devra être clairement indiquée.

3. Bibliographie et références dans le texte

Voir les exemplaires récents de *Lexikos*.