

Lexikos 19

Lexikos 19

Redakteur

Editor

J.C.M.D. du Plessis

Resensieredakteur

Review Editor

T. Harteveld



African Association for Lexicography

AFRILEX-REEKS 19:2009

AFRILEX SERIES 19:2009



BURO VAN DIE WAT

STELLENBOSCH

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

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

Uitgewer Publisher

BURO VAN DIE WAT
Posbus 245
7599 STELLENBOSCH

Kopiereg © 2009 deur die uitgewer
Alle regte streng voorbehou
Eerste uitgawe 2009

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

Gedruk en gebind deur Onyx Press
Stewartstraat 61 Goodwood

ISBN-13: 978-0-9584971-9-0
ISBN-10: 0-9584971-9-2
ISSN 1684-4904

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

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

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

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

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

Indekse Indexes

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

Adviesraad / Advisory Board

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

Redaksiekomitee / Editorial Committee

Prof. H.M. Batibo (Botswana)
Dr. H.L. Beyer (Namibië/Namibia)
Prof. W.A.M. Carstens (RSA)
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)
Dr. P.A. Mavoungou (Gaboen/Gabon)
Dr. J.S. Mdee (Tanzanië/Tanzania)
Prof. B.M. Mini (RSA)
Dr. V.M. Mojela (RSA)
Mnr. M.C. Mphahlele (RSA)
Mnr. M.H. Mpungose (RSA)
Dr. H.S. Ndinga-Koumba-Binza (Gaboen/Gabon)
Prof. D.J. Prinsloo (RSA)
Prof. P.H. Swanepoel (RSA)

Inhoud / Contents

Voorwoord	x
Foreword	xi
<i>J.C.M.D. du Plessis</i>	
'n Woord van AFRILEX	xii
A Few Words from AFRILEX	xiv
<i>Gilles-Maurice de Schryver</i>	
Redaksionele doelstellings	xvi
Editorial Objectives	xvii
Redaktionelle Ziele	xviii
Politique éditoriale	xix

Artikels / Articles

'n Teoretiese basis vir kontekstualisering in tweetalige woordeboeke	1
<i>Herman L. Beyer</i>	
Ensiklopedisiteit in die <i>Woordeboek van die Afrikaanse Taal</i> : 'n "Saak" van balans	23
<i>Willem Botha</i>	
The Lexicographic Treatment of Ideophones in Zulu	34
<i>Gilles-Maurice de Schryver</i>	
L'apport de la sociolinguistique à la lexicographie: l'exemple récent des belgicisms dans l'édition du <i>Nouveau Petit Robert</i> 2008	55
<i>Jean-Nicolas De Surmont</i>	
Geïntegreerde woordeboekgebruik van vakwoordeboeke vir aanleerders	72
<i>Rufus H. Gouws</i>	
Zum Stand und zu den Perspektiven der allgemeinsprachlichen Lexikographie mit Deutsch und Slowenisch	94
<i>Vida Jesenšek und Herbert Ernst Wiegand</i>	

Place et nécessité des travaux religieux, ethnohistoriques et linguistiques dans le paysage dictionnaire gabonais <i>P.A. Mavoungou</i>	144
Compiling the First Monolingual Lusoga Dictionary <i>Minah Nabirye</i>	177
The Representation of Vowel Duration in Civilian Dictionaries <i>Hugues Steve Ndinga-Koumba-Binza and Justus C. Roux</i>	197
The Evaluation of the Outside Matter in Dictionary Reviews <i>Sandro Nielsen</i>	207
A Perspective on the Lexicographic Value of Mega Newspaper Corpora — The Case of Afrikaans in South Africa <i>D.J. Prinsloo</i>	225
Function-related Secondary User Needs and Secondary Data? A Critical Examination of Some Central Concepts in the Modern Theory of Lexicographical Functions <i>Marcin Overgaard Ptaszynski</i>	242
How Effective is the Electronic Dictionary in Sense Discrimination? <i>Kim Hua Tan</i>	262
Reflections on Lexicographical User Research <i>Sven Tarp</i>	275
Die opvoedkundige waarde van woordeboeke: Voorstelle vir woordeboekonderrig in Suid-Afrika <i>Michele F. van der Merwe</i>	297
Volkslinguïstiek en dialectlexicografie in de zuidelijke Nederlanden <i>Jacques Van Keymeulen</i>	314

Leksikoprogrammatuur / Lexicosoftware

A Corpus-based Survey of Four Electronic Swahili–English Bilingual Dictionaries <i>Guy De Pauw, Gilles-Maurice de Schryver and Peter Waiganjo Wagacha</i>	340
--	-----

Projekte / Projects

- Towards a Corpus of South African English: Corraling the Sub-varieties 353
Leela Pienaar and Vivian de Klerk

Leksiko-opname / Lexicosurvey

- Lexikos* at Eighteen: An Analysis 372
Gilles-Maurice de Schryver
- Keeping in Touch. A Survey of Lexicography Periodicals 404
R.R.K. Hartmann

Leksikonotas / Lexiconotes

- South Africa's New African Language Dictionaries and their Use for the African Speech Communities 423
Juliane Klein
- Natural Science and Technology Terminology in the Sesotho sa Leboa Monolingual Dictionary 436
W.M. Mojapelo and V.M. Mojela
- Populating Sub-entries in Dictionaries with Multi-word Units from Concordance Lines 446
Thapelo J. Otlogetswe

Resensieartikels / Review Articles

- An Analysis of *Practical Lexicography: A Reader* (Ed. Fontenelle 2008) 458
Gilles-Maurice de Schryver
- The Metalexicographical Contribution of Pedro A. Fuertes-Olivera and Ascensión Arribas-Baño's *Pedagogical Specialised Lexicography: A Critical Review* 490
Dion Nkomo

Resensies / Reviews

- John Considine: *Dictionaries in Early Modern Europe: Lexicography and the Making of Heritage* 509
Loránd-Levente Pálfi
- A.P. Cowie (Editor): *The Oxford History of English Lexicography Volume I: General-purpose Dictionaries. Volume II: Specialized Dictionaries* 516
Loránd-Levente Pálfi
- ILALOK: *Dictionnaire vili-français / Mpisukulu bi kum' bi tshi vili ku tshi mputu* 524
P.A. Mavoungou
- Jean Pruvost (Éditeur): *Dictionnaires et mots voyageurs: les 40 ans du Petit Robert, de Paul Robert à Alain Rey* 528
Jean-Nicolas De Surmont

Lexikovaria / Lexicovaria

- The Dictionary of Lexicography and Dictionary Research 530
Rufus H. Gouws and Herbert Ernst Wiegand

Publikasieaankondigings / Publication Announcements

- Voorskrifte aan Skrywers 536
Instructions to Authors 537
Richtlinien für Autoren 538
Instructions aux Auteurs 539

Voorwoord

In die afdeling "Leksiko-opname" verskyn twee artikels, spesiaal geskryf vir hierdie nommer, wat veral toepaslik is vir *Lexikos* in hierdie stadium van sy publikasiegeskiedenis. Die eerste artikel deur Gilles-Maurice de Schryver gee 'n oorsig en ontleding van die bydraers en hulle bydraes tot *Lexikos* gedurende die afgelope agtien jaar. Dit toon onder andere hoe *Lexikos* geprobeer het om by te bly by die veranderinge en neigings in leksikografiese navorsing deur die toevoeging van nuwe afdelings om verskillende soorte artikels 'n plek te gee. 'n Belangrike vernuwing vanjaar kom in die vorm van 'n aparte supplement wat aan *Lexikos* groter moontlikhede bied vir die publikasie van versamelings artikels wat 'n besondere leksikografiese tema dek of langer monografieë wat 'n enkele leksikografiese onderwerp behandel. Vanjaar se supplement bevat artikels waarin personeellede van die African Languages Research Institute (ALRI) verskillende aspekte bespreek van die woordeboeke wat hulle tot dusver saamgestel en gepubliseer het.

Die tweede artikel deur R.R.K. Hartmann gee 'n oorsig van leksikografiese en tematies verwante tydskrifte. Alhoewel *Lexikos* die mondstuk is van AFRI-LEX, die vereniging vir die bevordering van leksikografie in Afrika, het dit ook 'n meer internasionale siening ingeneem deur artikels van 'n teoretiese aard te plaas wat sy lesers se leksikografiese agtergrond verbreed en hulle leksikografiese insig verdiep. Hartmann se artikel sal dus bydra om kontak met tydskrifte van 'n soortgelyke benadering as *Lexikos* uit te brei. Om hierdie rede word dit gevolg deur 'n Byvoegsel wat inligting verskaf van hoe kontak met hierdie tydskrifte bewerkstellig kan word.

Die redakteur wil graag bydraers bedank vir die voorlegging van artikels vir vanjaar se nommer van *Lexikos*. Aangesien dit 'n aansienlike aantal artikels bevat, het dit baie druk op die keurders geplaas. Die redakteur dra dus sy dank oor aan almal vir hul bereidwilligheid om artikels te beoordeel, dikwels op kort kennisgewing en binne 'n beperkte tyd.

'n Spesiale woord van dank moet gaan aan me. Riette Ruthven wat al die artikels professioneel en doeltreffend geset het en wat geredelik en geduldig veranderinge en korreksies aangebring het in die verskillende weergawes en herbewerkings waardeur die meeste artikels gegaan het. Soos altyd was me. Tanja Hartevelt beskikbaar om vinnig en vlot rekenaarkundige haakplekke en probleme uit te skakel en op te los.

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

Foreword

In the section "Lexicosurvey", appear two articles, especially written for this issue, which are particularly relevant to *Lexikos* at this stage of its publishing history. The first article by Gilles-Maurice de Schryver gives an overview and analysis of the contributors and their contributions to *Lexikos* during the past eighteen years. It indicates among others how *Lexikos* has tried to keep up with changes and trends in lexicographic research by introducing new sections to accommodate different kinds of articles. An important innovation this year comes in the form of a separate supplement giving *Lexikos* greater possibilities for publishing collections of articles covering a particular lexicographic theme or longer monographs dealing with a singular lexicographic subject. This year's supplement contains articles in which staff members of the African Languages Research Institute (ALRI) discuss different aspects of the various dictionaries they have compiled and published so far.

The second article by R.R.K. Hartmann gives a survey of lexicographic and thematically related journals. Although *Lexikos* is the mouthpiece of AFRI-LEX, the association for promoting lexicography in Africa, it has also taken a more international view by including articles of a theoretical nature which would expand the lexicographic background and deepen the lexicographic insights of its readers. Hartmann's article will therefore contribute to extend contact with journals of a similar approach to that of *Lexikos*. For this reason, it is followed by an Appendix giving information about how contact with these journals could be established.

The editor wishes to thank contributors for submitting articles to this year's issue of *Lexikos*. As it contains a considerable number of articles, it placed much pressure on the referees. The editor therefore extends his thanks to all for their willingness to judge articles, often at short notice and within a limited time.

A special word of thanks must go to Ms Riette Ruthven who professionally and efficiently typeset all the articles and readily and patiently inserted changes and corrections in the several versions and reworkings through which most articles went. As always Ms Tanja Harteveld was available to quickly and smoothly sort out and solve computational hitches and problems.

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

'n Woord van AFRILEX

Dit is 'n groot eer vir my as nuutgekoose President van AFRILEX om hierdie woord aan die lesers van *Lexikos* te rig. Indien daar gekyk word na die volkome toewyding waarmee my drie voorgangers — Rufus Gouws, Danie Prinsloo en Mariëtta Alberts — die taak verrig het, sal ek moeilik hulle buitengewone leierskap kan ewenaar. Wanneer 'n mens hulle inleidingswoorde lees wat sedert nommer 6 (1996) in *Lexikos* verskyn het, is dit interessant om die voorkeuronderwerpe van elkeen te herken: Rufus het nooit nagelaat om leksikografiese teorie ter sprake te bring nie, Danie het vanuit 'n meer praktiese en organisatoriese hoek gekyk, terwyl Mariëtta op die subdissipline terminologie gekonsentreer het. Elke akademikus het 'n liefhebber; myne is om ons strewes in publikasies te omskep wat oor die jare hul geldigheid sal bly behou. Wat dit betref, is ons besonder gelukkig om oor *Lexikos* te beskik.

Lexikos is nie die tydskrif van AFRILEX nie — omdat dit bydraes vanoor die wêreld trek, waarvan net die bestes gepubliseer word na 'n streng anonieme beoordeling deur ten minste twee keurders — maar dit is die voorkeuropublikasiebron en as sodanig dus die mondstuk van die lede van AFRILEX. *Lexikos* sal nie *Lexikos* wees sonder die toewyding van sy redakteur, dr. J.C.M.D. du Plessis, en sy assistent, me. Riette Ruthven nie. Hulle omskep gehaltemanuskripte in artikels wat dikwels byna volmaak is. Sulke aspekte gaan nie ongemerk verby nie, en dit is ongetwyfeld danksy die hoëgehaltevoorleggings, gekombineer met 'n tydige redigeerproses, dat *Lexikos* vanaf nommer 15 (2005) gekeur is vir insluiting in die volgende twee databasisse van die Thomson Reuters Web of Science Citation Index:

- Arts & Humanities Citation Index
- Current Contents/Arts & Humanities

Vanaf verlede jaar se nommer, nommer 18 (2008), neem die Web of Science *Lexikos* verder op in:

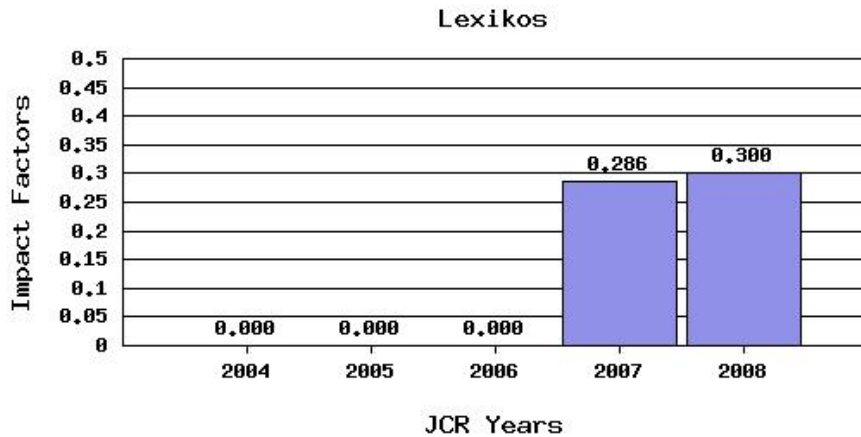
- Social Sciences Citation Index
- Social SciSearch
- Current Contents/Social & Behavioral Sciences
- Journal Citation Reports/Social Sciences Edition

Die laaste, insluiting in die Journal Citation Reports, is veral waardevol omdat dit beteken dat impakfaktore nou vir *Lexikos* bereken word. Die berekening van impakfaktore is elementêre wiskunde: Vir elke betrokke jaar verdeel 'n mens die aantal verwysings na die materiaal gepubliseer in die twee voorafgaande jare met die totale aantal artikels gepubliseer in hierdie twee jare. In die Journal Citation Reports van 2008 lewer dit die volgende vir *Lexikos* op:

Aanhalings in 2008 tot items gepubliseer	Aantal items gepubliseer
in:	in:
2007 = 11	2007 = 30
2006 = 4	2006 = 20
Totaal: 15	Totaal: 50

Berekening: $\frac{\text{Aanhalings tot onlangse items}}{\text{Aantal onlangse items}} \frac{15}{50} = 0.300$

Die huidige (2008) impakfaktor vir *Lexikos* is dus 0.300. Met gegewens geberg in die databasisse van die Web of Science kan 'n mens ook terugkyk, wat toon dat die impakfaktor vir *Lexikos* op die oomblik bestendig (en effens stygend) is.



Impakfaktore moet altyd binne konteks gesien word, en dit is derhalwe belangrik om rekening te hou met die plek wat 'n tydskrif inneem in die kategorie/-ieë waartoe dit behoort. *Lexikos* behoort tot die kategorie Linguistiek, waar dit 56ste (uit 68 tydskrifte) in die laaste (vierde) kwartiel geplaas is.

Gegee dat akademiese personelede van inrigtings van hoër onderwys toenemend "beoordeel" word, en dat hulle publikasies in hoë-impaktydskrifte geneig is om meer gewig te dra as enige ander oorwegings vir daardie beoordeeling, kan die waarde van die feit dat *Lexikos* nou 'n geïndekseerde tydskrif is — met 'n impakfaktor — nie genoeg beklemtoon word nie. In die vakgebied leksikografie, kom dié eer slegs twee tydskrifte toe: die *International Journal of Lexicography* en nou *Lexikos*. Wanneer gekyk word hoe impakfaktore bereken word, wil ek daarom graag almal van ons aanspoor om seker te maak dat ons altyd die toepaslike verwysings in al ons manuskripte insluit by materiaal gepubliseer gedurende die voorafgaande twee jaar in sowel die *International Journal of Lexicography* as *Lexikos*. Mag *Lexikos* se impakfaktor, en daarmee saam die lewendigheid van ons vakgebied, as gevolg daarvan styg!

Gilles-Maurice de Schryver
 President: AFRILEX

A Few Words from AFRILEX

It is a great honour for me, as the newly elected President of AFRILEX, to address the readers of *Lexikos*. Considering the complete dedication with which my three predecessors — Rufus Gouws, Danie Prinsloo and Mariëtta Alberts — performed the job, I will have a hard time matching their extraordinary leadership. When one reads through their introductory words that have appeared in *Lexikos* since Volume 6 (1996), it is pleasing to recognize the pet subjects of each: Rufus has never shied away from raising lexicographic theory, Danie took a more practical and organizational angle, while Mariëtta focused on the sub-discipline of terminology. Every scholar has a hobby-horse; mine is to turn our endeavours into publications that will stand up to the ravages of time. In that, we are particularly fortunate to have *Lexikos*.

Lexikos is not the journal of AFRILEX — as it attracts contributions from all over the world, with the best only being published following, for each, a rigorous blind adjudication process by at least two referees — but it is the favoured outlet and as such mouthpiece of the members of AFRILEX. *Lexikos* would not be *Lexikos* without the devotion by its editor, Dr J.C.M.D. du Plessis, and his right hand, Ms Riette Ruthven. They turn quality manuscripts into articles that are often very nearly perfect. Such aspects do not go unnoticed, and it is no doubt thanks to the high-quality submissions, combined to a timely editorial process, that *Lexikos* was selected, as of Volume 15 (2005), for inclusion in the following two databases of the Thomson Reuters Web of Science Citation Index:

- Arts & Humanities Citation Index
- Current Contents/Arts & Humanities

Starting with last year's volume, Volume 18 (2008), the Web of Science further includes *Lexikos* in:

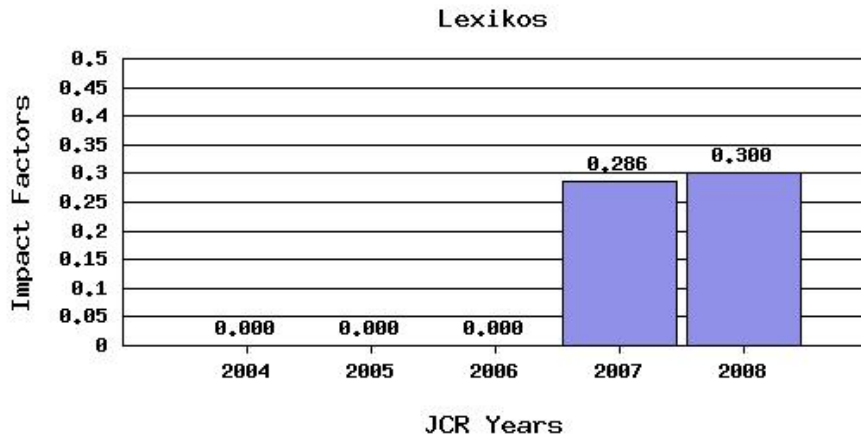
- Social Sciences Citation Index
- Social SciSearch
- Current Contents/Social & Behavioral Sciences
- Journal Citation Reports/Social Sciences Edition

The latter, inclusion in the Journal Citation Reports, is especially valuable, as it means Impact Factors are now being calculated for *Lexikos*. The calculation of Impact factors is child's mathematics: For any particular year, one divides the number of citations to material published in the two previous years, by the total number of articles published in those two years. In the 2008 Journal Citation Reports, this gives the following for *Lexikos*:

Citations in 2008 to items published		Number of items published	
in:	2007 = 11	in:	2007 = 30
	2006 = 4		2006 = 20
	Sum: 15		Sum: 50

Calculation: $\frac{\text{Citations to recent items}}{\text{Number of recent items}} \frac{15}{50} = 0.300$

The current (2008) Impact Factor for *Lexikos* is thus 0.300. With the data stored in the Web of Science databases, one can also look back, which shows that the Impact Factor for *Lexikos* is stable (and slightly rising) at present:



Impact Factors must always be viewed in context, and it is thus important to consider the position a journal takes in the category /-ies to which it belongs. *Lexikos* belongs to the Linguistics category, where it is ranked 56th (out of 68 journals), in the last (fourth) quartile.

Given that academic staff members at institutes of higher education are increasingly being 'rated', and that their publications in high-impact journals tend to have more weight than any other considerations for that rating, the value of the fact that *Lexikos* is now an indexed journal — with an Impact Factor — cannot be stressed enough. In the discipline of lexicography, this honour comes to only two journals: the *International Journal of Lexicography*, and now *Lexikos*. Looking at how Impact Factors are calculated, I would therefore like to urge all of us to make sure we always include the relevant references, in all our manuscripts, to material published during the preceding two years in both the *International Journal of Lexicography* and *Lexikos*. May *Lexikos's* Impact Factor, and with it the liveliness of our discipline, rise as a result!

Gilles-Maurice de Schryver
 President: AFRILEX

Redaksionele doelstellings

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

Die strewe van die AFRILEX-reeks is:

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

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

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

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

Editorial Objectives

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

The objectives of the AFRILEX Series are:

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

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

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

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

Redaktionelle Ziele

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

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

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

Gemäß den Zielsetzungen der AFRILEX-Serie werden:

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

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

Politique éditoriale

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

Les objectifs de la collection AFRILEX sont de :

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

Pour atteindre ces objectifs, la collection AFRILEX

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

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

'n Teoretiese basis vir kontekstualisering in tweetalige woordeboeke*

Herman L. Beyer, *Departement Taal- en Literatuurstudie, Universiteit van Namibië, Windhoek, Namibië en Departement Afrikaans en Nederlands, Universiteit Stellenbosch, Stellenbosch, Republiek van Suid-Afrika*
(hbeyer@unam.na)

Opsomming: Die hoofdoel van 'n tweetalige woordeboek kan beskou word as die kommunikatiewe bemagtiging van die woordeboekgebruiker deur die voorsiening van semanties-pragmatiese vertaalekwivalente ter wille van kommunikatiewe ekwivalensie. Wanneer 'n vertaalekwivalentparadigma aangebied word, moet die woordeboekgebruiker deur funksionele ekwivalent-diskriminerende inskrywings gehelp word om die kommunikatief gepaste vertaalekwivalent te selekteer vir die doel van teksresepisie, teksproduksie of vertaling binne die tersaaklike diskoers-situasie, d.w.s. die data-aanbod in die woordeboek behoort die gebruiker in die toepassing van ekwivalentdiskriminasie te ondersteun. In hierdie artikel word 'n teoretiese fundering van sodanige aanbod voorgestel, wat gebaseer word op die relevante literatuur en woordeboekdata. Die potensiaal van 'n kommunikatiewe benadering tot 'n gebruikersgerigte metaleksikografie word in die proses verken. Daar word vasgestel dat ekwivalentdiskriminerende inskrywings in 'n algemene tweetalige woordeboek binne die paradigma van *konteks*, soos leksikografies toegepas, aangebied word, en bepaalde klasse *kontekstualiserende inskrywings* word geïdentifiseer. Die verspreidingsstruktuur van dié inskrywingstipes kom ook kortliks aan bod. 'n Kommunikatiewe benadering benadruk dat die kommunikatiewe sukses van 'n woordeboek deur 'n goed beplande woordeboekspesifieke leksikografiese proses én woordeboekgebruik deur bevoegde teikengebruikers medebepaal word.

Sleutelwoorde: BETEKENISVERKLARING, BUITETEKST, DEKODERING, DISKRIMINATOR, EKWIVALENTDISKRIMINASIE, EKWIVALENTDISKRIMINERENDE INSKRYWING, ENKODERING, GEBRUIKERSGERIGTE METALEKSIKOGRAFIE, INGEVOEGDE TEKS, KOLLOKASIE, KOMMUNIKATIEWE BENADERING, KOMMUNIKATIEWE EKWIVALENSIE, KONTEKS, KONTEKSLEIDING, KONTEKSTUALISERENDE INSKRYWINGS, KONTEKSTUALISERING, KOTEKS, LEKSIKOGRAFIE, LEKSIKOGRAFIESE DATA, LEKSIKOGRAFIESE ETIKET, LEKSIKOGRAFIESE GLOS, SEMANTIES-PRAGMATIESE EKWIVALENSIE, TWEETALIGE WOORDEBOEK, VERSPREIDINGSTRUKTUUR, VRY KONSTRUKSIE

* Hierdie artikel is 'n aangepaste weergawe van 'n gedeelte van 'n hoofstuk uit 'n D.Litt.-proefskrif, *'n Metaleksikografiese ondersoek na kontekstleiding in Afrikaanse vertalende woordeboeke*, wat in April 2006 deur die Universiteit Stellenbosch, Stellenbosch, Republiek van Suid-Afrika, aanvaar is.

Abstract: A Theoretical Foundation for Contextualisation in Bilingual Dictionaries. The main aim of a bilingual dictionary may be regarded as the communicative empowerment of the dictionary user through the presentation of semantic-pragmatic translation equivalents for the purpose of communicative equivalence. When a translation equivalent paradigm is offered, the dictionary user should be assisted in selecting the communicatively appropriate translation equivalent for the purpose of text reception, text production or translation in a specific discourse situation by functional equivalent discriminating entries, i.e. the data presentation in the dictionary should support the user in the application of equivalent discrimination. This article proposes a theoretical foundation for such a presentation, which is based on the relevant literature and dictionary data. The potential of a communicative approach to a user-directed metalexigraphy is explored in the process. It is established that equivalent discriminating entries in a general bilingual dictionary are presented within the paradigm of *context*, as lexicographically applied, and specific classes of *contextualising entries* are identified. The distribution structure of these entries is briefly touched on. A communicative approach emphasises that the communicative success of a dictionary is co-determined by both a well-planned dictionary specific lexicographic process and dictionary use by competent target-users.

Keywords: BILINGUAL DICTIONARY, COLLOCATION, COMMUNICATIVE APPROACH, COMMUNICATIVE EQUIVALENCE, CONTEXT, CONTEXTUAL GUIDANCE, CONTEXTUALISATION, CONTEXTUALISING ENTRY, CO-TEXT, DECODING, DEFINITION, DISCRIMINATOR, DISTRIBUTION STRUCTURE, ENCODING, EQUIVALENT DISCRIMINATION, EQUIVALENT-DISCRIMINATING ENTRY, FREE CONSTRUCTION, INSERTED TEXT, LEXICOGRAPHIC DATA, LEXICOGRAPHIC GLOSS, LEXICOGRAPHIC LABEL, LEXICOGRAPHY, OUTER TEXT, SEMANTIC-PRAGMATIC EQUIVALENCE, USER-DIRECTED METALEXICOGRAPHY

1. Inleiding

Zgusta (1971: 294) vat die algemene doel van tweetalige woordeboeke soos volg saam:

The basic purpose of a bilingual dictionary is to coordinate with the lexical units of one language those lexical units of another language which are equivalent in their lexical meaning.

Leksikale items in die brontaal word met leksikale items in die doeltaal gekoördineer wat dieselfde betekenis verteenwoordig, oftewel wat semanties ekwivalent is. Sodanige doeltaalitems word *vertaalekwivalente* genoem. Volgens Wiegand (2002: 248) gaan dit egter nie uitsluitlik om die aanduiding van semantiese ekwivalensie nie: Die tweetalige woordeboek behoort na *semanties-pragmatiese ekwivalensie* te streef. Dit veronderstel dat die vertalende leksikograaf nie alleen die verantwoordelikheid het om een of 'n aantal semanties-pragmatiese vertaalekwivalente vir elke brontaallemma aan te bied nie, maar ook om die woordeboekgebruiker deur funksionele addisionele woordeboek-inskrywings benewens die vertaalekwivalente te lei om vir die tersaaklike dis-

koerssituasie die gepaste vertaalekwivalent te selekteer én om die geselekteerde vertaalekwivalent dan gepas en korrek in die geval van teksresepsie, teksproduksie of vertaling toe te pas, dit wil sê ter wille van *kommunikatiewe ekwivalensie* (vgl. Gouws 1996: 16). Hierdie eis geld veral indien as gevolg van die anamorfisme tussen tale daar slegs gedeeltelike ekwivalensie tussen brone doeltaalitem bestaan (vgl. Gouws en Prinsloo 2005: 154). Die doel van sodanige funksionele addisionele woordeboekinskrywings behoort dus te wees om die woordeboekgebruiker te help om ter wille van kommunikatiewe ekwivalensie tussen die aangebode vertaalekwivalente te diskrimineer, dit wil sê om *ekwivalentdiskriminasie* toe te pas.

2. Ekwivalentdiskriminasie

Daardie addisionele funksionele woordeboekinskrywings wat die funksie moet vervul om die woordeboekgebruiker te help met ekwivalentdiskriminasie word *ekwivalentdiskriminerende inskrywings* genoem. Die nut, en inderdaad essensiële behoefte aan ekwivalentdiskriminerende inskrywings in tweetalige woordeboeke word duidelik uit die volgende opmerkings deur Al-Kasimi (1977: 63):

- (1)(a) Nida expressed the fluid and inconsistent nature of meaning of words in his three fundamental semantic presuppositions:
 - (i) No word (or semantic unit) ever has exactly the same meaning in two different utterances.
 - (ii) There are no complete synonyms within a language.
 - (iii) There are no exact correspondences between related words in different languages.
- (b) Absolute equivalents which have exactly the same semantic and grammatical function in both languages [which are paired off in a bilingual dictionary] are rare.

Hieruit vloei die volgende resoluusie (Al-Kasimi 1977: 68):

- (2) The bilingual dictionary should provide meaning discriminations which enable the user to select the appropriate equivalent or the proper sense of an equivalent. And unless the problem of meaning discrimination is solved systematically, the bilingual dictionary cannot be a dependable guide to the proper equivalents.

Die gevolg is dat dit as een van die vertalende leksikograaf se primêre doelwitte beskou word om in sy/haar woordeboek 'n aanbod te bewerkstellig wat die ondubbelsinnige interpretasie van die aangebode data deur die teikengebruiker sal bewerkstellig (vgl. Gouws 2000: 40). Roberts (1990: 219) verwoord hierdie vereiste soos volg:

- (3) [I]t is no longer deemed sufficient to provide only translation equivalents for some language headwords, accompanied by an occasional example; it is now

considered highly desirable to provide semantic and stylistic discrimination of equivalents, detailed grammatical information, and collocational specifications for each headword and even for each sense division of the headword.

Die vereiste is egter meer as "highly desirable" vir 'n tweetalige woordeboek wat ten doel het om kommunikatiewe ekwivalensie te bewerkstellig — dit is onontbeerlik. Kromann et al. (1991: 2724) beskou die afwesigheid van ekwivalentdiskriminerende inskrywings in tweetalige woordeboeke tereg as "one of the ancient and deadly sins of translation lexicography".

Volgens Wiegand (1989: 251) lei die produksie van woordeboeke tot die kulturele praktyk van woordeboekgebruik. In 'n gebruikersgerigte metaleksikografie (vgl. o.m. Gouws en Prinsloo 2005: 39-44) kan die verband tussen woordeboekproduksie en woordeboekgebruik sinvol vanuit 'n kommunikatiewe benadering aan die hand van 'n leksikografiese kommunikasiemodel beskryf word (vgl. Beyer 2006). Sodanige model benadruk onder meer dat die leksikograaf tydens die woordeboekspesifieke leksikografiese proses geselekteerde leksikografiese data in woordeboekinskrywings enkodeer en dat die woordeboekgebruiker tydens woordeboekgebruik die aangebode woordeboekinskrywings dekodeer om spesifieke inligting te onttrek ten einde kommunikatief bemagtig te word.¹ Indien hierdie model op die tweetalige woordeboek toegepas word, is die implikasie dat kommunikatiewe ekwivalensie nie eensydig deur die leksikograaf bewerkstellig kan word nie, maar dat die woordeboekgebruiker ook 'n aktiewe rol in die proses speel. Met die fokus op die woordeboekspesifieke leksikografiese proses (d.i. die leksikograaf) kan gekonstateer word dat Afrikaanse tweetalige woordeboeke onvoldoende tot die medebewerkstelling van kommunikatiewe ekwivalensie bydra, met ander woorde dat onvoldoende leksikografiese data geënkodeer word en/of dat leksikografiese data onvoldoende geënkodeer word: Volgens die literatuur blyk die probleem te wees dat onvoldoende ekwivalentdiskriminerende inskrywings aangebied word wat die kommunikasie-georiënteerde funksies van die woordeboek (vgl. Tarp 2000) ondersteun. Met die fokus op woordeboekgebruik kan gekonstateer word dat die woordeboekgebruiker onvoldoende leksikografiese data moet dekodeer en/of leksikografiese data onvoldoende dekodeer: Wanneer die woordeboekgebruiker met 'n vertaalekwivalentparadigma gekonfronteer word, moet hy/sy ekwivalentdiskriminasie toepas ten einde die naslaanprosedure te voltooi, ongeag die aanbod al dan nie van ekwivalentdiskriminerende inskrywings. Die vraag is of die gebruiker suksesvol sal wees én in hoe 'n mate die aanbod in die woordeboek tot hierdie uitkoms sal bydra. Hierdie vraag het 'n tweesydigse implikasie. In die eerste plek impliseer dit dat 'n onderskeid getref kan word tussen *suksesvolle* en *onsuksesvolle ekwivalentdiskriminasie* tydens woordeboekgebruik. Dié konsepte kan soos volg omskryf word:

- (4) *Suksesvolle ekwivalentdiskriminasie* vind plaas wanneer die woordeboekgebruiker tydens woordeboekgebruik 'n kommunikatief gepaste vertaal-

ekwivalent (d.i. 'n semanties-pragmatiese vertaalekwivalent) in 'n aangebode vertaalekwivalentparadigma vir die doel van teksresepsie, teksproduksie of vertaling in 'n spesifieke diskoerssituasie selekteer.

- (5) *Onsuksesvolle ekwivalentdiskriminasie* vind plaas wanneer die woordeboekgebruiker tydens woordeboekgebruik 'n kommunikatief nie-gepaste vertaalekwivalent in 'n aangebode vertaalekwivalentparadigma vir die doel van teksresepsie, teksproduksie of vertaling in 'n spesifieke diskoerssituasie selekteer.

In die tweede plek kan 'n onderskeid getref word tussen (a) suksesvolle ekwivalentdiskriminasie wat gefasiliteer word deur die suksesvolle dekodering van funksionele leksikografiese data wat deur die leksikograaf in ekwivalentdiskriminerende inskrywings geënkodeer is, en (b) suksesvolle ekwivalentdiskriminasie wat plaasvind ten spyte van (die afwesigheid of gebrekkigheid van) die aanbod van ekwivalentdiskriminerende inskrywings. Geval (a) kan *funksioneel-gefasiliteerde suksesvolle ekwivalentdiskriminasie* genoem word, terwyl geval (b) *insidenteel-suksesvolle ekwivalentdiskriminasie* genoem kan word. Hierbenevens kan *disfunksioneel-gefasiliteerde onsuksesvolle ekwivalentdiskriminasie* onderskei word wanneer onsuksesvolle ekwivalentdiskriminasie gefasiliteer word as gevolg van (a) 'n gebrekkige aanbod van ekwivalentdiskriminerende inskrywings wat veroorsaak is deur die gebrekkige enkodering van funksionele leksikografiese data en/of die enkodering van disfunksionele leksikografiese data, en/of (b) die onsuksesvolle dekodering van funksionele leksikografiese data wat funksioneel in ekwivalentdiskriminerende inskrywings geënkodeer is.

Uiteraard behoort disfunksioneel-gefasiliteerde onsuksesvolle ekwivalentdiskriminasie ten alle koste voorkom te word. Om die woordeboekgebruiker aan sy/haar eie taalintuisie oor te laat om óf onsuksesvolle ekwivalentdiskriminasie óf insidenteel-suksesvolle ekwivalentdiskriminasie toe te pas is leksikografies onverdedigbaar. 'n Gebruikersgerigte leksikografie streef na die uitkoms van funksioneel-gefasiliteerde suksesvolle ekwivalentdiskriminasie ter wille van kommunikatiewe ekwivalensie.

Die vraag is binne watter paradigma ekwivalentdiskriminerende inskrywings in die woordeboek aangebied behoort te word. Uit Gouws (1989: 160-161) se beskrywing van die status en funksie van die vertaalekwivalent in die tweetalige woordeboek kan dié vraag beantwoord word. Die beskrywing lui soos volg:

- (6) Die status van 'n vertaalekwivalent is dat dit 'n leksikale item uit die doeltaal is wat in sekere vasgestelde kontekste gebruik kan word om die lemma mee te vertaal. Die semantiese funksie van 'n vertaalekwivalent is dus dat dit 'n aanduiding moet gee van die betekeniswaarde wat 'n gegewe lemma in 'n spesifieke konteks het, deur aan te toon hoe daardie spesifieke betekeniswaarde in die doeltaal geleksikaliseer word.

Indien die leksikograaf duidelikheid kan kry oor die aard en rol van die "sekere vasgestelde kontekste" waarin spesifieke vertaalekwivalente gebruik kan word om lemmata mee te vertaal, kan hy/sy verteenwoordigende data enkodeer in ekwivalentdiskriminerende inskrywings, met ander woorde die ekwivalent-diskriminerende inskrywings kan binne die paradigma van *konteks* aangebied word.

3. Konteks

Uit die beskrywing in (6) kan twee kenmerke van 'n vertaalekwivalent geïdentifiseer word: 'n Vertaalekwivalent gee naamlik (a) 'n aanduiding van die betekeniswaarde van die lemma waaraan dit geadresseer is (b) met betrekking tot 'n spesifieke diskoerssituasie.

3.1 Kenmerke van vertaalekwivalente

3.1.1 Kenmerk (a): Die semantiese funksie oftewel die *primêre konteks*

In terme van kenmerk (a) redeneer Gouws (1989: 161) dat vertaalekwivalente *konteksafhanklike* woordeboekinskrywings is aangesien hulle semantiese data weergee, dit wil sê 'n vertaalekwivalent geld as sodanig slegs binne die konteks waarin dit semanties ekwivalent aan die doeltaallemma is. Buite dié konteks kan 'n vertaalekwivalent nie bestaan nie. Semantiese ekwivalensie kan dus as die *primêre konteks* beskou word waarbinne vertaalekwivalente in die tweetalige woordeboek aangebied word. Dit is die tipe konteks wat geld in Zgusta (1971: 294) se formulering van die algemene doel van 'n tweetalige woordeboek soos aan die begin van hierdie artikel aangehaal, en dit verteenwoordig die semantiese aspek van Wiegand (2002: 248) se semanties-pragmatiese ekwivalensie.

3.1.2 Kenmerk (b): Die diskoerssituasie oftewel die *sekondêre konteks*

Kenmerk (b) verwys na die breër diskoerssituasie waarin kenmerk (a) geld, dus na die *sekondêre konteks*.

In Van Sterkenburg (2003: 393) word die term *konteks* vir leksikografiese doeleindes soos volg gedefinieer:

- (7) the words, sentences or texts that come before and after a particular lexeme considered relevant to make its meaning clear.

Hoewel die definieërs in (7) voldoende kan wees vir toepassing op verklarende woordeboeke, is dit te eng vir toepassing op tweetalige woordeboeke, aangesien dit by tweetalige woordeboeke in die eerste plek gaan oor die voorsiening van vertaalekwivalente, en nie die verklaring van betekenis nie. Crystal (1991:

78-79) voorsien vier definieë van die term *konteks*, waarvan twee bruikbaar is vir die doel van hierdie bespreking:

- (8)(a) A general term used in linguistics and phonetics to refer to specific parts of an utterance (or text) near or adjacent to a unit which is the focus of attention. The occurrence of a unit (e.g. a sound, word) is partly or wholly determined by its context, which is specified in terms of the unit's relations, i.e. the other features with which it combines as a sequence. The everyday sense of the term is related to this, as when one 'puts a word in context', in order to clarify the meaning intended, as in dictionary entries. Providing a context in this way is referred to as **contextualisation**. [...] Some scholars use the term *co-text* [...].
- (b) A term referring to the features of the non-linguistic world in relation to which linguistic units are systematically used. The term *situation* is also used in this sense, as in the compound term **situational context**. In its broadest sense, situational context includes the *total* non-linguistic background to a text or utterance, including the immediate situation in which it is used, and the awareness by the speaker and hearer of what has been said earlier and of any relevant external beliefs or presuppositions. Further distinctions are usually made in semantics and stylistics, distinguishing, for example, referential and emotive meaning from **contextual meaning**, i.e. information is signalled about the kind of use a linguistic unit has in its social context, e.g. whether it has a 'restricted' use (as in social pleasantries, or religious settings), or how it relates to such factors as age, sex or class of the speakers.

Die definieë in (8)(a) kom ooreen met dié in (7), maar in (8) word die belangrike onderskeid tussen linguistiese konteks (ook *koteks* genoem) en buitetalige of pragmatiese konteks getref. Kharma (1984: 199) onderskei tussen die vergelykbare "context of situation" en "linguistic or verbal context". Jarošová (2000: 18) omskryf *konteks* as "the environment in which the lemma is correlated with the particular equivalent", wat vergelykbaar is met Gouws se formulering oor die aard van die vertaalekwivalent in (6). Beide vorme van sekondêre konteks, te wete *koteks* en pragmatiese konteks, speel 'n rol in ekwivalentdiskriminasie.

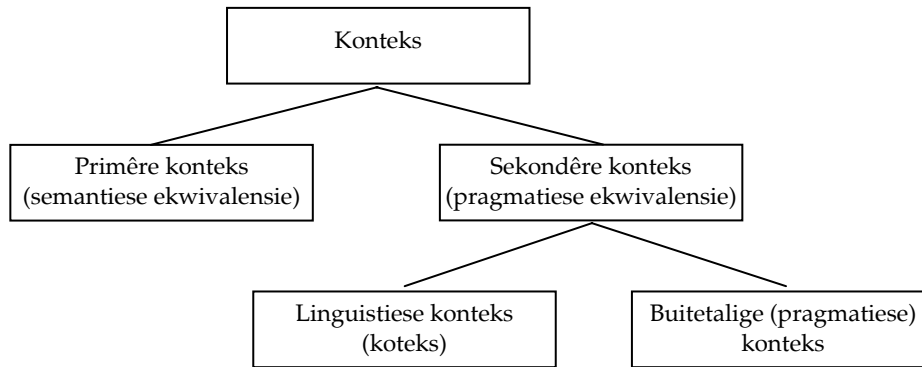
Die sekondêre konteks verteenwoordig die pragmatiese aspek van Wiegand (2002: 248) se semanties-pragmatiese ekwivalensie.

3.2 Die terme *konteks*, *kontekstualisering* en *konteksleiding* leksikografies toegepas

Die inligting in die voorgaande bespreking kan gekonsolideer word om die volgende werksdefinisie vir die term *konteks*, soos leksikografies toegepas, voor te stel:

- (9) Met betrekking tot die tweetalige woordeboek verwys die term *konteks* na beide die primêre en sekondêre kontekste (soos bo verklaar) waarin 'n bepaalde vertaalekwivalent met 'n brontaallemma gekoördineer word.

Die konsep van *konteks*, leksikografies toegepas, kan skematies soos volg weer-gegee word:



Leksikografiese data wat primêre of sekondêre konteks spesifiseer, kan *kontekstuele leksikografiese data* genoem word. Vervolgens kan 'n definisie vir *kontekstualisering* voorgestel word:

- (10) Wanneer die leksikograaf kontekstuele leksikografiese data in ekwivalentdiskriminerende inskrywings enkodeer, vind *kontekstualisering* plaas. Kontekstualisering word in die woordeboek in *kontekstualiserende inskrywings* gemanifesteer.

Dit is moontlik om ekwivalentdiskriminerende inskrywings binne 'n ánder paradigma as dié van konteks aan te bied, met ander woorde ekwivalentdiskriminerende inskrywings is nie by verstek ook kontekstualiserende inskrywings nie: Kontekstualiserende inskrywings verteenwoordig 'n subklas ekwivalentdiskriminerende inskrywings. In die *Superblokraiselwoordeboek* (Pansegrouw 1995) word die volgende artikel aangebied:

- (11) **ent** afstand, end, entplek, gedeelte, immuniseer, inokulasie, (in)okuleer, smetstof, stuk, uiteinde, vaksinasiemerk, ver

Dié woordeboek kan as 'n intratalige vertalende woordeboek beskou word (vgl. Louw 1991: 117). Ten einde die teikengebruiker (die blokraiselspeler) te help om (makliker) ekwivalentdiskriminasie toe te pas, sou die leksikograaf die artikel in (11) kon aanpas om soos volg daar uit te sien:

- (12) **ent** afstand [7], end [3], entplek [7], gedeelte [8], immuniseer [10], inokulasie [10], (in)okuleer [(9) 7], smetstof [8], stuk [4], uiteinde [8], vaksinasiemerk [14], ver [3]

Die leksikograaf sou selfs kon afsien van die oorhoofse alfabetiese ordening van ekwivalente en die woordeboekartikel soos volg struktureer:

- (13) **ent 3:** end, ver **4:** stuk **7:** afstand, entplek, okuleer **8:** gedeelte, smetstof, uiteinde **9:** inokuleer **10:** immuniseer, inokulasie **14:** vaksinasiemerk

Die ekwivalentdiskriminerende inskrywings in (12) en (13) val buite die paradigma van konteks soos leksikografies toegepas (omdat sodanige paradigma nie die teikengebruiker in die betrokke gebruikssituasie van hulp sal wees in ekwivalentdiskriminasie nie), maar binne die paradigma van *die aantal letters waaruit elke ekwivalent bestaan*. Die blokraaiselwoordeboekopsteller wat die doel van sy woordeboek, sy teikengebruiker se behoeftes en die tipiese gebruikssituasie volledig in ag neem, sal weet dat die keuse van 'n ekwivalent by die invul van 'n blokraaisel primêr bepaal word deur die aantal letters waaruit die ekwivalent bestaan. Dit wil nie sê dat die ekwivalente in (11), (12) en (13) onafhanklik van konteks, soos leksikografies toegepas, aangebied word nie — die ekwivalente word soos in 'n intertalige vertalende woordeboek binne hulle primêre kontekste met die lemma gekoördineer en in 'n enkele woordeboekartikel saamgegroeper (maar nie daarvolgens georden nie), want hulle verteenwoordig almal gedeeltelike sinonieme van die (verskillende homonimiese) leksikale items wat deur die lemmata verteenwoordig word (d.w.s. semantiese ekwivalensie geld wel). Die paradigma waarbinne die ekwivalentdiskriminerende inskrywings funksioneer, is egter nie dié van *konteks*, soos leksikografies toegepas, nie.

Die paradigma waarbinne ekwivalentdiskriminerende inskrywings in 'n woordeboek aangebied word, word dus nie arbitrêr nie, maar deur die ware doel en funksie(s) van die woordeboek bepaal. Gegee 'n tweetalige woordeboek se ware doel, sy funksies en die gevolglike strewe na kommunikatiewe ekwivalensie deur die aanbod van semanties-pragmatiese vertaalekwivalente, kan aangeneem word dat *konteks* die paradigma is waarbinne ekwivalentdiskriminerende inskrywings in sodanige woordeboek aangebied sal word.

Gouws (2000: 41) beskou kontekstualisering as 'n kritiese faktor wanneer die kommunikatiewe sukses van 'n tweetalige woordeboek bepaal word. Binne die raamwerk van 'n kommunikatiewe benadering tot 'n gebruikersgerigte leksikografie is dit belangrik om daarop te let dat kontekstualisering deur die leksikograaf nie outomaties tot funksioneel-gefasiliteerde suksesvolle ekwivalentdiskriminasie deur die woordeboekgebruiker lei nie: Kontekstualisering deur die leksikograaf kan ook tot disfunksioneel-gefasiliteerde onsuksesvolle ekwivalentdiskriminasie deur die woordeboekgebruiker lei indien (a) kontekstuele leksikografiese data disfunksioneel deur die leksikograaf in kontekstualiserende inskrywings geënkodeer is en/of (b) kontekstualiserende inskrywings disfunksioneel deur die woordeboekgebruiker gedekodeer word. Scenario (a) impliseer dat daar tydens die leksikografiese proses onderskei kan word tussen *funksionele kontekstualisering* en *disfunksionele kontekstualisering*:

- (14) *Funksionele kontekstualisering* vind plaas wanneer die leksikograaf funksionele kontekstuele leksikografiese data funksioneel in kontekstualiserende inskrywings enkodeer.
- (15) *Disfunksionele kontekstualisering* vind plaas wanneer die leksikograaf disfunksionele kontekstuele leksikografiese data in kontekstualiserende

inskrywings enkodeer en/of funksionele kontekstuele leksikografiese data disfunksioneel in kontekstualiserende inskrywings enkodeer.

Verder kan dit gebeur dat die woordeboekgebruiker aangebode kontekstualiserende inskrywings nie tydens ekwivalentdiskriminasie benut nie as gevolg van gebrekkige naslaanvaardighede, gebrekkige kennis van woordeboekstrukture en -funksies, gebrekkige vertrouwe in die woordeboek, 'n goed ontwikkelde taalintuïsie, ensovoorts, wat sal lei tot óf insidenteel-suksesvolle ekwivalentdiskriminasie óf onsuksesvolle ekwivalentdiskriminasie, maar in elk geval nie tot funksioneel-gefasiliteerde suksesvolle ekwivalentdiskriminasie nie. Vanuit die perspektief van die leksikografiese proses is funksionele kontekstualisering dié prosedure wat die kommunikatiewe sukses van die woordeboek medebepaal. Vanuit die perspektief van woordeboekgebruik is suksesvolle dekodeering van kontekstualiserende inskrywings (d.w.s. funksioneel-gefasiliteerde suksesvolle ekwivalentdiskriminasie) dié prosedure wat die kommunikatiewe sukses van die woordeboek medebepaal. Dit beteken dat die kommunikatiewe sukses van 'n tweetalige woordeboek nie onafhanklik van woordeboekgebruik bepaal kan word nie (hoewel onafhanklik van woordeboekgebruik daarvoor gehipoteiseer kan word). Daarbenewens moet die kommunikatiewe sukses van die woordeboek opnuut deur die uitkoms van elke individuele naslaanprosedure bepaal word. Indien konstant suksesvolle naslaanprosedures as die gevolg van funksioneel-gefasiliteerde suksesvolle ekwivalentdiskriminasie die kommunikatiewe sukses van 'n tweetalige woordeboek sou bevestig, sou dit daarop dui dat in terme van kontekstualisering 'n goed ontwerpte woordeboekspesifieke leksikografiese proses suksesvol toegepas is om die gewenste kenmerke van voorspelbaarheid, berekenbaarheid, ontleedbaarheid en kontroleerbaarheid (vgl. Wiegand 1997) aan die woordeboek te verleen.

Wanneer kommunikatiewe ekwivalensie deur funksionele kontekstualisering bewerkstellig word, het *konteksleiding* plaasgevind:

- (16) *Konteksleiding* vind plaas wanneer funksionele kontekstualisering deur die leksikograaf lei tot funksioneel-gefasiliteerde suksesvolle ekwivalentdiskriminasie deur die woordeboekgebruiker, dit wil sê die woordeboekgebruiker is funksioneel deur die aangebode kontekstualiserende inskrywings *gelei* tot suksesvolle ekwivalentdiskriminasie. Dus: *konteksleiding* = funksionele kontekstualisering + funksioneel-gefasiliteerde ekwivalentdiskriminasie

Vir *konteksleiding* om plaas te vind, is beide die ontwikkeling van 'n gebruikersvriendelike leksikografie en die vestiging van 'n woordeboekkultuur noodsaaklik (vgl. o.m. Gouws 1999).

Die res van hierdie artikel fokus op enkele aspekte van funksionele kontekstualisering.

4. Funksionele kontekstualisering

4.1 Tipes kontekstualiserende inskrywings

Iannucci (1967: 201-203) en Al-Kasimi (1977: 70-71) onderskei die volgende soorte betekenis- en ekwivalentdiskriminerende inskrywings wat in die tweetalige woordeboek kan voorkom:

- (17)(a) *Punktuasie* bied 'n soort negatiewe diskriminasie, deurdat betekenisonderskeidings binne die doeltaalekwivalentparadigma deur middel van kommapunttekens aangedui word, terwyl woordsinonieme binne 'n doeltaalsinoniemparadigma deur kommatekens geskei word.
- (b) *Woordklasaanduidings* differensieer vertaalekwivalente ten opsigte van die verskillende woordsoortelike funksies waarin hulle met die lemma gekoördineer word.
- (c) Leksikografiese definisies (*betekenisbeskrywings*) word aangewend om polisemiese waardes van lemmata te differensieer, dit wil sê om primêre konteks aan te dui.
- (d) *Kontekswoorde* of *-frases* lig die sekondêre konteks uit waarbinne vertaalekwivalente optree wat daardeur bepaal word.
- (e) *Leksikografiese etikette* word gebruik om lemmata en/of bepaalde vertaalekwivalente en/of kontekstinskrywings vir 'n bepaalde pragmatiese konteks te merk.
- (f) Voorbeeldsinne of *-frases* dui konteks aan.

Volgens Wiegand (1990: 34) moet die kommapunte wat vertaalekwivalente vir verskillende betekenisonderskeidings van die lemma onderskei en kommas wat woordsinonieme skei binne die teorie van leksikografiese tekste nie as leestekens beskou word nie, maar as homonieme van die betrokke leestekens, omdat die normale funksies van kommapunte en kommas nie geld nie. Hierdie siening maak die ongekwalfiseerde toepassing van die term *punktuasie* onbruikbaar. Die ";" wat vertaalekwivalente vir die verskillende polisemiese waardes van die lemmateken onderskei, word 'n *polisemie-aanduider* genoem (Wiegand 1996: 5), en dié ";" wat vertaalekwivalente binne 'n doeltaalsinoniemparadigma onderskei, word 'n *nie-tipografiese struktuurmerker*, spesifiek 'n *verdelingsmerker* genoem, waarvan die doel is om die woordeboekartikel se struktuur te ekspliseer deur duidelik identifiseerbare soeksones af te baken (Wiegand 1990: 34). Die merkers ";" en "," in die doeltaalekwivalentparadigmas van Afrikaanse tweetalige woordeboeke word dus gediskwalifiseer as kontekstualiserende inskrywings, aangesien die inskrywings se ware funksies is om die artikelteks te struktureer, en nie om kommentaar op vertaalekwivalente as sodanig te lewer nie. Wiegand se siening word in hierdie studie ondersteun,

wat impliseer dat "puntuasie" hoogstens insidenteel-suksesvolle ekwivalent-diskriminasie kan bewerkstellig.

In die *Groot Woordeboek/Major Dictionary* (Eksteen 1997 — voortaan GW), *Tweetalige Woordeboek/Bilingual Dictionary* (Bosman et al. 1984 — voortaan TW), *Pharos Afrikaans–Engels/English–Afrikaans Woordeboek/Dictionary* (Du Plessis 2005 — voortaan Pharos) en *Reader's Digest Afrikaans–Engelse Woordeboek/English–Afrikaans Dictionary* (Grobbelaar 1987 — voortaan RD) word woordklasaanduiding slegs aangebied indien 'n lemma multifunksioneel ten opsigte van woordklas is. Volgens Wiegand (1988: 567) vervul die aangawe van woordklas twee funksies: eerstens voorsien dit grammatiese data ten opsigte van die daaropvolgende ekwivalente, en tweedens beperk dit die bereik van voorafgaande inskrywings wat grammatikale data weergee oor die woordsoortlike funksie wat die nuwe woordklasaangawe voorafgaan. Die aanduiding van woordklas geld dus nie as 'n ekwivalentdiskriminerende datakategorie nie, hoewel dit as ekwivalentdiskriminerende inligtingskategorie tydens woordeboekgebruik kan funksioneer (vgl. Wiegand 1988: 548), soos in die volgende artikel in die GW:

(18) **subliminal**, (n) onderbewussyn; (a) onderbewus

Die woordklasaanduidende inskrywings "(n)" en "(a)" kan hier bloot insidenteel-suksesvolle ekwivalentdiskriminasie bewerkstellig; dit word duidelik as die volgende artikel uit die GW beskou word, waarin die woordklasaanduidende inskrywings geen ware ekwivalentdiskriminerende funksie kan vervul nie:

(19) **flake**³, (n) vlok; vonk; skilfer; flentertjie; snysel, stukkie; laag, blad; ~ of ICE, ysskots; ~ of SNOW, sneeuvlok; (v) in vlokke val; afskilfer; streep; in dun blaaië afbreek; skoonmaak; pluis; ~d FISH, vlokkies vis, gevlokte vis; ~ OFF, afskilfer [...]

Inskrywings wat woordklas aandui, word op grond van funksie derhalwe gediskwalifiseer as ekwivalentdiskriminerende en gevolglik kontekstualiserende inskrywings.

Die diskwalifikasie van "puntuasie" en woordklasaanduiding as kontekstualiserende inskrywings laat betekenisverklarings, kontekswoorde of -frases, leksikografiese etikette en voorbeeldsinne of -frases as potensiële kontekstualiserende inskrywings.

Die aanwending van betekenisverklarings verteenwoordig die eksplisering van die primêre konteks waarbinne vertaalekwivalente met lemmata gekoördineer word, naamlik semantiese ekwivalensie (i.t.v. Wiegand 2002: 246). Betekenisverklarings behoort dus as kontekstualiserende inskrywings beskou te word.

Volgens Iannucci (1967: 202) het kontekswoorde en -frases die spesifieke doel om as betekenisdiskriminators op te tree. Iannucci (1985: 61) onderskei tussen *betekenisdiskriminators* en *vertaalkomplemente*, waar eersgenoemde die

betekenis van die lemma waaraan dit geadresseer is deur 'n woord of frase weergee, en laasgenoemde bloot 'n bepaalde vertaalekwivalent grammaties voltooi. Vergelyk ter illustrasie die volgende uittreksel uit die GW:

(20) **depose'**, (onder eed) getuig, verklaar; afsit, ontslaan (uit 'n amp).

Die *Concise English Oxford Dictionary* (Pearsall 2001) ken die volgende betekenisonderskeidings aan die lemma **depose** toe:

- (21)(a) 1 remove from office suddenly and forcefully
 (b) 2 *Law* testify or give (evidence) on oath, especially in writing

Die vertaalekwivalent *getuig* word ten opsigte van die tweede betekenisonderskeiding van die lemma **depose** ((21)(b)) aangebied. Die tersaaklike betekenisonderskeiding van die leksikale item *getuig* word soos volg in die *Verklarende Handwoordeboek van die Afrikaanse Taal* (Odendal en Gouws 2005 — voortaan HAT) verklaar:

(22) 1 Getuienis aflê in 'n ondersoek of verhoor

Dit is duidelik dat terwyl die semantiese kenmerk [+ onder eed] 'n integrale deel van die betekenis van die Engelse leksikale item *depose* uitmaak, dit in terme van die GW oënskynlik slegs 'n opsionele element van die betekenis van die Afrikaanse leksikale item *getuig* verteenwoordig, en dat dié deel buite konteks geaktiveer moet word deur die byvoeging van die bepaling *onder eed*. Na regte behoort dié bepaling nie tussen hakies in (20) te verskyn nie, omdat die volledige vertaalekwivalent *onder eed getuig* behoort te wees (vgl. Pharos se akkurater hantering by die lemma **depose**). In Iannucci se terme sal die inskrywing "(onder eed)" as *vertaalkomplement* beskou word, en sal die woordeboekartikel in (20) geen inskrywings ter verklaring van betekenisonderskeidings (d.w.s. betekenisdiskriminatore) bevat nie. Hierdie soort aanbod is moontlik die produk van die woordgerigte neiging wat steeds in die Afrikaanse leksikografie gemanifesteer word.

Gouws (1989: 171) neem die term *vertaalkomplement* by Iannucci oor, maar gebruik dit "effens anders [...] as wat Iannucci voorstel deurdat dit eerder verwys na semantiese konteksleiding" (oftewel: kontekstualisering). Hiervolgens sal die inskrywing "(uit 'n amp)" teenoor die vertaalekwivalent *ontslaan* in (20) as *vertaalkomplement* beskou word, aangesien die inskrywing die gepaste betekenisonderskeiding van die vertaalekwivalent aan die gebruiker duidelik stel, soos wat blyk uit die betekenisverklaring van die lemma **ontslaan** in die HAT:

- (23)(a) 1 Los-, vrylaat uit gevangenskap
 (b) 2 Nie langer in diens hou nie; afdank, in die pad steek
 (c) 3 Vrystel, onthef van

Die vertaalkomplement "(uit 'n amp)" in (20) identifiseer tereg die tweede bete-

kenisonderskeiding ((23)(b)) van die leksikale item *ontslaan* as dié waarde in welke konteks *ontslaan* as vertaalekwivalent vir die brontaalitem *depose* gebruik kan word. Die bepaling *onder eed* by die vertaalekwivalent *getuig* sal hiervolgens bloot as 'n opsionele element van die vertaalekwivalent beskou word (indien die leksikograaf die bewerking as sodanig kan verdedig), en nie as vertaalkomplement nie.

Zgusta (1971: 329, 331, 336-340) onderskei tussen *glosse*, *etikette* en *voorbeeld* as kontekstualiserende inskrywings, en verdeel *glosse* later in *verklarende* en *beperkende glosse* (Zgusta 1987: 6). Roberts (1990: 226) gebruik ook die term *glos*. Svensén (1993: 146-148) onderskei tussen *gebruiksvoorbeelde*, *metalinguistiese inligting* en *ensiklopediese inligting* as tipes kontekstualiserende inskrywings.

Terwyl teoretici soos Iannucci (1967, 1985), Al-Kasimi (1977) en Svensén (1993) hulle terme uit die perspektief van *betekenisdiskriminasie* voorstel met die doel om kontekstleiding te bewerkstellig, wys Manley et al. (1988: 284) op die belangrike uitgangspunt dat een van die hoof funksies van die tweetalige woordeboek diskriminasie tussen *ekwivalente* is, en nie tussen betekenis nie. Volgens hulle moet 'n tweetalige woordeboek *ekwivalensiegestruktureerd* wees, en nie betekenisgestruktureerd nie (Manley et al. 1988: 299). Dit hou in dat sowel die inskrywings in die artikels in 'n tweetalige woordeboek as die struktuur van daardie artikels in die eerste plek daarop gerig moet wees om aan die gebruiker bruikbare vertaalekwivalente te voorsien, en nie om betekenis te beskryf nie. In hierdie verband meld Gouws (1996: 27-28) dat die beginsel van betekenisdiskriminasie wat in verklarende woordeboeke toegepas word, in tweetalige woordeboeke vervang behoort te word deur die beginsel van ekwivalentdiskriminasie, hoewel hierdie benadering uiteraard nie die aanbod van semantiese inligting in tweetalige woordeboeke negeer nie.

Manley et al. (1988) lewer verder kommentaar op die oënskyklik onaanvaarbare mengelmoes van terme wat in die literatuur voorgestel word om tipes kontekstualiserende inskrywings te beskryf en probeer om dit op te ruim deur hulle eie terme voor te stel: die terme is naamlik *glosse* (wat Gouws se *vertaalkomplemente*, Iannucci en Al-Kasimi se *kontekstwoorde en -frases*, Zgusta se *beperkende glosse*, Roberts se *glosse* en Svensén se *gebruiksvoorbeelde* en *ensiklopediese inligting* omvat) en *etikette*. Hierdie twee klasse inskrywings word *diskriminators* genoem (Manley et al. 1988: 290). Grammatiese inligting en voorbeelde word laastens onderskei. Manley et al. (1988) se terminologie maak nie vir betekenisverklarings voorsiening nie.

Vervolgens word bepaal watter tipes kontekstualiserende inskrywings in die GW, TW, Pharos en RD voorkom ten einde 'n klassifikasie vir Afrikaanse tweetalige woordeboeke daar te stel. Vergelyk die volgende verteenwoordigende uittreksels uit die GW in (24):

- (24)(a) **perceiv'ing**, skerpsienende; skerp; waarnemend; skrand, skerpsinnig
percen'tage, persentasie; persentsgewys(e); *on a* ~ *BASIS*, persentsgewys(e); *BY*
 ~s, persentsgewys(e); ~ **point**, persentasiepunt; ~ **error**, persentasiefout.

Percentage point refers to a (decimal) part of a percentage, e.g. if inflation rises by one percentage point (say from 10 to 11%), the actual percentage rise is 10%, while the rise in percentage points is 1%.

per'cept, die waarneembare; waargenomene; waarneming.

perceptibil'ity, waarneembaarheid, merkbaarheid, aanskoulikheid.

percep'tible, waarneembaar, merkbaar, duidelik, aanskoulik.

percep'tion, waarneming, persepsie, besef, gevoel; invordering; insig, begrip; inning, ontvangs, toe-eiening (*jur.*).

percep'tive, opmerkzaam; waarnemings=, gewaarwordings=; ~ **faculty**, waarnemingsvermoë.

percept'iveness, perceptiv'ity, waarnemingsvermoë, insig.

percept'ual, waarnemings=, perseptueel.

perch¹, (n) baars (vis).

perch², (n) **(-es)**, (dwars)stokkie (in 'n voëlkou); slaapstok (hoender); sitplek; veilige plek; roede (5½ jt.); langwa (voertuig); *HOP the ~*, bokveld toe gaan; *KNOCK off one's ~*, uitstof; kafloop; *the bird TAKES its ~*, die voël stryk neer (gaan sit); (v) op 'n stok sit (hoender); neerstryk (voël); hoog en droog sit; *~ed on a hill*, op 'n koppie geleë; **~er**, takslaper, voël wat op 'n tak slaap.

- (b) **habeas corp'us**, habeas corpus (bevelskrif om 'n gevangene voor die hof te bring).

In die artikeldeeltrajek in (24)(a) is dit duidelik dat die vertaalekwivalente wat vir die lemma **perceiving** aangebied word, van geen kontekstualiserende inskrywings voorsien is nie.² Die woordeboekartikel van die lemma **percentage** bevat twee voorbeeldfrases met vertaalekwivalente, naamlik "*on a ~ BASIS*, persentsgewys(e)" en "*BY ~s*, persentsgewys(e)". Hierdie konstruksies verteenwoordig kollokasies en word aangebied met die doel om die gebruiker te help met die aanwending van die lemma of 'n vertaalekwivalent tydens teksproduksie. Tussen die lemmata **percentage error** en **percept** verskyn 'n ingevoegde binneteks (vgl. Louw en Gouws 1996: 96-99), wat verdere ensiklopediese en gebruiksdatabasis oor die lemma **percentage point** verstrek. Die woordeboekartikels van die lemmata **percept**, **perceptibility** en **perceptible** bevat geen kontekstualiserende inskrywings nie. Die laaste vertaalekwivalent in die woordeboekartikel van die lemma **perception** word deur middel van 'n leksikografiese etiket "*jur.*" gekontekstualiseer, wat daarop wys dat die betrokke vertaalekwivalent (*toe-eiening*) binne die regs-konteks as 'n vertaalekwivalent van dié lemma geld. Die vertaalekwivalent wat vir die lemma **perch¹** aangebied word, en 'n aantal van die vertaalekwivalente in die woordeboekartikel van die lemma **perch²** word deur middel van glosse gekontekstualiseer. Die vertaalekwivalent (*dwars*)stokkie word van die glos in '*n voëlkou* voorsien om die pragmatiese konteks aan te dui waarbinne dié ekwivalent die geldende ekwivalent vir die lemma **perch** is, ensovoorts. Die artikel van **perch²** bevat ook twee voorbeeldfrases met vertaalekwivalente, naamlik "*the bird TAKES its ~*, die voël stryk neer (gaan sit)" en "*~ed on a hill*, op 'n koppie geleë". Eersgenoemde is 'n kollo-

kasie, en laasgenoemde kan as 'n "vry" sintagma beskou word waardeur 'n vertaalekwivalent (*geleë*) aangebied word. (Hier geld ooraadressering tussen die koteksafdeling en die vertaalekwivalentparadigma; vgl. Gouws 2000). Die ander inskrywings wat die voorkoms van voorbeeldfrases vertoon, is vaste uitdrukkings; hulle geld dus nie as kontekstualiserende of enige tipe ekwivalent-diskriminerende inskrywings nie en behoort na regte lemmastatus te geniet (vgl. Gouws 1989: 97-108). In die artikel van die lemma **habeas corpus** in (24)(b) word, naas die vertaalekwivalent wat kongruent aan die lemma is, die betekenisverklaring *bevelskrif om 'n gevangene voor die hof te bring* voorsien om die betekenis van die lemma (en die vertaalekwivalent) aan die gebruiker duidelik te maak. Die betekenisverklaring in die doeltaal verskyn in die tipiese artikelposisie van glosse.

(25)

WOORDE IN AKSIE**deist/theist**

Beide die *deist* en die *theist* glo in God. Eersgenoemde glo egter nie dat God Hom aan die mens openbaar nie, en ook nie dat 'n persoonlike verhouding met God moontlik is nie, terwyl laasgenoemde dit meestal wel glo.

deify v. **vergoddelik** (fâRcho'dâlâk) exalt to divine rank, personify as a god; **vereer** (fâReer') worship, idolise; **aanbid** (unbât') adore, idolise (e.g. a person).

dejected a. **neerslagtig** (neeRsluch'tâch) depressed, despondent, downcast, disconsolate, in low spirits; **bedruk** (bâdRâk') depressed, heavy-hearted, down in the dumps; **moedeloos** (môô'dâloows) crestfallen, despondent, disheartened; **terneergedruk** (târneeR'châdRâk) and (unus.) **terneergeslae** (târneeR'châsluâ) downhearted, depressed, downcast; **mistroostig** (mâstRoows'tâch) sick at heart, sad; **verslae** (fâRsluâ) downcast, downhearted, low-spirited; **bek-af**

dehorn' horings verwyder of afsaag, onthoring. **dehorn'ed cattle** onthoringde beeste; poenskoppe, poenskop beeste. **dehorn'er** onthoornier. **dehorn'ing** onthooring. **dehorn'ing paste** onthoorningsalf.

dehum'anise, **ize** ontmens, verdierlik, tot 'n dier maak. **dehumanisa'tion**, **zation** ontmens(lik)ing, verdierliking.

dehumid'ify ontvog(tig). **dehumidifica'tion** ontvogt(ig)ing. **dehumid'ifier** ontvogt(ig)er.

dehusk' (uit)dop; (af)dop, pel; ~ *maize* mielies afmaak. *Vid. husk v.*

dehy'drate ontwater, water onttrek, dehidr(a)-teer, anhidreer. **dehy'drant** wateronttrekkende middel. **dehydra'tion** ontwatering, wateronttrekking, dehidrasie, dehidr(at)ering, dehidra-(ta)sie. **dehy'drator**, **dehydrat'or** ontwateraar, ontwatertoestel

dehydro'genate, **dehydro'genise**, **ize** dehidrogeneer. **dehydrogen(is)a'tion**, **(iz)a'tion** dehidrogenering, dehidrogenasie.

de-ice' ontyser. **de-ic'er** ontyser, ysbestryder.

de'icide godsmoord; godsmoordenaar.

deic'tic aanwysend, deikties.

de'ify vergoed; vergoddelik; vereer, aanbid. **deifica'tion** vergoding; vergoddeliking. **de'ifier** vergoder. **de'iform** goddelik van gedaante.

deign (sig) verwerdig, toelaat.

Deir'dre Deirdre.

de'ist n. deïs. **de'ist a.**, **deis'tic(al)** deïsties. **de'ism** deïsme. **de'ity** godheid; goddelikheid.

deject' neerslagtig maak, ontmoedig. **deject'a** (mv.) uitwerpsels (*med.*). **deject'ed** neerslagtig, bedruk, moedeloos, terneergedruk, terneergeslae, mistroostig, verslae, bek-af, triestig, troosteloos, hangkop. **deject'edly** kop-onderstebo. **dejec'tion** neerslagtigheid, bedruktheid, terneergedruktheid, mistroostigheid, moedeloosheid, verslaen(t)heid; (terneer)geslaenheid; ont-

(bek'uf) down in the dumps, crestfallen; **triestig** (tRis'tách) gloomy, dismal, miserable; **troosteloos** (tRoows'táloows) disconsolate, inconsolable, forlorn; **hangkop** (hung'kop) hangdog, downcast, dejected.

lasting (*med.*).
de ju're (*Lat.*) de jure, van regsweë.
dek'a *vid. deca*.
dek'ko *dekkos, (sl.)* kykie.
delab'ialise, *size* ontrond (*fon.*). **delabialisa'tion**, *za'tion* ontronding.
delacta'tion spening; droog word (*van moeder*).

In (25) verteenwoordig die regterkantste kolom die primêre makrostruktuur in die RD wat die woordeboekartikels bevat. Leksikografiese etikette kom voor in die artikels van die lemmata **dejecta**, **dejection**, **de jure**, **dekko** en **delabialise**: die vertaalekwivalente *uitwerpsels* en *ontlasting* word gemerk as ekwivalente binne die pragmatiese konteks van mediese vaktaal; by die lemma **de jure** word aangedui dat dié lemma uit Latyn afkomstig is; die vertaalekwivalent *kykie* in die artikel van die lemma **dekko** word as behorende tot slengtaal gemerk; die vertaalekwivalent *ontrond* word gemerk as behorende tot die vaktaal van die fonologie. Die vertaalekwivalent *droog* word in die artikel van die lemma **delactation** van die glos *van moeder* voorsien om die pragmatiese konteks aan te dui waarbinne dit as vertaalekwivalent geld. Die linkerkantste kolom bevat ingevoegde tekste: eerstens, 'n ingevoegde teks wat addisionele semantiese en gebruiksdata oor die lemma **deist** onder die opskrif "WOORDE IN AKSIE" binne 'n tekskassie bevat (as deel van 'n sekondêre makrostruktuur); tweedens, elemente van 'n verdere sekondêre makrostruktuur waarin vertaalekwivalente soos *vergoddelik*, *vereer* en *aanbid* (wat in die primêre mikrostruktuur as vertaalekwivalente vir die lemma **deify** aangebied word) asook *neerslagtig*, *bedruk*, *moedeloos*, ensovoorts (wat in die primêre mikrostruktuur as vertaalekwivalente vir die lemma **dejected** aangebied word) van uitspraak-inligting en terugvertalings voorsien word (vgl. Louw en Gouws 1996: 94-96; Gouws 1996: 29-30). Hierdie ingevoegde tekste vorm nie deel van die mikrostruktuur van die woordeboek of onmiddellike stukke van die sentrale lys nie, maar het 'n makrostrukturele optrede as deel van die supplementêre teks (Louw en Gouws 1996: 97). In teenstelling hiermee vorm die sogenaamde "tekskassies" in die *Tweetalige Aanleerderswoordeboek/Bilingual Learner's Dictionary* (Du Plessis 1993) en *Basiswoordeboek van Afrikaans* (Gouws et al. 1994) telkens deel van die betrokke artikels en figureer dan as *ingevoegde binnetekste* (vgl. Gouws 1996: 23-24) — daar is dus nie in hierdie woordeboeke van supplementêre tekste sprake nie; vergelyk die onderstaande uittreksels uit *Basiswoordeboek van Afrikaans* in (26)(a) en *Tweetalige Aanleerderswoordeboek/Bilingual Learner's Dictionary* in (26)(b):

(26) (a)

oorhandig (**het oorhandig**) *wv*. Wanneer 'n mens iets oorhandig, gee jy dit vir iemand. *Die skoolhoof se vrou het die pryse by die prysuitdeling oorhandig. Hierdie dokumente moet persoonlik aan hom oorhandig word. Oorhandig 'n bietjie daardie briefie wat Susan nou net vir*

(b)

dogter daughter, girl *My eldest sister is married and has one daughter/girl and two sons/boys. My oudste suster is getroud en het een dogter en twee seuns.*
 dog-ter *selfstandige naamwoord (meervoud dogters)*

jou gegee het. Die oorhandiging van die nuwe seiljag aan sy bemanning vind vanaand om agtuur plaas. oorhandiging

Oorhandig is 'n taamlik formele woord. In die omgangstaal sal 'n mens nooit hoor: *Oorhandig daardie appel vir my, want ek is honger, nie, maar wel: Gee vir my die appel aan, want ek is honger.*

dogter slaan gewoonlik op iemand se kind en **meisie** op iemand wat jonger as 'n vrou is, maar die twee woorde is uitruilbaar wanneer jy praat van iemand wat nog op skool of wat kleiner en jonger as 'n meisie is: *skooldogter* of *skoolmeisie*, 'n oulike *dogtertjie* of *meisietjie*.

oorhê (het oor, sal oorhê, het oorgehad, om oor te hê) *ww.* Wanneer 'n mens iets oorhet, hou jy iets oor nadat jy gebruik het wat jy nodig gehad het. *Het jy dalk nog geld oor wat jy vir my kan leen, want teen die einde van die maand het ek gewoonlik niks meer oor nie. Sy het nog soveel materiaal oorgehad nadat sy die rok gemaak het dat sy nog 'n bloes ook daaruit kon maak.*

dogtertjie ① *little girl. They have two children: a boy of eight and a little girl of four.* Hulle het twee kinders: 'n seun van agt en 'n **dogtertjie** van vier. ② *baby girl. My sister had a baby girl last week. My suster het verlede week 'n dogtertjie gekry.*

□ **dog-ter-tjie** *selfstandige naamwoord (meer-voud dogtertjies)*

Op grond van die identifikasie van die tipes kontekstualiserende inskrywings wat in die data (GW, TW, Pharos en RD) voorkom, soos hierbo aangetoon, en die voorafgaande bespreking, kan kontekstualiserende inskrywings soos volg geklassifiseer word:

- (27)(a) *Diskriminators*, waaronder
- (i) *betekenisverklarings*,
 - (ii) *leksikografiese glosse*, en
 - (iii) *leksikografiese etikette*.
- (b) *Koteksinskrywings* (d.w.s. voorbeeldsinne en -frases), waaronder
- (i) *kollokasies*, en
 - (ii) *vry konstruksies*.
- (c) *Ingevoegde tekste*, waaronder
- (i) *ingevoegde binnetekste* (soos "tekscassies" as deel van artikels), en
 - (ii) *ingevoegde buitetekste* (soos illustrasieprente, wat buite artikelgrense funksioneer).

Die onderskeid tussen *diskriminators*, *koteksinskrywings* en *ingevoegde tekste* word gemotiveer op grond van leksikografiese funksie, wat die verspreidings- en adresseringstrukture van dié inskrywings bepaal.

4.2 Verspreidingstruktuur

Diskriminators tree hoofsaaklik in samehang met lemmata en vertaalekwivalente in die semantiese (sub)kommentaar op. Koteksinskrywings kan binne 'n nie-geïntegreerde mikrostruktuur 'n afsonderlike integraat konstitueer, maar

tree in elk geval binne die koteksafdeling op, wat nie sonder meer vir diskriminatorters geld nie. Volgens Gouws (2000a: 4) vind ekwivalentdiskriminasie, en dus kontekstualisering, primêr op artikelinterne vlak plaas. Dit veronderstel dat kontekstualiserende inskrywings hoofsaaklik in die sentrale lys sal voorkom en die kommunikasie-georiënteerde funksies van die woordeboek sal ondersteun (vgl. Gouws en Steyn 2005). Die data toon ook dat kontekstualiserende inskrywings op mikrostrukturele vlak nie onderling eksklusief funksioneer nie, maar dat meer as een tipe met identiese adresse binne 'n enkele woordeboekartikel en selfs binne 'n enkele integraat kan optree (vgl. Beyer 2006a).

Ingevoegde tekste tree óf binne, óf buite die primêre makrostruktuur en óf artikelintern, óf artikelekstern op. Ingevoegde buitetekste moet van buitetekste onderskei word. Terwyl ingevoegde buitetekste in die sentrale lys optree (d.w.s. by die sentrale lys *ingevoeg* is), tree buitetekste óf voor die sentrale lys as voortekste, óf na die sentrale lys as agtertekste op (vgl. Gouws 2004). Hoewel buitetekste ook die kommunikasie-georiënteerde funksies van die woordeboek kan ondersteun, vervul hulle nie direk en in situ 'n kontekstualiserende funksie nie, hoewel die potensiele rol wat hulle wel in konteksleiding kan speel nie genegeer moet word nie; sodanige rol kan veral deur 'n funksionele mediostruktuur uitgebou word. Vergelyk onder meer Gouws (2004) en Gouws en Steyn (2005) vir besprekings oor buitetekste.

Hoewel geen illustrasieprente in die GW, TW en Pharos voorkom nie, en in die RD tot die agtertekste beperk is, word hulle potensiele optrede in die vorm van ingevoegde tekste as kontekstualiserende inskrywings nie genegeer nie.

Iannucci (1967: 205) stel 'n ekstra-makrotekstuele vorm van kontekstualisering as moontlikheid voor. Dit behels die koördinasie van 'n tweetalige woordeboek met 'n eentalige verklarende woordeboek deur 'n stelsel van nommerverwysings waardeur die semantiese kommentaar in die verklarende woordeboek as kontekstualisering van die ooreenstemmende lemmata en vertaalekwivalente in die tweetalige woordeboek kan dien. Hoewel die implementering van so 'n voorstel nie onmoontlik is nie, erken Iannucci (1967: 205-206) self dat beduidende praktiese probleme met so 'n metode van kontekstualisering kan opduik. Sy voorstel word ook deur Al-Kasimi (1977: 74) as "cumbersome and impractical" gekritiseer, en word hier ook nie verder oorweeg nie.

Soos wat die klassifikasie in (27) suggereer, funksioneer diskriminatorters en koteksinskrywings op artikelinterne vlak, terwyl ingevoegde binnetekste binne of buite artikelgrense oftewel die mikrostruktuur funksioneer en ingevoegde buitetekste eksklusief daarbuite. Die klas ingevoegde binnetekste kan onderverdeel word in *artikelinterne ingevoegde binnetekste* en *artikeleksterne ingevoegde binnetekste*. Eersgenoemde tipe tree binne artikels (as deel van die mikrostruktuur) op, terwyl laasgenoemde tussen artikels optree. Ingevoegde buitetekste tree ook in die sentrale lys (en uitsluitlik artikelekstern) op, maar daar bestaan geen onmiddellik aantoonbare verband tussen sodanige buitetekste en enige woordeboekartikels in die sentrale teks nie, dit wil sê die ingevoegde buitetekste is aan geen spesifieke artikels geadresseer nie.

4.3 Adresseringstruktuur

Die adresseringstrukture wat ten opsigte van kontekstualiserende inskrywings kan geld, is 'n komplekse gegewe wat buite die bestek van hierdie artikel val. Vergelyk Gouws (2000) en Beyer (2006a) vir relevante besprekings.

5. Slot

Die wetenskaplike studie van kontekstualisering en konteksleiding in tweetaalige woordeboeke behoort te begin met 'n teoretiese fundering van dié verskynsel. Sodanige fundering behoort 'n raamwerk te skep vir sistematiese beskrywing en verdere teorievorming wat hopelik in die praktiese leksikografie in meer gebruikersvriendelike en derhalwe bemagtigende woordeboeke neerslag sal vind. Hierdie artikel is 'n poging om 'n bydrae tot sodanige teoretiese fundering te lewer. Die potensiaal wat 'n kommunikatiewe benadering in 'n gebruikersgerigte metaleksikografie kan inhou, word ook gedemonstreer.

Notas

1. Die konsepte *enkodering* en *dekodering* soos in hierdie artikel toegepas moet nie verwar word met die gebruik van die identiese terme deur o.m. Manley et al. (1988) as sinonieme vir onderskeidelik *teksproduksie* en *teksresepsie* in die beskrywing van woordeboekfunksies nie.
2. Kontekstualisering word nie in hierdie bespreking beoordeel nie.

Bronnelys

Primêre literatuur (woordeboeke)

- Bosman, D.B., I.W. van der Merwe en L.W. Hiemstra.** 1984⁸. *Tweetaalige Woordeboek/Bilingual Dictionary*. Kaapstad: Tafelberg.
- Du Plessis, M.** 1993. *Tweetaalige aanleederswoordeboek/Bilingual Learner's Dictionary*. Kaapstad: Tafelberg.
- Du Plessis, M. (Red.).** 2005. *Pharos Afrikaans–Engels/Engels–Afrikaans Woordeboek/Dictionary*. Kaapstad: Pharos.
- Eksteen, L.C.** 1997¹⁴. *Groot Woordeboek/Major Dictionary*. Kaapstad: Pharos
- Gouws, R, I. Feinauer en F. Ponelis.** 1994. *Basiswoordeboek van Afrikaans*. Pretoria: J.L. van Schaik.
- Grobbelaar, P. (Red.).** 1987. *Reader's Digest Afrikaans–Engelse Woordeboek/English–Afrikaans Dictionary*. Kaapstad: The Reader's Digest Association of South Africa.
- Odendal, F.F. en R.H. Gouws.** 2005⁵. *Verklarende Handwoordeboek van die Afrikaanse Taal*. Kaapstad: Pearson Education South Africa.
- Pansegrouw, L.L.** 1995. *Superblokraaiselwoordeboek*. Pretoria: J.L. van Schaik.
- Pearsall, J. (Red.).** 2001¹⁰. *The Concise Oxford English Dictionary*. Oxford: Oxford University Press.

Sekondêre literatuur

- Al-Kasimi, A.M.** 1977. *Linguistics and Bilingual Dictionaries*. Leiden: E.J. Brill.
- Beyer, H.L.** 2006. 'n Voorlopige model vir die sistematiese beskrywing van gebruikersvriendelike in woordeboeke. *Lexikos* 16: 46-66.
- Beyer, H.L.** 2006a. 'n *Metaleksikografiese ondersoek na konteksleiding in Afrikaanse vertalende woordeboeke*. Ongepubliseerde D.Litt.-proefskrif. Stellenbosch: Universiteit van Stellenbosch.
- Crystal, D.** 1991³. *A Dictionary of Linguistics and Phonetics*. Oxford: Blackwell.
- Gouws, R.H.** 1989. *Leksikografie*. Pretoria/Kaapstad: Academica.
- Gouws, R.H.** 1996. Bilingual Dictionaries and Communicative Equivalence for a Multilingual Society. *Lexikos* 6: 14-31.
- Gouws, R.H.** 1999. *Die maatskaplike gerigtheid van die metaleksikografie in 'n meertalige samelewing*. Intreerede. Stellenbosch: Universteit van Stellenbosch Drukkery.
- Gouws, R.H.** 2000. Doeltaalgerigtheid teenoor lemmagerigtheid in vertalende woordeboeke. *Tydskrif vir Geesteswetenskappe* 40(1): 39-47.
- Gouws, R.H.** 2000a. Strategies in Equivalent Discrimination. Zettersten, A. and V.H. Pedersen. (Eds.). 2000. *Symposium on Lexicography IX. Proceedings of the Ninth International Symposium on Lexicography, April 23–25, 1998 at the University of Copenhagen*: 1-23. Tübingen: Max Niemeyer.
- Gouws, R.H.** 2004. Outer Texts in Bilingual Dictionaries. *Lexikos* 14: 67-88.
- 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.
- Hausmann, F.J., O. Reichmann, H.E. Wiegand en L. Zgusta. (Reds.)**. 1989–1991. *Wörterbücher. Ein Internationales Handbuch zur Lexikographie/Dictionaries. An International Encyclopedia of Lexicography/Dictionnaires. Encyclopédie internationale de lexicographie*. Berlyn/New York: Walter de Gruyter.
- Iannucci, J.E.** 1967. Meaning Discrimination in Bilingual Dictionaries. Householder, F.W. en S. Saporta (Reds.). 1967. *Problems in Lexicography*: 201-216. The Hague: Mouton.
- Iannucci, J.E.** 1985. Sense Discriminations and Translation Complements in Bilingual Dictionaries. *Dictionaries* 7: 57-65.
- Jarošová, A.** 2000. Problems of Semantic Subdivisions in Bilingual Dictionary Entries. *International Journal of Lexicography* 13(1): 11-28.
- Kharm, N.N.** 1984. Contextualization and the Bilingual Learner's Dictionary. Hartmann, R.R.K. (Red.). 1984. *LEXeter '83 Proceedings. Papers from the International Conference on Lexicography in Exeter, 9–12 September 1983*: 199-206. Tübingen: Max Niemeyer.
- Kromann, H., T. Riiber en P. Rosbach.** 1991. Principles of Bilingual Lexicography. Hausmann, F.J. et al. (Reds.). 1989–1991: 2711-2728.
- Louw, J.P.** 1991. Semantiese norme in die leksikografie. *Suid-Afrikaanse Tydskrif vir Taalkunde* 9(4): 115-118.
- Louw, P.A. en R.H. Gouws.** 1996. Lemmatiese en nielemmatiese adressering in Afrikaanse vertalende woordeboeke. *Suid-Afrikaanse Tydskrif vir Taalkunde* 14(3): 92-100.

- Manley, J., J. Jacobsen en V.H. Pedersen.** 1988. Telling Lies Efficiently: Terminology and the Microstructure in the Bilingual Dictionary. Hyldgaard-Jensen, K. en A. Zettersten (Reds.). 1988. *Symposium on Lexicography III. Proceedings of the Third International Symposium on Lexicography, May 14–16, 1986, at the University of Copenhagen*: 281-302. Tübingen: Max Niemeyer.
- Roberts, R.P.** 1990. Organization of Information in a Bilingual Dictionary Entry. *EURALEX '90 Proceedings. IV International Congress*: 219-231. Barcelona: Bibliograf.
- Svensén, B.** 1993. *Practical Lexicography. Principles and Methods of Dictionary-Making*. Oxford: Oxford University Press.
- Tarp, S.** 2000. Theoretical Challenges to Practical Specialised Lexicography. *Lexikos* 10: 189-208.
- Van Sterkenburg, P. (Red.)**. 2003. *A Practical Guide to Lexicography*. Amsterdam/Philadelphia: John Benjamins.
- Wiegand, H.E.** 1988. "Shanghai bei Nacht." Auszüge aus einem metalexikographischen Tagebuch zur Arbeit beim Großen Deutsch–Chinesischen Wörterbuch. *Germanistische Linguistik* 89–90. Wiegand, H.E. (Red.). 1988. *Studien zur neuhochdeutschen Lexikographie* 6(2): 521-626.
- Wiegand, H.E.** 1989. Der gegenwärtige Status der Lexikographie. Hausmann, F.J., et al. (Reds.). 1989–1991: 246-280.
- Wiegand, H.E.** 1990. Printed Dictionaries and their Parts as Texts. An Overview of More Recent Research as an Introduction. *Lexicographica* 6: 1-126.
- Wiegand, H.E.** 1996. Das Konzept der semiintegrierten Mikrostrukturen. Ein Beitrag zur Theorie zweisprachiger Printwörterbücher. Wiegand, H.E. (Red.). 1996. *Wörterbücher in der Diskussion II. Vorträge aus dem Heidelberger Lexikographischen Kolloquium*: 1-81. Tübingen: Max Niemeyer.
- Wiegand, H.E.** 1997. Die gesellschaftliche Verantwortung der wissenschaftlichen Lexikographie. *Hermes* 18: 117-202.
- Wiegand, H.E.** 2002. Equivalence in Bilingual Lexicography: Criticism and Suggestions. *Lexikos* 12: 239-255.
- Zgusta, L.** 1971. *Manual of Lexicography*. The Hague: Mouton.
- Zgusta, L.** 1987. Translational Equivalence in a Bilingual Dictionary. *Dictionaries* 9: 1-47.

Ensiklopedisiteit in die *Woordeboek van die Afrikaanse Taal*: 'n "Saak" van balans*

Willem Botha, *Buro van die Woordeboek van die Afrikaanse Taal, Stellenbosch, en Departement Afrikaans en Nederlands, Universiteit Stellenbosch, Stellenbosch, Republiek van Suid-Afrika (wfb@sun.ac.za)*

Opsomming: Die *Woordeboek van die Afrikaanse Taal* (voortaan WAT) het vanaf die negende deel wegbeweeg van 'n meer ensiklopediese aanbod van inligting. Hiervoor het die woordeboek heelwat lof van teoretici ontvang. Die mening is egter ook uitgespreek dat die WAT wat leksikografiese funksie betref meer op die kognitiewe funksies, wat veral ensiklopediese inligting behels, moet konsentreer. In reaksie hierop word daar probeer om vas te stel of die WAT vanaf die negende deel 'n balans gevind het tussen ensiklopediese en semantiese inligting. Vir hierdie doel word daar gefokus op Wiegand se sienings rondom die hantering van ensiklopediese inligting in verklarende woordeboeke asook die meer tradisionele beginsel van komponentontleding.

Sleutelwoorde: ENSIKLOPEDIËSE INLIGTING, KOGNITIEWE FUNKSIE, KOMMUNIKATIEWE FUNKSIE, KOMPONENTONTLEDING, LEKSIKOGRAFIESE FUNKSIE, POLIFUNKSIONEEL, SEMANTIESE INLIGTING

Abstract: From the ninth volume, the *Woordeboek van die Afrikaanse Taal* (henceforth WAT) moved away from a more encyclopaedic presentation of information. Theorists reacted positively to this change of direction. The opinion was, however, also expressed that the WAT, as far as lexicographic function is concerned, should concentrate more on cognitive functions, which include especially encyclopaedic information. In response to this opinion, an effort is made to establish whether the WAT has from volume nine succeeded in finding a balance between encyclopaedic and semantic information. For this purpose the focus is on Wiegand's views on the presentation of encyclopaedic information in explanatory dictionaries as well as the more traditional principle of component analysis.

Keywords: COGNITIVE FUNCTION, COMMUNICATIVE FUNCTION, COMPONENT ANALYSIS, ENCYCLOPAEDIC INFORMATION, LEXICOGRAPHIC FUNCTION, POLIFUNCTIONAL, SEMANTIC INFORMATION

* Hierdie artikel is 'n verwerking van 'n hoofstuk uit 'n D.Litt.-proefskrif *Die impak van die leksikografieteorie op die samestelling van die Woordeboek van die Afrikaanse Taal*, wat in April 2003 deur die Universiteit Stellenbosch, Stellenbosch, Republiek van Suid-Afrika, aanvaar is.

1. Inleiding

Die eerste ag dele van die *Woordeboek van die Afrikaanse Taal* (voortaan WAT) is hewig gekritiseer vanweë die ensiklopediese aanbod van inligting in hierdie dele. Die WAT het egter op hierdie kritiek ag geslaan en vanaf Deel IX is daar nie weer na die ensiklopediese aard van artikels verwys nie, maar is die algemene bondigheid waarmee inligting aangebied word, aangeprys. Na aanleiding van 'n verrassende stelling van Feinauer (2007) dat die WAT meer op die kognitiewe (wat veral ensiklopediese en etimologiese inligting behels) moet konsentreer, word daar in hierdie artikel probeer om die mate van ensiklopedisiteit in dele IX tot XII van die WAT vas te stel.

2. Kritiek op die ensiklopediese element in die WAT

2.1 WAT I tot VI

Uit 'n ondersoek na die kritiek op die eerste ses dele van die WAT blyk dit dat die kritiek teen die ensiklopediese betekenisverklarings progressief feller geword het totdat dit met die verskyning van Deel VI 'n hoogtepunt bereik het (Botha: 2003: 83).

Die resensente is dit eens dat die aanbieding van saakbeskrywing in plaas van betekenisbeskrywing, die oorsaak van die WAT se ensiklopediese definisies is. Snijman (1995a: 210-211) erken dat ekstra-linguistiese faktore by die redaksie 'n rol speel in die bepaling van die hoeveelheid saakkenmerke wat vermeld word, en dat verbesondering oordryf kan word. Hy verstrekk egter ook insiggewende motivering vir die insluiting van skynbaar oortollige saakkenmerke by 'n simpleks, soos dat die saakkenmerke noodsaaklik is vir begrip van sommige samestellings of uitdrukkings waarin dit voorkom.

Snijman (1995a: 205-206) sluit hom aan by Zgusta (1971: 41) se beskouing dat daar dikwels 'n ooreenstemming is tussen die toepassingsveld van 'n woord en die kenmerkende eienskappe van die saak wat deur die woord benoem word.

Schoonees (1952: 289), hoofredakteur van Deel I tot IV, sê onder meer dat die redaksie "soveel as moontlik probeer gee het". Dit het "van Taalkundiges die verwyrt van ensiklopedisiteit gebring, van die kant van die praktiese gebruiker egter nog slegs ingenomenheid en aanmoediging — en dit tel ook!" Ook elders verklaar Schoonees (1958: 81-83) dat die WAT probeer gee wat die gebruikers verlang en as die woordeboek daardeur ensiklopedies word, "dan wil ons graag ensiklopedies wees". Schoonees (1957: 122) konstateer ook dat sekere woorde feitlik "ensiklopediese besonderhede vereis" om presies begryp en gehanteer te kan word. Dié woorde "is ryk aan individuele onderskeidings, soos kenmerke, en arm aan algemene onderskeidings ... Woorde wat sulke dinge benoem, (is) gebind aan die dinge met hulle individuele onderskeidings ... As die omskrywing presies wil wees, moet dit ook met die benoemde saak

self rekening hou." Met laasgenoemde twee stellings nader hy 'n teoretiese fundering van die ensiklopediese element in die WAT, maar die bevrediging van die behoeftes van die gebruiker bly sy primêre motivering vir die insluiting van ensiklopediese elemente.

Terwyl dit vir Schoonees om die behoeftes van die gebruiker gegaan het en hy die ensiklopediese op hierdie basis geregverdig het, het Snijman (1995a, 1995b, 1995c, 1995d) die ensiklopediese werkwyse van die WAT deurgaans teoreties probeer verantwoord. Hierin het hy egter konsekwent 'n teenoorgestelde standpunt ingeneem as die teoretici wat die kritiek uitgespreek het. So gebruik Eksteen (1968) Reichling se uitgangspunte om die werkwyse van die WAT te kritiseer terwyl Snijman (1995b) dit gebruik om die werkwyse te motiveer. In onlangse metaleksikografiese gesprekke word aangesluit by Snijman se standpunt dat die ensiklopediese in verklarende woordeboeke nie net benader kan word uit 'n blote kontrasstelling tussen leksikale betekenis en saak nie. Tarp (2000: 190) redeneer dat dit by 'n bespreking van kwessies soos saakbeskrywing en betekenisbeskrywing eerstens moet gaan oor die funksie van die spesifieke woordeboek en die bevrediging van die behoeftes van die gebruiker. Die pragmatiese moet dus voorop gestel word, maar dit moet terselfdertyd deur die teorie gelei word. Volgens Swanepoel (1995: 209) slaag Snijman daarin om verskeie van die kompleksiteite van die linguisties-teoretiese kwessies en die talle leksikografies-pragmatiese faktore wat by die ensiklopediese op die spel is, na vore te bring.

Alhoewel Snijman (1977: 17; 1995d: 185) na aanleiding van die kritiek op die ensiklopediese element in Deel VI verskeie praktiese oplossings voorgestel het, was dit duidelik dat hy nie ten gunste was van ingrypende veranderinge wat die eenheidskarakter van die WAT kon skend nie. Hierin is 'n mate van ironie aanwesig, aangesien die ensiklopediese karakter van Deel I geensins die afmetings aanneem van dié van Deel VI nie.

2.2 WAT VII tot VIII

Kritici is dit eens dat ten spyte van die WAT se poging om die ensiklopediese element te beperk, dele VII en VIII steeds gekenmerk word deur saakbeskrywing eerder as betekenisbeskrywing (Botha 2003: 107).

Volgens Gouws (1985 en 1992), Feinauer (1993), Raidt (1993) en Botha (1984) is daar 'n merkbare verbetering in deel VII en veral deel VIII ten opsigte van die ensiklopediese, maar is daar nog steeds gevalle waar gebruiks- en saakbeskrywing as betekenisbeskrywing aangebied word.

2.3 Herontwerpte WAT vanaf deel IX

Dit wil voorkom asof ensiklopedisiteit as verskynsel vanaf Deel IX finaal besweer is. Geen resensent noem ensiklopedisiteit as 'n negatiewe punt van kritiek teen enige van die dele vanaf WAT IX nie.

Volgens Gouws (1994) bewys WAT IX dat die redaksie op die kritiek ten opsigte van die ensiklopediese ag geslaan het. Oor die algemeen is die betekenisverklarings sins insiens nie net bondiger as vroeër nie, maar dit bevat ook baie minder ensiklopediese inligting. In hierdie opsig vertoon WAT IX baie beter as sy voorgangers.

In sy resensie van Deel X wys Gouws (1996) weer eens daarop dat die weergawe van 'n oormatige hoeveelheid ensiklopediese inligting 'n tradisionele punt van kritiek teen die WAT was. Hy konstateer voorts dat hierdie probleem in Deel X opgelos is en dat die artikels 'n deeglike betekenisomskrywing van die lemma bied. Hy maak egter die terloopse opmerking dat "die verbetering so ver gevoer (is) dat 'n mens soms selfs voel dat 'n bietjie meer inligting van waarde sou gewees het".

In 'n resensieartikel van WAT XII lewer Feinauer (2007: 269-270) 'n pleidooi vir meer ensiklopediese inligting in die WAT-definisies. Hiermee het die kritiek rondom ensiklopedisiteit in die WAT die kringloop voltooi: vanaf 'n hewige veroordeling van die ensiklopedisiteit na 'n aanprysing van die bondige definisies met deeglike betekenisbeskrywing sonder 'n oormaat ensiklopediese inligting en 'n uiteindelijke pleidooi vir 'n meer ensiklopediese benadering.

Feinauer (2007) beroep haar op Bergenholz en Tarp (2003: 176) wat 'n onderskeid tref tussen die kognitiewe en kommunikatiewe funksies van 'n woordeboek. Die behoefte van die gebruiker bepaal watter van hierdie funksies die sterkste in 'n bepaalde woordeboek behoort te figureer. Die kognitiewe of kennisgerigte funksies behels veral ensiklopediese en etimologiese inligting, spesifieke inligting oor die register of vakgebied van die lemma en grammatikale inligting. Een van die tipiese kommunikatiewe funksies is uitspraakleiding. Volgens Feinauer (2007) is die hoof funksie van die WAT as omvattende woordeboek waarskynlik meer kognitief as kommunikatief. Die WAT moet derhalwe meer op ensiklopediese en etimologiese inligting gerig wees as op byvoorbeeld uitspraakleiding. Vir laasgenoemde moet die gebruiker eerder standaard verklarende woordeboeke, aanleerderwoordeboeke of vertalende woordeboeke raadpleeg.

In teenstelling met bostaande standpunt van Feinauer dat die WAT meer kognitief as kommunikatief behoort te wees, kan dit waarskynlik eerder gekonstateer word dat 'n omvattende woordeboek soos die WAT nie kan kies om meer kognitief of kommunikatief te wees nie, maar moet probeer om op gebalanseerde wyse alle funksies te omvat. Uit die oogpunt van leksikografiese funksies behoort die WAT dus op gebalanseerde wyse polifunksioneel te wees. Feinauer (2009) kom dan ook tot die slotsom dat haar standpunt ten opsigte van die ensiklopediese eerder op kritiek op die oordadige aanbod van uitspraakleiding neerkom. Soms word meer as ses verskillende uitspraakmoontlikhede aangebied wat die indruk van 'n kommunikatiewe eerder as 'n kognitiewe of gebalanseerde benadering wek.

Ongelukkig word daar nie in die betrokke resensieartikel van Deel XII 'n analise van spesifieke betekenisverklarings gemaak om vas te stel of genoeg, te

min of te veel ensiklopediese inligting volgens die behoefte van die teikengebruiker verskaf word nie. Daar sal voorts probeer word om volgens die model van Wiegand vas te stel wat die stand van ensiklopedisiteit in die nuwe dele van die WAT vanaf WAT IX is.

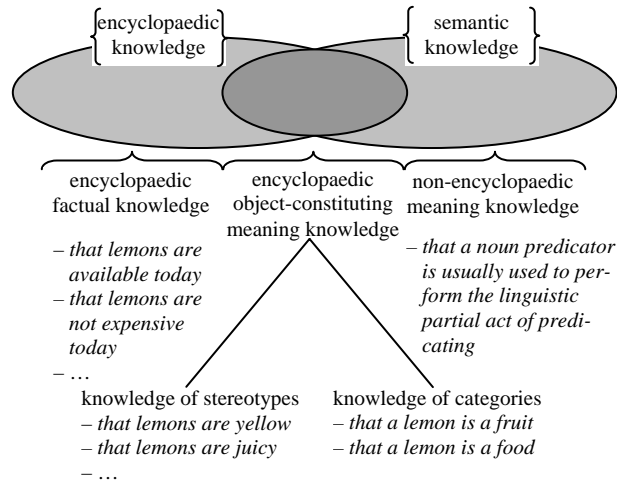
3. **Wiegand se model vir 'n balans tussen semantiese en ensiklopediese inligting in betekenisbeskrywing**

In hierdie gedeelte sal daar gefokus word op Wiegand se sienings rondom die hantering van ensiklopediese inligting in verklarende woordeboeke en sal dit ook teenoor die meer tradisionele beginsel van komponentontleding gestel word.

Wiegand (1984: 21-22 en 1994: 267-271) stel nie 'n absolute grens tussen betekenis- en saakbeskrywing nie. Volgens Wiegand kan 'n leksikografiese definisie ("lexical paraphrase"/"meaning paraphrase") altyd óf as betekenisbeskrywing óf as saakbeskrywing geïnterpreteer word. Indien dit as betekenisbeskrywing gesien word, karakteriseer die definisie steeds terselfdertyd die saak, en indien dit as saakbeskrywing gesien word, is dit steeds terselfdertyd 'n verklaring van die betekenis van die lemma. Daar is derhalwe 'n mate van oorevleueling tussen ensiklopediese inligting en semantiese inligting wat Wiegand ensiklopediese saakbeskrywende betekenis kennis ("encyclopaedic object-constituting meaning knowledge") noem. Gouws (1989: 186) sluit in 'n sekere sin by hierdie uitspraak aan wanneer hy konstateer dat 'n mate van ensiklopedisiteit in enige verklarende woordeboek onvermydelik is.

Volgens Wiegand (1994: 268) is die saakbeskrywende betekenis kennis daardie deel van die semantiese kennis wat as 'n kenmerkende deel van die ensiklopediese kennis beskou kan word. Hierdie kennis word daardeur gekenmerk dat dit gemeenskaplike kennis by die meerderheid van die sprekers is. In die onderstaande skets uit Wiegand (1994: 268) word aangetoon dat die ensiklopediese betekenis kennis uit kennis van stereotipes (*that lemons are yellow/juicy*) en kennis van kategorieë (*that a lemon is a fruit/food*) bestaan. Veral die kennis van stereotipes kan nie duidelik van ensiklopediese feitekennis onderskei word nie.

Onderstaande uiteensetting hou verband met die niegespesialiseerde woordeskat van 'n taal (Wiegand 1994: 269). Dit is die resultaat van 'n vraelys met 21 kenmerke van 'n suurlimoen wat aan 100 studente voorgelê is met die opdrag om aan te dui wat hulle as die belangrikste kenmerke beskou. Wiegand (1984: 27) stel dit duidelik dat hy nie van leksikograwe verwag om soortgelyke proefnemings te onderneem ten einde die gehalte van hul "lexical/meaning paraphrases" te verbeter nie. Die doel van die proefneming is eerder om aan te dui hoe die "byt" uit die vraag watter kenmerke in 'n betekenisparafrase opgeneem behoort te word, gehaal kan word. Eerstens moet die betekenisverklaring as 'n integrerende deel van die woordeboekartikel gesien word, en nie as 'n geïsoleerde entiteit nie.



Wiegand (1984: 27) sien die woordeboekartikel as 'n teks bestaande uit linguistiese voorligting waarmee die leksikograaf die woordeboekgebruiker voorlig oor die gebruik van 'n gelemmatiseerde leksikale item. Nie slegs die betekenisverklaring nie, maar ook ander tekselemente help om die betekenis van die lemma aan die gebruiker te verduidelik. Volgens Wiegand (1984: 27-28) is dit geen uitgemaakte saak dat die betekenis van 'n leksikale item die beste verstaan word deur na die sg. belangrikste kenmerke te verwys nie. Derhalwe moet ander relatief belangrike kenmerke ook in die woordeboekartikel in berekening gebring word. Sy voorstel is dat 'n leksikografiese voorbeeldsin geformuleer word waarin van die ander eienskappe figureer.

Met hierdie model wil Wiegand die begrip van 'n definisie wat op akademiese en wetenskaplike gebruik gebaseer is, vervang met 'n benadering wat ingestel is op algemene gebruik en die behoeftes van die teikengebruikers van 'n woordeboek, en wat derhalwe van woordeboek tot woordeboek aangepas kan word.

Volgens Wiegand (1994) kan 'n genus-differentia-definisie wat gebaseer is op 'n komponentontleding, nie die denotatiewe betekenis van 'n leksikale item toereikend weerspieël nie. Die gebruiker het meer kennis nodig om die leksikale item doeltreffend in die regte kontekste en volgens die gewone reëls in die omgangstaal te kan gebruik. Die leksikograaf moet die behoeftes van die gebruiker ken en die betekenisverklaring daarvolgens saamstel (Wiegand 1994: 265). Derhalwe sal die betekenisverklarings van leksikale items wat na sake verwys wat só vreemd aan die verwagte teikengebruiker is dat hy meer ensiklopediese kennis benodig ten einde dit uit te ken, geformuleer word om ensiklopediese feitekennis oor te dra. Alternatiewelik kan die ensiklopediese inligting wat buite die oorvleueling van die ensiklopediese en semantiese inligting val, ook deur die voorbeeldmateriaal of deur ander metaleksikografiese metodes oorgedra word ten einde die ongekontroleerde en oorbodige aanbod daar-

van in die betekenisbeskrywing teen te werk. Wiegand (1994: 271) is van mening dat taksonomiese name onder andere nie in 'n betekenisverklaring hoort nie. Die voorstel om ensiklopediese inligting ook deur middel van ander inligtingstipes as die betekenisverklaring te verskaf, sluit aan by Wiegand se model van die woordeboekartikel as 'n geïntegreerde teks.

Ten einde vas te stel of die WAT-artikels die gewenste balans tussen semantiese en ensiklopediese inligting volgens Wiegand se model handhaaf om die gebruiker in staat te stel om die lemma doeltreffend in normale taalgebruik aan te wend, sal enkele artikels uit Dele IX, X, XI en XII beskou word. Ter vergelyking sal die meer tradisionele metode van komponentontleding ook op dieselfde WAT-artikels toegepas word.

Die patroon waarvolgens *lemoen* in die WAT beskryf word, toon 'n groot ooreenkoms met die betekenisbeskrywing wat Wiegand (1984: 27) vir *lemon* formuleer:

lemon ... oval fruit of the lemon tree with juicy, sour pulp and yellow rind

lemoen ... 1 Vrug v.d. lemoenboom (LEMOENBOOM 1), met 'n oranje skil en soeterige sap wanneer dit ryp is

Beide betekenisverklarings gee "knowledge of categories" weer deur die saak waarna die definiendum verwys, as 'n vrug te beskryf. Beide gee "knowledge of stereotypes" weer deur kenmerke ten opsigte van kleur, inhoud en smaak te noem. Wiegand beskryf ook die vorm ("oval") terwyl die WAT-betekenisverklaring die kwalifikasie bevat dat die genoemde kenmerke manifesteer wanneer die lemoen ryp is.

Wiegand se "knowledge of categories" en "knowledge of stereotypes" toon ooreenstemming met die onderskeiding tussen gemeenskaplike en diagnostiese komponente by komponentontleding. Gemeenskaplike komponente plaas die lemma in die semantiese veld waartoe hy behoort deurdat 'n baie algemene betekeniskenmerk gespesifiseer word. Diagnostiese komponente is 'n weergawe van betekeniskenmerke wat daarop gemik is om die lemma van ander lede van dieselfde semantiese klas te onderskei (Gouws 1989: 190). Gouws wys daarop dat die gebruik van gemeenskaplike en diagnostiese komponente 'n regstreekse verband met die gebruik van genus-differentia-definisies vertoon. Die gemeenskaplike en diagnostiese komponente word as onderskeidelik die genus en die differentia in die definiens aangebied.

Uit die oogpunt van komponentontleding kan die kwalifikasie "wanneer dit ryp is" as 'n supplementêre komponent beskou word. Dit het nie betekenisonderskeidende waarde nie en is nie-noodsaaklike inligting wat neig na ensiklopediese feitekennis of saakbeskrywing.

Wiegand (1984: 28) sluit die rypheidskwalifikasie in onder die minder kenmerkende "stereotypes" wat in 'n leksikografiese voorbeeld (poëem) vervat is, nl.:

The lemon, which is a type of citrus fruit, has a thick rind and is green when unripe

In die WAT-artikel word soortgelyke ensiklopediese saakbeskrywende betekenisinligting ("encyclopaedic object-constituting meaning knowledge") in die voorbeeldmateriaal weergegee, soos dat die lemoen 'n sitrusvrug is, vitamien C bevat, afgeskil word en uit skyfies bestaan:

Vitamien C wat veral in koejawels, lemoene en ander sitrussoorte voorkom (P.C.J. Kuun: Mod. Landb. Vorm I², 1967, 146). 'n Vrou (sit) ... en aankyk hoe kinders ... lemoene afskil, skyfies daarvan onder mekaar verdeel en smaaklik eet (L. Maré: Lemoene, 1931, 6).

Samevattend kan verklaar word dat uit die oogpunt van komponentontleding die WAT-betekenisbeskrywing van *lemoen* in 'n baie geringe mate inligting bevat wat neig na die ensiklopediese. Ooreenkomstig Wiegand se model kan die inligting vervat in die frase "wanneer dit ryp is" net so goed deur 'n ander inligtingstipe, soos die voorbeeldmateriaal, weergegee word. Andersins stem die WAT-artikel in 'n baie hoë mate ooreen met die Wiegand-model.

'n Relevante vraag, in teenstelling met die vraag hoe vasgestel word of 'n betekenisverklaring te veel ensiklopediese betekenisinligting bevat, is hoe bepaal word of 'n verklaring genoeg ensiklopediese betekenisinligting bevat ten einde die gebruiker in staat te stel om die leksikale item te gebruik of om die saak waarna verwys word, uit te ken. *Lessenaar* word soos volg omskryf:

Tafel ontwerp vir skryf en studeer

Volgens beide Wiegand se model en komponentontleding is dit duidelik dat hier nie sprake van suiwer ensiklopediese feitekennis in die betekenisverklaring is nie. *Tafel* vorm die gemeenskaplike komponent, genus of "knowledge of categories" in die beskrywing, terwyl *ontwerp vir skryf* en (ontwerp vir) *studeer* die diagnostiese komponente, "knowledge of stereotypes" of differentiae vorm.

Uit die sitate kan die gebruiker egter ensiklopediese inligting herwin, soos dat 'n lessenaar laaie kan hê:

Daar is altesaam ag laaitjies in die lessenaar om die orde te help handhaaf (Boerd. S.A., Sept. 1971, 86).

In die geval van *lessenaar* word die minimum "encyclopaedic object-constituting meaning knowledge" weergegee, maar kan daar, soos Wiegand inderdaad bepleit, ensiklopediese inligting uit die voorbeeldmateriaal geabstraheer word. Of die inligting voldoende is vir die gebruiker om die lemma doeltreffend in die regte kontekste en volgens die gewone reëls in die omgangstaal aan te wend, sal afhang van die bepaalde gebruiker se behoeftes. As die WAT se teikengebruiker, nl. "alle sprekers en gebruikers van Afrikaans, maar veral dié spektrum gebruikers wat strek van senior hoërskoolleerlinge tot Afrikaansakademici" egter in gedagte gehou word, behoort die inligting wel toereikend te wees met die oog op doeltreffende kommunikasie.

Sommige items met 'n hoë gebruiksfrekwensie uit die niegespesialiseerde

woordeskat het soms ook 'n gespesialiseerde betekenis as deel van die gespesialiseerde woordeskat van 'n taal. Wiegand is van mening dat 'n betekenisomskrywing ook vir hierdie "uitbreiding" of ensiklopediese aspekte voorsiening moet maak. Die woordeboekgebruiker sal nie die item korrek kan gebruik indien hy nie oor die nodige ensiklopediese feitekennis beskik nie. Die artikel van *olie* kan as illustrasie dien:

¹olie ...

I s.nw.

1 a 'n Kleurlose tot swart vloeibare koolwaterstofverbinding met 'n donkerbruin of groenerige skynsel, wat in 'n natuurlike toestand i.d. aardkors voorkom, gew. uit boorgate verkry word, en waaruit verskeie ontvlambare brandstowwe, petrochemikalieë en smeermiddels d.m.v. distillings- of raffineringprosesse verkry word; sin. *petroleum*, *ruolie*; *aardolie* (minder gebruiklik), *rotsolie* (ongewoon), *steenolie* (ongewoon).

b 'n Ontvlambare vloeistof wat veral verkry word deur olie (¹OLIE I 1 a) te distilleer of te raffineer, oplosbaar in alkohol en eter, maar onoplosbaar in water is, 'n hoë viskositeit besit, en hfs. as brandstof vir aandrywing, verligting of verhitting, en as smeermiddel gebruik word.

In niegespesialiseerde gebruik kan die betekenisonderskeidings van *olie* waarskynlik verskraal word tot onderskeidelik "donker vloeistof wat uit die aarde verkry word en waaruit veral brandstof vervaardig word" en "verwerkte olie (¹OLIE I 1 a) wat as brandstof en smeermiddel gebruik word". Uit die oogpunt van komponentontleding kan alle bykomende inligting waarskynlik as supplementêre komponente of ensiklopediese inligting beskou word.

Ooreenkomstig Wiegand se model kan geredeneer word dat die leksikograaf die behoeftes van die gebruiker in ag moet neem en die betekenisverklaring daarvolgens moet saamstel (Wiegand 1994: 265). Derhalwe moet die relatief gesofistikeerde gebruiker van 'n omvattende woordeboek soos die WAT van genoeg feitekennis voorsien word ten einde 'n gedetailleerde begrip te verkry van die saak waarna die leksikale item verwys. Gevolglik sou die aangehaalde WAT-betekenisomskrywings van *olie* as voldoende en as sonder 'n oorbodige aanbod van ensiklopediese feitekennis beskryf kan word.

Petrol is op dieselfde wyse as *olie* hanteer:

petrol s.nw. ...

1 Vlughtige, ontvlambare vloeistofmengsel van koolwaterstowwe wat deur distillasie uit petroleum, deur kraking van mineraalolie, deur die samepersing van aardgas of die sintese van watergas uit steenkool verkry word, en wat veral as brandstof vir binnebrandenjins, as 'n oplosmiddel, 'n verligtingsmiddel of 'n verdunningsmiddel gebruik word; soms sin. met *brandstof* en *petrolbrandstof* (minder gebruiklik).

Dieselfde argumente sou mutatis mutandis kan geld vir die betekenisbeskrywing van *Manicheïsme*, 'n item wat in 'n hoë mate vreemd aan die gebruikers van die WAT behoort te wees, nl.:

Manicheïsme ... (*na Manes, Persiese wysgeer*) Sinkretistiese geestelik-filosofiese leerstelling wyd aangehang i.d. derde en vierde eeu n.C. i.d. Romeinse Ryk, wat voorhou dat daar 'n kosmiese konflik bestaan tussen lig en duisternis, en wat leer dat dit die mens se plig is om die magte v.d. goeie te bevorder, veral deur die beoefening van asketisme en die onthouding van sinlike en materiële genot.

Uit die voorafgaande blyk dit dat indien die WAT-betekenisbeskrywings van leksikale items uit die niegespesialiseerde woordeskat getoets word aan die metode van komponentontleding, daar weinig of geen saakbeskrywing verskaf word nie. In die geval van items uit die gespesialiseerde woordeskat of items wat na sake verwys wat in 'n hoë mate vreemd aan die gemiddelde gebruiker is, kan daar wel 'n ensiklopediese element onderskei word. Hier moet dit egter ook in gedagte gehou word dat 'n mate van ensiklopedisiteit onvermydelik is in enige verklarende woordeboek (Gouws 1989: 186). Gemeet uit die oogpunt van Wiegand se model waarvolgens daar 'n oorvleueling is van ensiklopediese en semantiese betekenis by items uit die niegespesialiseerde woordeskat, en waarvolgens ensiklopediese inligting vereis word by items uit die gespesialiseerde woordeskat en items wat na "vreemde" sake verwys, is die gegewe ensiklopediese inligting in die WAT-betekenisbeskrywings noodsaaklik.

4. Slot

In hierdie artikel is daar probeer om op objektiewe wyse vas te stel of die WAT vanaf Deel IX 'n balans tussen semantiese en ensiklopediese inligting handhaaf. Uit die toepassing van die metode van komponentontleding en ook Wiegand se model op enkele artikels sou die gevolgtrekking gemaak kon word dat sodanige balans wel aanwesig is. By implikasie is die gebruiker se behoeftes dus in gedagte gehou by die skryf van die artikels. Die redaksie moet dus voortgaan om hom by die samestelling van die woordeboek te laat lei deur die vraag of die gebruikers van genoeg semantiese en ensiklopediese inligting voorsien word om aan hulle behoeftes te voldoen.

Bibliografie

- Bergenholtz, Henning en Sven Tarp. 2003. Two Opposing Theories: On H.E. Wiegand's Recent Discovery of Lexicographic Functions. *Hermes. Journal of Linguistics* 31: 171-196.
- Botha, W.J. 1984. WAT Deel VII. *Buurman* 15(2): 33 en 35.
- Botha, W.F. 2003. *Die impak van die leksikografieteorie op die samestelling van die Woordeboek van die Afrikaanse Taal*. Ongepubliseerde D.Litt.-proefskrif. Stellenbosch: Universiteit van Stellenbosch.
- Eksteen, L.C. 1968. Ensiklopediese beskrywing in die leksikografie. *Tydskrif vir Geesteswetenskappe* 8(3-4): 236-249.

- Feinauer, I.** 1993. Die agste deel van die *Woordeboek van die Afrikaanse Taal*. *Tydskrif vir Geesteswetenskappe* 33(4): 286-298.
- Feinauer, I.** 2007. Doen die *Woordeboek van die Afrikaanse Taal XII* dit vir die Afrikaanse taal en die Suid-Afrikaanse leksikografie? *Lexikos* 17: 259-277.
- Feinauer, I.** 2009. Rondetafelgesprek op 4 Februarie 2009 met prof. I. Feinauer en die WAT-redaksie as gespreksgenote na aanleiding van haar resensie van WAT XII in *Lexikos* 17: 259-277.
- Goows, R.H.** 1985. Die sewende deel van die *Woordeboek van die Afrikaanse Taal*. *Standpunte* 38(4): 13-25.
- Goows, R.H.** 1989. *Leksikografie*. Kaapstad: Academica.
- Goows, R.H.** 1992. *Woordeboek van die Afrikaanse Taal, Deel VIII*. SAUK, Skrywers en Boeke.
- Goows, R.H.** 1994. *Woordeboek van die Afrikaanse Taal, Deel IX*. SAUK, Skrywers en Boeke.
- Goows, R.H.** 1996. *Woordeboek van die Afrikaanse Taal, Deel X*. SAUK, Skrywers en Boeke.
- Raidt, E.H.** 1993. Die agste deel van die *Woordeboek van die Afrikaanse Taal*. *Lexikos* 3: 215-226.
- Schoonees, P.C.** 1952. Bespreking van die Afrikaanse Woordeboek (A-C): Repliek van Hoofredakteur. *Tydskrif vir Wetenskap en Kuns Nuwe Reeks* 12(2): 289-290.
- Schoonees, P.C.** 1957. Op los skroewe. *Standpunte* 11(5-6). 1957: 122-126.
- Schoonees, P.C.** 1958. *Rondom die Woordeboek*. Kaapstad: Nasionale Boekhandel Bpk.
- Snijman, F.J.** 1977. Aangehaal in die berig "Word die WAT dalk eers in 2120 voltooi?", *Die Burger*, 1 November: 17.
- Snijman, F.J.** 1995. *Die leksikograaf as eksegeet. Opstelle en lesings*. Stellenbosch: Buro van die WAT.
- Snijman, F.J.** 1995a. Bespreking van WAT VI deur J. Combrink — 'n paar opmerkings. Snijman, F.J. 1995: 201-211.
- Snijman, F.J.** 1995b. Ponerings sonder voldoende fundering. Snijman, F.J. 1995: 189-200.
- Snijman, F.J.** 1995c. Is illustrasies in 'n woordeboek funksioneel? Snijman, F.J. 1995: 157-168.
- Snijman, F.J.** 1995d. Oor die ensiklopediese. Snijman, F.J. 1995: 169-186.
- Swanepoel, P.H.** 1995. 'n Eksegese van 'n leksikografiese metode: Die WAT en die probleem van ensiklopedisiteit in die verklarende linguistiese woordeboek. *Lexikos* 5: 187-210.
- Tarp, S.** 2000. Theoretical Challenges to Practical Specialised Lexicography. *Lexikos* 10: 189-208.
- Wiegand, H.E.** 1984. On the Structure and Contents of a General Theory of Lexicography. Hartmann, R.K.K. (Red.). 1984. *LEXeter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9-12 September 1983*: 13-30. Tübingen: Max Niemeyer.
- Wiegand, H.E.** 1994. Elements of a Theory Toward a So-called Lexicographic Definition. Wiegand, H.E. 1999. *Semantics and Lexicography. Selected Studies (1976-1996)*: 203-282. Redaksioneel versorg deur Antje Immken en Werner Wolski. Tübingen: Max Niemeyer.
- Zgusta, L.** 1971. *Manual of Lexicography*. Den Haag: Mouton.

The Lexicographic Treatment of Ideophones in Zulu

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

Abstract: The ideophone, a word class not unique to but highly characteristic of the Bantu languages, presents particular challenges in both monolingual and bilingual lexicography. Not only is this part of speech without a counterpart in most other languages, the meaning of ideophones is highly elusive. In this research article these challenges are studied by means of an analysis of the treatment of ideophones in a corpus-driven Zulu–English school dictionary project.

Keywords: LEXICOGRAPHY, DICTIONARY, BILINGUAL, CORPUS, FREQUENCY, BANTU, ZULU (ISIZULU), ENGLISH, IDEOPHONE, SEMANTIC IMPORT, PARAPHRASE, PART-OF-SPEECH MISMATCH

Samenvatting: De lexicografische behandeling van ideofonen in Zoeloe. De ideofon, een woordklasse die niet uniek maar wel heel karakteristiek is voor de Bantoetalen, is een echte uitdaging in zowel de monolinguale als bilinguale lexicografie. Niet enkel heeft deze woordklasse geen equivalent in de meeste andere talen, de betekenis van ideofonen is heel moeilijk vast te leggen. In dit onderzoeksartikel worden deze uitdagingen onderzocht aan de hand van een analyse van de behandeling van ideofonen in een corpus-gedreven Zoeloe–Engels schoolwoordenboekproject.

Slutelwoorden: LEXICOGRAFIE, WOORDENBOEK, TWEETALIG, CORPUS, FREQUENTIE, BANTOE, ZOELOE, ENGELS, IDEOFOON, SEMANTISCHE LADING, PARAFRASE, VLOEKENDE WOORDKLASSEN

1. The coinage and meaning of the term 'ideophone'

Although the term 'ideophone' has not been entered as a lemma sign in the second edition of the *Oxford English Dictionary* (OED Online 2009), it may be found within the article of the combining form 'ideo-', as shown in (1):

- (1) ideo- combining form of Gr. *ἰδέα* **IDEA**, as in [...] **ideophone** (-fəʊn) [Gr. *φωνή* voice, sound], (a) term used by A. J. Ellis (in contradistinction to *ideograph*) for a sound or group of sounds denoting an idea, i.e. a spoken word; (b) a term used principally in Bantu linguistics to refer to particular classes of onomatopoeic and sound-symbolic words found in these languages; so **ideophonetics**, the subject of 'ideophones'; hence **ideophonic a.**; **ideophonous** (-dəfəʊnəs) *a.*, relating to spoken words as sounds denoting ideas; [...]

The OED's citation evidence starts with: "1881 [A.J. ELLIS](#) *Synops. Lect. Lond. Dialectical Soc.* 2 Nov., Mimetics, ideographics, and *ideophonetics. Fixed ideograph, variable *ideophone, and their connection." Unfortunately, the reference to the Ellis source is a dead cross-reference. However, according to the OED lexicographer Jesse Sheidlower: "The OED is citing a printed card announcing two of the London Dialectical Society's November meetings, mailed by Ellis to James Murray (they were friends), and subsequently deposited by Murray in the OED archives" (*Language Hat* 2008), and he points out that: "We'll be clarifying our bibliography to show that this is not a published item" (*The Ideophone* 2008).

First recorded in 1881, then, the meaning assigned to the term 'ideophone' under (1, sense a), was redefined as (2) in the *Century Dictionary Supplement* of 1909 (*Century Dictionary Online* 2009):

- (2) ideophone (ī - d ē ' ō - f ō n), *n.* [Gr. *ἰδέα*, idea, + *φωνή*, sound.] In *phonetics*, the auditory symbol of a word or phrase that is perceived as a whole and thus constitutes a single idea. Ideophones are distinguished as *sensory* or *motor*, according as the sound or group of sounds corresponding to the word or phrase is heard or spoken. See **ideogram*, 2. First used by A.J. Ellis. *Scripture, Exper. Phonetics*, p. 132.

Since Doke's (1935) publication of *Bantu Linguistic Terminology*, however, the term has been considerably expanded, as seen in the oft-quoted opening section of Doke's definition:¹

- (3) IDEOPHON (Idéophone) [*Ideophon*]. A vivid representation of an idea in sound. A word, often onomatopoeic, which describes a predicate, qualificative or adverb in respect to manner, colour, sound, smell, action, state or intensity. The ideophone is in Bantu a special part of speech, resembling to a certain extent in function the adverb, together with which it is classified as a descriptive.

In addition to the expansion in meaning compared to the earlier definitions, Doke's more important addition to the concept of ideophones is that he underscores the special status accorded to ideophones in Bantu, namely that they are a word class distinct from the other parts of speech (POSS). No wonder, then, that the OED's definition (1, sense b) singles out Bantu linguistics, even though ideophones (or at least aspects of them) are found in many (if not all) of the world's languages.

In reviewing the Bantu literature, Doke noted that various authors had suggested many a term for the ideophone. He lists: radical, descriptive adverb, descriptive complement, indeclinable verbal particle, intensive interjection, interjectional adverb, onomatopoeic vocable, onomatopoeic adverb, onomatopoeia, onomatopoeic substantive, mimic noun, indeclinable adjective, etc. (Doke 1935: 119). Ironically, Weakley (1973: 2) correctly concludes:

Of all the above terms onomatopoeia is probably one of the least suitable. It would not even be appropriate I should imagine, to say the ideophone is usually onomatopoeic. As Fortune points out: "The fact among others, that ideophones can be used to indicate complete silence makes the term onomatopoeic an unrepresentable term for ideophones as a whole."

In another review of the Bantu literature, Samarin (1971), summarized in Weakley (1973: 7), collected the following 'characteristics' of ideophones:

- Ideophones represent or express actions or ideas.
- Ideophones echo, express, or mirror sense impressions or perceptions.
- Ideophones express or reflect emotions and feelings.
- Ideophones complete thought or create images.
- Ideophones vivify speech.

So far, the focus has been on semantic (and pragmatic) aspects of ideophones — a logical by-product of looking at dictionary definitions. Even more striking, however, are the unique phonological, morphological and syntactic aspects of ideophones — all of which have received considerable attention in the scientific literature. For a global perspective on these aspects, see the collection edited by Voeltz and Kilian-Hatz (2001), and especially the 'Bibliography of ideophone research' therein. The most recent and continuously updated source on ideophones is doubtless the blog *The Ideophone* by Mark Dingemans, who includes the function of ideophones in his research. His most recent working definition (June 2009) for the ideophone is: "marked words that vividly evoke sensations and perceptions" (Dingemans 2009).

2. Ideophones in (Zulu) metalexigraphy

With now well over a century of linguistic research into the world's ideophones, one can fairly say that this class of words starts to be rather well described linguistically. Lexicographic aspects, however, have unfortunately largely been ignored. To the best of our knowledge, only two serious attempts have been made, both in *Lexikos*, by Childs (1993: 21-23) and Khumalo (2002: 270-271).

Childs starts his discussion of ideophones in Kisi as follows: "Ideophones pose enormous problems to the lexicographer because of their monumental variation and semantic indeterminacy." He then goes on to pose some important questions, including: Is it the lexicographer's task to faithfully record the fact that ideophones exploit prosodic resources? What does one do when the ideophone seems to have no independent meaning? Given the close relationship between ideophones and gestures, how must gestures be represented? Unfortunately, Childs does not give any answers to these questions. Furthermore, while phonological and pragmatic aspects are indeed important, lexi-

graphy is primarily concerned with meaning, so of these questions an answer to the second one is most needed.

In that sense, Khumalo's discussion of ideophones is more revealing. Describing the compilation of a monolingual (Zimbabwean) Ndebele dictionary, he shows how three distinct defining formats were developed for ideophones, formats patterned on the COBUILD style of writing definitions (cf. Hanks 1987).

With specific reference to bilingual lexicography, one only finds trivial statements, with no solutions whatsoever. For instance, all that is said by Gauton (2008: 112) on ideophones is: "Languages differ in their parts of speech. For example, a language such as Zulu distinguishes the word category 'ideophone' which does not exist in a language such as English." Or by Jadezweni (1998: 323): "The difficulty of providing English equivalents for Xhosa ideophones can be a nightmare to learners [...] It requires a lot of imagination to be able to come up with explanations of ideophones."

Of course, ideophones have been entered into dictionaries, and in dictionaries for the Bantu languages, they have typically been termed 'ideophone' in the slots for the word classes too. What is missing is a proper metalexigraphic analysis, especially one that bridges a Bantu language (where ideophones are a distinct part of speech) with a non-Bantu language.

This is exactly what is done in the present contribution. As a case study, the ideophones in a Zulu to English school dictionary are looked into. As such, this study forms part of a series of studies in which each of the various Zulu word classes is analysed from a lexicographic point of view. Earlier instalments in the series looked into the possessive pronouns (De Schryver and Wilkes 2008), the adjectives (De Schryver 2008), and the quantitative pronouns (De Schryver 2008a).

While the number of ideophones in each (Bantu) language varies, it is generally assumed that there are 'many', or that they at least "represent a sizeable proportion of a language's lexicon" (Childs 1994: 179). That this is indeed the case is not disputed, but current dictionaries — for Bantu languages generally, and for Zulu particularly — may well over-represent ideophones to the detriment of words in other word classes. There is likely a sociolinguistic underpinning for this, as Weakley (1973: 9) observes: "It seems that the proper use of ideophones can be correlated with a mastery of the language concerned."² In earlier times, lexicographers who worked without access to large electronic corpora may thus very well have overdone it by including in their dictionaries as many ideophones as they could, just as they tended to stock their dictionaries with (often rare) idioms and proverbs. Conversely, it may also be the case that ideophones are used more often in spoken than in written language, and thus that the modern corpus-driven approach to dictionary making will actually under-represent the ideophones, given corpora are mostly built up from written sources.

Fivaz (1963) proceeded to count all ideophones in the largest dictionary

available for Zulu, viz. Doke and Vilakazi's (1953) *Zulu–English Dictionary*, and arrived at 2 600 ideophones. With about 30 000 lemmas in that dictionary,³ this corresponds to a massive 8.67%. Based on the occurrence frequencies in our 8.5-million-word written Zulu corpus, however, we conclude that ideophones in (written) Zulu are not particularly frequent: Only five make it into the top 1 500 lemmas, with exactly 100 in a dictionary covering the 5 000 most frequent lemmas. In our project, and expressed in per cent, the ideophones thus make up 2.00% of the lemmas (100 out of 5 000, compared to 2 600 out of 30 000 or 8.67% in Doke and Vilakazi).

With the present study, then, the Zulu word classes considered from a lexicographic point of view so far are shown in Figure 1, with the possessive pronouns representing 1.98%, the adjectives 2.52%, the quantitative pronouns 0.66%, and the ideophones 2.00%.

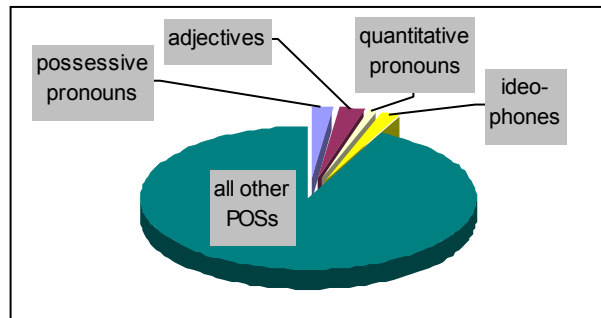


Figure 1: Zulu POS categories studied from a lexicographic point of view

Although another 'small' category in dictionary terms, the metalexicographic description of ideophones is highly relevant. For one, in our Zulu to English dictionary project, their compilation took an average three times longer than the compilation of entries in any other word class. It is no exaggeration, then, that in Bantu lexicography ideophones are a lexicographer's worst nightmare. This is so, not because of their peculiar linguistic properties — be these phonological, morphological or syntactic (see for example, with specific reference to Zulu: Fivaz (1963), Voeltz (1971), Von Staden (1974, 1977), Taljaard and Bosch (1993: 162), Childs (1996), Poulos and Msimang (1998: Chapter 8), or Msimang and Poulos (2001)) — but because of their semantic import that is hard to pinpoint, describe and represent lexicographically.

Although our Zulu–English dictionary project is a bidirectional one, the focus will be on the Zulu to English side, as that is the side where the Zulu ideophones are lemmatized. As will be clear from the discussion below, attempting to 'reverse out' ideophones, and thus attempting to force ideophones into the reverse side of the dictionary as lemmas on the English to Zulu side, is a futile attempt. This does not mean that there are no Zulu ideophones to be found in the English to Zulu side of the dictionary. When translating into Zulu,

mother-tongue speakers often feel the need to introduce them, as may be seen from entries (4) and (5), where highlights have been added for ease of reference:

- (4) **retailer** *noun* [^{pl.} **retailers**] [**ECONOMICS**] ► **isitolo** 𑀩 a business that sells goods directly to consumers • *ibhizinisi elidayisela ngqo abathengi* ♦ Consumers buy clothes and food from retailers, but the retailers buy from wholesalers. • *Abathengi bathenga izingubo nokudla ezitolo, kodwa izitolo zithenga kumawulisela.*
- (5) **endangered** *adjective* [^{comp.} **more endangered** • ^{supl.} **most endangered**] [**BIOLOGY**] ► [**RC+**]sengozini 𑀩 in danger of dying out or ceasing to exist • *ukuba sengozini yokuphela nya* ♦ The government protects animals and plants that are endangered. • *Uhulumeni ulondoloza izilwane nezitshalo ezisengozini.*

In the definition for *retailer* under (4), the meaning of the English adverb 'directly' has been rendered with the Zulu ideophone *ngqo* 'of straightness, of directness', and in the definition of *endangered* under (5), the English phrasal verb 'cease to exist' has been rendered in Zulu by combining the verb *-phela* 'come to an end; get finished; run out' with the ideophone *nya* 'of complete absence'.

Actually, given that our Zulu corpus also contains sources translated from other languages into Zulu, several attested ideophonic uses found their way into the corpus, and from there into our dictionary, through languages other than Zulu. To give one example: When translating the sentence 'Dawn was breaking and they could just make out several villages dotted about on the open hillslopes', taken from Jenny Seed's (1968) *The Voice of the Great Elephant*, the translator, N.S. Ntuli (1988), introduced an ideophone, as shown in (6):

- (6) **gqwa** *ideophone* ► (*of being sparse, of being dotted*) ♦ *Kwase kusa, isibonakala imiziithe gqwa gqwa* laphaya emaqeleni. • *Dawn was breaking and they could just make out several villages dotted about on the open hillslopes.*

This bidirectionality is non-trivial, as it clearly indicates that even though there is no word class 'ideophone' in English, in order to produce idiomatic Zulu, certain English concepts and expressions 'require' the use of ideophones when they are translated into Zulu.

3. The semantic import of (Zulu) ideophones

The last statement in Section 2 automatically leads to two areas in need of further investigation in bilingual lexicography. On the one hand one needs to be able to 'map' the meaning of an ideophone in one language onto a relevant concept or expression in another language, and vice versa. On the other hand one needs to be able to 'map' the part of speech ideophone in one language onto a relevant part of speech in another, and vice versa. Although the part of speech

ideophone is rarely problematic in monolingual (Bantu) lexicography (various linguistic tests can and have been designed to pinpoint this word class), delineating the meaning remains an arduous task. In this regard, Childs (1993: 22) rightly points out:

Determining the meaning of ideophones can prove incredibly frustrating to the lexicographer since the meaning of an ideophone requires a context for interpretation much more than other words. In addition, ideophones require for their understanding an intensive knowledge of the language, a knowledge often inaccessible to an outsider (Samarin 1967).

3.1 Stacked paraphrases (rather than translation equivalents)

A concern that recurs in the literature on ideophones is thus that "it is extremely difficult to characterize in a simple way, the meanings of ideophones" (Weakley 1973: 8). After having analysed and described the 100 most frequent Zulu ideophones lexicographically, the golden rule that emerges for dictionary makers is that meanings assigned to ideophones should not be too broad neither too specific: The art is to establish the right level of generalization as far as the semantic import is concerned, with the examples functioning (true to their core function) as possible instances only. Selected ideophones to illustrate this follow in (7) to (10):

- (7) **do** ideophone ► (of nothing) ♦ Wampaya naso sonke isibhedlela kodwa **do** ukumthola. • He looked all over the hospital but **he didn't find him at all**. ♦ Ngaquba kwa-Oom Joe, umsebenzi **do**. • I temporarily stayed at Uncle Joe's place, but **there was no work**.
- (8) **khumu** ideophone ► (of stopping, of retiring) ♦ Hhayi madoda, **khumu!** Ngeke simthole lo muntu. • No guys, **it's enough!** We'll never find this person. ♦ Usezothi **khumu** ebholeni uZidane. • Zidane will now **be retiring** from soccer.
- (9) **mpo** ideophone 1 ► (of being erect) ♦ Wavuka wama **mpo** uGovu wathi: ... • Govu woke up, **stood straight up**, and said: ... 2 ► (of extreme action) ♦ Sekubanda, **mpo**, eNingizimu. • It is now cold, **extremely cold**, in the South. ♦ Kuthe ilanga selithe, **mpo**, ibandla lihlezi phansi nje emthunzini. • When it is **extremely hot**, the group of men simply sits on the ground in the shade. 3 ► (of being full) ♦ Ligcwaliseni isaka lithi **mpo**. • Fill up the bag **to the brim**.
- (10) **ngqi** ideophone 1 ► (of tightness, of security, of holding firm) ♦ Wase ezivalela yena endlini ethi **ngqi**. • He then **locked himself securely** inside the house. ♦ Yalibamba iqakala. Yalibamba yalithi nje **ngqi**. • It grabbed his ankle. It **grabbed it very tightly**. 2 ► (of getting stuck) ♦ UVika wayengasakwazi nakunyakaza. Nengqondo yayisithe nje **ngqi**. • Vika could no longer move. His mind too **got stuck**.

The task of the lexicographer, when presented with page-fulls of instances of the use of a particular ideophone, is to try to deduce the meaning from the col-

lective evidence. *In context*, meanings may be distilled on a generic level, roughly led by the highlights in (7) to (10). The problem with ideophones is that it is sheer impossible to demarcate where the meaning of the context starts and stops, and by consequence, to clearly pinpoint the true semantic import of the ideophone itself, that is, the ideophone *in isolation*. (The highlights in (7) to (10) are purposely generous in this regard.)

Nonetheless, with enough evidence, one or more generic meanings do appear, and in monolingual dictionary making, these then form the basis for the write-up of the definitions for each sense. To the bilingual dictionary compiler, however, no foolproof translation equivalents are available (except when mapping cognate languages onto one another, or when dealing with languages that each have ideophones as word classes). Rather, each sense of an ideophone in translation dictionaries is merely a *paraphrase* of the semantic import, with the examples not random but hand-picked instances to substantiate the *range* of possible uses.

As a result, and as seen in (7) to (10), as well as (6), instead of translation equivalents, paraphrases are provided (in italics, and between brackets), using the convention to start those paraphrases with 'of', typically followed by a verb in the *-ing* form or an abstract noun. Paraphrases may be stacked, as in (10, sense 1), in order to best cover the meaning seen in the majority of the corpus lines. The example sentences in (6) to (10) are not merely illustrative material (there are no translation equivalents to illustrate anyway); rather, these examples truly support the meanings.

3.2 Part-of-speech mismatch

Compared to monolingual dictionary making, the absence of a corresponding word class in the non-Bantu language, forces the bilingual dictionary maker to come up with various strategies to 'translate/transpose' the examples. When the ideophones are onomatopoeic-like, English *sound words* can be inserted or recourse can be taken to English *exclamations*. Instances of the former may be seen in (11, sense 3) and (12), instances of the latter in (13) and (14, sense 1):

- (11) **nsi** ideophone 1 ► (of tightness, of security) ♦ Yathatha imbeleko, yabopha umntwana yamuthi nsi emhlane. • She took the baby-carrying cloth and fastened the child securely on her back. 2 ► (of extreme action) ♦ Imoto yema nsi ingasavumi ukuhamba. • The car came to a complete standstill, and didn't want to go any further. 3 ► (of laughing) ♦ Bahleke baginqike, nsi, nsi, nsi, nsi, nsi, nsi, nsi, nsi. • And then they rolled over laughing: hi, hi, hi, hi, hi, hi, hi, hi.
- (12) **pho**² ideophone ► (of dripping, of crying heavily) ♦ Ngalesi sikhathi zase zehla izinyembezi ku-Alice zithi pho pho pho, engazi kodwa ukuthi ukhalelani. • At this time the tears came down from Alice's face – drip drip drip – while she did not know why she was crying.

- (13) **nci** ideophone ► (of happiness, of surprise) ♦ "Nci, nci, nci ...!" kuqhuba uBhekokwakhe ngokusola. • "Yippee, yippee, yippee ...!" continued Bhekokwakhe in disbelief.
- (14) **qhu** ideophone 1 ► (of exploding/bursting sound) ♦ Ngelinye ilanga bezihambela phakathi kwemihlanga emide ezansi nomfula; **qhu** umsindo omkhulu engathi kuqhuma isibhamu. • One day while they were walking between the long reeds down by the river – "bang!" – a loud sound went off as if coming from a rifle. 2 ► (of burning, of being dry) ♦ Ngoma, ngasha **qhu**, emphinjeni. • I was thirsty and had a burning feeling in my throat.

In most cases, however, the Zulu ideophones need to be translated with English *verbs*, as in (8) above, or with English *phrasal verbs*, as in (9, sense 1) or (10, sense 2) above.

At times, Zulu ideophones may also conveniently be translated with *noun phrases*, as in (9, sense 3) above, or (15) below:

- (15) **qu** ideophone ► (of brief period of time) ♦ Ekupheleni konyaka lowo wake wathi **qu** ekhaya ebuyele uKhisimusi. • At the end of that year he went home for a short while, returning for Christmas.

English *adjectives* and *adverbs* too, are particularly frequent as stand-ins for Zulu ideophones, as seen in (16, sense 2) and (17, senses 1 and 3), respectively (18) and (19, sense 1):

- (16) **klwi** ideophone 1 ► (of scribbling) ♦ Nempela yalithatha iphepha yalithi **klwi**, **klwi**, ibhalela umnumzane yethula isikhalo sayo. • He really took a paper and scribbled on it, while writing to the gentleman and submitting his complaint. 2 ► (of being slim and tall) ♦ Kwakuyinsizwa nje ethe **klwi** esicela emashumini amathathu eminyaka. • It was a tall young man approaching his thirties.
- (17) **nse** ideophone 1 ► (of extreme action) ♦ Imoto yama **nse**. • The car came to a complete stop. ♦ Kanti bengihlakaniphile **nse** ngisenza. • In fact I was very clever when I did it. 2 (of healthiness) ♦ Ngaphandle kokucinana okuncane wayesephile **nse**. • Apart from a slight chest complaint he was very healthy. 3 ► (of straightness, of directness) ♦ Mina ngakhuluma into eqonde **nse**. • I said something straightforward. • -fana **nse** ► be exactly equal ♦ Siyabona-ke ukuthi imiqondo kayifani **nse**. • We note that the ideas are not exactly the same.
- (18) **gelekeqe** ideophone 1 ► (of completeness) ♦ Ngale nkathi sekusile **gelekeqe**. • At this time it is already entirely daylight. 2 ► (of sudden action) ♦ UMvelase waxuma wathi **gelekeqe** waphhepha. • Mvelase suddenly jumped up and escaped.
- (19) **ngqo** ** ideophone Compare **nqo** 1 ► (of straightness, of directness) ♦ Buka **ngqo** manje ezinhlosweni zakho. • Face your goals directly now. ♦ Uthe akazi ukuthi kungani bengamtshelanga **ngqo**, uma kunezinsolo eziqondene naye. • He said he doesn't know why they haven't told him directly, if there are suspicions that con-

cern him. 2 ► (of knocking) ♦ Ngqo-ngqo-ngqo. Vula mnumzane, singama-phoyisa. • Knock-knock-knock. Open up sir, we are the police.

▪ -qonda ngqo ► go straight (to) ♦ Wasuka eMbali waqonda ngqo ekhaya, kawachezuka ndawo. • He left Mbali and went straight home, without branching off anywhere.

The use of other English parts of speech and constructions to translate Zulu ideophones is also possible but infrequent.

Finally, there are instances where the ideophone simply *disappears*, as it is rendered by a paraphrase in idiomatic English, as seen in (20) and (21):

(20) **juqu** ideophone ► (of cutting off, of snapping) ♦ UMahommed wayithi **juqu** ngomemese intambo. • Mahommed cut the string with a knife. ♦ Uqwashe kuze kuse **juqu!** • You must be alert **until dawn!**

(21) **shu** ideophone ► (of going right into) ♦ Aphenye izingubo, azithi **shu** kuzona ikhanda livele kancane. • He turned over the blankets **and covered himself, with only his head sticking out a little bit.**

▪ [NEGATIVE +] ... **shu** 1 ► (of no meaning, of no use) ♦ Udumo aluthi **shu**, singayibamba i-Chiefs. • Fame doesn't mean anything, we can stop Chiefs. ♦ Ungalindeli amaphilisi ama-antibiotics owathola kudokotela noma ekliniki, awathi **shu** egciwaneni lomkhuhlane. • You must not wait for the antibiotic pills you get from the doctor or the clinic; they are of no use when it comes to the flu bug. 2 ► (of silence) ♦ Wabadonsa ngendlebe ukuthi bangayithi **shu** kumuntu leyo ndaba. • He warned them not to say a word to anyone.

3.3 Ideophones as intensifiers of meanings

Ideophones also tend to 'stress' an aspect that was already mentioned, meaning that their function at that point in the sentence is to give more weight to the action expressed by the verb, the pronoun, etc. In (22), for instance, the ideophone *qhwaba* 'of being alone' intensifies the exclusive quantitative pronoun of class 1 *yedwa* 'alone; on her/his own', here prefixed by the relative concord of class 1: *oyedwa* '(only) one; (only) a single' (cf. De Schryver 2008a: 99). Rather than just '(only) one' on the one hand, and 'of being alone' on the other, put together the meaning in English becomes 'one and only'.

(22) **qhwaba** ideophone ► (of being alone) ♦ Umntanami **oyedwa qhwaba** uTholakele. • My one and only child is Tholakele. ♦ Ngangingedwa **qhwaba**. • I was entirely on my own.

Given that the prime target of the Zulu–English school dictionary consists of junior users, such examples are, where possible, avoided for exemplification purposes — but not skipped altogether, and in any case always accompanied by more straightforward examples.

3.4 Ideophones in combinations and fixed expressions

While some ideophones could be said to stand on their own, corpus evidence clearly shows that one out of ten typically combines with specific words in addition, as seen in (17) and (19) above, or in (23) below:

- (23) **phaqa** *ideophone* 1 ► (of reality, of truth) ♦ Mina ngingumZulu phaqa futhi ngiyaziqhenya ngalokho. • I am a true Zulu and, furthermore, I am proud of that. 2 ► (of snapping) ♦ Bathi ingalo yakhe yephuke phaqa. • They are saying that his arm was completely severed.
 • **umuntu phaqa** ► ordinary/uncivilized/uneducated person ♦ Yebo Nkosi, ngalwa ngoba ngingakufuni ukudelelwa umuntu phaqa. • Yes my King, I fought because I didn't want to be despised by an ordinary person.

Although it is still possible to uncover chains of meaning — from the so-called basic meaning of the ideophone in isolation, to the meaning of the combination — the combined meaning is not transparent enough to leave the combination untreated in a corpus-driven lexicographic description.

For another one out of ten ideophones, the *only* occurrence is in so-called fixed expressions, as seen in (24) and (25):

- (24) **nhlo** *ideophone*
 • **selokhu kwathi nhlo** ► right from the beginning; right from the onset ♦ Selokhu kwathi nhlo umndeni wakithi uhlala kuleli pulazi. • Right from the beginning my family lived on this farm.
- (25) **qwa** *ideophone*
 • **-mhlophe qwa** ► bright-white; snow-white; very white ♦ Washo ehleka kuvela *amazinyo* akhe amhlophe qwa. • He said so laughing, revealing his bright-white teeth.

Unlike ideophones in isolation, which can only be provided with one or more paraphrases, examples (17), (19), (23), (24) and (25) show that combinations and fixed expressions that include ideophones may successfully be translated with (phrasal) verbs, nouns, adjectives, etc. in English. For these, then, proper translation equivalents may be provided (hence the different typography).

3.5 Ideophones in constructions

Corpus evidence further enables one to record typical patterns or constructions, as seen in (21) above, or (26) below:

- (26) **quthu** *ideophone*
 • **[NEGATIVE +] ... -thi quthu ubuthongo** ► no longer get sleepy ♦ Akasabuthi quthu uZeb ebusuku, udla amathambo engqondo ngesimo sika-Agnes. • Zeb no longer gets sleepy at night; he's thinking very hard about the condition of

Agnes. ♦ Kasisabuthi quthu ubuthongo ubusuku nemini. • *We are no longer sleeping, night and day.*

The corpus evidence indicates that for most patterns ideophones are embedded in a negative construction (as is also the case in (21) and (26)).

4. Corpus vs. non-corpus-driven dictionary compilation of (Zulu) ideophones

Starting with the publication of *Looking Up* (Sinclair 1987), a considerable amount of scholarly lexicographic literature has been devoted to corpus-driven dictionary making. That the results as well as the corpus-driven dictionaries themselves are (very) different compared to those from the pre-corpus era is evident on all levels — be these levels macrostructural, mediostructural or microstructural, and even with regard to the treatment of the extra (front, middle and back) matter. An example of how differently the extra matter may be approached in corpus-driven Bantu lexicography can be found in De Schryver and Taljard (2007).

The corpus-driven lexicographic treatment of the Zulu ideophones will now be compared with the treatment of ideophones in the two most widely used pre-corpus era dictionaries for Zulu, viz. Doke and Vilakazi's (1953) *Zulu-English Dictionary*, and Dent and Nyembezi's (1995) *Scholar's Zulu Dictionary*.

4.1 The macrostructure

Given that Doke and Vilakazi's dictionary contains as many as six times more lemmas (30 000 lemmas vs. 5 000 in our project), and Dent and Nyembezi's dictionary (with 13 500 lemmas⁴) nearly three times more, one could assume that each

of the 100 ideophones in our project is also covered in the existing two dictionaries. This should especially be the case given that those 100 are the hundred most-frequently used ones. This assumption, however, is not corroborated, as may be deduced from the data in Table 1.

Table 1 brings together the 100 ideophones lemmatized in our project, together with their corpus frequencies (in 8.5 million words) and frequency bands (i.e. star-ratings in the dictionary), and contrasts this with the information in the dictionaries by Doke and Vilakazi (D&V), and Dent and Nyembezi (D&N). Whenever at least one of the senses mentioned by D&V, respectively D&N, is also attested in the corpus, that ideophone has been marked with a tick (✓). Both D&V and D&N only cover four fifths of the top hundred Zulu ideophones. In D&V a total of 8 are missing (-), and for a further 12 the meanings provided are not seen in the corpus (≠). Note that, since the publication of D&V, the spelling has been adapted — the spelling as found in D&V is therefore also shown next to the ticks. In D&N a total of 11 ideophones are missing

Table 1: Top 100 Zulu ideophones from a corpus and dictionary perspective

Dictionary	Freq.	Band	D&V	D&N	Dictionary	Freq.	Band	D&V	D&N
bha	285		✓ baa	✓	nci	51		✓	-
bhe	198		✓ bee	✓	ngci	276		✓	✓
bhu	125		✓ bu	✓	ngqa	564	*	✓	✓
bhuqe	200		✓ buqe	✓	ngqi	206		✓	✓
cababa	77		✓ cábaba	✓	ngqo	959	**	✓	✓
chaphasha	54		✓	✓	nhla	184		✓	✓
chithi saka	91		- chithi	✓	nhlo	124		✓	✓
co	129		✓	✓	njo	339		✓	✓
cosololo	146		✓	✓	nqo	70		✓	✓
cosu	162		≠	≠	nse	152		✓ nsee	✓
cwaka	478	*	✓	✓	nsi	101		✓	✓
cwe	160		✓	✓	nya	1129	**	✓	✓
damu	74		✓	✓	nyaka	274		✓	✓
dedelele	51		✓	✓	phama	59		✓	✓
dekle	67		✓	✓	phaqa	123		≠	✓
dengwane	102		-	-	phecelezi	141		-	-
dinsi	54		✓	✓	psheshe	59		✓	✓
dlengelele	92		✓	✓	pho	182		≠	≠
do	100		✓	-	phuhle	56		✓	✓
du	751	*	✓	✓	qakala	188		-	-
dwi	61		✓	✓	qalaqala	60		- qala	- qala
fahla	50		✓	✓	qatha	74		✓	✓
fahlafahla	50		-	✓	qathatha	71		✓	✓
gaga	70		✓	✓	qekelele	136		✓	✓
gelekeqe	69		✓	-	qha	226		≠	≠
gidi	141		✓	✓	qhamu	105		✓	✓
gozololo	81		✓	✓	qho	53		✓	✓
gqi	85		≠	✓	qhu	51		✓	? (qho)
gqwa	51		≠	✓	qhwa	63		≠	-
gubhu	73		✓ gúbu	✓	qhwaba	60		✓ qhwába	✓
ha	50		✓ ha or haa	? (interj, ≠)	qingqo	52		✓	✓
heqe	88		≠	≠	qithi	70		✓	✓
hluthu	51		≠	≠	qu	92		✓	✓
hlwi	88		✓	✓	quthu	58		-	≠
ja	85		✓ jaa	✓	qwa	233		✓	✓
jeqe	175		✓	✓	saka	314		✓	✓
ji	72		✓ jii	✓	shelele	131		≠	✓
juqu	55		✓	✓	shi	261		✓	✓
khaxa	58		✓	✓	shu	119		✓	-
khimilili	105		✓	✓	shwi	69		✓ shwii	✓
khumu	86		≠	(≠)	tebhhu	68		✓ tébu	✓
klabe	85		✓ klábe	✓	thushu	127		✓	✓
klwi	61		✓ klwii	✓	thuthu	69		✓	✓
lacu	96		✓	✓	thwi	62		✓	✓
lungu	98		✓	✓	tu	50		✓	✓
mbo	167		✓	✓	vo	365		✓	✓
memfu	98		✓	✓	wu	64		✓ wuu	✓
mpo	68		✓	✓	xaxa	194		≠	≠
mpu	66		✓	-	xhifi	119		✓	✓
ncamashi	114		-	-	zwi	100		✓	✓

(-), a further 8 have been assigned meanings unattested in the corpus (\neq), and 2 more are errors (?). Regarding these errors: The ideophone *ha* 'of extreme action' has been lumped with the homonymous interjection *ha* 'ha!; gosh!', but the ideophonic meanings given are unattested; and the ideophone *qhu* 'of exploding/bursting sound; of burning, of being dry' has been misspelled *qho* (while nonetheless being lemmatized between *-qhova* and *-qhuba*).

With regard to a dictionary's macrostructure, one thus notes that even though a dictionary like D&V contains as many as 2 600 ideophones, one fifth of the most frequently occurring hundred ideophones are either missing or have been entered with an unattested meaning. For two of these, however, an incomplete form (*chithi* rather than *chithi saka* 'of scattering all over') or the unreduplicated form (*qala* rather than *qalaqala* 'of looking left and right') have been lemmatized. The ideophone *chithi saka* was lemmatized correctly in D&N. One other ideophone missing in D&V has been entered correctly in D&N: *fahlafahla* 'of speaking briefly'. Conversely, five ideophones entered in D&V are missing from D&N: *shu* 'of going right into', *do* 'of nothing', *gelekeqe* 'of completeness; of sudden action', *mpu* 'of looking around in search of something', and *nci* 'of happiness, of surprise'.

Although it may not immediately be apparent from the statistics presented in Table 1, the semantic characterization found in D&N (\checkmark) is on the whole better than that found in D&V (\checkmark). Indirectly, this may be deduced from the fact that four ideophones with unattested meanings in D&V (\neq) have attested meanings in D&N (\checkmark). Of course, one should not rule out the possibility of meaning shifts and changes. Even though the corpus used spans sources from the 1930s to the present (cf. De Schryver 2008: 69), it is entirely plausible to assume that Doke's data collection included ideophones that were common before the 1930s: After all, Doke's D.Litt. dissertation, *The Phonetics of the Zulu Language*, was published in 1926. A shift in meaning may for example be seen for the ideophone *khumu*:

- D&V (unchanged from the first edition in 1948): 'of taking off, loosening, unharnessing; of crumbling, breaking into small particles'
- D&N (updated edition in 1995): 'of loosening; of becoming unhitched; crumbling; of disengaging in a fight'
- corpus evidence (2009, cf. (8) above): 'of stopping, of retiring'

The absence from both dictionaries of common ideophones, such as *qakala*, used in the fixed expression *-gqizi qakala* 'not care at all', or *phecelezi* 'of saying differently', is regrettable. While the latter may be found in two monolingual Zulu dictionaries, Nyembezi's (1992) *Isichazamazwi sanamuhla nangomuso*, as well as Mbatha's (2006) *Isichazamazwi sesiZulu*, other ideophones are again absent there. A corpus-driven approach to the macrostructure, then, truly helps the lexicographer in decisions on what to include and what to omit.

4.2 The mediostructure

With regard to a dictionary's mediostructure, corpus (frequency) information enables one to connect related aspects dispersed throughout the dictionary text. With regard to ideophones, cross-references may be employed to refer the dictionary user from lesser frequent variants to more frequent ones (e.g. *qhwa* (freq. 63) > *qwa* (freq. 233), used in *-mhlophe qwa* 'bright-white; snow-white; very white'); or to cross-refer closely related forms, both in terms of orthography and meaning (e.g. *ngqo* 'of straightness, of directness; of knocking' vs. *nqo* 'of knocking; of being right on top; of precision'); or to indicate how one ideophone is derived from another, often through reduplication (e.g. *qathatha* 'of falling down on a certain spot; of arriving at an exact time' < *qatha* 'of falling down, of dropping; of arriving; of being solid'); or finally to show how one ideophone may form part of another (e.g. *saka* 'of intensity, of emphasis' vs. *chithi saka* 'of scattering all over').

4.3 The microstructure

Less trivial than the macrostructural differences discussed in Section 4.1 are the microstructural ones. In this regard, the ticks in Table 1 do not imply a perfect correspondence between the meaning(s) uncovered in our corpus-driven study, and the meaning(s) recorded in the existing dictionaries. Two examples follow to illustrate this; first the treatment of *qho* as found in D&V in (27), in D&N in (28), and according to our analysis in (29):

- (27) **qho** (8–9) ideo. [> **qhoza** ; **qhoqhoza** ; **uḃuqhoqhoqho** ; **uqhoqhoqho**.]
 1. of the noise of heavy boots, heavy steps. *ukuthi qho qho* or *ukuthi qho qho qho*.
 2. of the bubbling sound of a smoking-horn.
 3. of thoroughness, completeness, perfection. *Lomuntu unawe qho* (This man keeps fast by you, or is shadowing you). *ḃakuḃoné qho* (They saw you clearly). *Ngithé qho naye emseḃenzini wami* (I am competing with him on equal terms in my business).
- (28) **qho** (ideo) of determination; of sharp footsteps (as one walking on wooden floor).
- (29) **qho** ideophone **1** ► (of realness, of trueness) ♦ Amabhubesi yizona **qho** izitha zezi-ndlulamithi. • *Lions are the real enemies of giraffes.* **2** ► (of directness) ♦ Manje uvalo lokucwila selunaye **qho**. • *The fear of drowning is now directly upon him.* **3** ► (of completeness, of perfection) ♦ Ngikuphicaphica ngesilwane sami esinemilenze emine, kepha esehluleka **qho** ukuhamba ngemilenze yaso. [isiphicaphicwano; impendulo = itafula] • *I let you guess my animal that has four legs but is entirely unable to walk on its legs. [riddle; answer = table]* ♦ Lo muthi ufana **qho**, nomhlwazi. • *This tree is completely the same as the Bushman's tea tree.* **4** ► (of repetitive sound) ♦ Walibeka emlonyeni igudu wadonsa, wadonsa, lasho pha-

kathi lathi: "qho, qho, qho." • *He put the smoking horn in his mouth and pulled and pulled, until it sounded "puff, puff, puff" inside.*

The analysis seen in (29) summarizes the corpus evidence, with the so-called 'senses' ordered in order of corpus frequencies. In our analysis, a total of four general meanings were uncovered. Each meaning is typically supported by one real example from the corpus, sometimes more when the evidence indicates a wide semantic range (as for the second sense in (29)). Looking back from (29) to (27), one notices that D&V's first two senses are actually two *instances* of a more generic meaning, 'of repetitive sound', and should thus actually have been lumped. That sense is not the most frequent one. The more frequent meanings, 'of realness, of trueness' and 'of directness', have however been missed by D&V, with only the third meaning overlapping in both studies. When one now considers D&N in (28), one notices that the first two meanings seen in the corpus are again absent, with the next two variations in (non-generic) formulation.

As a second example, see the treatment of *ngqa* as found in D&V in (30), in D&N in (31), and according to our analysis in (32):

- (30) **ngqa** (3–8) *ideo.* [> **umngqaka** ; **ngqábalazi.**]
 1. of looking straight in the face; of seeing for the first time. *Lomuntu ngimuthi ngqa ukumbona* (I see this person for the first time).
 2. of brightness, brilliance. *Indlu ekhanyiswa ngogesi ithi ngqa* (A room lit by electricity is brilliant).
 3. of finishing off at a stroke. *Wayithi ngqa ingilazana yotshwala* (He drunk off at a single draught the glass of beer).
- (31) **ngqa** (*ideo*) of seeing for the first time (used after verb *qala*.)
- (32) **ngqa** * *ideophone*
 ▪ **-qala ngqa** ► **do for the very first time** ♦ *Umdlalo wokuqala ngqa adlala kuwo weshashalazi kwaba ngokaMbongeni Ngema obizwa ngokuthi Asinamali 2.* • *The very first stage play in which she played was one by Mbongeni Ngema called Asinamali 2.*

For the ideophone *ngqa* corpus evidence overwhelmingly points to one specific fixed expression, as seen in the screenshot reproduced in Figure 2. With a frequency of 564 and a rank of 1 272, the ideophone *ngqa* belongs to the top 1 500 lemmas in our project, hence the star rating (*) in (32). When this highly frequent ideophone is preceded by the verb *-qala* 'begin; start; commence', the meaning of *-qala ngqa* is 'do for the very first time'. This, then, is also the evidence summarized in the dictionary article shown in (32). D&N in (31) got the treatment almost right: They still focus on 'seeing' for the first time rather than the more generic 'doing' for the first time. D&V in (30), however, only vaguely approach the evidence with their first sense. The other senses offered by D&V are *very rare*: Only about 30 (out of 564!) corpus lines have a meaning that is different from the one seen in (32).

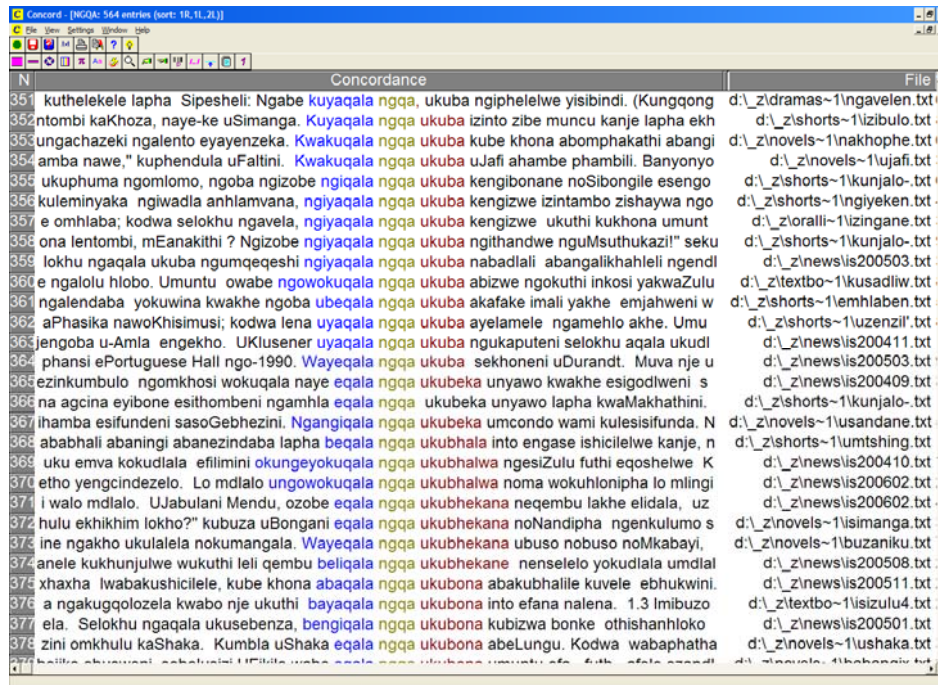


Figure 2: Corpus evidence for *ngqa* as *-qala ngqa* 'do for the very first time'

Without wanting to discredit the great contribution to Zulu lexicography by D&V (a dictionary praised for its detailed linguistic description), as well as D&N (a dictionary praised for its inclusion of numerous relevant combinations and fixed expressions), it should be clear that an electronic Zulu corpus enables the dictionary compiler to take especially the semantic aspects of the Zulu lexicon to the next level. Large amounts of real evidence enable a far more precise delineation of the semantic import; not only of single items (cf. Sections 3.1 and 3.2), but also of items frequently collocating with other items (cf. Sections 3.3 and 3.4), or patterns colligating (cf. Section 3.5).

5. Discussion

In his linguistic study of the Zulu ideophone, Von Staden (1977: 195) summarized the semantic aspects as follows:

Semantically, ideophones have a binary function. On the one hand, ideophones tend to be more explicit than corresponding non-ideophonic forms; an implication of this feature is that they also intensify meanings. On the other hand, ideophones also differentiate more precisely within a specific semantic field. Many ideophones have a great number of semantic variants, whilst quite a number of synonyms and homonyms are also found. Certain voice quality features and ges-

tures play an important supporting role especially in respect of the semantic aspect of differentiation.

Although it is tempting to accept Von Staden's dual view of ideophones, the second aspect — i.e. that ideophones would also differentiate more precisely within a specific semantic field — is not corroborated by large amounts of corpus data for Zulu. Rather, ideophones acquire a *specific* meaning only when combined with other parts of speech, especially verbs. Or, as Childs (1993: 22) puts it: "Semantically, ideophones can do as little as simply underscore the meaning of the verb with which it has a close collocational association." The first part of the dual view is attested by corpus data, and is apparent from the lexicographic data presented throughout this article.

That there are many semantic variants, synonyms and homonyms is mostly confirmed. One particularly 'popular' ideophonic sense is 'of extreme action'. In the text so far, this meaning has been encountered for the ideophones *mpo* (9), *nsi* (11), *nse* (17), and *ha* (Section 4.1). It is also attested for the ideophones *bhe*, *hluthu*, *mbo*, *ngci*, *qingqo*, *wu*, and *zwi*. (Note that these eleven ideophones are not (necessarily) freely interchangeable for this meaning: There is a need for real examples that show typical environments.)

Our lexicographic carving up of the homonymy-polysemy (dis)continuum is rather uncommon, in that we decided to group all so-called 'senses' of a particular orthographic form under that form, even when one would be able to recognize different homonyms based on phonetic, morphological or even semantic grounds.⁵ This decision was taken with the target user group in mind (for whom a single listing is easier to process), and was also inspired by the fact that the assigned meanings are generic, which excludes polysemy by definition.

Lastly, although the supporting role played by voice quality and gestures was not studied (these are aspects that are absent from the now common but limited text-only corpora), given the 'vivid' aspect of ideophones, they are perfectly acceptable as correct. In a paper dictionary such aspects could be covered in the extra matter, while ideophones are of course prime candidates for audio and especially video illustrations in an electronic dictionary (cf. De Schryver 2003: 165-167).

In conclusion, then, although seen as a lexicographer's worst nightmare, a careful corpus-driven study of ideophones enables the dictionary compiler to present dictionary articles for ideophones that 'look' similar to the articles for lemmas in other word classes. In reality, however, the traditional (monolingual) definitions and (bilingual) translation equivalents are actually carefully crafted, generic paraphrases. These paraphrases are supported by hand-picked authentic examples, examples meant to substantiate the possible range of each set of paraphrases. Where relevant, the ideophonic import may be and has to be stemmed by means of the inclusion of common combinations, fixed expressions, and patterns colligating, at which point generic meanings morph into highly precise vivid language. Specifically for bilingual lexicography with a

Bantu language as the source language, the presumed problem of the non-existence of a word class 'ideophone' in the target language becomes irrelevant from the moment paraphrases rather than translation equivalents are employed.

Acknowledgements

The author would like to thank *Ghent University* for its continued support of his field trips in Africa.

Endnotes

1. Although oft-quoted, hardly any scholars seem to go back to the original source, rather copying mistakes in typography and spelling from one another, and even missing out on crucial words (like 'sound'). Typographical errors may be found in the OED citation, missing and misspelled words in, for example, Childs (1994: 180) and Allan (2001: 139).
2. For more on sociolinguistic aspects of (Zulu) ideophones, see the highly revealing study by Childs (1996).
3. To determine the total number of lemmas in Doke and Vilakazi's (1953) *Zulu-English Dictionary*, every fiftieth page was sampled starting with page 50, and the page average (32.67 lemmas per page) was multiplied with the total number of pages (918 pages), resulting in an extrapolated 29 988 lemmas overall.
4. To determine the total number of lemmas in Dent and Nyembezi's (1995) *Scholar's Zulu Dictionary*, the first page and every subsequent fifteenth page was sampled, and the page average (62.13 lemmas per page) was multiplied with the total number of pages (218.5 pages), resulting in an extrapolated 13 576 lemmas overall.
5. For example, the two so-called 'senses' in (19) — 'of straightness, of directness' and 'of knocking' — are clearly derived from different verbal stems.

References

- Allan, K. 2001. *Natural Language Semantics*. Oxford: Blackwell Publishers.
- Century Dictionary Online (initially published between 1889 and 1909). 2009. Dictionary Entry for 'Ideophone' [online]. <http://www.global-language.com/CENTURY/>
- Childs, G.T. 1993. Lexicography in West Africa: Preparing a Bilingual Kisi-English Dictionary. *Lexikos* 3: 13-28.
- Childs, G.T. 1994. African Ideophones. Hinton, L., J. Nichols and J.J. Ohala (Eds.). 1994. *Sound Symbolism*: 178-204. Cambridge: Cambridge University Press.
- Childs, G.T. 1996. Where Have All the Ideophones Gone? The Death of a Word Category in Zulu. *Toronto Working Papers in Linguistics* 15: 81-103.
- Dent, G.R. and C.L.S. Nyembezi. 1995³ [1988², 1969]. *Scholar's Zulu Dictionary*. Pietermaritzburg: Shuter & Shooter.

- 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. A New Way to Lemmatize Adjectives in a User-friendly Zulu-English Dictionary. *Lexikos* 18: 63-91.
- De Schryver, G.-M.** 2008a. The Lexicographic Treatment of Quantitative Pronouns in Zulu. *Lexikos* 18: 92-105.
- De Schryver, G.-M. and E. Taljard.** 2007. Compiling a Corpus-based Dictionary Grammar: An Example for Northern Sotho. *Lexikos* 17: 37-55.
- De Schryver, G.-M. and A. Wilkes.** 2008. User-friendly Dictionaries for Zulu: An Exercise in Complexicography. Bernal, E. and J. DeCesaris (Eds.). 2008. *Proceedings of the XIII EURALEX International Congress, Barcelona, 15-19 July 2008*: 827-836. Sèrie Activitats 20. Barcelona: Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada.
- Dingemans, M.** 2009. *How to Do Things with Ideophones: Observations on the Use of Vivid Sensory Language in Siwu*. Presentation at the SOAS Research Seminar, June 3. London: The School of Oriental and African Studies.
- Doke, C.M.** 1926. *The Phonetics of the Zulu Language*. D.Litt. Dissertation. *Bantu Studies: A Journal Devoted to the Scientific Study of Bantu, Hottentot, and Bushman*; Volume 2 (July 1926, Special number). Johannesburg: Witwatersrand University Press.
- Doke, C.M.** 1935. *Bantu Linguistic Terminology*. London: Longmans, Green & Co.
- Doke, C.M. and B.W. Vilakazi.** 1953² [1948]. *Zulu-English Dictionary*. Johannesburg: Witwatersrand University Press.
- Fivaz, D.** 1963. *Some Aspects of the Ideophone in Zulu*. M.A. Thesis. Hartford Studies in Linguistics 4. Hartford: Hartford Seminary Foundation.
- Gauton, R.** 2008. Bilingual Dictionaries, the Lexicographer and the Translator. *Lexikos* 18: 106-118.
- Hanks, P.** 1987. Definitions and Explanations. Sinclair, J.M. (Ed.). 1987: 116-136.
- Jadezweni, M.W.** 1998. Book Review: B. Kirsch, S. Skorge and N. Matsiliza. 1996. *An English-Xhosa Companion for Health-Care Professionals*. *Lexikos* 8: 322-325.
- Khumalo, L.** 2002. Defining Formats and Corpus-based Examples in the General Ndebele Dictionary, *Isichazamazwi SesiNdebele*. *Lexikos* 12: 264-274.
- Language Hat* (Stephen Dodson's Blog). 2008. A Couple of Blogs — January 10, 2008 [online]. <http://www.languagehat.com/archives/002995.php>
- Mbatha, M.O.** 2006. *Isichazamazwi sesiZulu*. Pietermaritzburg: New Dawn Publishers.
- Msimang, C.T. and G. Poulos.** 2001. The Ideophone in Zulu: A Re-examination of Conceptual and Descriptive Notions. Voeltz, F.K.E. and C. Kilian-Hatz (Eds.). 2001: 235-249.
- Ntuli, N.S.** 1988. *Izwi Lendlovu Enkulu*. Johannesburg: Scholastic Publishers.
- Nyembezi, C.L.S.** 1992. *Isichazamazwi sanamuhla nangomuso*. Pietermaritzburg: Reach Out Publishers.
- OED Online* (based on the second edition of 1989). 2009. *Oxford English Dictionary* [online]. <http://dictionary.oed.com/>
- Poulos, G. and C.T. Msimang.** 1998. *A Linguistic Analysis of Zulu*. Cape Town: Via Afrika.
- Samarin, W.J.** 1967. Determining the Meanings of Ideophones. *Journal of West African Languages* 4(2): 35-41.
- Samarin, W.J.** 1971. Survey of Bantu Ideophones. *African Language Studies* 12: 130-168.
- Seed, J.** 1968. *The Voice of the Great Elephant*. New York: Pantheon Books.

- 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: Collins ELT.
- Taljaard, P.C. and S.E. Bosch.** 1993² [1988]. *Handbook of isiZulu*. Pretoria: J.L. van Schaik.
- The Ideophone* (Mark Dingemans's Blog). 2008. On the History of the Term 'Ideophone' — January 10, 2008 [online]. <http://ideophone.org/on-the-history-of-ideophone/>
- Voeltz, F.K.E.** 1971. Toward the Syntax of the Ideophone in Zulu. Kim, C.-W. and H. Stahlke (Eds.). 1971. *Papers in African Linguistics*. Edmonton: Linguistic Research.
- Voeltz, F.K.E. and C. Kilian-Hatz (Eds.)**. 2001. *Ideophones*. Amsterdam: John Benjamins.
- Von Staden, P.M.S.** 1974. *Die ideofoon in Zulu*. Unpublished Ph.D. Dissertation. Johannesburg: Rand Afrikaans University.
- Von Staden, P.M.S.** 1977. Some Remarks on Ideophones in Zulu. *African Studies* 36(2): 195-224.
- Weakley, A.J.** 1973. *An Introduction to Xhosa Ideophone Derivation and Syntax*. B.A. Hons. Thesis. Communication 2. Grahamstown: Rhodes University, Department of African Languages.

L'apport de la sociolinguistique à la lexicographie: l'exemple récent des belgicisms dans l'édition du *Nouveau Petit Robert* 2008¹

Jean-Nicolas De Surmont, *Université de Leeds, Leeds, Royaume-Uni*
(*jdesurmont@yahoo.fr*)

Résumé: Cet article souhaite analyser le contenu de l'inclusion des 48 nouveaux belgicisms dans le *Nouveau Petit Robert* 2008 (dorénavant NPR 2008). Il retrace les antécédents de collaborations entre différents linguistes et les Éditions Robert pour l'inclusion de belgicisms et évoque le processus qui a conduit à la décision de mettre à jour la nomenclature des belgicisms dans le NPR 2008. Il fait aussi le point sur ces belgicisms en comparant, tableau à l'appui, la présence des belgicisms récemment ajoutés à l'aide d'autres dictionnaires français. À partir de cette comparaison et des données sociolinguistiques fournies dans le cadre du projet de recherche Valibel, il s'interroge enfin sur la pertinence de l'inclusion de ces belgicisms.

Mots-clés: FRANÇAIS DE BELGIQUE, METALEXICOGRAPHIE, SOCIOLINGUISTIQUE

Abstract: The Contribution of Sociolinguistics to Lexicography: The Recent Example of Belgicisms in the Edition of the *Nouveau Petit Robert* 2008. This article aims to analyse the nature of the 48 new Belgicisms included in the *Nouveau Petit Robert* 2008 (henceforth NPR 2008). It traces the history of the collaboration between different linguists and the Robert editions regarding the inclusion of Belgicisms and touches on the process leading to the decision to include the list of Belgicisms in the NPR 2008. It also takes stock of these Belgicisms by comparing them with the aid of a table to the Belgicisms recently added by means of other French dictionaries. Lastly, proceeding from this comparison and the sociolinguistic evidence furnished by the research project Valibel, it questions the relevance of the inclusion of these Belgicisms.

Keywords: BELGIAN FRENCH, METALEXICOGRAPHY, SOCIOLINGUISTICS

Introduction

C'est en 1968² et non en 1972 comme on l'affirme souvent, que les dictionnaires de langue française ont commencé à inclure des francophonismes. Parmi ceux-ci mentionnons les belgicisms qui retiendront notre attention dans ce texte. C'est plus précisément la contribution récente du NPR 2008 dans la mise à jour de ces belgicisms qui sera analysée. L'ajout de 48 belgicisms dans le NPR 2008 à la suite de notre collaboration au projet Valibel s'avère être une contribution historique puisque depuis une vingtaine d'années le Robert n'incluait

plus de belgicisms. Afin de voir les critères qui ont été retenus pour sélectionner ces belgicisms et la dimension plus ou moins novatrice du choix lexicographique, nous avons dépouillé un corpus de quelques dictionnaires de langue française récents et avons comparé les résultats d'enquêtes en Belgique francophone avec les mots sélectionnés. Au terme de ces comparaisons, différents points de conclusion émergent, permettant d'évaluer les forces et les faiblesses de ces choix.

Antécédents de collaboration des éditions Robert pour l'inclusion des belgicisms

C'est par l'entremise d'Émile Seutin (*Petit Robert* (dorénavant PR) en 1983³) et de celle de Maurice Piron⁴ (*Grand Robert* (dorénavant GR) pour l'édition de 1985) que les Éditions Robert ont ajouté à la nomenclature du PR et du GR de nombreux belgicisms. Le *Dictionnaire du français vivant*, l'un des premiers dictionnaires reconnus pour avoir tenu compte du français régional, comptait déjà 270 lexèmes belges en 1973.⁵ Lors d'une conversation en 2003 avec Danièle Morvan,⁶ celle-ci avait mentionné que la nomenclature déjà présente suffisait à ces yeux et une proposition avait été faite de combler les lacunes du GR et du PR.⁷ Entre septembre 2003 et juin 2004, nous nous sommes attelés à recenser la présence de l'ensemble de la nomenclature des belgicisms du Centre de recherche Valibel (Variétés Linguistiques du français en Belgique⁸) dans les dictionnaires suivants: *Grand Robert*, *Banque de données de langue française panfrancophone*, dans la base textuelle Frantext et dans le *Trésor de la langue française informatisé*. Ce travail de longue haleine et minutieux devait pour ainsi dire conduire à l'établissement des choix que Valibel prendrait ensuite sur la pertinence ou non d'inclure certains belgicisms dans le PR permettant aussi la confirmation ou le retrait de plusieurs belgicisms déjà existant (l'enquête sociolinguistique et l'enquête métalexigraphique comparée étant complémentaires). Le linguiste ardennais Michel Francard nous enjoignait à noter tout type d'information que ne possédait pas déjà le Centre de recherche Valibel et d'y ajouter ces informations sur une fiche de contrôle. Toute connaissance encyclopédique était évidemment la bienvenue, aidant en même temps à la constitution des données métalinguistiques du *Dictionnaire du français en Belgique* (dorénavant DFB), projet de recherche en cours au Centre de recherche Valibel de l'Université Catholique de Louvain et qui conduira à la publication d'un dictionnaire (le titre du projet étant celui du dictionnaire à paraître) sur le français de Belgique dont la principale caractéristique novatrice sera la prise en compte de la sociolinguistique par la technique lexicographique.

Le Dictionnaire du français en Belgique

L'ajout des belgicisms dans le NPR 2008 s'inscrit dans un projet plus vaste,

celui du Centre de recherche Valibel et plus précisément au sein de la composante lexicographique de ce centre. La composante lexicographique de l'équipe (Régine Wilmet, Geneviève Geron) accueille régulièrement des étudiants et des chercheurs étrangers pour la collaboration au dictionnaire. Or, l'activité de publications métalexographiques par le centre de recherche nous semble moins importante⁹ que celle émanant de la partie de l'équipe dont les activités sont centrées sur la phonologie du français contemporain (PFC) et sur la sociolinguistique du français en Belgique. Nous avons voulu remédier à cette faiblesse en commentant de façon plus attentive le travail métalexographique de l'équipe, tout en portant notre attention sur l'ajout des belgicisms dans le NPR 2008. Il convient néanmoins d'évoquer à grands traits les dimensions novatrices qu'apportera le DFB à la lexicographie du français contemporain tant sur le plan méthodologique que sur le plan du contenu.

La majorité des dictionnaires spécialisés sont des encyclopédies (dictionnaires de médecine, d'informatique, de marine) et «beaucoup plus rarement de véritables dictionnaires de langue proprement dit».¹⁰ Issu de dépouillements importants de corpus linguistiques, de métalinguistique (belges et français surtout), d'enquêtes sociolinguistiques, le DFB, à visée linguistique et encyclopédique, accordera une large part de ses informations au français des médias, écrits et audio-visuels. Le futur DFB semble constituer, selon Marc Van Campenhoudt, un dictionnaire complet. L'ouvrage paraîtra chez De Boeck et suivra vraisemblablement le modèle du dictionnaire de Pierre Rézeau (bilan lexicographique, carte géolinguistique¹¹), même si de légères différences sont à signaler, comme le fait que l'enquête de vitalité précède l'établissement de la nomenclature et l'adjonction d'images offrant ainsi des développements encyclopédiques importants. Signalons à titre de rappel historique que *l'Atlas linguistique de France* (1902–1920) d'Edmond Edmont et Jules Gilliéron comportait lui-même 23 points d'enquête en Wallonie; il figure comme précurseur de *l'Atlas Linguistique de la Wallonie* d'après les enquêtes du dialectologue liégeois Jean Haust,¹² qui conduit son enquête dans 400 communes, un réseau de points très serrés. L'introduction en lexicographie régionale de cartes géolinguistiques nous semble une grande innovation et témoigne surtout de l'apport théorique et méthodologique de la sociolinguistique sur l'activité lexicographique du français en Belgique.¹³ À ce titre, notre perspective s'est aussi alignée sur celle de Rézeau pour ce qui concerne le dépouillement d'une documentation qui ne s'appuie pas uniquement sur des dictionnaires de référence mais aussi sur des recherches d'attestations provenant de l'Hexagone.¹⁴ La publication du *Dictionnaire des régionalismes de France* (dorénavant DRF) sous la direction de Rézeau constitue non seulement une étape dans l'étude du français régional mais aussi l'achèvement d'une entreprise largement inspirée des *Matériaux pour l'étude des régionalismes du français* publiés en onze volumes entre 1983 et 2003; elle semble témoigner de la volonté de l'administration française de décentraliser (loi de la décentralisation datant de 1982) ses efforts en vue d'un regain pour les études du français régional.¹⁵

Les belgicisms du *Nouveau Petit Robert* 2008

L'analyse de la nomenclature des belgicisms permet de constater que les belgicisms retenus sont pour la plupart des belgicisms lexicaux;¹⁶ seul quelques belgicisms sémantiques ont été retenus. Ces particularités lexicales constituent comme beaucoup de topolectismes des emprunts, interférences ou calques du wallon ou du flamand, moins souvent de l'anglais, ainsi que la dérivation suffixale et les changements de sens.¹⁷ Bien que les topolectismes belges contiennent également des faits syntaxiques, leur description est souvent bien minime dans les dictionnaires et les travaux qui leur sont consacrés sont encore moindres. Parmi ces belgicisms lexicaux certains sont des belgicisms de partage c'est-à-dire, selon De Surmont (2004), des unités lexicales également employées dans d'autres variétés de français que celle de Belgique (Congo, Suisse, Québec). L'extension géographique d'un emploi constitue ce que Pierre Rézeau décrit comme de la *dérégionalisation*.¹⁸ La majorité des belgicisms figurent déjà dans le *Petit Larousse illustré* (dorénavant PLI) et dans le dictionnaire de Georges Lebouc (1998) et cela depuis parfois dix ans ce qui n'est guère étonnant étant donné que le PR n'avait guère ajouté de belgicisms depuis la collaboration du dialectologue liégeois Maurice Piron.

La majorité des belgicisms ont surtout été retenus en fonction des résultats de l'enquête de vitalité sur la base d'un taux de reconnaissance (vitalité passive) supérieure à 90% parmi les informateurs.¹⁹ L'enquête de vitalité, comme nous l'avons déjà mentionné,²⁰ constitue l'une des particularités du dictionnaire de Valibel en ce qu'elle précède l'établissement de la nomenclature du DFB. À cet égard l'établissement de la vitalité des emplois s'inscrit en partie dans une tradition inaugurée par le *Trésor de la langue française* (dorénavant TLF), premier dictionnaire ouvert au quantitativisme, même s'il avait négligé les attestations essentiellement littéraires présentant un net décalage avec le français oral. Les enquêtes ont été menées en Wallonie et à Bruxelles tenant compte de l'âge de la scolarité et de la localisation des informateurs et sont plus contrôlées que celle pratiquées par Baetens Beardsmore. Dans le cadre de notre analyse comparative nous avons aussi questionné quelques personnes sur la nomenclature sélectionnée afin de tester les choix opérés par l'équipe de Valibel (Jean-Marie Le Clézio, Jacques De Decker et des étudiants de Bruxelles de provenances variées). Certaines acceptions consignées des belgicisms ont été écartées lorsque le taux de vitalité était estimé trop bas. L'écart entre la forte vitalité d'emploi de certains lexèmes et l'absence de représentation dans les sources métalinguistiques françaises nous semble aussi un facteur récurrent qui a justifié la sélection de certains lexèmes. Il est évident que ce jugement est porté en fonction du dictionnaire français qui sera publié chez De Boeck, ainsi que des acceptions non retenues pour le NPR 2008 étant donné les contraintes d'espace. Cette décision de retenir les belgicisms en fonction des résultats d'une enquête sociolinguistique effectuée sur le territoire francophone de Belgique nous permet d'affirmer que c'est en fonction du lectorat belge du PR qu'ont été retenus les belgicisms et non en fonction des lecteurs français.

NOMENCLATURE DES BELGICISMES AJOUTÉS DANS LE NOUVEAU PETIT ROBERT 2008³⁰

NPR 2008	Delcourt 1998-1999	PLI 2008	Lebouc 1998	Lebouc 2006	Doppagne 1979	Massion 1987	TLF 1971-1994	Bal et al. 1994
Avant plan	✓	✓ (?)	✓	✓	✓ réf. au Larousse XIXe s.	✓	∅	Signalé comme rare en France
babelutte	∅	✓ (2006)	✓ sous bablute	✓	✓	✓	✓	Le signale aussi à Lille, etc.
barémique	✓	✓ (1998)	✓	✓	∅	∅	∅	✓
blinquer	✓	✓ (?)	✓	✓	∅	✓ (1933)	✓	✓
broil	✓	✓ (2000)	✓	✓	∅	✓	∅	✓
3. buse	✓	✓ (?)	✓	✓	∅	✓	✓ (sans mention du fr. de Belgique)	✓
Buser	✓	✓ (?)	✓	✓	∅	✓	✓	✓
cacaille	✓	✓ (1998)	✓	✓	∅	∅	∅	✓
caillant, ante	✓ (1983)	✓ (2000)	✓	✓	∅	✓	∅	✓
Carabistouille	✓	✓ (?)	✓	✓	∅	✓	✓ citant Beardsmore-	✓
castard ou castar, arde	✓	✓ (1998)	✓	✓	∅	✓	∅	✓
chantoir	✓	✓ (?)	∅	∅	Long commentaire en réf. Maurice Piron	✓	✓	✓
Copion	✓	✓ (1998)	✓	✓	∅	✓	∅	✓
croille	✓	✓ (?)	✓	✓	∅	✓ le donne comme dialectal	∅	✓

NPR 2008	Delcourt 1998-1999	PLI 2008	Lebouc 1998	Lebouc 2006	Doppagne 1979	Massion 1987	TLF 1971-1994	Bal et al. 1994
crollé,ée	✓	✓ (?)	✓	✓	∅	∅	∅	✓
croller	✓	∅	✓	✓	∅	✓	∅	✓
Cumulet	✓	✓ (?)	✓	✓	∅	✓	∅	✓
Doubleur	✓	✓ (?)	✓	✓	✓	∅	∅	✓
Endéans	✓	✓ (?)	✓	✓	✓le mentionne comme étant dans le PR 1977 marqué «Région.».	✓(étymon 1387)	✓	✓
Essuie	✓	✓ (?)	✓	✓	∅	✓	∅	✓
Fancy-fair	✓	✓ (1998)	✓	✓	∅	✓	∅	✓
Fricadelle	✓	✓ (?)	✓	✓	∅	✓	✓(sans men- tion de la Belgique)	✓
Frigolite	∅	∅	∅	∅	∅	∅	∅	∅
Guindaille	✓	✓ (?)	✓	✓	∅	✓	∅	✓
Guindailleur	✓	✓ (2000)	✓	✓	∅	✓	✓	✓
Guindailleur, euse	✓	∅	✓	✓	∅	✓	✓	✓
Kicker	✓	✓ (1998)	✓	✓	∅	✓	∅	✓(sous la forme Kieker)
Kinésiste	✓	✓ (1998)	✓	✓	∅	∅	✓	✓
Kot	✓	✓ (?)	✓	✓	∅	✓(donné 1937 première at- test.)	∅	✓
Koter	∅	∅	✓	∅	∅	✓	∅	✓
Mofler	✓	✓ (?)	✓	✓	∅	✓	✓	✓
mouchette	∅	∅	✓	✓	∅	∅	∅	✓

NPR 2008	Delcourt 1998-1999	PLI 2008	Lebouc 1998	Lebouc 2006	Doppagne 1979	Massion 1987	TLF 1971-1994	Bal et al. 1994
nic-nac ou nicnac	∅	✓ (2005)	✓	✓	∅	✓	∅	✓
nonantaine	∅	✓ (?)	✓	✓	∅	✓	✓	✓
Péquet ou péket	∅	✓ (1998)	✓	✓	∅	✓	∅	✓
Pils	✓	✓ (1998)	✓	✓	∅	✓	∅	✓
Plasticine	∅	✓ (1998)	∅	∅	∅	∅	∅	✓
Prester	✓	✓ (?)	✓	✓	∅	✓	✓	✓
Ring	∅	✓ (?)	✓	✓	∅	✓	∅	✓
Septantaine	∅	✓ (?)	✓	✓	∅	✓	✓	✓
Singlet	∅	✓ (?)	✓	✓	∅	∅	✓	✓
socket ou soc- quet	✓	✓ (1998)	Sous soquet ✓	✓	∅	∅	✓	✓
Subsidier	✓	✓ (?)	✓	✓	∅	∅	∅	✓
Taque	✓	✓ (?)	✓	✓	∅	✓ (première attest. 1793)	✓	✓
Taximan	✓	Sorti du PLI en 1998	✓	✓	∅	∅	∅	✓
Tévé	∅	∅	✓	∅	∅	∅	✓ (sans men- tion de la Belgique)	✓
toute-boîte ou toutes-boîtes	✓	✓ (?)	✓	✓	∅	✓	∅	✓
trouiller	∅	∅	∅	✓	∅	∅	∅	∅
zoning	✓	✓ (?)	∅	✓	∅	∅	∅	∅

P.S.: Le point d'interrogation signifie que nous n'avons pas pu contrôler la source de la première attestation.

Ces belgicisms, absents du TLF, du *Grand Robert* de la BDLP²¹ et de la base de données textuelles Frantext et de quelques sources que j'avais contrôlées, se trouvent également dans les corpus des 48 belgicisms du NPR 2008. On peut de même observer, sur la base d'une analyse comparative avec la nomenclature des belgicisms consignée dans le *Dictionnaire universel* d'Hachette, que les exemples mentionnés, voire les définitions, sont souvent plus courts dans le NPR 2008, même si certains mots apparaissant dans le NPR 2008 ne figurent pas dans le *Dictionnaire universel*. Les quelques toponymies de partage avec les helvétismes sont tirés du *Dictionnaire suisse romand* (dorénavant DSR) d'André Thibault et Pierre Knecht (1997), contrairement au dictionnaire de belgicisms de Christian Delcourt qui consigne des helvétismes sur la base d'autres sources non reprises par l'équipe de Francard, les deux ayant pourtant travaillé en collaboration jadis pour l'élaboration de la base de données textuelles Beltext.²²

Les améliorations apportées à l'édition 2008 sont évidentes quant à l'harmonisation du marquage diatopique qui a tenu compte de nos propositions émises lors du congrès EURALEX de 2004 (Voir De Surmont 2004). Nous avons effectivement remarqué que le GR, comme de nombreuses sources lexicographiques, souffrait d'un manque d'homogénéité des marques diatopiques. Ainsi, l'ensemble des belgicisms est marqué «Région. (Belgique)». C'est là, il nous semble, vouloir faire preuve d'une plus grande clarté dans la présentation des informations métalinguistiques de l'article lexicographique. Notons que l'information étymologique consignée de plusieurs des belgicisms comporte des données intraterritoriales (flamand, d'origine wallonne), tout autant qu'extraterritorial (anglais) ou d'origine incertaine. Étant donné la nature des enquêtes effectuées, nous pensons que les belgicisms d'origine wallonne comme d'origine dialectale sont le résultat à la fois de la redéfinition du dialecte²³ comme pratique patoisante liée à la disparition même des pratiques patoisantes, en même temps que des échanges féconds entre les sociétés traditionnelles et modernes et la mobilité des locuteurs. Le nombre de locuteurs patoisant diminue ce qui n'empêche évidemment pas le fonds dialectal d'alimenter le français régional. Il est intéressant de signaler à cet effet que l'origine wallonne ou flamande relevé par le GR3 n'est pas toujours consignée dans le NPR 2008. C'est le cas de *buse* /échec à un examen/ dont le sens étymologique wallon est /haut-de-forme/, de *crolle* /boucle de cheveux/.²⁴ Dans le cas de *cumulet*, si le GR3 affirme qu'il s'agit d'un «mot wallon d'orig. obscure», la première syllabe représente probablement *cul*, le NPR 2008 affirme quant à lui qu'il est issu du moyen néerlandais *tumulen* /se tourner avec/, peut-être, sous l'influence du français *cul* (Picardie; Belgique). Ceci confirme la position de Jacques Pohl, dont l'ouvrage, paru en 1962, fait office de premier livre «purement descriptif» pour le français de Belgique.²⁵ Le mot *endéans* fait l'objet d'une marque diatopique associée à une marque diachronique régressive dans le GR3 (vx ou régional (Belgique)), et didact. (droit, admin.) à une information étymologique exhaustive dans le NPR 2008: fin XIV (Tournai) de 1 *en*, 1 *de* et de l'ancien français *enz* «dans», du latin *intus* /à l'intérieur/. Un réel travail a été fait au niveau du

contrôle des sources et de l'information étymologique en s'inspirant des données des rubriques «Histoire et Bilan métalexigraphique» qui apparaîtront dans le DFB permettant d'améliorer ou de préciser l'information étymologique voire de créer la dispersion topolectale des emplois.²⁶ Les bilans métalexigraphiques de François Massion ont été particulièrement utilisés sur le plan diachronique, ce qui est en l'occurrence le cas de *endéans* où les auteurs ont préféré indiquer «fin XIV, 1387» comme l'avait fait François Massion en 1987.²⁷ Pour les attestations plus récentes et la dispersion géographique, notamment en Afrique, le dictionnaire de Christian Delcourt est particulièrement significatif, bien qu'il a été peu exploité par l'équipe Valibel si l'on en juge le fait que les topolectismes de partage ne sont pas mentionnés. Certains traitements nous semblent néanmoins étonnants, comme *guindaille* /sortie joyeuse et bien arrosée/ dont le NPR 2008 fait remonter la première attestation à 1889. Il s'agit probablement d'une altération de *godaille* /ripaille/ qui remonterait au moyen néerlandais *goed ale*, mot à mot «bonne (*goed*) bière (*ale*)». Or le GR3, dont la publication est pourtant antérieure, antidate la première attestation de 9 ans: «mot wallon, p.-ê. de *godaille* d'après *guinse* «beuverie», du francique *winst* /profit/, ou d'après 2. *guindal*». C'est aussi le cas de *mofler* dont le GR3 affirme qu'il provient du wallon (liégeois) *mofe* /moufle, gros gants/; d'où «frapper d'un coup au visage» alors que le NPR 2008 affirme *mofler*: 1930 est un mot d'origine incertaine. Le mot *taque* apparaît comme non marqué dans le PLI et dans le GR3 ce qui peut laisser soupçonner la nécessité de le marquer dans le NPR 2008. On constate aussi une extension topologique de la zone d'emploi du mot *taximan* qui passe d'un emploi marqué français d'Afrique dans le GR3 à un topolectisme aussi bien employé en Afrique qu'à Madagascar et en Belgique dans le NPR 2008.²⁸ Le mot *Tévé* présente une autre particularité. Dans le GR3 aucune marque diachronique régressive apparaît pour l'emploi en français de France alors que le NPR 2008 le mentionne avec la marque Belgique. Une divergence de datation apparaît aussi pour *zoning* où le GR3 mentionne 1934 alors que le NPR 2008 mentionne 1994. Le GR3 ne mentionne par ailleurs aucune restriction géographique d'emploi alors que le NPR 2008 le prétend être un belgicisme.²⁹ Bien qu'ici le bilan métalexigraphique soit mitigé c'est surtout la divergence de résultat entre le GR3 et le NPR 2008 qui nous semble douteux. Une antidatation nous semble tout à fait normale mais de post-dater un emploi de 60 ans d'un produit des dictionnaires Robert antérieur à un autre nous paraît étonnant.

Conclusion

Il faut constater que la représentation de la Belgique dans les dictionnaires Robert a laissé à désirer depuis une vingtaine d'années. On remarque en outre que, contrairement au français québécois, le français de Belgique suscite moins d'intérêt que n'en a suscité en tout les cas la dialectologie wallonne. La publication de dictionnaires universitaires du français de Belgique ne suscite pas non plus un intérêt politique comme c'est le cas au Québec, cela étant peut-être dû

au fait qu'au Québec le désir d'émancipation socio-économique des élites canadiennes françaises et la libération culturelle qui s'en est suivie, ont toujours été intimement liés aux questions linguistiques et s'inscrit dans une politique bilatérale mise en place dans les années 1960.

Les problèmes communautaires, caractéristiques pour la Belgique sont nuisibles pour l'unité belge et l'on ne saurait nier le fait que la Belgique (pays de réformes, de chantiers et de compromis ...) viennent derrière la Suisse et le Québec dans l'élaboration d'un dictionnaire des particularités lexicales incluant des informations de natures historique, philologique et métalinguistiques, comme le *Dictionnaire historique du français québécois* (initié par Marcel Juneau et continué par Claude Poirier) et le DSR et le DRF.

Les ajouts de belgicisms dans le NPR 2008 veulent être plus représentatifs de la vitalité des usages sur la base des acquis des enquêtes sociolinguistiques. Les résultats de ces enquêtes ont donné lieu à la constitution de corpus importants et servent en quelque sorte à mieux contrôler la dispersion géographique et sociolinguistique des usages. Cela permet de confirmer la pertinence des propositions d'Emile Seutin dont les propositions se faisaient sur la base des ouvrages passés et d'intuitions personnelles, et non pas forcément acceptées par les éditions Robert. La question qui se pose dorénavant n'est plus la question de la validité des intuitions personnelles de linguistes mais la représentation des ressources textuelles qui sont à la disposition des linguistes de l'école néolouvainiste de sociolinguistique. L'échantillon de ressources textuelles est-il représentatif des différentes productions orales observables sur le territoire de la Communautés Wallonie-Bruxelles?³¹ Tout au moins ce renouvellement de la technique lexicographique sur la base des acquis de la sociolinguistique tend à vouloir s'écarter de la tradition puriste qui procédait par le couple *dites/ne dites pas* et donnait une vision morcelée de la langue. Néanmoins la composante géolinguistique, en particulier en situation de contact et en tenant compte de l'évolution historique des contacts entre Flamands et francophones devrait aussi orienter la rubrique historique.³²

Dans la mesure où la nomenclature des ajouts de belgicisms du NPR 2008 préfigurant au DFB, ou tout au moins répondant au cadre méthodologique du groupe de recherche Valibel, constituent selon nous un élément objectivable de réflexion sur les méthodes lexicographiques en cours actuellement en Belgique francophone. En outre cela nous aura permis de mieux connaître les pratiques récentes de propositions topoclectales aux dictionnaires de langue générale du français.

À la lumière de ces nouvelles réflexions il serait intéressant dorénavant de mieux cerner le vaste domaine des topoclectismes de partage en relativisant les frontières érigées en barrières entre le français de référence et le français régional. En outre la question des mots attestés en Belgique avant de l'être en France et passés dans l'usage courant et les naissances indépendantes³³ devraient nous conduire à une réflexion plus globale sur la définition du belgicisme prenant en compte non seulement les questions géolinguistiques comme avant mais aussi les questions sociolinguistiques.

Notes

1. Nos remerciements vont à Michel Francard d'avoir mis à notre disposition le matériel de l'enquête géolinguistique et surtout Geneviève Geron de nous avoir assisté régulièrement dans la conduite de nos recherches.
2. C'est en effet le PLI sorti à l'été 1967, dont la sortie coïncide avec le passage du général De Gaulle au Québec, qui consigne quelques québécismes (*bleuet, coureur des bois, débarbouillette, millage, poudrerie, tuque* et *verge*) et non le *Dictionnaire du français vivant* (1972) comme on a l'habitude de le mentionner. Cette même édition du PLI voit disparaître la formule calquée sur celle de Voltaire: «un dictionnaire sans exemples est un squelette», *exemple* ayant été substitué à *citation*.
3. Conversation personnelle Emile Seutin, le 30 octobre 2007. Parti de la Belgique vers le Québec en 1963, Seutin devient professeur de l'Université de Montréal. Il proposa 180 belgicisms dans le PR. Environ 80 pour la première fois en 1984 et ensuite progressivement pendant une dizaine d'années à Josette Rey Debove. Piron et Seutin ne se sont jamais concertés dans leur travail ce qui fait que le PR et le GR pouvant être très divergeant dans leur nomenclature des belgicisms.
4. Notons que Piron a aussi collaboré au *Grand Larousse encyclopédique* et au *Trésor de la langue française*.
5. Voir Jacques Pohl (1979: 136). Il ne comptait pourtant alors que 230 québécismes alors que dans le PLI la présence de québécisme a toujours été plus importante que celles des francophonismes des auteurs communautés de langue française. Marion Spickenbom (1996) mentionne qu'en 1953, Maurice Piron évalue à 150 le nombre de belgicisms consignés dans les dictionnaires généraux de langue française, chiffre porté à 240 vingt ans plus tard avec une tendance à ne considérer alors que les belgicisms lexicaux. En 1972, le *Dictionnaire du français vivant* de Bordas répertorie 270 belgicisms et en 1973, 292. La liste ira en augmentant, Albert Doppagne en recense 405, Heinz Jürgen Wolf 500.
6. Conversation personnelle, octobre 2003.
7. Conversation personnelle, mars 2007. Mentionnons, qu'en août 2006, le site Internet de Valibel indiquait que l'édition 2007 du PR allait ajouter à sa nomenclature des belgicisms ou laissait sous-entendre déjà cela alors que ce n'est en fait que fin juin 2007 à la sortie de l'édition 2008 que les nouveaux belgicisms ont été ajoutés.
8. Localisé à l'Université Catholique de Louvain, ce projet de recherche constitue l'un des maillons importants de la mise sur pied par Bernard Quemada d'un trésor des variétés du français panfrancophone et présente à ce titre des liens importants avec le réseau de Claude Poirier.
9. On notera cependant une partie des travaux issus de notre séjour à Valibel: De Surmont (2004, 2005) et à paraître et les contributions de Michel Francard, Geneviève Geron et Régine Wilmet plutôt axée sur l'apport sociolinguistique à la métalexicographie qu'à la métalexicographie comme telle.
10. Marc Van Campenhoudt (2000: 127).
11. Les cartes géolinguistiques présentent les zones d'ombrage en dégradé variant, au dixième près, selon le taux de reconnaissance pour tout le territoire francophone (Wallonie et Bruxelles incluant aussi Comines) et ont pour certaines déjà été établies entre 1999 et 2002 (cela a été systématiquement utilisé par Pierre Rézeau (en 2001) pour la première fois). (Nous remer-

cions Geneviève Geron de nous avoir «gracieusement» fait part la carte de *aboutonner*.) D'autres cartes sont parues dans les contributions collectives de Francard, Geron et Wilmet. Ajoutons que la rubrique «Citation» comprend des exemples extraits de la toile et d'émissions de radio ce qui constitue en lexicographie régionale de langue française un élément très novateur. En outre, la rubrique «Emploi» confronte l'oral et l'écrit et distribue les taux de reconnaissance et d'emploi des lexèmes sur le territoire francophone de la Belgique. À ce jour, la rubrique historique recense essentiellement des sources métalinguistiques et la forme de ces rubriques rappelle celle instituée par le TLF. Le fond dialectal bien que connu par certains membres de l'équipe (Geneviève Geron, Régine Wilmet, Michel Francard) n'est pas repris systématiquement dans la rubrique historique comme le faisait François Massion (1987) de manière détaillée. Bien que peu mise en évidence, la dimension géolinguistique des *Chasses aux Belgicisms* de Bal et al. a été pris en compte puisque chacun des co-auteurs enquêtait dans sa région: Albert Doppagne couvrant la région de Huy et de Bruxelles, Joseph Hanse celle de Florenne, Bal celle de Charleroi, André Goosse les Ardennes notamment près de Laroche, L. Warnant couvrait la région de Liège et madame Lenoble-Pinson celle de Libramont. Conversation personnelle Michèle Lenoble-Pinson, 29 novembre 2007.

12. Mentionnons que Francard (1995–1996) s'est intéressé à la technique lexicographique du *Dictionnaire liégeois* de Jean Haust. L'ensemble de sa production antérieure nous laisse penser qu'il devrait être plutôt considéré comme héritier de la dialectologie wallonne que de la lexicographie française et qu'il marque à sa manière la transition de l'hégémonie de l'école dialectologie wallonne vers la prise en compte du français de Belgique, trop longtemps considéré que par les puristes. L'enquête de l'ALF cartographiait surtout la variation phonétique au détriment de la variation sociolinguistique. Plusieurs facteurs ont entraîné la définition de nouveaux projets intégrant une composante sociolinguistique variationniste: «l'évolution économique et sociale qui a réduit l'importance démographique des zones rurales aux dépens des zones urbaines, la mise en chantier d'atlas dans les domaines nouveaux à la structuration linguistique particulière comme ceux de la 'Romania Nova', la constitution de la sociolinguistique comme discipline [...]» (Jean-Paul Chauveau 2003: 78). Voir aussi Jean-Paul Chauveau (1995–1995) sur l'entreprise de Jean Haust, et Guylaine Brun-Trigaud, Yves Le Berre et Jean Le Dù (2005) sur l'ALF.
13. On consultera à ce propos le site du réseau EFF (Étude du français en francophonie) à l'adresse suivante: http://www.eff.auf.org/article.php?id_article=25 et Michel Francard, Geneviève Geron et Régine Wilmet (2003: 140). Comme le soulignent avec raison ces mêmes auteurs (p. 141), l'apport de cette discipline à la constitution de la nomenclature reste très limité, malgré les intentions déclarées dans la plupart des ouvrages déjà publiés en langue française. Notons néanmoins les travaux fondateurs de Heinz Fuchs.
14. Voir Francard, Geron et Wilmet (2003: 146) et De Surmont (2004). Notre travail initial à Valibel consistait notamment à rechercher des attestations que les autres membres de l'équipe n'avaient pas trouvées en s'appuyant sur les pages Internet récentes, sur la base Frantext et sur des recherches à partir du moteur de recherche Google.
15. Le dictionnaire écrit en collaboration avec Pierre Rézeau a reçu une bonne critique de toute part, ses mérites ont aussi été vantés par A. Goosse (2007: 87). Voir Pierre Rézeau (2007: 7).
16. La lexicographie variationniste dispose à cet égard de plus d'outils que la grammaire. Ainsi dans l'édition 2007 du *Bon usage*, le grammairien André Goosse peut s'appuyer sur un corpus linguistique fort important pour la mise à jour de sa grammaire mais en revanche le corpus

métalinguistique d'articles, de dictionnaires ou d'études portant sur la syntaxe du français régional a fait l'objet de très peu de travaux. C'est donc à la lexicographie qui décrit parfois des faits syntaxique que la grammaire de référence s'en remet encore en 2007. André Goosse s'est ainsi approvisionné aux *Dictionnaires des régionalismes de France*, dictionnaire que nous n'avions pas dépouillé en 2004 et qui en toute logique ne semble pas avoir été pris en compte dans les belgicisms ajoutés à l'édition 2008 du PR, puisqu'aucun topolectisme de partage figurant dans le NPR 2008 ne figure dans le DRF dirigé par Pierre Rézeau alors que celui-ci fait pourtant référence régulièrement au français de Belgique.

17. La classification des québécoisismes par Claude Poirier (1995) a été adoptée à plusieurs reprises par Michel Francard par la suite, bien que sur le plan méthodologique Francard a pris ses distances vis-à-vis du point de vue restrictif de Poirier. Dans le cas du néerlandais il est permis de se demander comme A. Goose qui du français ou du néerlandais est le receveur ou le donneur. (Voir Goosse 2007: 85.)
18. 2003: 131. Le problème de la régionalisation et de la dispersion géographique des emplois et de la dérégionalisation apportent à réfléchir sur la variation linguistique francophone en Afrique noire. Ainsi Claude Frey propose le terme de «sous-région» pour désigner l'ensemble géographique homogène comprenant plusieurs états africains (pour le concept de «sous-région» voir Klinkenberg 2008: 85). Voir Claude Frey (2003: 167), voir aussi Piron (1973: 295-304). Piron considère comme belgicisms les particularités communes à plusieurs provinces de Belgique.
19. Sur les 48 belgicisms consignés, seuls 2 d'entre eux nous semblent poser un problème, comme *chantoir* qui n'est pas consigné dans la majorité des sources métalinguistiques consultées et inconnu des personnes que nous avons consultées y compris Jacques De Decker, ainsi que et des informateurs de Francard lui-même. Il est fort probable que Francard ait subi l'influence de Maurice Piron malgré les données de l'enquête et des sources métalinguistiques (voir à cet égard le commentaire du lexicographe ardennais et de ses deux collaboratrices 2003: 144-145). Le taux de reconnaissance de *mofler* est moins élevé que les autres lexèmes malgré sa continuelle présence dans les sources métalinguistiques belges y compris dans le GR3. L'enquête de Valibel s'appuie sur un questionnaire écrit adressé à 86 informateurs en plus d'une banque de données textuelles orales importantes obtenues à partir d'enregistrements réalisées avec quelques 530 francophones wallons et bruxellois. (Voir Francard, Geron et Wilmet 2002a: 21; 2003: 142.) Notons que la vitalité selon l'âge, le sexe ou le degré de scolarité est mélangée aux variables régionales et la vitalité selon l'origine géographique est également mélangée aux autres variables mentionnées. L'ensemble de ces variables est corrélée avec les différences de réalisations linguistiques au sein du même système linguistique.
20. Voir De Surmont (À paraître).
21. Mentionnons que même si la BDLP (Base de données lexicographiques panfrancophone) ne serait que peu employée pour la nomenclature de Valibel en novembre 2003, la nomenclature du Delfrabel (Dictionnaire électronique du français de Belgique) comportait 5124 entrées.
22. Source: Conversation personnelle Geneviève Geron, août 2003.
23. Dans le volume de *l'Encyclopédie de la Pléiade* consacré au langage dirigé par André Martinet, Jean Fourquet (cité par Jean Séguy 1973: 65) prétend «que c'est l'expérience de diversités à l'intérieur de ce qu'on considère comme une langue, et les dénominations de choses n'étant pas identiques, ont des ressemblances évidentes, et que, de façon générale, les différences observées ne sont pas telles qu'elles empêchent la compréhension». En 1978 Patrick Charau-

- deau (1978: 16) écrit que le dialecte est considéré comme une variété régionale d'une langue. Or, selon la définition plus récente de Marina Yaguello (1988: 154), le mot *dialecte* possède un sens linguistique et un sens sociolinguistique. Au sens linguistique il signifie la variété d'une langue et au sens sociolinguistique le parler vernaculaire non normé et le plus souvent non écrit n'ayant pas le statut de langue. On pourra aussi se reporter à Léon Warnant (1973: 100).
24. Même si dans ce cas la vitalité de l'emploi a diminué étant donné la disparition progressive du référent, il aurait pu être intéressant de mentionner l'origine wallonne du mot puisque le sens /haut de forme/ est une acception connue du mot en français de Belgique. Plus encore le lien entre le chapeau et l'échec aux examens aurait pu être évoqué.
 25. 1962: 17. Voir A. Goosse (2007: 87).
 26. On notera que l'apport de la sociolinguistique, peut-être mieux documenté pour la dialectologie wallonne, l'est encore fort peu pour le français de Belgique. Voir Albert Doppagne (1978: 50) pour une critique sur la dispersion du français régional. Voir aussi Michel Francard, Geneviève Geron et Régine Wilmet (2002: 13, 14).
 27. Notons que la recherche du DFB ne s'inscrit pas dans une perspective diachronique ce qui peut expliquer les lacunes sur ce plan (Voir Francard, Geron et Wilmet 2002a: 15).
 28. Il aurait été pertinent ici de préciser la zone géographique entendue par la marque topolectale «Afrique», sachant que la zone qu'elle couvre est plutôt vaste, témoignant peut-être d'une vision fédéraliste de l'Afrique. Voir Claude Frey (2003: 173). Ici la marque topolectale diffère en quelque sorte d'une marque synthétique alors que pour le français régional d'Europe, le caractère régional est beaucoup plus limité géographiquement.
 29. La question de la dispersion des emplois a été discutée par De Surmont (à paraître) et Cleante (2000: 9-10).
 30. Nous n'avons pas retenu Albert Doppagne (1978) puisque sa nomenclature est plus réduite. La date figurant à côté du PLI est la date d'entrée dans le dictionnaire lorsqu'elle nous est connue.
 31. Voir Michel Francard, Geneviève Geron et Régine Wilmet (2002b: 74).
 32. On consultera ici avec profit l'ouvrage de Joep Kruijssen (1995) sur les emprunts romans dans les dialectaux limbourgeois de la Hesbaye qui montre qu'avant 1962 on a un maintien des éléments français empruntés avant la loi de 1962.
 33. Voir André Goosse 2007: 97.

Bibliographie

- Baetens Beardsmore, Hugo.** 1971. *Le français régional de Bruxelles*. Bruxelles: Presses Universitaires de Bruxelles.
- Brun-Trigaud, Guylaine, Yves Le Berre et Jean Le Dù.** 2005. *Lectures de l'Atlas linguistique de la France de Gillérion et Edmont, Du temps dans l'espace*. Paris: Éditions du Comité des travaux historiques et scientifiques.
- Charaudeau, P.** 1978. *Les conditions linguistiques d'une analyse du discours*. Thèse soutenue à l'Université Charles-de-Gaulle Lille III.
- Chauveau, Jean-Paul.** 2003. Histoire des langues romanes et géographie linguistique, Romanische Sprachgeschichte und Sprachgeographie. Ernst, Gerhard, Martin-Dietrich Glessgen, Christina Schmitt et Wolfgang Schweickard (Éds.). 2003. *Romanische Sprachgeschichte. Ein Internatio-*

nales Handbuch zur Geschichte der romanischen Sprachen/Histoire linguistique de la Romania. Manuel international d'histoire linguistique de la Romania. Tome 1: 72-90. Berlin/New York: Walter de Gruyter.

- De Surmont, Jean-Nicolas.** 2004. Caractéristiques et variations du marquage diatopique. Williams, Geoffrey et Sandra Vessier (Éds.). 2004. *Proceedings of the Eleventh Euralex International Congress, Euralex 2004, Lorient July 6–10, 2004: 779-786.* Lorient: Université de Bretagne Sud.
- De Surmont, Jean-Nicolas.** 2005. Quelques remarques sur les belgicisms métalinguistiques. *Présence francophone* 65: [198]-211.
- De Surmont, Jean-Nicolas.** À paraître. Contourner la marque diatopique: vitalité des emplois et présence des topolectismes dans les dictionnaires de variétés de français. Actes du colloque de Chypre 2006, sur la marque lexicographique.
- Francard, Michel.** 1995–1996. La technique lexicographique du Dictionnaire liégeois. *Les dialectes de Wallonie*, t/M 23–24: 221-239.
- Francard, Michel, Geneviève Geron et Régine Wilmet.** 2002a. Diffusion et vitalité des particularités lexicales du français en Belgique: une enquête sociolinguistique. Wakely, Richard (Éd.). 2002. *Les Belges: enregistreurs de tous les usages. Actes d'un colloque qui s'est tenu à l'Université d'Édimbourg les 7 et 8 avril 2000 et publiés en bénéficiant de l'Ambassade de France en Grande Bretagne: 11-32.* Édimbourg: Université d'Édimbourg.
- Francard, Michel, Geneviève Geron et Régine Wilmet.** 2002b. La banque de données VALIBEL: des ressources textuelles orales pour l'étude du français et à Bruxelles. Pusch, Claus D. et Wolfgang Raible (Éd.). 2002. *Romanistische Korpuslinguistik. Korpora und gesprochene Sprache/Romance Corpus Linguistics. Corpora and Spoken Language: 71-80.* Tübingen: Gunter Narr.
- Francard, Michel, Geneviève Geron et Régine Wilmet.** 2003. Les 'belgicisms' sont-ils 'belges'? La nomenclature du *Dictionnaire du français en Belgique*. Nobel, Pierre (sous la dir. de). 2003. *Variations linguistiques Koinè, dialectes, français régionaux: 137-150.* Besançon: Presses Universitaires de Franche-Comté.
- Frey, Claude.** 2003. Identités lexicales et variétés de français en France et hors de France: tendance centripètes et centrifuges des formes et des cultures. Nobel, Pierre (sous la dir. de). 2003. *Variations linguistiques Koinè, dialectes, français régionaux: 165-190.* Besançon: Presses Universitaires de Franche-Comté.
- Goosse, André.** 2007. Belgicisms exportés. *Cahiers de lexicologie* 91(2): 85-102.
- Klinkenberg, Jean-Marie.** 2008. À propos du traitement lexicographique des belgicisms. Bavoux, Claudine (Éd.). 2008. *Le français des dictionnaires: L'autre versant de la lexicographie française: 78-88.* Bruxelles: De Boeck.
- Kruijssen, Joep.** 1995. *Geografische patronen in taalcontact, Romaans leengoed in de Limburgse dialecten van Haspengouw.* Amsterdam: P. J. Meertens-Instituut.
- Piron, Maurice.** 1973. Les belgicisms lexicaux: essai d'un inventaire. *Mélanges Paul Imbs: 295-304.* Strasbourg.
- Pohl, Jacques.** 1962. *Témoignages sur la syntaxe du verbe dans quelques parlers français de Belgique.* Bruxelles: Palais des Académies.
- Pohl, Jacques.** 1979. *Les variétés régionales du français, études belges (1945–1977).* Bruxelles: Éditions de l'Université de Bruxelles.
- Poirier, Claude.** 1995. Les variantes topolectales du français. Francard, Michel et Daniel Latin (Éds.). 1995. *Le régionalisme lexical: 13-56.* Louvain-la-Neuve: De Boeck, AUPELF-UREF.

- Rézeau, Pierre.** 2003. De la cancoillotte à la tartiflette (et à la boîte chaude) avec le *Dictionnaire des régionalismes de France*. Nobel, Pierre (sous la dir. de). 2003. *Variations linguistiques Koinè, dialectes, français régionaux*: 128-136. Besançon: Presses Universitaires de Franche-Comté.
- Rézeau, Pierre.** 2007. Introduction. Rézeau, Pierre (études rassemblées par). 2007. *Richesses du français; Géographie linguistique. Vol. 1*. Bruxelles: De Boeck.
- Séguy, Jean.** 1973. Les Atlas linguistiques de la France par régions. *Langue française* 18: 65-90.
- Spickenbom, Marion.** 1996. *Belgizismen in französischen Wörterbüchern und Enzyklopädien seit Anfang dieses Jahrhunderts*. Münstersche Beiträge zur Romanischen Philologie 16. Münster: Nodus.
- Van Campenhoudt, Marc.** 2000. De la lexicographie spécialisée à la terminographie: vers un métadictionnaire? Thoiron, Philippe et Henri Béjoint. 2000. *Le sens en terminologie*: 127-152. Traavaux du CRTT. Lyon: Presses Universitaires de Lyon.
- Warnant, Léon.** 1973. Dialectes du français et français régionaux. *Langue française* 18: 100-125.
- Wolf, Heinz Jürgen.** 1991. *Französische Sprachgeschichte*. Heidelberg/Wiesbaden: Quelle & Meyer.
- Yaguello, Marina.** 1988. *Catalogue des idées reçues sur la langue*. Paris: Éditions du Seuil.

Corpus métalinguistique

- Bal, Willy, Albert Doppagne, André Goosse, Joseph Hanse, Michèle Lenoble-Pinson, Jacques Pohl et Léon Warnant.** 1994. *Belgicisms, inventaire des particularités lexicales du français de Belgique*. Louvain-la-Neuve: Éditions Duculot.
- Cleante [pseud. de Louis Chalon].** 2000. *Tours et expressions de Belgique. Prononciation, grammaire, vocabulaire*. Bruxelles: Duculot.
- Davau, Maurice, Marcel Cohen et Maurice Lallemand.** 1973. *Dictionnaire du français vivant*. Londres: Harrap.
- Delcourt, Christian.** 1998–1999. *Dictionnaire du français de Belgique*. 2 vol. Bruxelles: Éditions Le Cri.
- Doppagne, Albert.** 1979. *Belgicisms de bon aloi*. Bruxelles: Fondation Charles Plisnier, Office du Bon langage.
- [EN COLL.]. *Dictionnaire universel*. Paris: Hachette.
- [EN COLL.]. *Petit Larousse illustré*. Paris: Larousse. [Différentes éditions consultées]
- Fuchs, Heinz.** 1988. *Untersuchungen zu Belgizismen, Zu Ursprung und Verbreitung lexikalischer Besonderheiten des belgischen Französisch*. Bonner Romanistische Arbeiten 29. Frankfurt/Bern/New York/Paris: Peter Lang.
- Hanse, Joseph.** 1971. *Chasse aux belgicisms*. Bruxelles: Fondation Charles Plisnier.
- Knecht, Pierre (sous la direction de) conçu et dirigé par André Thibault, avec la collaboration de Gisèle Boeri et Simone Quenet.** 1997. *Dictionnaire suisse romand. Particularités lexicales du français contemporain*. Carouge/Genève: Éditions Zoé.
- Lebouc, Georges.** 1998. *Le Belge dans tous ses états. Dictionnaire de belgicisms, grammaire et prononciation*. Paris: Éditions Bonneton.
- Lebouc, Georges.** 2006. *Dictionnaire de belgicisms*. Bruxelles: Éditions Racine.
- Massion, François.** 1987. *Dictionnaire de belgicisms*. Frankfurt am Main/Bern/New York/Paris: Peter Lang.
- Rézeau, Pierre.** 2001. *Dictionnaire des régionalismes de France. Géographie et histoire d'un patrimoine linguistique*. Bruxelles: De Boeck & Larcier/Éditions Duculot.

Trésor de la langue française. Dictionnaire de la langue du XIX^e et du XX^e siècle (1789–1960). 1971–1994.
16 vol. Paris: Éditions du Centre national de la Recherche Scientifique (vol. 1–10)/Gallimard
(vol. 11–16).

Informateurs

Jean Marie Le Clezio, 17 octobre 2007.

Jacques De Decker, 17 octobre 2007.

Geïntegreerde woordeboekgebruik van vakwoordeboeke vir aanleerders

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

Opsomming: Na 'n kort bespreking van aanleerderwoordeboeke en vakwoordeboeke word die fokus verskuif na vakwoordeboeke vir aanleerders. Die aanleerders wat hier ter sprake kom, kan aanleerders van 'n taal en/of aanleerders van 'n vakgebied wees. Albei hierdie tipes leerders word bespreek asook daardie vakwoordeboeke wat op aanleerders van sowel 'n taal as die vak gerig is. Met verwysing na die Suid-Afrikaanse leksikografiese omgewing en die opdrag van die nasionale woordeboekeenhede, word aangevoer dat vakwoordeboeke vir aanleerders 'n wesenlike rol kan speel in die vestiging van 'n woordeboekkultuur, maar ook die voorkoming van 'n verdere leksikografies verlore generasie. Die konsep van geïntegreerde woordeboekgebruik word ingevoer en bespreek, met spesifieke verwysing na vakwoordeboeke vir aanleerders. Die voordeel van multiveldvakwoordeboeke vir aanleerders word aangetoon en aspekte van verskillende strukture van sulke woordeboeke word bespreek. Daar word aangevoer dat vir sulke woordeboeke 'n geamalgameerde sentrale teks verkieslik is bo 'n geamalgameerde makrostruktuur.

Sleutelwoorde: AANLEERDERWOORDEBOEK, GEAMALGAMEERDE MAKROSTRUKTUUR, GEAMALGAMEERDE SENTRALE TEKS, GEÏNTEGREERDE BINNETEKS, GEÏNTEGREERDE WOORDEBOEKGEBRUIK, KOGNITIEWE FUNKSIE, KOMMUNIKATIEWE FUNKSIE, LAAISTRUKTUUR, LEKSIKOGRAFIES VERLORE GENERASIE, LEKSIKOGRAFIESE FUNKSIES, MULTIVELDVAKWOORDEBOEK, SISTEMATIESE INLEIDING, TOEGANGSTRUKTUUR, VAKWOORDEBOEK, VAKWOORDEBOEK VIR AANLEERDERS, WOORDEBOEKGEBRUIKSPAKKET, WOORDEBOEKKULTUUR, WOORDEBOEKTIPOLOGIE

Abstract: Integrated Dictionary Use of Specialised Dictionaries for Learners. Following a brief discussion of learner's dictionaries and specialised dictionaries the attention is focused on specialised dictionaries for learners. The relevant learners targeted by these dictionaries can either be learners of a language or learners of a subject field. Both these types of learners are discussed as well as specialised dictionaries aimed at learners of both a language and the subject field. With reference to the South African lexicographic environment and the assignment of the national lexicographic units it is argued that specialised dictionaries for learners can play a significant role in establishing a dictionary culture but also in preventing a further lexicographically lost generation. The concept of integrated dictionary use is introduced and discussed, with specific reference to specialised dictionaries for learners. The benefits of multifield technical dictionaries for learners are indicated and aspects of different structures of these dictionaries are discussed. It is argued that for such dictionaries an amalgamated central text is preferred to an amalgamated macrostructure.

Keywords: ACCESS STRUCTURE, AMALGAMATED CENTRAL LIST, AMALGAMATED MACROSTRUCTURE, COGNITIVE FUNCTION, COMMUNICATIVE FUNCTION, DICTIONARY CULTURE, DICTIONARY TYPOLOGY, DICTIONARY USAGE PACKAGE, DRAWER STRUCTURE, INTEGRATED DICTIONARY USE, INTEGRATED INNER TEXT, LEARNER'S DICTIONARY, LEXICOGRAPHIC FUNCTIONS, LEXICOGRAPHICALLY LOST GENERATION, MULTIFIELD SPECIALISED DICTIONARY, SPECIALISED DICTIONARY, SPECIALISED DICTIONARY FOR LEARNERS, SYSTEMATIC INTRODUCTION

1. Inleidend

Een van die opvallende kenmerke van die hedendaagse leksikografie is die uitgebreide tipologiese verskeidenheid. In hierdie verband is dit belangrik om vertrouwd te wees met 'n aantal gevestigde onderskeidings ten opsigte van die versameling woordeboeke. Een so 'n onderskeiding is dié tussen linguistiese en ensiklopediese woordeboeke. In linguistiese woordeboeke is die klem tradisioneel op die taalkundige kenmerke van woorde en in ensiklopediese woordeboeke eerder op die saak waarna die betrokke woord verwys. 'n Belangrike verdere tipologiese tweedeling is dié tussen woordeboeke wat met die algemene taal werk en daardie woordeboeke wat met die taal van spesifieke deelgebiede van die leksikon, byvoorbeeld vaktaal, werk. Dit lei tot die onderskeid tussen algemene en beperkte woordeboeke. Binne die kategorie beperkte woordeboeke bestaan 'n verskeidenheid subtipes waaraan daar nie nou aandag gegee word nie. Vakwoordeboeke, ook bekend as gespesialiseerde woordeboeke, is een so 'n sub tipe wat vir die doel van hierdie artikel van belang is.

Woordeboeksoorte verteenwoordig nie afgeslote kategorieë en 'n oorvleueling van kenmerke kom algemeen voor. Hierdie soort oorvleueling geld tussen woordeboeke wat tot dieselfde breë kategorie behoort, byvoorbeeld die kategorie linguistiese woordeboeke, maar ook tussen woordeboeke wat tot verskillende breë kategorieë behoort, byvoorbeeld linguistiese woordeboeke wat sekere ensiklopediese kenmerke mag vertoon.

In die beplanning van enige woordeboek is dit belangrik dat bestaande woordeboektipologie nie die leksikograaf mag dwing om slegs kenmerke wat vir die bepaalde woordeboeksoort goedgekeur is in die beplande woordeboek in te sluit nie. Die vryheid moet bestaan en benut word om 'n woordeboek so saam te stel dat dit ook tipologies dit weerspieël wat die geïdentifiseerde teikengebruiker nodig het. Hartmann (1989) het dit duidelik gestel dat die samestelling van enige woordeboek voorafgegaan moet word deur 'n indringende behoeftebepaling, en so 'n behoeftebepaling is slegs moontlik na 'n ondubbel-sinnige vasstelling van die teikengebruiker van die beplande woordeboek.

In navolging van Gouws (2007a) kan aangevoer word dat 'n formele tipologiese klassifikasie 'n veel beperkter rol moet speel in besluite oor die aard en inhoud van nuwe woordeboeke. Na aanleiding van die gebruikersprofiel moet funksies van 'n beplande woordeboek vasgestel word, en op grond van die vasgestelde funksies moet die inhoud en woordeboekstrukture bepaal word. Die

aard van die woordeboek moet nie bepaal word deur 'n tipologiese klassifikasie wat vooraf gemaak word nie. Dit moet eerder volg as 'n reaksie ter beskrywing van die voltooid woordeboek. So 'n vasstelling wat eers gemaak word nadat die woordeboek voltooi is, kan nie altyd binne die bestaande tipologiese kategorieë plaasvind nie, en dit mag nodig wees dat 'n nuwe en unieke tipologiese beskrywing vir 'n gegewe woordeboek aangewend moet word. Een van die maniere waarop daar aan 'n te streng tipologiese klassifikasie ontsnap kan word, is deur die doelbewuste vermenging van kenmerke van verskillende en uiteenlopende woordeboeksoorte ter wille van die daarstelling van 'n produk wat in 'n hoër mate geskik is vir die behoeftes van 'n gegewe gebruikersgroep. Die afgelope dekade het 'n daadwerklike vestiging gesien van hierdie beginsel van tipologiese hibridisering in die leksikografie. Hierdie benadering bied aan leksikograwe die moontlikheid om werklik gebruikergerigte woordeboeke te kan saamstel.

Bepaalde aspekte van onder meer tipologiese hibridisering word in hierdie artikel gebruik om voorstelle te maak vir 'n verbeterde aanbod en 'n nuwe soort gebruik van een tipe gespesialiseerde woordeboek, naamlik vakwoordeboeke wat vir aanleerders geskryf word.

2. Die vestiging van bepaalde tipologiese kategorieë

2.1 Aanleerderwoordeboeke

Aanleerderwoordeboeke kan gesien word as een van die duidelikste voorbeelde van die vestiging van 'n nuwe tipologiese kategorie. 'n Vroeë bydrae in hierdie verband was A.S. Hornby se *Oxford Advanced Learner's Dictionary* wat in 1948 verskyn het. Dit was egter veral sedert die sewentigerjare van die vorige eeu dat die ontwikkeling van hierdie kategorie op dreef gekom het, veral in die Britse woordeboekpraktyk. Die uiteindelige publikasie van die "Groot Vyf", dit is die eentalige Engelse aanleerderwoordeboeke van die uitgewers Oxford University Press, MacMillan, Longman, Cambridge University Press en Collins, was 'n aanduiding van die verwesenliking van 'n nuwe tipologiese kategorie. Naas die eentalige aanleerderwoordeboeke het daar algaande ook 'n verskeidenheid tweetalige aanleerderwoordeboeke verskyn. Daar was egter reeds vroeg 'n dispuut onder leksikograwe en taalkundiges oor watter tipe aanleerderwoordeboek, een- of tweetalig, die beste hulp aan 'n aanleerder kan bied (vergelyk in hierdie verband byvoorbeeld Atkins 1985). Van Britse Engelse aanleerderwoordeboeke het daar gereeld nuwe uitgawes verskyn sodat hierdie woordeboeke voortdurend 'n toonaangewende rol gespeel het, ondanks die toetrede van ander tale tot die aanleerderkategorie. Die vestiging en ontwikkeling van aanleerderwoordeboeke as tipologiese kategorie het ook ruimskoots aandag gekry in die teoretiese leksikografie. In talle navorsingspublikasies is daar aandag gegee aan 'n groot verskeidenheid aspekte van aanleerderleksikografie (vergelyk in hierdie verband onder meer Hausmann 1974, Cowie 1987,

Otto 1989, Dolezal en McCreary 1999, Herbst en Popp 1999, Wiegand 1998; 2002 en Tarp 2008). Interessant in hierdie ontwikkeling is die uitvoerige aandag wat daar gewy is aan die verskyning van twee Duitse aanleerderwoordeboeke (vergelyk Wiegand 1998; 2002).

Aanleerderwoordeboeke word saamgestel met inagneming van die behoeftes en naslaanvaardighede van 'n duidelik en ondubbelsinnig omskrewende teikengebruiker. Eentalige aanleerderwoordeboeke is gerig op aanleerders wat die spesifieke taal wil aanleer maar maak nie noodwendig voorsiening vir die eiesoortige probleme wat uit die moedertaal van die voornemende gebruiker spruit nie. Daarteenoor is leksikograwe van tweetalige aanleerderwoordeboeke bewus van die spesifieke moedertaal van die tipiese teikengebruiker en kan hulle vir bepaalde probleme van die teikengebruiker op 'n meer gerigte manier voorsiening maak.

As tipologiese kategorie het aanleerderwoordeboeke 'n duidelike werklike doel. Daar is wel verskillende benaderings ten opsigte van die funksies van aanleerderwoordeboeke. In sommige gevalle is die teksresepsiefunksie oorheersend terwyl die teksproduksiefunksie in ander gevalle meer aandag kry. In baie aanleerderwoordeboeke word die kognitiewe funksie nie genoegsaam vervul nie maar in ander aanleerderwoordeboeke wel, veral waar die voornemende gebruikers kultuurgrense moet oorskry in hulle kennismaking met die taal wat aangeleer en die omgewing waar die betrokke taal gepraat word (vergelyk in hierdie verband die sterk kognitiewe funksie van die *Kenkyusha Lighthouse Dictionary* (Takebayashi en Kojima 1990²), 'n tweetalige aanleerderwoordeboek met Engels en Japannees as taalpaar, maar wat uitsluitlik op die behoeftes van Japanse aanleerders van Engels afgestem is). Die werklike doel van 'n aanleerderwoordeboek is gerig op die vervulling van die somtotaal van die leksikografiese funksies wat vir die betrokke aanleerderwoordeboek geïdentifiseer is. Die aanleer van 'n vreemde taal met die moontlikheid om tekste te kan verstaan en tekste te kan skep, asook om iets van die betrokke kultuur te leer, staan sentraal in die opdrag aan aanleerderwoordeboeke.

2.2 Vakwoordeboeke

Binne die woordeboektipologie bestaan 'n lang tradisie van woordeboeke wat daarop gerig is om die leksikon van 'n bepaalde vaktaal weer te gee, die sogenaamde vak- of gespesialiseerde woordeboeke. Soos in die geval van die algemene leksikografie het die praktyk van die maak van vakwoordeboeke ook in 'n preteoretiese milieu ontwikkel. Die ontwikkeling van die teoretiese leksikografie het aanvanklik veral teoretiese modelle geskep vir die daarstelling van algemene woordeboeke. Algaande het die teoretiese gesprek ook die vakleksikografie betrek (vergelyk in hierdie verband onder meer Schaeder en Bergenholtz 1984 en Bergenholtz en Tarp 1995).

Ook in die kategorie vakwoordeboek kan verskillende subtypes onderskei word. Daar is eerstens die algemene onderskeid tussen een- en twee- of meer-

talige vakwoordeboeke. Bergenholtz en Tarp (1995) maak voorsiening vir belangrike verdere subtipologiese verdelings, onder meer die onderskeid tussen kultuurafhanklike en kultuuronafhanklike vakwoordeboeke, subveld-, enkelveld- en multiveldvakwoordeboeke, minimerende en maksimerende vakwoordeboeke en vakwoordeboeke vir die leek, die semideskundige en die deskundige. Hierdie verdelings is deurgaans gerig op 'n aanbod wat in 'n hoër mate die behoeftes van spesifieke gebruikers probeer bevredig.

Die aandag wat daar die afgelope twee dekades aan ontwikkelinge in die vakleksikografie gegee is en die teoretiese basis wat vanuit die metaleksikografie daarvoor ontwikkel is, is moontlik gemaak deur, maar is ook nogmaals 'n bewys van die status van die leksikografie as 'n onafhanklike dissipline wat nie meer 'n onderafdeling van die taalkunde is nie (vergelyk in hierdie verband onder meer Tarp 2000, Wiegand 1998a en Gouws 2005). Die aanbod en bewerking in vakwoordeboeke is naamlik nie in die eerste plek gerig op die weergawe van taalkundige data nie. Dit gaan primêr om vakkundige inhoud. Wat belangrik is, is dat dieselfde teoretiese model waarvolgens die funksies, inhoud en struktuur van algemene woordeboeke bepaal word, ook van toepassing is in die beplanning van vakwoordeboeke. Die bestek van die algemene leksikografieteorie strek veel verder as net linguistiese woordeboeke. Die bewerking wat deur hierdie teorie moontlik gemaak word, is nie in die eerste plek gerig op die aanbod van taalkundige data nie maar wel op die aanbod van data wat volgens die aard en behoeftes van die geïdentifiseerde gebruikersprofiel nodig is.

Hierdie ontwikkeling in die algemene leksikografieteorie het 'n bepaalde kentering in die aard van vakwoordeboeke meegebring. Dit is so dat talle vakwoordeboeke, veral vroeër, nie die leksikografiese behoeftes van hulle teikengebruikers bevredig het nie, onder meer omdat daar in die beplanning van die woordeboeke te min aandag gegee is aan die gebruikersprofiel of aan die teoretiese vereistes van produkte wat as woordeboeke geproduseer word. Talle ouer vakwoordeboeke is deur deskundiges op die betrokke vakgebied saamgestel sonder enige leksikografiese as medewerkers. Dit het dikwels gelei tot produkte wat vanuit 'n leksikografiese perspektief as onbevredigend beskou moet word, deurdat hulle, vanweë die beperkte bewerking, dikwels nie die grense van 'n blote woordelys oorskry het nie, meermale ook met die gee van slegs 'n enkele vertaalekwivalent. Deur die vakleksikografie as 'n formele deel van die algemene leksikografieteorie te betrek, word die beplanning van vakwoordeboeke in 'n veel hoër mate met inagneming van die betrokke teoretiese riglyne en die noodsaaklike betrokkenheid van leksikografiese kundiges gedoen. 'n Goeie voorbeeld uit die Suid-Afrikaanse leksikografiese omgewing is die huidige werk aan 'n nuwe uitgawe van die *Woordeboek van Afrikaanse Geneeskundeterme* (Lochner, in voorbereiding). In die beplanningsfase van die herbewerking het die Hoofredakteur, prof. Jan Lochner, die hulp van 'n teoretiese leksikograaf ingeroep om te verseker dat die uiteindelijke woordeboek 'n voldoende teoretiese basis het. In die samestelling van die woordeboek word die Hoofredakteur bygestaan deur 'n span deskundiges uit die mediese vakgebied. Die betekenisverklarings van die terme word deur die deskundiges

gedoen maar die manuskrip word dan aan 'n teoretiese leksikograaf voorgelê om te verseker dat die aanbod nie net vakkundig nie maar ook leksikografies in orde is. Samewerking tussen vakkundige en leksikograaf lei tot die daarstelling van 'n beter woordeboek.

3. Die leksikografiese aanbod van vakterme

Die ontwikkeling van die vakleksikografie op teoretiese vlak en die voortspruitende opbloeï in die leksikografiese praktyk het gelei tot 'n opname en bewerking van 'n groot verskeidenheid vakterme uit 'n groot verskeidenheid vakgebiede in 'n groot verskeidenheid vakwoordeboeke. Desondanks moet 'n vakwoordeboek steeds gesien word as 'n gesofistikeerde leksikografiese instrument wat blyke gee van 'n goed gevestigde en uitgebreide woordeboekversameling in die betrokke taalgemeenskap. Omdat vakwoordeboeke gerig is op die behoeftes van 'n relatiewe klein gebruikersgroep is hierdie woordeboeksoort van die latere toevoegings tot die leksikografiese versameling van 'n bepaalde taal. Wanneer daar gepraat word oor die leksikografiese bewerking van vakterme, moet daar 'n onderskeid gemaak word tussen die bewerking van sulke terme in vakwoordeboeke en die opname en bewerking daarvan in algemene woordeboeke. Die opname van vaktalitems is naamlik nie beperk tot vakwoordeboeke nie. Veral in taalgemeenskappe met beperkte toegang tot leksikografiese bronne bestaan daar 'n goed gemotiveerde neiging ten gunste van multifunksionele hibriediese woordeboeke, en een aspek van hierdie hibriediese aard is juis die insluiting van items uit verskillende vakgebiede in woordeboeke wat primêr op die algemene leksikon gerig is.

Teenoor die uiters beperkte bewerking in talle vakwoordeboeke is die bewerking van vakterme in algemene woordeboeke gedoen in ooreenstemming met die kriteria wat gegeld het vir die bewerking van die ander items uit die algemene leksikon in daardie woordeboek. Vir die leek is daar gevolglik in 'n algemene woordeboek 'n veel duideliker bewerking van die beperkte en diverse groep vakterme aangesien hierdie gespesialiseerde items op 'n niegespesialiseerde manier verklaar is. Dit geld veral in omgewings waar woordeboeke vir ontluikende tale saamgestel is en waar daardie tale nie oor 'n verskeidenheid woordeboeksoorte beskik nie.

In Afrikaans het vroeëre dele van die WAT (vergelyk byvoorbeeld deel IV en V) talle vakterme opgeneem en van 'n redelik uitvoerige bewerking voorsien. Dit was veral in 'n tyd toe vakwoordeboeke nog nie vryelik beskikbaar was nie. Hierdie soort werkswyse is in elk geval nie vreemd in omvattende verklarende woordeboeke, die kategorie waartoe die WAT behoort, nie. Die omvang van die bewerking word dikwels bepaal deur die beskikbaarheid al dan nie van ander leksikografiese bronne waar inligting oor die betrokke vakterm verkry kan word.

Vakterme word egter ook in algemene woordeboeke van 'n beperkter omvang, byvoorbeeld hand- en standaardwoordeboeke, opgeneem. Hier word die

opname egter bepaal deur die optrede in tipiese gespreksituasies waarby die teikengebruikers van die woordeboek betrokke is. Dit gaan veral om daardie vakterme wat voorkom in gesprekke tussen deskundige of semideskundige en leek. Die gewone taalgebruiker, ook in 'n taalgemeenskap met 'n gevestigde woordeboekkultuur, beskik nie noodwendig oor 'n vakwoordeboek met byvoorbeeld die mediese of die juridiese leksikon as teikengebied nie. In hulle daaglikse kommunikasie word gewone taalgebruikers, leke ten opsigte van byvoorbeeld die mediese of die juridiese terrein, van tyd tot tyd blootgestel aan vakterme uit hierdie gebiede, byvoorbeeld tydens 'n besoek aan 'n medikus of 'n prokureur. Die opname van 'n beperkte keuse vakterme in 'n algemene woordeboek is 'n standaardpraktyk. Die bewerking van sulke terme is gerig op die behoeftes en naslaanvaardighede van die teikengebruiker asook die funksies van die betrokke woordeboek, en nie op die behoeftes van die deskundige op die bepaalde vakgebied nie. Selfs in aanleerderwoordeboeke word vakterme van hierdie aard opgeneem, soos byvoorbeeld blyk uit die verklaring van die term *tumour* in COBUILD: "A *tumour* is a mass of diseased or abnormal cells that has grown in a person's or animal's body." So 'n betekenisparafraze is verstaanbaar vir die leek en waarskynlik feitelik aanvaarbaar vir die deskundige. Al is die formulering eenvoudig, moet die wetenskaplike of gespesialiseerde waarde van 'n term steeds op 'n gepaste en ondubbelsinnige manier oorgedra word.

Die werklike probleem wanneer gebruikers hulle tot 'n algemene woordeboek wend om 'n leksikografiese bewerking van vakterme te vind, is dat daar geen versekering is dat die spesifieke term in die betrokke woordeboek gevind sal word nie. Die keuse van vakterme vir opname in algemene woordeboeke is te beperk. Gebruikers moes dus daaraan gewoond raak dat vakterme in vakwoordeboeke nageslaan moet word. Dit het 'n bykomende las op die betrokke leksikografe geplaas want raadpleging van woordeboeke deur gebruikers wat eerder algemene woordeboeke as vakwoordeboeke sou verkies, beteken dat sekere vakwoordeboeke nie net deskundiges en semideskundiges nie maar ook leke as teikengebruikers het. Die opstellers van vakwoordeboeke moet 'n baie duidelike beeld van hulle gebruikersprofiel hê: Is die teikengebruikersgroep deskundiges, semideskundiges en/of leke?

Alhoewel dit in hierdie artikel aanvaar word dat vakterme ook in algemene woordeboeke, insluitende algemene aanleerderwoordeboeke, opgeneem en bewerk word, is die fokus voortaan op die opname en bewerking van sulke leksikale items in vakwoordeboeke. Dit gaan egter nie oor alle vakwoordeboeke nie maar oor vakwoordeboeke wat op 'n bepaalde gebruikersgroep gerig is, soos in die hieropvolgende afdelings toegelig word.

4. Vakwoordeboeke vir aanleeders

Een van die tipologiese hibridiserings wat spruit uit ontwikkelinge in sowel die aanleerder- as die vakleksikografie is die skep van vakwoordeboeke vir aan-

leerders (vergelyk Fuertes-Olivera en Arribas-Baño 2008 vir 'n bespreking van bepaalde aspekte van hierdie woordeboek tipe). Ook hier word 'n verdere onderskeid gemaak, te wete tussen eentalige en tweetalige vakwoordeboeke vir aanleerders. Dit spreek vanself dat vakwoordeboeke ook op spesifieke gebruikersgroepe gerig moet wees en die woordeboekkonseptualiseringsplan moet met duidelike inagneming van die gebruikersprofiel saamgestel word.

In die tipologiese klassifikasies eentalige vakwoordeboek vir aanleerders en tweetalige vakwoordeboek vir aanleerders skep die deelterm eentalige vakwoordeboek en tweetalige vakwoordeboek nie begripsprobleme nie aangesien die gewone interpretasie van die terme *eentalige*, *tweetalige* en *vakwoordeboek* hier geld. Dit is weliswaar nie duidelik of dit om 'n sub-, enkel- of multiveldwoordeboek gaan, of dit 'n kultuurgebonde of niekultuurgebonde woordeboek is en of die woordeboek 'n maksimerende of 'n minimerende benadering het nie. Die leksikografiese funksies van 'n woordeboek is ook nie duidelik uit die deelterm nie maar dit is kwessies wat alles in die betrokke woordeboekkonseptualiseringsplan uiteengesit word en wat nie uit die tipologiese klassifikasie hoef te blyk nie. Die deelterm *aanleerder* in die bogenoemde tipologiese klassifikasies is egter 'n bron van groter onduidelikheid. Hierdie deelterm kan in die bepaalde klassifikasie geïnterpreteer word as sou dit na die aanleerder van 'n taal of na die aanleerder van 'n bepaalde vakgebied verwys. 'n Vinnige oorsig oor bestaande vakwoordeboeke vir aanleerders wys dat albei hierdie lesings van die deelterm *aanleerder* wel voorkom. In alle gevalle is die betrokke woordeboek 'n vakwoordeboek wat aan die eise van dié tipologiese klassifikasie moet voldoen. As vakwoordeboek vir aanleerders het die spesifieke waarde van *aanleerder* 'n regstreekse invloed op wat 'n voornemende gebruiker van so 'n woordeboek te wagte kan wees. Waar 'n bepaalde term 'n polisemiese leksikale item verteenwoordig, is dit noodsaaklik dat daar in die gebruik van daardie term ondubbelsinnig leiding gegee moet word oor watter lesing in die betrokke gebruiksomgewing geaktiveer word. Dit geld ook woordeboeke wat as vakwoordeboeke vir aanleerders of gespesialiseerde aanleerderwoordeboeke bekend staan.

Die *Oxford Dictionary of Computing for Learners of English* (Pyne en Tuck 1996) is 'n voorbeeld van 'n vakwoordeboek vir aanleerders waarin die fokus op die vakwoordeboekgebruiker is wat Engels as taal aanleer. Dit word duidelik in die voorwoord van hierdie woordeboek gestel. Hier gaan dit gevolglik om die eersgenoemde betekenisonderskeiding van die begrip *aanleerder*, naamlik die aanleerder van die betrokke taal, en waar so 'n woordeboek saamgestel word, vereis dit dat die gebruikersprofiel ook voorsiening vir die taalvermoë van die teikengebruiker moet maak. 'n Vakwoordeboek vir aanleerders waar die tweede betekenisonderskeiding van die begrip *aanleerder*, dit is die fokus op die aanleerder van die vakgebied, sentraal staan, is die *SASOL Science and Technology Resource* (Hartmann-Petersen 2001). In hierdie woordeboek wat leerders in die betrokke vakgebied inlei, is nie net die inhoud en aanbieding nie maar ook die metataal op die kennis-, kommunikasie- en opvoedkundige vlak

van die teikengebruikers gerig. Die metataal en die aanbieding in hierdie woordeboek gee duidelik blyke van die opvoedkundige opdrag van die woordeboek.

Die beginsel van tipologiese hibridisering in die leksikografie laat ruimte vir talle moontlikhede, onder andere ook dat 'n vakwoordeboek vir aanleerders op die aanleerder van die vak sowel as die aanleerder van die taal gerig kan wees. Dit kan lei tot 'n veeldoelige produk wat in bepaalde gebruiksomgewings van groot waarde kan wees. Hierdie hibridisering vra egter 'n kundige aanpassing en integrasie van aanleerder- en vakwoordeboek. Daar moet ook deeglik besin word oor die behoefte aan sulke woordeboeke alvorens daar met die beplanning en samestelling begin word.

5. 'n Kennismaking met enkele aspekte van bestaande vakwoordeboeke vir aanleerders

Om woordeboeke, ook vakwoordeboeke vir aanleerders, tot effektiewe gebruiksinstrumente te verhef, is dit belangrik dat daar duidelikheid oor die leksikografiese funksies en die werklike doel van so 'n woordeboek moet wees. In hierdie verband kan daar gekyk word na die *Ungarisch-Deutsches Deutsch-Ungarisches Fachwörterbuch zur Rentenversicherung* (Ildokó 2005). In die voorwoord van hierdie woordeboek word dit genoem dat die woordeboek saamgestel is met inagneming van bevindinge uit die metaleksikografie, pedagogiese leksikografie, tweetalige vakleksikografie asook navorsing in woordeboekgebruik. Dit word ook gestel dat hierdie woordeboek vir vertalers en tolke wat met Hongaars en Duits as taalpaar werk, bruikbaar wil wees in die spesifieke vakgebied wat hier ter sprake is. Dit word verder beskou as 'n aanleerderwoordeboek vir inrigtings vir vertaling en tolking en waar Duits as gespesialiseerde vreemde taal onderrig word. Volgens die voorwoord is die woordeboek geskik as vakwoordeboek vir sowel teksproduksie as teksresepsie. Hier gaan dit duidelik om gebruikers wat nie net behoefte aan 'n vakwoordeboek het nie maar ook heelwat inligting oor die betrokke taalpaar benodig. Albei lesings van die begrip *aanleerder* geld vir hierdie woordeboek alhoewel die taalgerigte benadering nie net vir beginneraanleerders van waarde is nie maar veral ook vir gevorderde aanleerders.

Vergelykbaar hiermee is die reeds genoemde *Oxford Dictionary of Computing for Learners of English*. Synde 'n vakwoordeboek gaan dit in hierdie woordeboek uiteraard ook oor die bepaalde vakgebied maar veral oor aanleerders van Engels wat leiding nodig het met betrekking tot die vakgebied van rekenaars. Volgens die voorwoord is die woordeboek "especially for learners of English" bedoel. Die makrostrukturele keuse van die woordeboek bied 'n verteenwoordigende versameling items uit die betrokke vakgebied maar die bewerking van hierdie items fokus veral op die taalbehoefte van die teikengebruikers. Dit

blyk onder meer uit die volgende verduideliking in die voorwoord oor die bewerking van die lemmas:

Clear explanations of the grammar and meaning of words, along with authentic examples showing how words are really used make this dictionary a useful tool for helping to read, write, speak and understand the English used in computing today.

En in die gebruikersleiding:

If more detailed information on grammar or pronunciation is needed or if cross-references, collocates, plurals, synonyms, parts of the verb or spelling and American variants are required they have all been convincingly placed together in the separate language column.

Die verwysing daarna dat hierdie woordeboek vir aanleerders van Engels bedoel is wat tot die middel- en gevorderde vlak behoort, wys nogmaals op die manier waarop daar hier met die begrip *aanleerder* gewerk word. Vakgebiedkennis is egter nie op die agtergrond nie. Die woordeboek is 'n waardevolle naslaanwerk "for people who use computers and people who study computing" maar die vakgerigtheid vorm saam met die taalgerigtheid die fokus van die woordeboek. In die voorwoord word dit genoem dat die woordeboek steun op die gekombineerde kundigheid van kenners op die gebied van rekenaars en dié gemoeid met die onderrig van Engels as vreemde taal. Die aanleerders van die taal is hier nie die moedertaalsprekers nie maar wel die vreemdetaalsprekers.

Teenoor hierdie vakwoordeboeke vir aanleerders waar sowel vak as taal 'n rol speel, staan die Duitse uitgewer Duden se vakwoordeboeke vir aanleerders in die reeks *Schüler Duden*. Hier gaan dit in 'n veel hoër mate om die vakgebied as om die taal. In die voorwoord van een van die woordeboeke in hierdie reeks, die *Schüler Duden: Informatik* (Claus en Schwill 1997) met die subtitel *Ein Sachlexikon für die Schule* ('n vakwoordeboek vir die skool), word dit duidelik gestel dat die woordeboek gerig is op leerders wat die veld van die informatika betree. 'n Vergelykbare benadering geld in nog 'n lid van hierdie reeks, die *Schüler Duden: Die Musik* (Kwiatkowski 1989). Geen verwysing word egter gemaak na enige poging om aanleerders met Duits te help nie.

Binne die Suid-Afrikaanse woordeboekomgewing kan vakwoordeboeke vir die aanleerder 'n baie belangrike rol speel in sowel die gerigte daarstelling van vakkundige data vir aanleerders en hulp met die aanleer van taal as in die vestiging van 'n woordeboekkultuur wat nie in isolasie gesien moet word nie maar deel van 'n omvattender kennis- en naslaankultuur uitmaak. Wat die aanleer van taal betref, hoef die begrip *aanleerder* nie uitsluitlik na gebruikers te verwys wat die betrokke taal as vreemde taal aanleer nie. Ook moedertaalsprekers wat in die aanleerfase is, kan binne die bestek van die term val. Leksikografe sal goed doen deur aan die beplanning en samestelling van sulke woordeboeke oorweging te gee.

6. **Woordeboeke en die Suid-Afrikaanse nasionale leksikografiese eenhede**

Met die vestiging van die Suid-Afrikaanse nasionale leksikografiese eenhede (NLE's) was dit nodig om 'n bestekopname te maak van watter woordeboeke in watter tale beskikbaar was. In 'n veel vroeëre opname (vergelyk Alberts en Mtintsilana 1988) is reeds bevind dat daar 'n uitgebreide behoefte aan aanleerderwoordeboeke bestaan. Die begin van die NLE-era is daardeur gekenmerk dat sommige eenhede voortgegaan het met die werk aan bestaande woordeboekprojekte in die betrokke taal terwyl ander eenhede op soek was na 'n gepaste woordeboekprojek. Dit was al gou duidelik dat die wens van PanSAT, naamlik om 'n meerdelige omvattende verklarende woordeboek vir elk van die elf ampstale van Suid-Afrika daar te stel, nie spoedig verwesenlik sou word nie. In hierdie verband sê Gouws (2001: 77):

The compilation of a comprehensive monolingual dictionary presupposes a fully standardised language, a typological infrastructure in the given language as well as advanced lexicographic expertise and a sound metalexicographic basis. This is not the type of dictionary to be compiled as a first project. Although it should be the ultimate goal for every language to compile a comprehensive dictionary, it also has to be regarded as the lexicographic crown jewel, which is acquired once a network of smaller dictionaries has been completed.

Aan die opstel van kleiner woordeboeke en die vestiging van 'n tipologiese infrastruktuur sou die ontluikende standaardtale nog uitvoerig aandag moes gee in die keuse en uitvoering van woordeboekprojekte.

Dit is egter nie net die talle leemtes in die tipologiese infrastruktuur wat in ag geneem moet word in die leksikografiese beplanning van die NLE's en kommersiële woordeboekuitgewers nie. Die gebrek aan 'n woordeboekkultuur het 'n stremmende uitwerking op die potensiele gebruik van nuwe en bestaande woordeboeke. Die vestiging van 'n woordeboekkultuur mag ook nie geïsoleer word van die breër onderrig- en opvoedingstaak nie. Sedert die vestiging van die NLE's het enkele jare verloop. Sommige eenhede het die projekte voltooi waaraan daar vroeër reeds gewerk is, ander het kleinere nuwe projekte afgehandel terwyl die Buro van die WAT steeds suksesvol voortgaan met die omvattende projek waaraan daar sedert 1926 gewerk word. Vir al die eenhede was dit aanpassingsjare maar die werk wat tot dusver gedoen is, het die geleentheid geskep vir die leksikografiese basis wat nodig is vir verdere projekte. Alhoewel die samestelling van omvattende woordeboeke steeds die einddoel is, moet leksikograwe van die verskillende woordeboekeenhede bedag wees op die leksikografiese behoeftes van hulle onderskeie taalgemeenskappe. Iets wat tot dusver in 'n te beperkte mate verwerklik is, is die vestiging van 'n woordeboekkultuur. Die skep van omvattende woordeboeke sal in hierdie verband nie die gewenste uitwerking hê nie. In verskeie taalgemeenskappe word Suid-Afrika met 'n leksikografies verlore generasie gekonfronteer. Vir hierdie gene-

rasie sal nuwe woordeboeke nie veel waarde hê nie. Wat wel binne die visie van die NLE's moet wees, is om te verseker dat die volgende generasies nie met dieselfde probleem te kampe het nie. Alhoewel die nasionale taalliggame van die onderskeie tale en onderskeie provinsies asook die kurrikulumbepanners van die verskillende onderwysdepartemente 'n wesenlike rol kan en moet speel in die vestiging van 'n woordeboekkultuur en die verwerwing van woordeboekvaardighede mag die rol van die NLE's in hierdie verband nie onderskat word nie. Juis hier kan vakwoordeboeke vir die aanleerder 'n bydrae maak wat verder strek as die vestiging van 'n woordeboekkultuur.

7. Geïntegreerde woordeboekgebruik

Binne 'n opvoedkundige omgewing moet daar 'n onderskeid gemaak word tussen geïsoleerde en geïntegreerde woordeboekgebruik. In die eersgenoemde geval gaan dit om die toevallige gebruik van 'n woordeboek of die gebruik van 'n woordeboek binne 'n bepaalde leeropdrag maar nie as 'n formele deel van die leeropdrag nie. Daarteenoor impliseer geïntegreerde woordeboekgebruik dat dit 'n wesenlike deel van 'n bepaalde leeropdrag is. Geïntegreerde woordeboekgebruik voorvereis die beplanning en samestelling van 'n bepaalde woordeboek as deel van 'n groter pakket van naslaanbronne. In die leksikografie bestaan daar reeds die gevestigde gebruik van sogenaamde woordeboekwerkboeke. Dit gaan om werkboeke met oefeninge wat die gebruiker help om óf 'n spesifieke woordeboek beter te kan gebruik óf om woordeboeknaslaanvaardighede oor die algemeen te verhoog. Wat die eersgenoemde aspek betref, is die *Oxford Tweetalige Werkboek/Workbook* (Hiles 2007) 'n goeie voorbeeld. Hierdie werkboek vergesel die *Oxford Afrikaans-Engels English-Afrikaans Skoolwoordeboek School Dictionary* (Louw 2007). Oefeninge in die werkboeke help die gebruiker om die stelsel van die betrokke woordeboek te verstaan en om die gehalte van die onttrekking van inligting te verhoog. Die oefeninge is so saamgestel dat die teikengebruikers van die betrokke woordeboek nie probleme met die begrip en benutting daarvan het nie. Wat algemene woordeboekwerkboeke betref (vergelyk byvoorbeeld Kipfer 1984), word woordeboek oefeninge in die werkboek ingesluit maar die werkboek is nie aan een spesifieke woordeboek gekoppel nie (vergelyk ook die bespreking van woordeboekwerkboeke in Stark 1990).

Die kombinasie van woordeboek plus werkboek is gerig op 'n beter ontsluiting van die betrokke woordeboek maar dit is nog nie 'n voorbeeld van geïntegreerde woordeboekgebruik nie; daarvoor is die mede-optrede van 'n verdere bron noodsaaklik. 'n Tipiese voorbeeld van geïntegreerde woordeboekgebruik vind 'n mens waar 'n woordeboek gekoppel is aan 'n bepaalde nieleksikografiese bron en die gebruik van die woordeboek onder meer daarop gerig is om die inhoud van die betrokke bron beter te ontsluit. Die mede-optrede van 'n woordeboekwerkboek kan 'n versterkende bonuswaarde hê maar is nie 'n verpligte komponent van die pakket nie.

As 'n voorbeeld van 'n woordeboekgebruikspakket kan daar verwys word na die kombinasie handboek, woordeboek en woordeboekwerkboek wat 'n mens kry met betrekking tot die skoolhandboekreeks *Nuwe Afrikaans sonder grense* (Gouws en Latti 1996–2002), die woordeboek *Nuwe woordeboek sonder grense* (Gouws, Stark en Gouws 2004) en die gepaardgaande woordeboekwerkboeke *Nuwe woordeboek sonder grense Werkboek* (Gouws en Latti 2004–2005) waarvan vier vir die verskillende grade (4, 5, 6 en 7–9) voorsiening maak (vergeelyk in hierdie verband ook die bespreking in Steyn 2004 en Steyn en Gouws 2005). Die handboekreeks is bedoel vir leerders met Afrikaans as addisionele taal en die betrokke woordeboek is saamgestel vir dieselfde teikengebruikersgroep, d.w.s. aanleerders van Afrikaans, en die doel is onder meer dat die woordeboek saam met die handboek gebruik word, ook ter verbetering van die aanleer van die werk wat in die handboek aan die orde kom. Die korpus wat vir die woordeboek gebruik is, is onder meer saamgestel uit studiemateriaal waarmee leerders wat die betrokke handboekreeks gebruik daaglik in aanraking kom. Die werkboek slaan 'n brug tussen die woordeboek en die handboek met oefeninge waar oplossings vir probleme in die handboek in die woordeboek gevind moet word. In hulle omgang met Afrikaans as vak word die leerders bekendgestel aan die handboek asook die woordeboek en word hulle geleer om die woordeboek as praktiese gebruiksmiddel in te span in die oplossing van probleme uit die betrokke vakgebied wat in die handboek bespreek word. Hier lei geïntegreerde woordeboekgebruik tot groter insig in die betrokke vakgebied maar terselfdertyd tot die verwerwing van woordeboekgebruiksvaardighede en die vestiging van 'n woordeboekkultuur. Gegee die vroeë verwysing na die verskillende waardes van die begrip *aanleerder* gaan dit hier om 'n woordeboek wat op die aanleerder van die taal gerig is maar die gebruik van die woordeboek is geïntegreer in die aanleer van die betrokke vakgebied. *Nuwe woordeboek sonder grense* is 'n algemene aanleerderwoordeboek vir leerders op skoolvlak maar maak ook voorsiening vir vakterminologie wat in die betrokke taalkurrikulum voorkom. Dit bied egter 'n wyer leksikale versameling as bloot die terminologie van 'n bepaalde vakgebied.

Hierdie soort benutting van 'n woordeboek kan 'n voorbeeld wees vir maniere waarop woordeboeke vir die Afrikatale saamgestel kan word sodat die gebruiker wat sowel taalaanleerder (ook moedertaalaanleerder) as vakaanleerder is daardeur bevoordeel word terwyl dit ook tot die vestiging van 'n woordeboekkultuur lei. Sodoende kan die gevaar van verdere leksikografies verlore generasies bestry word.

8. Vakwoordeboeke vir aanleerders in Suid-Afrika

Die meeste vakwoordeboeke wat in die Suid-Afrikaanse tale geskryf word, het deskundiges en semideskundiges as teikengebruikers. Tot dusver is die vakleksikografie nog nie deur die NLE's bedryf nie. In die verlede was daar ook 'n meningsverskil daaroor of vakleksikografie binne die opdrag van die NLE's val

al dan nie. Hierdie aspek word nie in hierdie artikel bespreek nie. Daar word van die standpunt uitgegaan dat dit die opdrag van die NLE's is om vir die leksikografiese versorging van hulle onderskeie tale verantwoordelikheid te neem, wat die beplanning van gespesialiseerde woordeboeke insluit.

Wat vakwoordeboeke betref, is dit reeds genoem dat 'n subtipologiese onderskeid gemaak word op grond van die kennisvlak van die geïdentifiseerde teikengebruiker van die woordeboek, naamlik vakwoordeboeke vir deskundiges, semideskundiges en leke. In aanleerderwoordeboeke word daar eweneens 'n subtipologiese onderskeid gemaak op grond van die kennisvlak van die gebruikers. Dit lei tot woordeboeke gerig op beginner-, medium- en gevorderdevlakgebruikers. In vakwoordeboeke vir aanleerders kan hierdie twee stelle gebruikerstipes tot 'n verskeidenheid moontlikhede lei, nl. *beginnervlakdeskundige*, *beginnervlaksemideskundige*, *beginnervlakleek*; *mediumvlakdeskundige*, *mediumvlaksemideskundige*, *mediumvlakleek*; *gevorderdevlakdeskundige*, *gevorderdevlaksemideskundige*, *gevorderdevlakleek*. Al hierdie gebruikersmoontlikhede word nie in hierdie artikel bespreek nie. Aandag word gegee aan die rol van vakwoordeboeke vir aanleerders op skoolvlak en veral op laerskoolvlak. Hierdie teikengebruikers is wat die vakgebied betref, leke, en wat die taal betref, op beginner- tot mediumvlak. Die tipiese vakwoordeboek vir aanleerders wat hier te sake is, is dié wat gerig is op leke in die betrokke vakgebied en in die beginner- tot mediumfase van taalaanleer.

9. 'n Geïntegreerde benadering en die skep van 'n woordeboekkultuur

'n Opvallende verskynsel in veral die aanleerder- maar ook die vakwoordeboekversameling is dat hierdie woordeboeksoorte in 'n beperkte aantal tale saamgestel word. Albei tipes word as gesofistikeerde leksikografiese onderneemings beskou. Aanleerderwoordeboeke kom veral, alhoewel nie uitsluitlik nie, voor in daardie tale wat groot internasionale verkeerstale is. Vakwoordeboeke word in veel meer tale aangetref maar die samestelling van vakwoordeboeke voorveronderstel die bestaan van 'n gevestigde vakwoordeskat in die betrokke taal. Die beginsel van 'n vakwoordeboek vir aanleerders waar dit om 'n aanleerder van die vakgebied gaan, impliseer nie noodwendig dat dit 'n taal behels wat 'n terminologies leidende rol speel nie maar wel dat dit 'n taal is waarin die vakterminologie bestaan. Vakwoordeboeke word egter ook gesien en gebruik as instrumente om die bepaalde vakterminologie te vestig. Soos reeds genoem, behoort die beplanning en samestelling van 'n vakwoordeboek 'n gesamentlike projek van vakkundige en leksikograaf te wees. Dit is nie 'n geleentheid vir die leksikograaf om terme te skep nie maar wel vir die vakkundige om sowel gevestigde as nuwe terme vir leksikografiese blootstelling en bewerking voor te lê. Die nege Afrikatale wat ampstale in Suid-Afrika is, is nog nie almal ten volle gestandaardiseer nie, soos blyk uit die talle voortgaande gesprekke oor byvoorbeeld ortografiese pre- en proskripsie. Ook die vakterminologie is nie grondig gevestig in al dié tale nie. Op hierdie gebied kan daar met reg van 'n

ontluikende vakterminologie gepraat word. Om met vakwoordeboeke vir die deskundige en semideskundige te begin, kan in 'n hoë mate frustrerend wees aangesien talle terme in hierdie stadium nog ontbreek. Ook in hierdie opsig kan die inwerkingstelling van 'n benadering van geïntegreerde woordeboekgebruik veel daartoe bydra om die ontwikkeling van die ontluikende vakterminologie te bevorder. Vakwoordeboeke vir laerskoolleerders kan 'n ideale beginpunt wees. 'n Pakket bestaande uit handboek, woordeboek en dalk selfs ook woordeboekwerkboek kan vir die vroeë fases van leerarea-onderdig saamgestel word. Vir die leerders wat aan hierdie geïntegreerde woordeboekgebruik blootgestel word, kan daar voorsiening gemaak word vir 'n vergelykbare pakket in die daaropvolgende fase. Deur die loop van hulle skoolloopbaan kan geïntegreerde woordeboekgebruik 'n geïntegreerde deel van hulle opleiding wees. Algaande kan hulle in hulle woordeboekgebruik vorder na algemene woordeboeke.

Die NLE's, in samewerking met die onderskeie onderwysdepartemente en ten beste ook met bestaande uitgewersmaatskappye, kan hulle in die bemagtiging van hulle onderskeie taalgemeenskappe daarop toelê om vakwoordeboeke vir aanleerders te skryf. Deur 'n benadering van geïntegreerde woordeboekgebruik na te streef met die woordeboeke wat help om die handboek te ontsluit en, verkieslik, met 'n gepaardgaande woordeboekwerkboek wat leerders kan help met die verwerwing van woordeboekgebruiksvaardighede, kan die NLE's 'n wesentliche rol speel in 'n veelfasettige kennisvestiging. Vir woordeboeke wat ter wille van so 'n benadering van geïntegreerde woordeboekgebruik saamgestel word, is dit belangrik dat die leksikografe indringend aandag gee aan die tersaaklike leksikografiese funksies. Vakwoordeboeke het primêr 'n kognitiewe funksie terwyl aanleerderwoordeboeke ook 'n sterk kommunikatiewe funksie het. 'n Vakwoordeboek vir aanleerders wat veral op die gebruiker as aanleerder van die vak gerig is, het 'n sterker kognitiewe funksie terwyl 'n sterker kommunikatiewe funksie geld in gevalle waar die woordeboek ook vir die aanleer van taal voorsiening maak. In 'n benadering van geïntegreerde woordeboekgebruik sal kognitiewe en kommunikatiewe funksies gelykwaardige aandag moet kry, en vir die kommunikatiewe funksies sal sowel teksproduksie as teksresepsie ter sake wees. Dit gaan om 'n polifunksionele woordeboek. Wat die kommunikatiewe funksie betref, speel teksresepsie al dan teksproduksie normaalweg 'n bepalende rol daarin of die woordeboek in die makrostrukturele keuse 'n maksimerende al dan 'n minimerende benadering volg. 'n Maksimerende benadering, dit wil sê waar die makrostrukturele keuse so uitvoerig moontlik is, is veral ter sake in die geval van teksresepsie. In die geval van teksproduksie is 'n minimerende benadering, dit wil sê 'n beperkter makrostrukturele keuse maar 'n uitvoeriger mikrostrukturele bewerking, eerder die keuse. Waar albei kommunikatiewe funksies aanwesig is, moet 'n balans getref word tussen maksimerend en minimerend. In die geval van die voorgestelde vakwoordeboeke sal die behoeftes van die teikengebruikers bepalend wees. Aangesien dit ten opsigte van die vakgebied om leke gaan, is slegs

'n beperkte makrostrukturele keuse nodig. Die gebruikers sal wel baat by 'n uitvoerige mikrostrukturele bewerking van hierdie items, ook ter wille van die taalaanleerfunksie. Die metataal van hierdie bewerking moet versoenbaar wees met die taalvlak van die teikengebruikers.

In die beplanning van die voorgestelde vakwoordeboeke vir aanleerders is dit belangrik om kennis te neem van 'n bepaalde subtipologiese onderskeid wat Bergenholtz en Tarp (1995) maak, te wete dié tussen sub-, enkel- en multiveldwoordeboeke. Vir laerskoolleerders is subveldwoordeboeke nie so ter sake nie. Die ideaal is om enkelveldwoordeboeke vir elke afsonderlike leerarea te hê. Op pad na hierdie ideaal sou egter voorlopig multiveldwoordeboeke gebruik kon word waarin die leksikon van verwante leerareas gedek word.

10. Aspekte van die struktuur van die voorgestelde vakwoordeboek vir aanleerders

10.1 'n Verskeidenheid aangepaste strukture

Die algemene leksikografieteorie maak voorsiening vir 'n verskeidenheid woordeboekstrukture. In die woordeboekkonseptualiseringsplan van elke afsonderlike woordeboek is dit die taak van die leksikograaf om te besluit watter strukture vir die betrokke woordeboek ter sake is en hoe die algemene strukture 'n woordeboekspesifieke aanpassing en toepassing moet kry. Dit geld ook vakwoordeboeke en aanleerderwoordeboeke, en waar dit om vakwoordeboeke vir aanleerders gaan, is 'n verdere aanpassing nodig. Dit is belangrik dat die relevante strukture so ontwikkel moet word dat die leksikografiese funksies van die betrokke woordeboek daardeur vervul kan word. Gevolglik mag geen struktuur sonder meer op 'n woordeboek afgedwing word nie. As dit gaan oor die aard van onder meer die dataverspreiding-, raam-, makro-, mikro-, toegangs- en adresseringstruktuur, is dit belangrik dat hierdie strukture telkens aangepas word met inagneming van die funksies van die woordeboek maar ook die gebruikersprofiel, en meer spesifiek die behoeftes en naslaanvaardighede van die potensiele teikengebruiker. Die beginsel van geïntegreerde woordeboekgebruik is nuut en die implikasies daarvan vir woordeboekstrukture is nog glad nie bespreek nie. Slegs enkele voorstelle met betrekking tot enkele struktuurtypes word vervolgens aan die hand gedoen. 'n Veel omvattender bespreking is nodig.

10.2 'n Geamalgameerde sentrale teks versus 'n enkele geamalgameerde makrostruktuur

Geïntegreerde woordeboekgebruik voorveronderstel die interaktiewe gebruik van minstens een woordeboek en minstens een ander bron. In die geval van vakwoordeboeke vir aanleerders is hierdie ander bron by voorkeur 'n hand-

boek of ander vorm van studiemateriaal vir die betrokke teikengebruikersgroep. Daar kan egter ook meer as een woordeboek en meer as een handboek of niewoordeboekbron wees. In die geval van 'n multiveldvakwoordeboek vir aanleerders in die laerskool sal daar waarskynlik nie een ooreenstemmende multiveldhandboek wees nie. Die beplanning van 'n multiveldvakwoordeboek vir geïntegreerde gebruik moet liefers so geskied dat handboeke vir soveel as moontlik van die vakke wat in die woordeboek aangebied word, as geïntegreerde woordeboeksterne buitetekste betrek kan word. Die sentrale komponent van so 'n woordeboek kan óf 'n geamalgameerde sentrale teks, óf 'n enkele geamalgameerde makrostruktuur wees (vergelyk Gouws en Leroyer 2009). In die geval van 'n geamalgameerde sentrale teks, dit is 'n teks met 'n laaistruktuur, sal elke vakgebied van die multiveldwoordeboek 'n afsonderlike en selfstandige teks in die sentrale komponent van die woordeboek wees, soos 'n reeks laaie binne die woordeboekstruktuur. As "groot teks" (vergelyk Kammerer en Wiegand 1998) bestaan 'n multiveldwoordeboek met 'n laaistruktuur uit opeenvolgende tekste wat elk 'n selfstandige enkelveldwoordeboek is waarin die terminologie van 'n gegewe vakgebied aangebied en bewerk word. Hierdie afsonderlike tekste is gelykwaardige komponente van die groot teks en almal tree as selfstandige binnetekste op, dit is die afsonderlike tekste in die sentrale komponent van 'n woordeboek met 'n raamstruktuur, hetsy 'n volle hetsy 'n halwe raam. Binnetekste waartussen daar deur middel van kruisverwysings verbande gelê word, staan as geïntegreerde binnetekste bekend. In 'n multiveldvakwoordeboek word geïntegreerde binnetekste veral aangetref waar dieselfde terme in meer as een vakgebied voorkom en daar dienooreenkomstige kruisverwysings verskyn. Tussen hierdie tekste bestaan daar dus 'n geïntegreerde verband.

Waar 'n multiveldwoordeboek 'n geamalgameerde makrostruktuur het, bestaan die sentrale komponente slegs uit een alfabeties geordende teks wat al die lemmata uit al die verskillende vakgebiede bevat. Vir elke lemma word daar per etiket aangedui tot watter veld of velde die betrokke term behoort. Vanuit sowel 'n sistematiese as 'n didaktiese oogpunt is 'n geamalgameerde sentrale teks, die laaistruktuur, aan te beveel bo 'n geamalgameerde makrostruktuur vir 'n multiveldvakwoordeboek vir aanleerders. Vir gebruikers bied die laaistruktuur vinnige toegang tot die lemmas wat vir 'n bepaalde vakgebied verstrek word.

10.3 Die dataverspreidingstruktuur

In die beplanning van enige woordeboek moet daar uitvoerig aandag gegee word aan die dataverspreidingstruktuur (vergelyk in hierdie verband Bergenholtz en Tarp 1995 asook Bergenholtz, Tarp en Wiegand 1999). As 'n draer van tekstipes (vergelyk Kammerer en Wiegand 1998) bevat 'n woordeboek 'n verskeidenheid tekste wat in die sentrale komponent asook die voor- en agtertekste-afdelings, die sogenaamde buitetekste van die woordeboek aangebied word. Die beplanning van die dataverspreidingstruktuur is nie net op die sen-

trale teks gerig nie maar ook op die voor- en agtertekste-afdelings wat aan leksikograwe 'n geleentheid verskaf om daardie data in 'n woordeboek aan te bied waarvoor die verstekartikel nie voorsiening maak nie. In sowel vak- as aanleerderwoordeboeke kan die buitetekste 'n wesenlike rol speel in die data-aanbod. In aanleerderwoordeboeke kan hulle onder meer gebruikers vertrouwd maak met die grammatiese stelsel van die taal wat aangeleer word, asook met bepaalde kultuurgebruike aangesien dit nie net oor die aanleer van 'n taal gaan nie maar ook oor die aanleer van 'n nuwe kultuur — iets wat binne die kognitiewe funksie van die woordeboek val. In vakwoordeboeke is daar talle aspekte van die vakgebied wat in buitetekste deur middel van byvoorbeeld tabelle weergegee kan word om die gebruiker te help met 'n beter verwerwing van kennis van die betrokke vakgebied (vergelyk in hierdie verband die buitetekste in die *SASOL Science and Technology Resource*). Woordeboeke van die toekoms behoort op 'n vernuwende manier buitetekste te gebruik. Dit is belangrik dat die buitetekste ook geïntegreerd moet wees in die funksies van die woordeboek om sodoende 'n transtekstuele benadering tot woordeboekfunksies te verseker (vergelyk Gouws 2007).

In die vakleksikografie word daar baie klem geplaas op die optrede van 'n sogenaamde sistematiese inleiding as een van die buitetekste van 'n vakwoordeboek. Dit is 'n teks waarin 'n inleiding gegee word tot die vakgebied waarvan die terminologie in die betrokke woordeboek behandel word (vergelyk in hierdie verband byvoorbeeld die *Encyclopedic Dictionary of Gene Technology* (Kaufmann en Bergholtz 1998)). As 'n geïntegreerde buitetekste (vergelyk Bergholtz, Tarp en Wiegand 1999 en Bergholtz en Nielsen 2006) word daar vanuit die sentrale teks van die woordeboek kruisverwysings na die sistematiese inleiding gemaak sodat gebruikers bykomende inligting kan onttrek. Dieselfde benadering word gevolg in die uitvoerige vakwoordeboek vir deskundiges en semideskundiges *Fachwörterbuch zur Lexikographie und Wörterbuchforschung/Dictionary of Lexicography and Dictionary Research* (Wiegand et al. 2009). Ook wat geïntegreerde buitetekste betref, sal baie van die data wat in sulke tekste in 'n vakwoordeboek vir aanleerders wat as geïsoleerde produk verskyn, aangebied word (vergelyk byvoorbeeld die *SASOL Science and Technology Resource* en sy buitetekste), nie deel van die vakwoordeboek se data-opdrag wees nie maar eerder binne die bestek van die handboek se opdrag val.

Waar 'n stelsel van geïntegreerde woordeboekgebruik gevolg word, is daar werklike gevolge vir die struktuur van die woordeboek. In die dataverspreidingstruktuur moet die leksikograaf die data wat reeds in die handboek verskyn, in ag neem. Dit bring mee dat so 'n woordeboek byvoorbeeld nie 'n sistematiese inleiding hoef te hê nie; die handboek dien hierdie doel. Belangrik is dat daar dan vanuit die woordeboek regstreekse kruisverwysings na die handboek gemaak moet word om die bewerking van bepaalde terme in die woordeboek in 'n wyer besprekingskonteks te plaas. Terme wat in die handboek gebruik en in die woordeboek verklaar word, kan in die handboek gemerk word, byvoorbeeld deur dit in vet druk aan te bied. Sodoende sal gebruikers weet dat die pakket voorsiening maak vir 'n leksikografiese bewer-

king van die betrokke terme. Die wisselwerking tussen hand- en woordeboek deur middel van kruisverwysings verhef die handboek tot 'n woordeboekeksterne geïntegreerde buitetekst.

Ten opsigte van die kognitiewe funksie kan die woordeboek wel aanvullend tot die handboek optree deur per geïntegreerde buitetekste voorsiening te maak vir relevante data wat nie in die handboek aangebied word nie. Al is die dataverspreidingsstruktuur van 'n beperkter omvang, is daar steeds ruimte vir 'n raamstruktuur. In hierdie verband is dit nogmaals nodig dat vakkundige en leksikograaf saam aan die inhoud van die woordeboek beplan.

10.4 Die toegangstruktuur

Die sukses van enige woordeboekraadplegingshandeling word in die eerste plek daardeur bepaal dat gebruikers toegang verkry tot die data waarna hulle soek. Daaruit kan inligting onttrek word ter bevrediging van die tersaaklike gebruikersbehoefte. Tans word in die metaleksikografie klem gelê op die omvattende toegangsproses wat veel meer as net die toegangstruktuur insluit (vergelyk in hierdie verband Bergenholtz en Gouws 2008). Die toegangsproses maak onder meer daarvoor voorsiening dat verskillende gebruikers verskillende toegangsroetes kan volg wat uiteenlopende beginpunte het. In die geval van geïntegreerde woordeboekgebruik kan die beginpunt van die toegangsproses van 'n gegewe woordeboekraadplegingshandeling byvoorbeeld in die teks van 'n woordeboekeksterne bron lê, soos die handboek wat as deel van die pakket aangebied word. In die beplanning van vakwoordeboeke vir aanleerders wat op geïntegreerde woordeboekgebruik gerig is, moet daar in die woordeboekkonseptualiseringsplan voorsiening gemaak word vir 'n verskeidenheid toegangsmoontlikhede; dit geld sowel die eksterne as die interne toegangsroetes.

Politoeganklikheid word daardeur in die hand gewerk dat data nie net via die makrostrukturelement wat die gidselement van 'n bepaalde artikel is, bereik kan word nie. Kruisverwysings vanuit 'n geïntegreerde buitetekst van woordeboekeksterne teks skep bykomende toegangsgeleenthede. Waar so 'n vakwoordeboek 'n raamstruktuur vertoon, kan die agtertekste-afdeling een of meer buitetekste insluit wat alfabetiese registers is van die terme wat vir die verskillende vakgebiede in die multiveldwoordeboek aangebied word. 'n Termregister kan vir elke afsonderlike vakgebied verskaf word. Sodoende word die gehalte van die woordeboek as politoeganklike bron nog verder verhoog.

Belangrik in die gebruik van 'n multiveldvakwoordeboek vir aanleerders, veral waar geïntegreerde woordeboekgebruik nagestreef word, is dat die leksikograaf vertrouwd moet wees met verskillende moontlikhede in die preraadplegingsfase van die toegangsproses (vergelyk Bergenholtz en Gouws 2008).

11. Die medium van vakwoordeboeke vir aanleerders

'n Besluit wat 'n leksikograaf reeds in die fase van die woordeboekkonseptualiseringsplan moet neem, geld die medium waarin die woordeboek aangebied word, naamlik in gedrukte en/of elektroniese formaat. In die laasgenoemde geval is 'n verdere besluit of dit as 'n CD ROM-weergawe of as internetweergawe beskikbaar moet wees. Wat die keuse van die medium betref, speel die gebruikersprofiel asook die gebruik- en gebruikersituasie 'n belangrike rol. Hierdie aspek word nie in hierdie artikel uitvoerig bespreek nie.

Wat die situasie van aanleerders in die Suid-Afrikaanse skoleomgewing betref, mag leksikograwe nie te gou uitsluitlik op gedrukte woordeboeke terugval nie. Talle skole, ook vir leerders uit die voorheen benadeelde groepe, het redelike toegang tot rekenaargeriewe en die internet. Vakwoordeboeke vir aanleerders moet op die huidige situasie maar ook op die onmiddellike toekoms gerig wees. Dit sal goed wees indien hierdie bronne in sowel elektroniese as gedrukte formaat beskikbaar kan wees. Dit geld nie net die woordeboeke nie maar die volle pakket wat die geïntegreerde gebruik moontlik maak.

'n Groot voordeel van 'n vakwoordeboek wat in die elektroniese medium beskikbaar is, veral as internetwoordeboek, is die politoeganklikheid daarvan, die aktiewe benutting van kruisverwysings asook die maklike toegang tot inskrywings wat nie lemmastatus het nie. Dit is aspekte waaraan leksikograwe indringend in hulle beplanning aandag moet gee.

12. Ten slotte

Vakwoordeboeke vir aanleerders is 'n woordeboek tipe wat in die Suid-Afrikaanse woordeboekgebruiksomgewing met vrug benut kan word. Die bereiking van die werklike doel van die betrokke woordeboeke is belangrik maar sulke woordeboeke kan ook 'n beter woordeboekkultuur help vestig. Die geïntegreerde gebruik van hierdie woordeboeksoort kan daadwerklik bydra tot die verbetering van kennisoordrag en woordeboekgebruiksvaardighede. Dit kan 'n aandeel hê in die voorkoming van verdere leksikografies verlore generasies.

Bibliografie

- Alberts, M. en P. Mtintsilana.** 1988. *Leksikografiese ontleding van Afrikaans en Afrikatale-leksikons*. RGN-verslag LEXI-2. Pretoria: Raad vir Geesteswetenskaplike Navorsing.
- Atkins, B.T.** 1985. Monolingual and Bilingual Learners' Dictionaries: A Comparison. Ison, R. (Red.). 1985. *Dictionaries, Lexicography and Language Learning*: 15-24. Oxford: Pergamon Press.
- Bergenholtz, H. en R.H. Gouws.** 2008. The Access Process for Fixed Expressions. *Lexicographica* 23: 236-260.
- Bergenholtz, H. en S. Nielsen.** 2006 Subject-field Components as Integrated Parts of LSP Dictionaries. *Terminology* 2(12): 281-303.

- Bergenholtz, H. en S. Tarp (Reds.).** 1995. *Manual of Specialised Lexicography: The Preparation of Specialised Dictionaries*. Amsterdam: John Benjamins.
- Bergenholtz, H., S. Tarp en H.E. Wiegand.** 1999. Datendistributionsstrukturen, Makro- und Mikrostrukturen in neueren Fachwörterbüchern. Hoffmann, L. et al. (Reds.). 1999. *Fachsprachen. Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft/Languages for Special Purposes. An International Handbook of Special-Language and Terminology Research: 1762-1832*. Berlyn: De Gruyter.
- Claus, V. en A. Schwill.** 1997³. *Schüler Duden: Informatik*. Mannheim: Dudenverlag.
- Cowie, A.P. (Red.).** 1987. *The Dictionary and the Language Learner*. Tübingen: Max Niemeyer.
- Dolezal, F.T. en D.R. McCreary.** 1999. *Pedagogical Lexicography Today. A Critical Bibliography on Learners' Dictionaries with Special Emphasis on Language Learners and Dictionary Users*. Tübingen: Max Niemeyer.
- Fuertes-Olivera, P.A. en A. Arribas-Baño.** 2008. *Pedagogical Specialised Lexicography. The Representation of Meaning in English and Spanish Business Dictionaries*. Amsterdam/Philadelphia: John Benjamins.
- Gouws, R.H.** 2001. Lexicographic Training: Approaches and Topics/Formation lexicographique: approches et thèmes. Emejelu, J.D. (Red.). 2001. *Éléments de lexicographie gabonaise. Tome 1: 58-94*. New York: Jimacs-Hillman.
- Gouws, R.H.** 2005. Meilensteine auf dem historischen Weg der Metalexikographie. *Lexicographica* 21: 158-178.
- Gouws, Rufus H.** 2007. A Transtextual Approach to Lexicographic Functions. *Lexikos* 17: 77-87.
- Gouws, R.H.** 2007a. Op pad na 'n nuwe woordeboektipologie. *Southern African Linguistics and Applied Language Studies* 25(3): 319-331.
- Gouws, R.H. en P. Leroyer.** 2009. Verhoogde leksikografiese toeganklikheid in die oorgang van 'n toeristewoordeboek na 'n toeristegids as naslaanbron. *Tydskrif vir Geesteswetenskappe* 49(1): 145-149.
- Gouws, R.H., M. Stark en L. Gouws.** 2004. *Nuwe woordeboek sonder grense*. Kaapstad: Maskew Miller Longman.
- Gouws, S. en M. Latti.** 1996–2002. *Nuwe Afrikaans sonder grense*. Kaapstad: Maskew Miller Longman.
- Gouws, S. en M. Latti.** 2004. *Nuwe woordeboek sonder grense Werkboek Gr. 4*. Kaapstad: Maskew Miller Longman.
- Gouws, S. en M. Latti.** 2004. *Nuwe woordeboek sonder grense Werkboek Gr. 6*. Kaapstad: Maskew Miller Longman.
- Gouws, S. en M. Latti.** 2004. *Nuwe woordeboek sonder grense Werkboek Gr. 5*. Kaapstad: Maskew Miller Longman.
- Gouws, S. en M. Latti.** 2005. *Nuwe woordeboek sonder grense Werkboek Senior Fase Gr. 7–9*. Kaapstad: Maskew Miller Longman.
- Hartmann, R.R.K.** 1989. Sociology of the Dictionary User: Hypotheses and Empirical Studies. Hausmann, F.J. et al. (Reds.). 1989–1991. *Wörterbücher. Ein internationales Handbuch zur Lexikographie/Dictionaries. An International Encyclopedia of Lexicography/Dictionnaires. Encyclopédie internationale de lexicographie: 102-111*. Berlyn: De Gruyter.
- Hartmann-Petersen, P. et al.** 2001. *SASOL Science and Technology Resource*. Claremont: New Africa Education Publishing.
- Hausmann, F.J.** 1974. Was ist und was soll ein Lernwörterbuch? *Zeitschrift für französische Sprache und Literatur* 84: 97-129.

- Herbst, T. en K. Popp (Reds.).** 1999. *The Perfect Learners' Dictionary (?)*. Tübingen: Max Niemeyer.
- Hiles, L.** 2007. *Oxford Tweetalige Werkboek / Workbook*. Kaapstad: Oxford University Press.
- Hornby, A.S.** 1948. *Oxford Advanced Learner's Dictionary*. Oxford: Oxford University Press.
- Ildokó, F.** 2005. *Ungarisch–Deutsches Deutsch–Ungarisches Fachwörterbuch zur Rentenversicherung*. Szeged: Grimm Kiadó.
- Kammerer, M. en H.E. Wiegand.** 1998. Über die textuelle Rahmenstruktur von Printwörterbüchern. Präzisionen und weiterführende Überlegungen. *Lexicographica* 14: 224-238.
- Kaufmann, U. en H. Bergenholtz (Reds.).** 1998. *Encyclopedic Dictionary of Gene Technology / Diccionario Enciclopédico Ingeniería Genética*. Vol. I: English (With Spanish Equivalents); Vol. II: Español (Con equivalentes en inglés). Toronto: Lugus Libros Latin América.
- Kipfer, B.A.** 1984. *Workbook on Lexicography: A Course for Dictionary Users with a Glossary of English Lexicographical Terms*. Exeter: University of Exeter Press.
- Kwiatkowski, G. (Red.).** 1989². *Schüler Duden: Die Musik*. Mannheim: Dudenverlag.
- Lochner, J.** In voorbereiding. *Woordeboek van Afrikaanse Geneeskundeterme*. Kaapstad: Pharos.
- Louw, P.A. (Red.).** 2007. *Oxford Afrikaans–Engels / English–Afrikaans Skoolwoordeboek / School Dictionary*. Kaapstad: Oxford University Press.
- Otto, A.N.** 1989. *Kriteria vir 'n Afrikaanse aanleerderwoordeboek*. Ongepubliseerde D.Litt.-verhandeling. Stellenbosch: Universiteit van Stellenbosch.
- Pyne, S. en A. Tuck.** 1996. *Oxford Dictionary of Computing for Learners of English*. Oxford: Oxford University Press.
- Schaeder, B. en H. Bergenholtz (Reds.).** 1994. *Fachlexikographie. Fachwissen und seine Repräsentation in Wörterbüchern*. Tübingen: Gunter Narr.
- Sinclair, J. (Red.).** 1987. *Collins COBUILD English Language Dictionary*. Londen/Glasgow: Collins.
- Stark, M.P.** 1990. *Dictionary Workbooks. A Critical Evaluation of Dictionary Workbooks for the Foreign Language Learner*. Exeter: University of Exeter Press.
- Steyn, M.** 2004. The Access Structure in Learner's Dictionaries. *Lexikos* 14: 275-298.
- Steyn, M. en L. Gouws.** 2005. *Nuwe Woordeboek sonder Grense: A Typological and Communicative Bridge*. *Lexikos* 15: 164-178.
- Takebayashi, S. en Y. Kojima (Reds.).** 1990². *Kenkyusha's Lighthouse English–Japanese Dictionary*. Tokio: Kenkyusha.
- Tarp, S.** 2000. Theoretical Challenges to Practical Specialised Lexicography. *Lexikos* 10: 189-208.
- Tarp, S.** 2008. *Lexicography in the Borderland between Knowledge and Non-Knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Tübingen: Max Niemeyer.
- Wiegand, H.E. (Red.).** 1998. *Perspektiven der pädagogischen Lexikographie des Deutschen. Untersuchungen anhand von Langenscheidts Großwörterbuch Deutsch als Fremdsprache*. Tübingen: Max Niemeyer.
- Wiegand, H.E.** 1998a. *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexicographie*. 1. Teilband. Berlin: De Gruyter.
- Wiegand, H.E. (Red.).** 2002. *Perspektiven der pädagogischen Lexikographie des Deutschen II. Untersuchungen anhand des de Gruyter Wörterbuch Deutsch als Fremdsprache*. Tübingen: Max Niemeyer.
- Wiegand, H.E. et al. (Reds.).** 2009. *Fachwörterbuch zur Lexikographie und Wörterbuchforschung / Dictionary of Lexicography and Dictionary Research*. Berlin: De Gruyter.
- Woordeboek van die Afrikaanse Taal*. Stellenbosch: Buro van die WAT.

Zum Stand und zu den Perspektiven der allgemeinsprachlichen Lexikographie mit Deutsch und Slowenisch

Vida Jesenšek, *Abteilung für Germanistik, Philosophische Fakultät, Universität Maribor, Maribor, Slowenien (vida.jesensek@siol.net)*

und

Herbert Ernst Wiegand, *Germanistisches Seminar, Universität Heidelberg, Heidelberg, Bundesrepublik Deutschland (herbert.ernst.wiegand@gs.uni-heidelberg.de)*

Zusammenfassung: Nach einer kurzen Übersicht über die allgemeinsprachlichen zweisprachigen Wörterbücher mit Deutsch und Slowenisch werden folgende Wörterbücher näher untersucht: Plet, DebN, DebS, PonsN und PonsS. Im ersten Teil der Wörterbuchanalyse werden der Benutzerbezug, die Wörterbuchbasis, die primär gebuchte Lexik sowie die Umtexte und eingelagerten Binnentexte berücksichtigt. Nur in Plet ist der Benutzerbezug relativ deutlich erkennbar. Bei den anderen Wörterbüchern bleibt er weitgehend undeutlich. In den vier neueren Wörterbüchern weisen alle betrachteten Komponenten erhebliche Mängel auf: z.B. fehlende mediostrukturelle Vernetzung von Umtexten und Wörterverzeichnis, fehlende Datenakzessivität bei den eingelagerten Binnentexten, ungeschickte Wahl der Wörterbuchbasis und unausgewogene Lemmaselektion. Im zweiten Teil der Wörterbuchanalyse werden anhand von Wörterbuchartikeln ausgewählte Aspekte des Wörterbuchgegenstands und der Wörterbuchform der genannten Wörterbücher untersucht; dabei werden die für das jeweilige Wörterbuch typischen Artikelmikrostrukturen dargestellt und Vorschläge gemacht, wie man die Artikelstrukturen verändern kann, damit die Gestaltung benutzerfreundlicher wird, und zwar auch dann, wenn das Datenangebot reichhaltiger ist. Dadurch werden einige Perspektiven für die Verbesserung der lexikographischen Versorgung im Sprachenpaar Deutsch und Slowenisch erkennbar.

Stichwörter: ANGABENADRESSIERUNG, BASALER WÖRTERBUCHARTIKEL, BENUTZERBEZUG, GLATTALPHABETISCHE HAUPTZUGRIFFSSTRUKTUR, HYBRIDE MIKROSTRUKTUR, KURZARTIKEL, MIKROARCHITEKTUR, MONOALPHABETISCHE MAKROSTRUKTUR, REINE MIKROSTRUKTUR, UMTEXT, ZWEISPRACHIGE LEXIKOGRAPHIE

Abstract: About the Position and Perspectives of General Lexicography with regard to German and Slovenian. A brief overview of general bilingual dictionaries with German and Slovenian is followed by a closer investigation of the following dictionaries: Plet, DebN, DebS, PonsN and PonsS. In the first part of the dictionary analysis the user reference, the dictionary basis, the primary recorded lexical items, as well as outer texts and phased-in inner texts are investigated. Only in Plet the user reference is relatively clearly recognisable. In the other

dictionaries it remains quite unclear. In all four more recent dictionaries all the components that have been investigated display serious shortcomings, e.g. lacking mediostructural linking of outer texts and central word list, the lack of data accessivity in the phased-in inner texts, inappropriate selection of the dictionary basis and the unbalanced lemma selection. In the second part of the dictionary analysis selected aspects of the dictionary subject matter and the dictionary form of the mentioned dictionaries are investigated by means of dictionary articles. In addition the typical article microstructures of each dictionary are given and proposals are made to change the article structures to produce a more user-friendly presentation, also when the data presentation is expanded. This reveals some perspectives for the improvement of the lexicographic treatment in the language pair German and Slovenian.

Keywords: BASIC DICTIONARY ARTICLE, BILINGUAL LEXICOGRAPHY, HYBRID MICROSTRUCTURE, ITEM ADDRESSING, MICROARCHITECTURE, MONOALPHABETIC MACROSTRUCTURE, OUTER TEXT, PURE MICROSTRUCTURE, SHORT ARTICLE, STRAIGHT ALPHABETICAL MAIN ACCESS STRUCTURE, USER REFERENCE

1. Vorbemerkung

Systemveränderungen und allgegenwärtige Forderungen nach mehr Zusammenarbeit und Integration im zusammenwachsenden Europa (EU) hatten in den letzten Jahren zur Folge, dass das Forschungs- und Arbeitsfeld der zweisprachigen Lexikographie an Interesse und Intensität gewonnen hat. Dazu trägt wesentlich auch die sprachpolitische Forderung nach Sprachenvielfalt und Mehrsprachigkeit in Europa bei. Das Bedürfnis nach zweisprachigen Lexika steigt mit dem wachsenden Interesse an dem Fremdsprachenlernen und mit immer intensiverer Übersetzungstätigkeit.

Durch die europäische politische Integration wurden einer Reihe von den in Europa gesprochenen Sprachen, und darunter auch dem Slowenischen, neue Positionen und Rollen zugeteilt. Das Slowenische ist immer öfter in der Position einer Fremd- oder Zweitsprache, die man erlernen will/soll und immer mehr die Ausgangs- oder Zielsprache der Übersetzung. Für beides sind qualitätsvolle und zuverlässige sprachliche Referenzwerke unentbehrlich. Da aber die gegenwärtige zweisprachige Lexikographie mit Slowenisch quantitativ eher bescheiden und qualitativ relativ mangelhaft ist, will der vorliegende Artikel zu einer besseren lexikographischen Praxis in der Erarbeitung der künftigen Sprachwörterbücher und zugleich zur Weiterentwicklung der metalexikographischen Forschung im slowenischen Sprachraum beitragen.

2. Die Wörterbücher mit Deutsch und Slowenisch: eine Übersicht

Ein Meilenstein in der Geschichte der slowenisch–deutschen zweisprachigen allgemeinen Lexikographie (die Anfänge reichen in das 16. Jh. zurück) war die Herausgabe des ersten slowenisch–deutschen Wörterbuchs größeren Umfangs

von Maks Pleteršnik Ende des 19. Jhs. (Plet 1894–1895). Seitdem sind relativ viele allgemeine Wörterbücher unterschiedlichen Umfangs und verschiedener Komplexität erschienen. Im Cobiss, Slowenischen bibliographischen Informationssystem (<http://www.cobiss.si/>) kann man dies leicht überprüfen. In zwei chronologisch geordneten und hinsichtlich der Ausgangssprache differenzierten Listen (a) und (b) wird im Folgenden eine Auswahl präsentiert. Berücksichtigt wurden Wörterbücher mittleren und größeren Umfangs, nicht beachtet wurden dagegen kleinere Hand- und Schulwörterbücher.

(a) Zweisprachige Wörterbücher mit Slowenisch als Ausgangssprache

- Pleteršnik, Maks (Hrsg.). 1894–1895. Transliterierte Ausg. 2006. *Slovensko–nemški slovar I–II* [*Slowenisch–deutsches Wörterbuch I–II*]. Ljubljana: ZRC SAZU.
- Kramarič, France (Hrsg.). 1905. *Slovensko–nemški slovar* [*Slowenisch–deutsches Wörterbuch*]. Wien/Leipzig: A. Hartleben.
- Hubad, France (Hrsg.) 1908. *Anton Janežičev slovensko–nemški slovar* [*Slowenisch–deutsches Wörterbuch von Anton Janežič*]. Celovec/Klagenfurt: Tiskarna Družbe sv. Mohorja.
- Bradač, Fran, Ivan Pregelj (Hrsg.). 1930. *Slovensko–nemški slovar* [*Slowenisch–deutsches Wörterbuch*]. Ljubljana: Jugoslovanska knjigarna.
- Tomšič, France (Hrsg.). 1958, 1961, 1966, 1973, 1977, 1983, 1988, 1991, 1993. *Slovensko–nemški slovar. Slowenisch–deutsches Wörterbuch*. Ljubljana: Državna založba Slovenije.
- Kotnik, Janko (Hrsg.). 1963, 1974, 1975, 1978, 1981, 1990, 1992, 1994. *Langenscheidts Universal-Wörterbuch Slowenisch: Slowenisch–Deutsch, Deutsch–Slowenisch*. Berlin usw.: Langenscheidt.
- Gradišnik, Janez (Hrsg.). 1966, 1971, 1981, 1986, 1991/1992, 1998. *Priročni slovensko–nemški slovar. Slowenisch–deutsches Handwörterbuch*. Maribor: Obzorja.
- Debenjak, Doris, Božidar Debenjak, Primož Debenjak (Hrsg.). 1995, 1997, 1999, 2003, 2008, 2009. *Veliki slovensko–nemški slovar* [*Großes slowenisch–deutsches Wörterbuch*]. Ljubljana: DZS. Print- und CD-ROM-Ausgaben.
- PONS *Splošni slovensko–nemški slovar. Kompaktwörterbuch Slowenisch–Deutsch*. 2008. Ljubljana: Rokus Klett.

(b) Zweisprachige Wörterbücher mit Deutsch als Ausgangssprache

- Kramarič, France (Hrsg.). 1907. *Deutsch–slowenisches Wörterbuch*. Wien, Leipzig: A. Hartleben.
- Bartel, Anton (Hrsg.). 1905, 1921. *Deutsch–slowenisches Hand-Wörterbuch*. Von Anton Janežič. Klagenfurt, Prevalje: St. Hermagoras Bruderschaft.
- Tominšek, Josip (Hrsg.). 1924. *Nemško–slovenski slovar* [*Deutsch–slowenisches Wörterbuch*]. Ljubljana: Ig. Kleinmayr & Fed. Bamberg.

- Bradač, Fran, Janko Šlebinger (Hrsg.). 1928. *Nemško-slovenski slovar* [Deutsch-slowenisches Wörterbuch]. Ljubljana: Jugoslovanska knjigarna.
- Tomšič, France (Hrsg.). 1938, 1944, 1954, 1974, 1959, 1980, 1985, 1986, 1989, 1990. *Nemško-slovenski slovar. Deutsch-slowenisches Wörterbuch*. Ljubljana: Državna založba Slovenije.
- Gradišnik, Janez (Hrsg.). 1996, 1997, 2000, 2004, 2006, 2007. *Nemško-slovenski, Slovensko-nemški slovar* [Deutsch-slowenisches, Slowenisch-deutsches Wörterbuch]. Maribor: Obzorja.
- Debenjak, Doris, Božidar Debenjak, Primož Debenjak (Hrsg.). 1992, 1993, 1994, 1996, 1998, 1999, 2001, 2003, 2005, 2007, 2008. *Veliki nemško-slovenski slovar. [Großes deutsch-slowenisches Wörterbuch]*. Ljubljana: DZS. Print- und CD-ROM-Ausgaben.
- Antič, Igor (Hrsg.) 1999, 2002. *Nemško-slovenski slovar* [Deutsch-slowenisches Wörterbuch]. Ljubljana: Mladinska knjiga.
- PONS *Splošni Nemško-slovenski slovar. Kompaktwörterbuch Deutsch-Slowenisch*. 2006. Ljubljana: Rokus Klett.

Mit Rücksicht auf Umfang, Aktualität, Verbreitung aufgrund zahlreicher Neuauflagen und somit auch hinsichtlich der Relevanz und Repräsentativität für die zweisprachige Lexikographie mit Deutsch und Slowenisch bieten sich für eine detaillierte Betrachtung folgende Wörterbücher an:

- Plet = *Slovensko-nemški slovar* [Slowenisch-Deutsches Wörterbuch] von Pletersnik.
- DebN = *Veliki nemško-slovenski slovar* [Großes deutsch-slowenisches Wörterbuch] von Debenjak.
- DebS = *Veliki slovensko-nemški slovar* [Großes slowenisch-deutsches Wörterbuch] von Debenjak.
- PonsN = *Splošni nemško-slovenski slovar* [Allgemeines deutsch-slowenisches Wörterbuch] von PONS.
- PonsS = *Splošni slovensko-nemški slovar* [Allgemeines slowenisch-deutsches Wörterbuch] von PONS.

Die Auswahl wird folgendermaßen begründet: Bei Plet handelt es sich um das erste umfangreiche Wörterbuch zu diesem Sprachenpaar, DebN und DebS sind die Wörterbücher mit dem umfangreichsten Lemmabestand, PonsN und PonsS gehören zu einer Wörterbuch-Serie der zweisprachigen Wörterbücher, die einen relativ großen Bekanntheitsgrad weltweit erreicht hat. Schließlich musste die Auswahl der zu betrachtenden Wörterbücher auch aus Platzgründen reduziert werden.

Im Verlauf der zwei weiteren Kapitel erfolgt eine zusammenfassende Darstellung der ausgewählten Wörterbücher. Sie werden in folgenden Hinsichten charakterisiert: Benutzerbezug, Wörterbuchbasis, primär gebuchte Lexik, Umtexte, Wörterbuchgegenstand und Wörterbuchform.

3. Wörterbuchanalyse I: Benutzerbezug, Wörterbuchbasis, primär gebuchte Lexik, Umtexte und eingelagerte Binnentexte

In diesem Kapitel wird dargestellt, wie die ausgewählten Wörterbücher sich auf ihre Adressaten beziehen, ihre Benutzerbedürfnisse berücksichtigen und welche Funktionen sie damit erfüllen möchten. Betrachtet werden die Wörterbuchbasis, der Lemmabestand hinsichtlich der Selektion und des Umfangs sowie entsprechende Aussagen in den Umtexten. Die Darstellungen in diesem Kapitel haben zum Ziel, Aussagen über die Benutzerangemessenheit in Bezug auf ihre funktionalen Aspekte zu formulieren und somit Grundlagen für die nachfolgende Analyse zum Wörterbuchgegenstand und zur Wörterbuchform (Kap. 4) zu schaffen.

3.1 Plet

Das slowenisch–deutsche Wörterbuch von Pleteršnik entstand aufgrund der gesellschaftlichen Notwendigkeiten und Wünschen, die im slowenischen Sprachraum um die Mitte des 19. Jhs. festgestellt und geäußert wurden und im Einklang mit den sprachorientierten nationalen Forderungen der europäischen revolutionären Zeit um das Jahr 1848 standen. Das Wörterbuchprogramm entstammt nämlich den Tätigkeitsplänen des 1848 gegründeten Slowenischen Vereins (Slovensko društvo), dessen primäre bildungskulturelle Aufgabe die Erarbeitung und Herausgabe eines repräsentativen Wörterbuchs des Slowenischen gewesen war. Ein solches Wörterbuch sollte in zwei Teilen, einem deutsch–slowenischen und einem slowenisch–deutschen Teil erscheinen. Während der deutsch–slowenische Teil in der Tat relativ schnell erarbeitet und herausgegeben wurde (1860), dauerte es trotz der landespolitischen finanziellen und personellen Unterstützung bis zum Jahr 1894 bzw. 1895, bis die beiden Bände des slowenisch–deutschen Teiles unter der Redaktion von Pleteršnik erschienen sind.

Plet ist ein zweibändiges allgemeines zweisprachiges Wörterbuch und enthält genau 102 522 Lemmata der slowenischen Ausgangssprache (Furlan 2006: IX). Über den Wörterbuchplan und die Wörterbuchkonzeption von Plet ist inzwischen an mehreren Stellen ausführlich geschrieben worden (u.a. Jakopin 1994 und 1997, Jesenšek 2004, Furlan 2008); aus Platzgründen werden an dieser Stelle nur die für die nachfolgenden Ausführungen wesentlichen Informationen darüber gegeben. Das sprachliche Ausgangsmaterial gründete auf einem systematisch durchdachten Wörterbuchkonzept und entstammte zahlreichen und relativ vielfältigen Quellen, die im Vorspann (Abkürzungsverzeichnis) akribisch verzeichnet sind. Zu den primären Textquellen zählten vorrangig theologisch-kirchliche und profane Literatur sowie publizistische und fachliche Texte, die vom Anfang des slowenischen Schriftentums im 16. Jh. bis zu der 2. Hälfte des 19. Jhs. veröffentlicht wurden. Anhand des Abkürzungsverzeichnisses sind ebenso sekundäre, tertiäre und weitere Quellen der Wörter-

buchbasis identifizierbar. Bei der Materialerhebung und -übernahme wurden die bedeutendsten Wörterbücher jener Zeit systematisch konsultiert (vor allem ältere (auch handschriftliche) zweisprachige Wörterbücher mit Slowenisch und Deutsch, aber auch einsprachige Wörterbücher des Russischen, Kroatisch-Serbischen, Deutschen (Grimm) und dialektale (bayrische, italienische) Wörterbücher. Die Letzteren wurden insofern konsultiert, als zwischensprachliche lexikalische Zusammenhänge gesucht wurden. Unter den tertiären Quellen wurden vorhandene slowenische Grammatiken und sprachwissenschaftliche Abhandlungen (vor allem zur Phonetik, Dialektologie und Terminologie) weitgehend berücksichtigt. Die Wörterbuchbasis wurde weiterhin durch mehr oder weniger umfangreiche Wortsammlungen der gebildeten Muttersprachler sowie der dialektsprechenden Informanten erweitert.

Gemäß der Wörterbuchkonzeption, wie sie in dem relativ umfangreichen Einleitungstext in das Wörterbuch von Pleteršnik selbst beschrieben wurde, bestand der Hauptzweck des Wörterbuchs darin, als repräsentatives Wörterbuch des Slowenischen gelten zu können. Die gebuchte Lexik sollte die gedankliche Welt der Slowenen, den Geist der Sprache und der Nation im möglichst vollen Umfang darstellen (Pleteršnik 2006: VII). Das Ziel der Lexikographen bestand also darin, eine Art Inventarisierung der slowenischen Lexik durchzuführen und somit ein historisches Wörterbuch des Slowenischen zu erstellen. Aus finanziellen und vor allem aus personellen Gründen erwies sich die Einbeziehung der historischen Dimension jedoch bereits in der Planungsphase als wenig realistisch, so dass man sich entschloss, mit dem Wörterbuch eher „praktische Zwecke“ zu verfolgen. Diese bestanden vor allem in der (möglichst ausführlichen) Dokumentation der slowenischen usuellen standard-sprachlichen und dialektalen Lexik. Es wurde eine möglichst hoher Vollständigkeitsgrad bei der äußeren Selektion angestrebt, so dass nicht nur die schriftlich belegte allgemeinsprachliche Lexik im Wörterbuch enthalten ist, sondern in relativ hohem Maße ebenso der dialektale und fachliche Wortschatz sowie eine beträchtliche Anzahl der Lexik aus älteren Perioden des slowenischen Schriftentums seit dem 16. Jh. Die regional, dialektal oder fachsprachenspezifisch verbreitete Lexik ist im Plet weitgehend diatopisch und diatechnisch markiert. Von der Buchung ausgeschlossen blieben lediglich die gelegentlich gebildete bzw. nicht usuelle Lexik sowie Fremdwörter, insofern slowenische Äquivalente belegt worden waren. Die potentiellen Benutzer sind somit primär Slowenischsprechende, die sich über die muttersprachliche Lexik ausführlich informieren wollen.

Nach der Betrachtung der Wörterbuchbasis, des Umfangs der gebuchten Lexik und ihrer Selektion sowie aufgrund der Aussagen zu den geplanten Benutzern in den Umtexten kann festgestellt werden, dass die Konzipierung und Ausarbeitung von Plet in hohem Maße auf den spezifischen gesellschaftlichen und nationalorientierten Forderungen in der Mitte des 19. Jhs. basierte. Diese bestanden darin, eine ganzheitliche lexikographische Darstellung der slowenischen usuellen Lexik in ihrer allgemeinsprachlichen (schriftsprachli-

chen und regional-dialektalen) sowie fachsprachlichen und historischen Ausprägung zu erarbeiten. In Bezug auf den primären muttersprachlichen Wörterbuchbenutzer lässt sich Plet als hochgradig benutzeradäquat charakterisieren. Es ist bisher noch immer das umfangreichste deutsch-slowenische Wörterbuch mit slowenischer Ausgangssprache und bleibt auch nach dem Erscheinen des großen fünfbändigen einsprachigen Wörterbuchs des Slowenischen (*Slovar slovenskega knjižnega jezika* 1980–1991, 93 142 Lemmata) eine wichtige thesaurische Sammlung der slowenischen Lexik, die die gesamte Lexikographie mit Slowenisch im 20. Jh. beeinflusst hat (Toporišič 1998: 5). Die Wahl des Wörterbuchtyps (zweisprachiges Wörterbuch mit Slowenisch als Ausgangs- und Deutsch als Zielsprache) ist wohl gesellschaftspolitisch begründet und ist vor allem in Abhängigkeit mit der damaligen Zugehörigkeit des slowenischen Sprachraumes zur Österreichischen Monarchie zu betrachten. Da bei ca. 10 % der aufgenommenen Lexik jedoch Slowenisch (und nicht Deutsch) als Sprache der semantischen Kommentare vorkommt (Furlan 2006: IX), lassen sich beim zweisprachigen Plet zugleich Ansätze eines einsprachigen slowenischen Bedeutungswörterbuchs erkennen. Dies mag seine oben besprochene Funktionalität zusätzlich rechtfertigen.

3.2 PonsN und PonsS

Die neuesten Wörterbücher mit Deutsch und Slowenisch sind Lizenzausgaben der Pons-Wörterbücher des Klett-Verlages. Das PonsN und PonsS sollen jeweils ca. 70 000 „Stichwörter, Phrasen und Satzbeispiele“ umfassen, so laut Angaben auf dem Umschlag, in der Einleitung und in Werbetexten. Nach einer genaueren Untersuchung verzeichnen PonsN und PonsS allerdings weniger Lemmata als angedeutet, nämlich ca. 41 000 bzw. 35 000. Die Einschätzung gründet auf folgender Überprüfung: Im jeweiligen Wörterbuch wurden Lemmata auf 12 zufällig ausgewählten Seiten gezählt. Ihre Durchschnittszahl pro Seite (47, 5 im PonsN und 41 im PonsS) und die Einbeziehung des Gesamtumfangs des Wörterverzeichnisses (das Wörterverzeichnis im PonsN umfasst 874 und im PonsS 856 Seiten) ergeben die genannte Anzahl der aufgenommenen Lemmata, die als realistischer Schätzwert gelten kann.

Die Lemmaliste im PonsN basiert auf der deutschen Lemmaliste des Klett-Verlages. Das Wörterbuch solle den aktuellsten deutschen Wortschatz verzeichnen (samt Anglizismen und Neologismen), ebenso eine Auswahl an österreichischer und Schweizer Lexik sowie eine (unbestimmte) Anzahl der Fachwörter und stilistisch markierter Lexik. Laut Aussagen in der Einleitung (PonsN 2006: V) ist die Auswahl der Lexik in der Zusammenarbeit mit „vielen Übersetzern und Germanisten“ auf den slowenischen Benutzer zugeschnitten worden, es bleibt jedoch völlig unklar, was gemeint ist.

Auch das PonsS will höchstaktuell sein, es soll „die aktuellste Lexik des gegenwärtigen Slowenisch“ verzeichnen (PonsS 2008: V). Die Lemmaliste der slowenischen Ausgangssprache wurde laut Einleitung in mehreren Selektions-

schritten gewonnen. Die Wörterbuchbasis bildete das slowenische Textkorpus Nova beseda (http://bos.zrc-sazu.si/s_beseda.html), ein frei zugängliches Akademiekorpus. Dieses enthält gegenwärtig (Januar 2009) 240 Millionen Wörter, wobei einen weit überwiegenden Anteil des Korpus publizistische Texte darstellen. Im Vergleich mit dem ebenso seit 2006 frei zugänglichen slowenischen Korpus FidaPLUS (www.fidaplus.net) im gegenwärtigen Umfang von 621 Millionen Wörtern gilt Nova beseda als bedeutend weniger repräsentativ für den gegenwärtigen slowenischen Sprachgebrauch, was natürlich die Frage aufwirft, warum die Selektion „der aktuellsten slowenischen Lexik“ gerade anhand dieses Korpus erfolgte. Einzelne Selektionskriterien werden allerdings nicht erläutert, so dass man nicht weiß, wie die Aktualität der gegenwärtigen slowenischen Lexik gemessen und was darunter überhaupt verstanden wurde. Die Fachlexik sollte in den beiden Wörterbüchern reichlich vertreten sein, zumal im jeweiligen Abkürzungsverzeichnis Abkürzungen zur diatechnischen Markierung von 52 verschiedenen Sachgebieten verzeichnet sind. Allerdings lässt sich aus den Aussagen in den Umtexten beider Wörterbücher Näheres über die Auswahl und Anzahl der Fachlexik nicht nachvollziehen.

Über den Benutzerbezug erfährt man in der Einleitung zum PonsN, dass es „für möglichst breiteste Benutzerkreise“ vom Nutzen sein will. Angesprochen ist sowohl der slowenisch sprechende Benutzer, „der mit der deutschen Sprache in Kontakt kommt“ als auch der deutschsprachige Benutzer, der „Slowenisch lernt“ (PonsN 2006: V). Zumal die knappe Beschreibung der Benutzerzielgruppen in der Einleitung vom PonsS mit der entsprechenden Passage im PonsN völlig identisch ist, ist anzunehmen, dass bei der Erarbeitung beider Wörterbücher die verschiedenen Ausgangsprachen und folglich spezifische Benutzerbedürfnisse, Benutzungsanlässe und -situationen falls überhaupt, eher wenig, eine Rolle gespielt haben. So rechnen beide Wörterbücher mit folgenden Benutzungssituationen: (1) die Benutzung in konfliktbedingten rezeptiven und produktiven (vor allem translatorischen) Situationen, in denen sich, nach den Wörterbüchern, vorrangig ein slowenisch sprechender Benutzer befindet, wenn er in Kontakt mit der deutschen Sprache kommt und (2) die Benutzung in rezeptiv- und produktiv ausgerichteten Lernsituationen, in denen sich allerdings nicht nur ein deutschsprachiger Slowenisch-Lerner befindet, wie die Wörterbücher suggerieren, sondern ebenso ein slowenischsprachiger Deutsch-Lerner.

Zum Benutzerbezug der beiden Pons-Wörterbücher ist Folgendes festzustellen: Die Auswahl und der Umfang der gebuchten Lexik richten sich vorrangig nach dem anvisierten Lerner-Benutzer. Dafür sprechen das Streben nach der Aktualität der aufgenommenen Lexik, die Textkorpora als Wörterbuchbasis und ebenso Umtexte im Nachspann beider Wörterbücher. Im PonsN findet man eine relativ umfangreiche (28 Seiten) auf Slowenisch verfasste Wörterbuchgrammatik des Deutschen, im PonsS dagegen eine vergleichbare auf Deutsch verfasste Wörterbuchgrammatik des Slowenischen. Bei den Flexionstabellen in den beiden Grammatiken fällt auf, dass sie in keiner direkten Verbindung mit dem Wörterbuchtext stehen, was u.a. mit einer entsprechenden

Nummerierung der Flexionsklassen möglich gewesen wäre. Die übrigen Umtexte im Nachspann beider Wörterbücher (Liste der Zahlwörter, Maße und Gewichte, Listen geographischer Namen in deutsch- und slowenischsprachigen Gebieten, Nützliche Redewendungen (i.e. Kommunikationsphrasen), Musterbriefe) sind inhaltlich identisch und parallel zweisprachig; verschiedene Sprachen nehmen allerdings jeweils eine unterschiedliche Ausgangsposition ein. Darüber hinaus enthalten beide Wörterbücher eine kleinere Anzahl von eingelagerten Binnentexten (Land und Leute), in denen kulturelle, nationale und/oder sprachliche Spezifika erläutert werden. Ihr Nachteil besteht allerdings darin, dass der Benutzer weder über die alphabetische Hauptzugriffsstruktur noch über eine andere äußere Zugriffsstruktur (wie z.B. über eine Registerzugriffsstruktur) auf sie zugreifen kann und so der Eindruck entsteht, dass sie eher beliebig in den beiden Wörterbüchern verstreut sind.

3.3 DebN und DebS

Nach Angaben im Vorwort verzeichnet das DebN ca. 120 500 und mit sog. Dubletten ca. 123 000 „Lemmata des Gegenwartsdeutsch“ (DebN 2001: V) und das DebS ca. 100 000 „Lemmata der gegenwärtigen slowenischen Standardsprache“. Bei der Auswahl der Lexik wurde „eine besondere Aufmerksamkeit der Sprache der täglichen allgemeinen Kommunikation sowie der Fachsprache geschenkt“ (DebS 2003: V), während bei der Materialerhebung die Lexik der Literatur weniger berücksichtigt und dialektale und umgangssprachliche Ausdrücke prinzipiell gemieden wurden. Dasselbe gilt auch für das DebN, zumal die Aussagen im Vorwort inhaltlich fast identisch sind. Prinzipien, nach denen die Auswahl der aufzunehmenden Lexik erfolgte, lassen sich aus den Umtexten beider Wörterbücher jedoch nicht genauer nachvollziehen.

Aus dem Quellenhinweis im DebS geht hervor, dass unter den primären Quellen deutschsprachige (!) Fachtexte (vorrangig zur Botanik und Zoologie) und ihre slowenischen Übersetzungen dominieren. Unter den sekundären Quellen haben neben den allgemeinen einsprachigen und zweisprachigen slowenisch-deutschen Wörterbüchern einen sehr großen Anteil wiederum deutsche Fachlexika und enzyklopädische Referenzwerke. Dementsprechend sind fachsprachliche Ausdrücke im DebS zahlenmäßig überproportional stark vertreten. Das Abkürzungsverzeichnis notiert 22 Abkürzungen zur diatechnischen Markierung der Fachlexik. Tertiäre Quellen vermisst man in der Wörterbuchbasis vom DebS. Über die Wörterbuchbasis von DebN ist Ähnliches zu berichten, da das Literaturverzeichnis dem aus dem DebS im Wesentlichen gleicht (vgl. Jesenšek 2005).

Die Dominanz der deutschsprachigen Wörterbuchbasis mag verwundern; sie kommt aber offensichtlich durch die starke Ausrichtung der beiden Wörterbücher auf die Bedürfnisse bei der Übersetzung zustande. Denn auf die Translationssituationen, die zur Wörterbuchbenutzung führen können und auf die Probleme der Äquivalentfindung wird in den einleitenden Umtexten mehrere Male hingewiesen, während die Wörterbuchbenutzung bei der gestörten

Textrezeption und Textproduktion sowie die in Lernsituationen nicht thematisiert wird. Mit der Ausnahme einer kurzen slowenischen Wörterbuchgrammatik im DebS enthalten DebN und DebS keine weiteren Umtexe. — Zusammenfassend ist anzumerken, dass die Benutzerbezüge in DebN und DebS nur schwer nachvollziehbar sind.

4. Wörterbuchanalyse II: Wörterbuchgegenstand und Wörterbuchform

In diesem Kapitel werden Teile des Wörterbuchgegenstandes und Ausschnitte der Wörterbuchform der ausgewählten Wörterbücher betrachtet. Der Wörterbuchgegenstand eines bestimmten Sprachwörterbuchs ist die Menge der in diesem Wörterbuch lexikographisch bearbeiteten Eigenschaftsausprägungen von wenigstens einer, höchstens aber von endlichen vielen sprachlichen Eigenschaften bei einer bestimmten Menge von im Wörterbuch lemmatisch und nichtlemmatisch genannten sprachlichen Ausdrücken, die zu mindestens einem Wörterbuchgegenstandsbereich gehören (vgl. Wiegand 1998: 301f). Zu zweisprachigen Wörterbüchern gehören zwei Wörterbuchgegenstandsbereiche: Im vorliegenden Fall sind dies das Deutsche und sein korrekter Gebrauch sowie das Slowenische und sein korrekter Gebrauch. Bei unseren Bemerkungen zum Wörterbuchgegenstand beschränken wir uns auf Wörter der drei Hauptwortarten Substantiv, Verb und Adjektiv sowie auf Mehrwortbenennungen. Der Schwerpunkt der nachfolgenden Wörterbuchanalyse liegt auf der Wörterbuchform; dennoch kann (aus Platzgründen) nur ein relativ schmaler Ausschnitt aus der Wörterbuchform der ausgewählten Wörterbücher Berücksichtigung finden: Betrachtet werden vor allem reine und hybride Artikelmikrostrukturen; ein kurzer Blick wird auch auf artikelinterne Angabeadressierungsstrukturen, auf die alphabetischen Hauptzugriffsstrukturen sowie die monoalphabetischen Makrostrukturen geworfen. Bei der Wörterbuchanalyse wird vor allem auf Wiegands Theorie der Wörterbuchform zurückgegriffen (vgl. u.a. Hausmann und Wiegand 1989, Kammerer und Wiegand 1998/99; Konerding und Wiegand 1994; Meyer und Wiegand 2000; Pan Zaiping und Wiegand 1995; Schierholz und Wiegand 2005; Wiegand 1989, 1989a, 1989b, 1990, 1990a, 1991, 1994/95, 1995, 1995a, 1996, 1996a, 1996b, 1997, 1998, 1998a, 1998b, 1999, 1999a, 2000, 2000a, 2000b, 2000c, 2000d, 2001, 2001a, 2002, 2002a, 2002b, 2002c, 2002d, 2003, 2003a, 2004, 2005, 2005a, 2005b, 2005c, 2005d, 2006, 2006a, 2006b, 2007, 2007a, 2007b, 2007c, 2007d, 2007e, 2008, 2008a, 2009, 2009a). Das Ziel der nachfolgenden Wörterbuchanalyse besteht vor allem darin, die wichtigsten der für die ausgewählten Wörterbücher charakteristischen Typen von textuellen Strukturen darzustellen sowie Ausschnitte aus dem Wörterbuchgegenstand und damit das Datenangebot kritisch zu beleuchten.

4.1 Plet

Gegeben seien die Wörterbuchartikel in Abb. 1.

wa₁:

glěštati, -am, *vb. impf.* 1) warten, pflegen, *C., M.*; živino g, *Kr.-Erj. (Torb.)*; g se, sich wohl sein lassen, *Jan.*; prim. glajštati; – 2) vermögen, haben, *Mur., Jan.*; ne gleštam krajcarja, *Št.*; pri tej hiši dober kruh gleštajo, *Gor.*

wa₂:

óče, -éta, *m.* der Vater; stari o., der Großvater; duhovni o., der geistliche Vater (eines primicerierenden Priesters), *Št.*; krušni o., der Pflegevater; cerkveni očetje, die Kirchenväter, *Cig., Jan., nk.*; veliki o. = starejšina, *C.*; krstni o. = boter, *ogr.-C.*

wa₃–wa₁₉

bakrár, -rja, *m.* der Kupferschmied, *Z.*
bakrárnica, **f.* die Kupferschmiede, *Z.*
bakrárstvo, *n.* das Kupferschmiedgewerbe, *Z.*
bákrast, *adj.* kupferartig, kupferfärbig, *Z.*
bakrĕn, *adj.* kupfern, Kupfer-, *Mur., Cig., Jan., Cig. (T.), nk.*; bakrena ruda, das Kupfererz, *Cig. (T.)*.
bakrĕnast, *adj.* kupfericht, *Cig., Jan.*
bakrĕnec, -nca, *m.* die Kupferlasur, *Z.*; modri b. *Erj. (Min.)*; rdeči b., das Rothkupfererz, *Erj. (Min.)*.
bakrenica, *f.* das Cementwasser (eine kupferhaltige Flüssigkeit), *Cig., Erj. (Min.)*; das Kupferwasser, *Jes.*
bakreník, *m.* das Kupferbergwerk, *Z.*
bakrenina, *f.* die Kupferwaare, *Z.*
bákrenje, *n.* die Verkupferung, *Z.*
bákriti, -im, *vb. impf.* verkupfern, *Z.*
bakrorĕstvo, *n.* die Kupferstecherei, (-štvo, *Cig.*).
bakrorĕž, -rĕža, *m.* der Kupferstich, *Cig., Jan., nk.*
bakrorĕžac, -zca, *m.* der Kupferstecher, *Cig., Jan., nk.*
bakrorĕžan, -zna, *adj.* Kupferstech-: bakrorežna šola, *Navr. (Let.)*.
bakrorĕžje, *n.* = bakrorestvo, *Jan.*

wa₂₀–wa₃₃

glínit, *adj.* thonhaltig, *Cig. (T.)*.
glíniti, -im, *vb. impf.* mit Thon überziehen, *Cig.*
 1. **glínja**, *f.*, *Mur., Cig., Jan.*, *pogl. glina.*
 2. **glínja**, *f.* die grüne Nusschale, *C.*; – prim. klenovec, klinovina.
glínjavica, *f.* = 2. glínja, *C.*
glínjavka, *f.* = glínjavica, *Fr.-C.*
glínje, *n.* der Thon, *Cig., Nov. -C.*
glínjevica, *f.* *Cig., Jan.*, *pogl. glinovica.*
glínka, *f.* = glínica, *Z., C.*
glínokóp, -kópa, *m.* der Thongraber, *Cig.*
 1. **glínov**, *adj.* iz gline, *Trub.-C.*
 2. **glínov**, *adj.* glinova ovca, *Nov.*, *pogl. glenav.*
glínovát, -áta, *adj.* thonhaltig, *Z.*; glínovati železovec, *Cig. (T.)*; *pogl. glinovnat.*
glínovac, -vca, *m.* der Thonstein, *Cig., Jan.*; – prim. glinavec.

wa₃₄–wa₃₅

sočúťje, *n.* das Mitgefühl, die Sympathie, *Cig., Jan., C., nk.*
sočúvstvĕn, -stvĕna, *adj.* sympathetisch, *C.*; sočúvstveni živec, *Cig. (T.)*.

wa₃₆–wa₃₇

glísta, *f.* 1) gliste, die Eingeweidewürmer (entozoa, helminthes): navadna g., der Spulwurm (ascaris lumbricoides), otročja g., der Kindenwurm, die Kindermade (oxyuris vermicularis), *Erj. (Ž.)*; konjska g., der Palissadenwurm (strongilus), ovčja g., der Schafwurm (strongilus filaria), svinjska g., der Kratzer (echinorynchus gigas), ribja g., der Inger (myxine), *Erj. (Z.)*; – der Regenwurm, *Cig. C., jvzhŠt.*; – 2) gliste, = glisti, *SIGradec-C., Notr., Levst. (Rok.)*.
glístast, *adj.* 1) wurmförmig, *Mur., V.-Cig., Jan.*; – 2) = glistav, *Mur.*

Abb. 1: Wörterbuchartikel wa₁ bis wa₃₇ aus Plet

Im Folgenden betrachten wir zunächst hierarchische reine Artikelmikrostrukturen von Kurzartikeln aus Plet in abstrakter und konkreter Ausprägung. Textuelle Strukturen dieses Typs sind die wichtigsten Teilstrukturen von hierarchischen (abstrakten und isomorphen konkreten) reinen Artikelkonstituentenstrukturen. Bei der Bildung der Letztgenannten werden die nichttypographischen Mikrostrukturanzeiger (wie z.B. Kommata, Punkte, Strichpunkte und senkrechte Striche als Textkonstituenten berücksichtigt (vgl. z.B. Wiegand 2000a: 235ff)). Dagegen ist dies bei der Konstruktion von hierarchischen Artikelmikrostrukturen nicht der Fall. Alle Artikel in Abb. 1 sind vollständig kondensierte Wörterbuchartikel (i.S.v. Wiegand 2003a: 205ff); dies bedeutet,

dass sie keine Angabetexte aufweisen. Der Standardisierungsgrad der Artikel ist relativ hoch: Bei gleichem Lemmazeichentyp sind die Artikel in Plet häufig gleichartig gestaltet, so dass auch die lexikographische Bearbeitung der jeweiligen artikelspezifischen Ausschnitte des Wörterbuchgegenstandes relativ einheitlich ist; daher sind die konkreten hierarchischen Artikelmikrostrukturen der Wörterbuchartikel häufig isomorph. Dies gilt vor allem für Kurzartikel, von denen viele Einzeilenartikel sind, wie z.B. wa₃–wa₆ in Abb. 1. Beispielsweise sind die konkreten hierarchischen reinen Artikelmikrostrukturen von wa₁₁–wa₁₃ isomorph, und die drei konkreten Mikrostrukturen der Artikel sind zur zugehörigen abstrakten hierarchischen Mikrostruktur isomorph. Dies ist in Abb. 2 dargestellt.

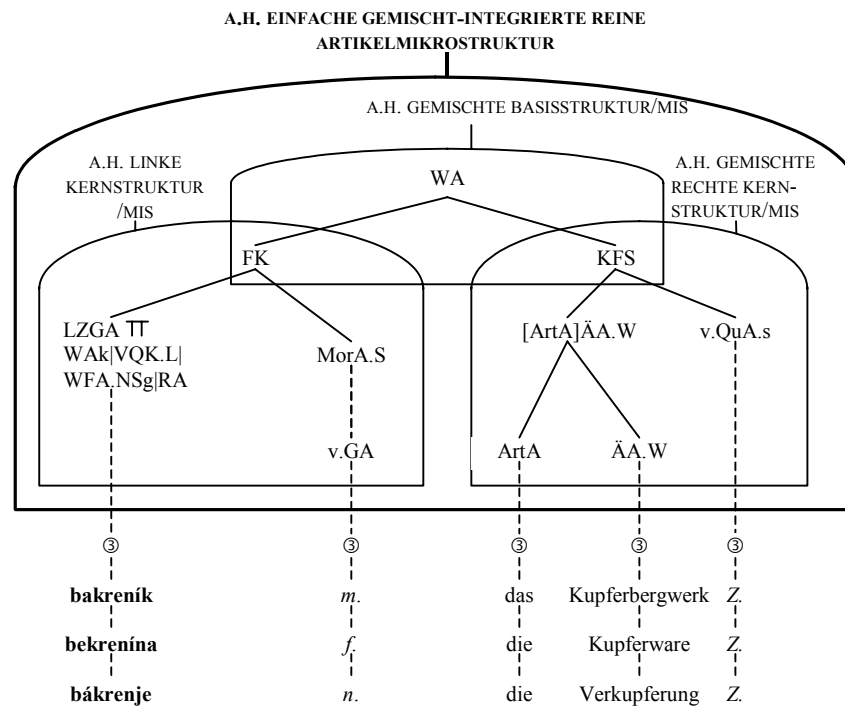


Abb. 2: Einfach kommentierter verdichteter Strukturgraph zur abstrakten hierarchischen (und den isomorphen konkreten hierarchischen) reinen Artikelmikrostrukturen, die zu wa₁₁–wa₁₃ in Abb. 1 gehören. Darstellungskonventionen: „x – y“ bedeutet (von unten nach oben gelesen) soviel wie x ist eine Teilangabe von y; „x – – y“ bedeutet (von unten nach oben gelesen) soviel wie x ist ein Element von y (= x ∈ y); „x – – ③ – – y“ bedeutet soviel wie die Element-Klassen-Beziehung besteht 3 mal. Abkürzungen: WA = Wörterbuchartikel; alle anderen Abkürzungen sind solche für Angabeklassen: FK = Formkommentar; KFS = Kommentar zur Form und Semantik; LZGA = Lemmazeichengestaltangabe; „TT“ = oben erweitert um; WAK|VQK.L = Wortakzent- zugleich Vokalquantität-

tätskennzeichnung zur Länge; „|“ = zugleich; WFA.NSg = Wortformangabe Nominativ Singular; RA = Rechtschreibangabe; MorA.S = Morphologieangabe beim Substantiv; v.GA = verdichtete Genusangabe; [Art A]ÄA.W = um eine Artikelangabe linkerweiterte Wortäquivalentangabe; ArtA = Artikelangabe; ÄA.W = Wortäquivalentangabe; v.QuA.s = verdichtete Angabe einer sekundären Quelle

Reine hierarchische Artikelmikrostrukturen sind Ordnungsstrukturen. In konkreter Ausprägung sind sie dadurch erhältlich, dass auf einer Trägermenge, deren Elemente alle Angaben und der ganze Wörterbuchartikel sind, zwei strukturprägende Relationen definiert werden, und zwar eine Relation vom Typ der Präzedenzrelation (mit dem Relationsterm *x geht y voraus*) sowie eine Relation vom Typ der partitiven Relation (mit dem Relationsterm *x ist eine Teilangabe von y*). Bei der abstrakten Ausprägung sind die Elemente der Trägermenge Angabeklassen eines Wörterbuchs und die Klasse der Wörterbuchartikel. Die elementaren und nichtelementaren Angaben jedes Wörterbuchartikels sind durch eine Anwendung der Methode der nichtexhaustiven funktional-positionalen Segmentation erhältlich (vgl. z.B. Wiegand 2000a: 235ff u. 2005: 217ff). In den Wörterbüchern zu Sprachen, die mit einer Alphabetschrift verschriftet sind und in denen auf Zeilen von links nach rechts geschrieben wird, treten deutlich mehr als 200 Typen von hierarchischen Artikelmikrostrukturen auf. Der Typ der einfachen hierarchischen Artikelmikrostruktur gehört stets zu einem basalen Wörterbuchartikel; Artikel dieses Typs bestehen immer aus einem Formkommentar und einem unmittelbar folgenden Zweitkommentar, der unterschiedlich ausgeprägt sein kann. In Plet sind alle Zweitkommentare Kommentare zur Form und Semantik, weil nach dem Mikrostrukturenprogramm von Plet, das anhand der gegebenen textuellen Strukturen der Wörterbuchartikel vollständig rekonstruiert werden kann, im Zweitkommentar sowohl Angaben zur Form als auch solche zur Bedeutung vorgesehen sind. In den Kurzartikeln wa₁₁–wa₁₃ sind z.B. die nichtadjazent linksadressierten Wortäquivalentangaben Angaben zur Bedeutung und die adjazent rechtsadressierten Artikelangaben Angaben zur Form. Einfache Artikelmikrostrukturen heißen *integriert* — um es sehr allgemein und stark vereinfacht auszudrücken —, wenn allein die angesetzte Bedeutungsstruktur des Lemmzeichens die Struktur des Zweitkommentars bestimmt, so dass alle Angaben im textuellen Skopus einer bestimmten Bedeutungsangabe stehen (genauere Bestimmung z.B. in Meyer und Wiegand 2000: 91). *Gemischt-integriert* heißt eine Artikelmikrostruktur dann, wenn der zugehörige Wörterbuchartikel als Zweitkommentar keinen semantischen Kommentar, sondern einen Kommentar zur Form und Semantik aufweist.

Zu jedem Mikrostrukturtyp lässt sich ein allgemeines Mikrostrukturbild angeben. Allgemeine Mikrostrukturbilder sind nach festen Vorschriften erzeugte Ausschnitte aus Strukturgraphen für abstrakte hierarchische reine Artikelmikrostrukturen. Sie sind formale Darstellungsmittel und dienen dazu, den jeweiligen Mikrostrukturtyp, der zu einem Artikel gehört, schnell zu erkennen.

Das allgemeine Mikrostrukturbild, das zu den Artikeln wa₁₁–wa₁₃ in Abb. 1 und zu Hunderten von weiteren Einzeilenartikeln in Plet gehört, die isomorphe Mikrostrukturen aufweisen, findet sich in Abb. 3.

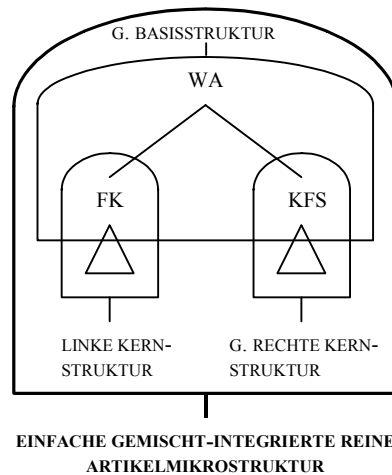


Abb. 3: Allgemeines Mikrostrukturbild zu gemischt-integrierten reinen Artikelmikrostrukturen, die zu basalen Wörterbuchartikeln mit monosemen Lemmazeichen gehören wie z.B. wa₁₁–wa₁₇ in Abb. 1. Abkürzungen: G = GEMISCHTE

Zu jedem allgemeinen Mikrostrukturbild kann ein einfaches Artikelstrukturschema angegeben werden. Solche Schemata erleichtern den Strukturvergleich der Mikrostrukturen unterschiedlicher Artikel. Das einfache Artikelstrukturschema, das zum allgemeinen Mikrostrukturbild in Abb. 3 und damit zu wa₁₁–wa₁₃ und zu allen Artikeln angegeben werden kann, deren Mikrostruktur (MikS) zu den Mikrostrukturen von wa₁₁–wa₁₃ isomorph sind, lautet wie folgt: WA: _(MikS) FK < KFS (mit „:“ für *besteht aus* und „<“ für *geht voraus*; der Index MikS kann in Kotexten weggelassen werden, in denen klar ist, dass es um Mikrostrukturen geht).

Zu Wörterbuchartikeln, die mindestens einen funktionalen Angabezusatz aufweisen, kann nicht nur eine reine, sondern stets auch eine hybride Artikelmikrostruktur in konkreter und abstrakter Ausprägung angegeben werden; *hybrid* heißen konkrete Artikelmikrostrukturen dann, wenn ihre Trägermengen nicht elementenhomogen sind, also ausschließlich Textkonstituenten enthalten, sondern wenn ihre Trägermengen elementenheterogen sind, weil ihre Elemente von verschiedener Herkunft sind, da sie sowohl durch eine Anwendung der Methode der funktional-positionalen als auch durch eine Anwendung der Methode der nichtfunktional-positionalen Segmentation sowie durch eine Anwendung der Methode der segmentativen Isolierung gegeben sind. — Im Folgenden werden anhand von wa₁₁ in Abb. 1 alle methodischen Schritte erläutert, die ausgeführt werden müssen, um eine hybride Artikelmikrostruktur zu erhalten.

Mit funktionalen Angabezusätzen werden elementare Angaben oben oder unten oder binnen erweitert. Durch Angabenerweiterung mittels funktionaler Angabezusätze entstehen erweiterte elementare Angaben, die deswegen als elementar gelten, weil sie nicht durch eine Anwendung der funktional-positionalen Segmentation in mindestens zwei Angaben (restfrei) segmentiert werden können. Während Angaben funktional-positional isolierbare Textkonstituenten sind, also Textsegmente mit eigener Form, eigener genuiner Funktion sowie eigener textueller Position, gilt dies für funktionale Angabezusätze nicht, weil sie entweder keine eigene Position — wie die oben und unten erweiterten funktionalen Angabezusätze — aufweisen, oder weil sie — wie die binnerweiternden Glossate — zwar eine eigene textuelle Position aufweisen, aber funktional-positional nicht isolierbar sind. Durch die Angabenerweiterung erhalten elementare Angaben eine interne Angabestruktur, die systematisch analysiert werden kann, wie nachfolgend gezeigt wird, und zwar am Beispiel von wa₁₁. In diesem Kurzartikel ist die Lemmazeichengestaltangabe, nämlich „**bakreník**“, eine elementare Angabe, die durch einen bifunktionalen Angabezusatz, nämlich eine Wortakzentkennzeichnung, oben erweitert ist, die zugleich eine Vokalquantitätskennzeichnung zur Länge ist. Der bifunktionale Angabezusatz ist durch einen Akut realisiert, der an das Angabeformsegment **i** hinadressiert ist. Als elementare Angabe ist die als lemmatische Substantivangabe ausgeprägte oben erweiterte Lemmazeichengestaltangabe „**bakreník**“ nicht funktional-positional segmentierbar. Um die Elemente ihrer internen Struktur zu ermitteln, muss sie durch eine Anwendung der Methode der segmentativen Isolierung segmentiert werden. Dabei werden horizontale Segmentationsschnitte so gelegt, dass die oben und die unten erweiternden funktionalen Angabezusätze vom Rest der erweiterten elementaren Angabe abgetrennt werden. Das Segmentationsergebnis besteht demgemäß aus dem Segment **bakreník** und dem Akut, so dass zwei Teile einer erweiterten Angabe vorliegen. Um die zwar visuell deutlich erkennbare Position des Akuts über dem Angabeformsegment **i** auch auf einem methodischen Weg einwandfrei und eindeutig festlegen zu können, wird „**bakreník**“ mittels einer Anwendung der Methode der nichtfunktional-positionalen Segmentation so segmentiert, dass sich die folgenden drei Angabeformsegmente ergeben: **bakren | í | k**: hierbei sind die senkrechten Striche Segmentationsfugenmarkierungen. Das Segment **bakren** gehört zur Klasse der vorderen Angabeformsegmente (vAFSeg), so dass gilt: **bakren** ∈ vAFSeg; **i** gehört zur Klasse der mittleren Angabeformsegmente (mAFSeg), so dass gilt: **i** ∈ mAFSeg. Schließlich gehört **k** zur Klasse der hinteren Angabeformsegmente (hAFSeg), und es gilt demgemäß die Aussage **k** ∈ hAFSeg. Weiterhin gelten die folgenden Aussagen: (i) **bakren** < **i** < **k** sowie (ii) vAFSeg < mAFSeg < hAFSeg mit „<“ für *geht voraus*.

Führt man die erläuterten Segmentationsoptionen in entsprechender Weise bei den Lemmazeichengestaltangaben von wa₁₂ und wa₁₃ aus, dann erhält man folgende beiden Segmentationsergebnisse: **bakren | í | na** und **b | á | krenje**. Für die Angabeformsegmente gelten die anhand von **bakren | í | k** gemachten Aussagen *mutatis mutandis*.

Um die zur konkreten internen Angabestruktur isomorphe abstrakte interne Angabenstruktur von „**bakreník**“ in wa₁₁ in Abb. 1 angeben zu können, wird nachfolgend zunächst eine Trägermenge — sie heiÙe $M_{Ar}^a(\mathbf{bakreník})$ — für die zu bildende Angabestruktur gebildet; sie kann wie folgt angegeben werden: $M_{Ar}^a(\mathbf{bakreník}) = \{mAFSeg, WAK | VQK.L\}$.

Diese Trägermenge enthält mithin zwei Elemente, die Klassen sind; sie gehört zu den elementenheterogenen Trägermengen, da nicht beide Klassen solche von Textkonstituenten sind. Im nächsten Schritt wird auf der Trägermenge eine zweistellige irreflexive und damit asymmetrische sowie transitive Relation — sie heiÙe $R_{ob}^a(\mathbf{bakreník})$ — vom Typ der textarchitektonischen *oberhalb*-Relation definiert, zu der der Relationsterm *x ist oberhalb von y* gehört, mit „*x*“ als Variable für oben erweiternde funktionale Angabezusätze und „*y*“ als Variable für Angabeformsegmente, die deren Bezugsadresse bilden; $R_{ob}^a(\mathbf{bakreník})$ ist eine Teilmenge des Kartesischen Produktes der Trägermenge mit sich selbst, so dass gilt: $R_{ob}^a(\mathbf{bakreník}) \subseteq M_{Ar}^a(\mathbf{bakreník}) \times M_{Ar}^a(\mathbf{bakreník})$. Für das einfache Kreuzprodukt einer Menge *M* mit sich selbst wird im Folgenden M^2 geschrieben. Die Relation prägt auf der Trägermenge eine Struktur, die zu den vertikalen Angabearchitekturen gehört. Strukturen, deren strukturprägende Relationen entweder *oberhalb*- oder *unterhalb*- oder *links*- oder *rechts*-Relationen sind, heißen in der Theorie der Wörterbuchform *Architekturen*. Vertikale Angabearchitekturen von oben oder unten erweiterten elementaren Angaben lassen sich mit den gleichen formalen Darstellungsmitteln darstellen wie die vertikalen Textarchitekturen von Wörterbuchartikeln (vgl. z.B. Wiegand 2001a: 191ff). In Abb. 4 findet sich das allgemeine Angabearchitekturbild für die oben um eine Wortakzentkennzeichnung, die zugleich eine Vokalquantitätskennzeichnung zur Länge ist, erweiterten Lemmazeichengestaltangaben in wa₁₁–wa₁₃.

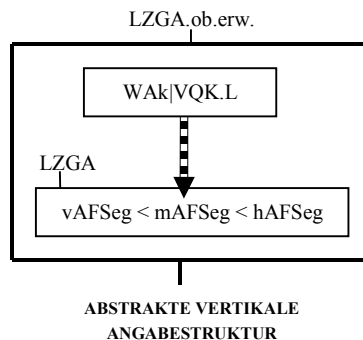


Abb. 4: Allgemeines Angabearchitekturbild zu den oben erweiterten Lemmazeichengestaltangaben in wa₁₁–wa₁₃ in Abb. 1; *Darstellungskonventionen:* „*x* \dashrightarrow *y*“ bedeutet soviel wie *x ist oberhalb von y*, wenn der Pfeil nach unten zeigt; *Abkürzungen:* vAFSeg = vorderes Angabeformsegment; mAFSeg = mittleres Angabeformsegment; hAFSeg = hinteres Angabeformsegment; WAK|VQK.L = Wortakzent-, zugleich Vokalquantitätskennzeichnung zur Länge; LZGA.ob.erw. = oben erweiterte Lemmazeichengestaltangabe

Die abstrakte vertikale Angabearchitektur, deren allgemeines Angabearchitekturmodell in Abb. 4 zu sehen ist, ist eine Teilstruktur der abstrakten (a) hierarchischen hybriden (hy) Angabestruktur (Ans), die zur Lemmazeichengestaltung „**bakreník**“ in wa_{11} gehört. Um diese Angabestruktur zu erhalten, benötigen wir zunächst eine geeignete Trägermenge, die alle Elemente der Angabestruktur enthält; diese heiÙe $M_{hyAns}^a(\mathbf{bakreník})$ und kann mit einer Mächtigkeit von $|6|$ wie folgt angegeben werden:

$$M_{hyAns}^a(\mathbf{bakreník}) = \{LZGA \text{ TT } WAK | VQK.L | WFA.NSg | RA, \\ LZGA | WFA.NSg | RA, WAK | VQK.L, vAFSeg, mAFSeg, h.AFSeg\}.$$

Im nächsten Schritt werden auf dieser Trägermenge die folgenden drei strukturprägenden Relationen definiert:

- (i) $R_p^a(\mathbf{bakreník})$, eine irreflexive (damit asymmetrische) und transitive Relation vom Typ der Präzedenzrelation mit dem Relationsterm x geht y voraus (mit „ x “ und „ y “ als Variablen für Klassen von Textsegmenten); es gilt: $R_{ap}^a(\mathbf{bakreník}) \subseteq (M_{hyAns}^a(\mathbf{bakreník}))^2$
- (ii) $R_{part}^a(\mathbf{bakreník})$, eine reflexive, asymmetrische und transitive Relation vom Typ der partitiven Relation mit dem Relationsterm u ist ein Teil von v (mit „ u “ und „ v “ als Variablen für Klassen von Textsegmenten); es gelten: $R_{part}^a \subseteq (M_{hyAns}^a(\mathbf{bakreník}))^2$ sowie $R_{part}^a(\mathbf{bakreník}) \cap R_p^a(\mathbf{bakreník}) = \emptyset$.
- (iii) $R_{ob}^a(\mathbf{bakreník})$, eine textarchitektonische Relation vom Typ der oberhalb-Relation mit dem Relationsterm r ist oberhalb von s (mit „ r “ als Variable für oben erweiternde funktionale Angabezusätze und „ s “ als Variable für deren unten situierte Bezugsadressen); es gelten: $R_{ob}^a(\mathbf{bakreník}) \subseteq (M_{hyAns}^a(\mathbf{bakreník}))^2$; $R_{ob}^a(\mathbf{bakreník}) \cap R_p^a(\mathbf{bakreník}) = \emptyset$; $R_{ob}^a(\mathbf{bakreník}) \cap R_{part}^a(\mathbf{bakreník}) = \emptyset$.

Als Ergebnis der erläuterten Operationen ist die abstrakte hierarchische architektonisch angereicherte Angabestruktur der um eine Wortakzent- und Vokalquantitätskennzeichnung oben erweiterten Lemmazeichengestaltung „**bakreník**“ aus wa_{11} erhältlich, die zugleich eine Wortformangabe für den Nominativ Singular und eine Rechtschreibangabe ist.

Die in Abb. 5 dargestellte abstrakte hierarchische architektonisch angereicherte Angabestruktur ist isomorph zur konkreten hierarchischen architektonisch angereicherten Angabestruktur, die die Lemmazeichengestaltung „**bakreník**“ in wa_{11} in Abb. 1 aufweist. Diese konkrete Struktur ist auf dem gleichen methodischen Weg erhältlich wie die in Abb. 5 dargestellte abstrakte Struktur; lediglich die Elemente der Trägermenge sind keine Klassen, sondern konkrete Textsegmente oder Individuennamen für diese.

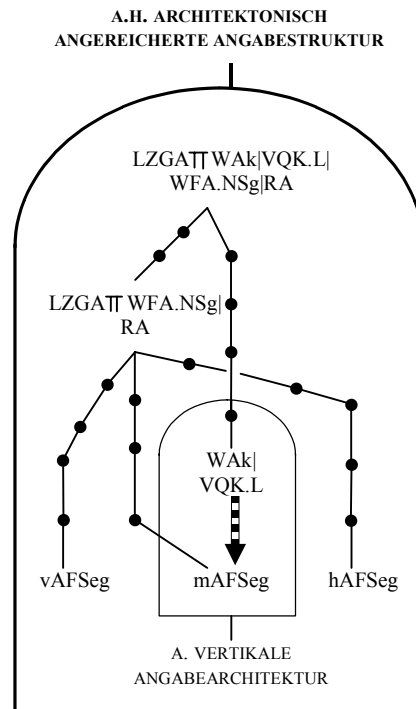


Abb. 5: Einfach kommentierter und mit architektonischen Komponenten angereicherter Strukturgraph zur abstrakten architektonisch angereicherten Angabestruktur, die zur Lemmazeichengestaltangabe in wa_{11} gehört; Darstellungskonventionen: „ $u \text{---} \bullet \text{---} v$ “ bedeutet (von unten nach oben gelesen) soviel wie u ist ein Teil von v ; Abkürzungen: A = ABSTRAKTE

Die in Abb. 5 präsentierte Angabestruktur ist eine Teilstruktur der abstrakten hierarchischen einfachen gemischt-integrierten hybriden Artikel mikrostruktur, die zu wa_{11} in Abb. 1 gehört. Diese Artikel mikrostruktur ist auf der Darstellungsebene dadurch erhältlich, dass man die in Abb. 5 dargestellte abstrakte hierarchische architektonisch angereicherte Angabestruktur in die in Abb. 2 dargestellte Artikel mikrostruktur integriert; anders ausgedrückt heißt das: Die beiden Strukturgraphen werden strukturadäquat kontaminiert. Der klassische Weg für die Bildung der abstrakten hierarchischen einfachen gemischt-integrierten hybriden Artikel mikrostruktur, die zu wa_{11} gehört, besteht allerdings darin, dass man die Trägermenge der in Abb. 2 dargestellten reinen Artikel mikrostruktur (die sich — nebenbei bemerkt — aus der Abb. 2 leicht ablesen lässt) um die folgenden fünf Klassen von Textsegmenten erweitert: LZGA TT WFA.NSg|RA, WAK|VQK.L, vAFSeg, mAFSeg und hAFSeg. Dies sind die Textsegmentklassen, zu denen die Textsegmente gehören, die erhältlich sind, wenn man — wie oben gezeigt — die Lemmazeichengestaltangabe

„**bakreník**“ zuerst mittels einer Anwendung der Methode der segmentativen Isolierung und daraufhin mittels einer Anwendung der Methode der nicht-funktional-positionalen Segmentation in Textsegmente zerlegt. Die elementen-heterogene Trägermenge für die abstrakte hybride Artikel mikrostruktur — sie heiße $M_{\text{hyMiS}}^{\text{a}}$ (**bakreník**) — kann entsprechend mit einer Mächtigkeit von $|15|$ wie folgt angegeben werden:

$$M_{\text{hyMiS}}^{\text{a}}(\text{bakreník}) = \{\text{WA, FK, KFS, LZGA TT Wak} | \text{VQK.L} | \text{WFA.NSg} | \text{RA, MorA.S, [ArtA]ÄÄ.W, v.QuA.s, v.GA, ArtA, ÄÄ.W, LZGA} | \text{WFFA.NSg} | \text{RA, Wak} | \text{VQK.L, vAFSeg, mAFSeg, hAFSeg}\}.$$

Auf $M_{\text{hyMiS}}^{\text{a}}$ (**bakreník**) wird dann eine Relation vom Typ der Präzedenzrelation, eine vom Typ der partitiven Relation mit dem Relationsterm *x ist eine Teilangabe von y* (mit „x“ als Variable für Angaben und mit „y“ als Variable für Angaben und Wörterbuchartikel), eine weitere Relation vom Typ der partitiven Relation mit dem Relationsterm *u ist ein Textsegment (ohne Textkonstituentenstatus) von v* sowie eine textarchitektonische Relation vom Typ der oberhalb-Relation definiert. Nach Ausführung dieser Operationen ergibt sich die in Abb. 6 auszugsweise dargestellte abstrakte hierarchische einfache gemischt-integrierte hybride Artikel mikrostruktur, die zu wa_{11} in Abb. 1 gehört. Die Struktur ist deswegen nur auszugsweise dargestellt, weil die zum KFS gehörige gemischte rechte Kernstruktur als Teilstruktur der Artikel mikrostruktur mit der in Abb. 2 dargestellten Kommentarstruktur, nämlich der abstrakten hierarchischen gemischten rechten Kernstruktur, identisch ist.

Weit über die Hälfte der Substantiv-, Adjektiv- und Verbartikel in Plet gehören zum Typ des basalen Wörterbuchartikels mit Kommentar zur Form und Semantik (i.S.v. Wiegand 2003: 263f), so dass das mikrostrukturelle Artikelstrukturschema $\text{WA} : \text{FK} < \text{KFS}$ und das allgemeine Mikrostruktur bild in Abb. 3 für sie gilt. Werden Homonyme als Lemmata angesetzt, erhält der zugehörige Kurzartikel eine topikalisierte Homonymenangabe; da die Homonymie eine Bedeutungsbeziehung zwischen n sprachlichen Zeichen ist (mit $n \geq 2$), ist eine Homonymangabe keine Formangabe und somit keine Teilangabe des Formkommentars: Demgemäß stehen die Homonymenangaben in wa_{22} und wa_{23} sowie in wa_{30} und wa_{31} in Abb. 1 im Präkommentar (PräK), so dass links-erweiterte Artikel vorliegen, zu denen das mikrostrukturelle Artikelstrukturschema $\text{WA} : \text{PräK} < \text{FK} < \text{KFS}$ und das allgemeine Mikrostruktur bild in Abb. 7 gehören. — In relativ vielen Kurzartikeln steht nach dem nichttypographischen Mikrostruktur anzeiger „—“ eine lemmatisch orientierte vollständige Verweisangabe, wie z.B. in wa_{23} die polyadressierte Verweisangabe „prim. klenovec, klinovina“ und in wa_{33} die monoadressierte Verweisangabe „prim. glinavec“. Da die zugehörigen Verweisungen sich jeweils auf das gesamte sprachliche Zeichen und damit auf die Form und Bedeutung beziehen, sind diese Verweisangaben Teilangaben des Kommentars zur Form und Semantik und stehen nicht im Postkommentar eines rechtserweiterten Artikels, sondern in einer finalen Verweisposition des KFS eines basalen Wörterbuchartikels.

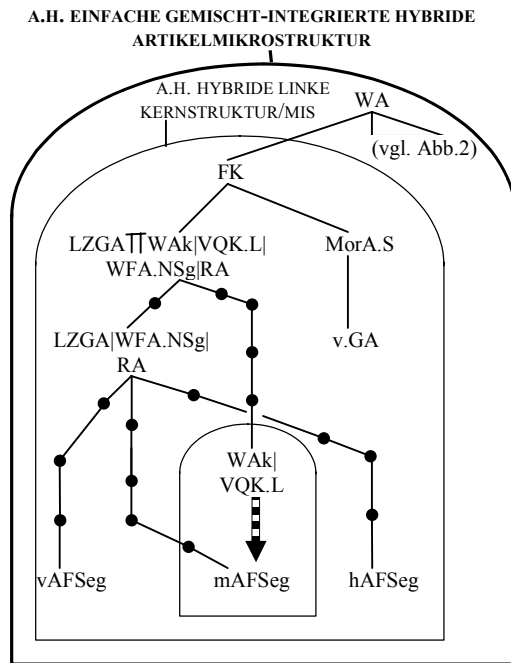


Abb. 6: Nicht vollständig ausgeführter Strukturgraph zur abstrakten hierarchischen einfachen gemischt-integrierten hybriden Artikelmikrostruktur, die zu wa₁₁ gehört

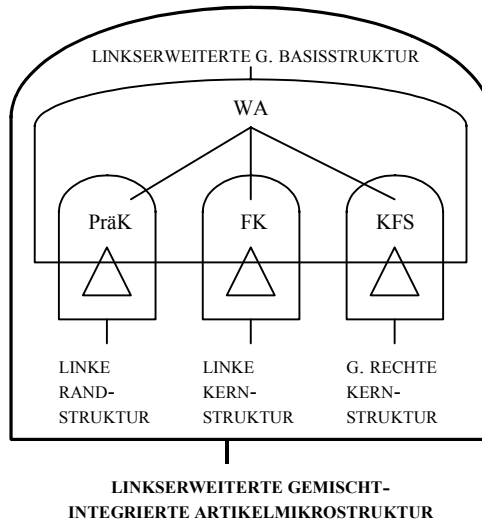


Abb. 7: Allgemeines Mikrostrukturbild zu gemischt-integrierten reinen Artikelmikrostrukturen, die zu linkserweiterten Wörterbuchartikel (wie z.B. wa₂₂ u. wa₂₃ in Abb. 1) gehören

In Abb. 8 werden die abstrakte (und isomorphe konkrete) linkerweiterte gemischt-integrierte reine Artikel mikrostrukturen dargestellt.

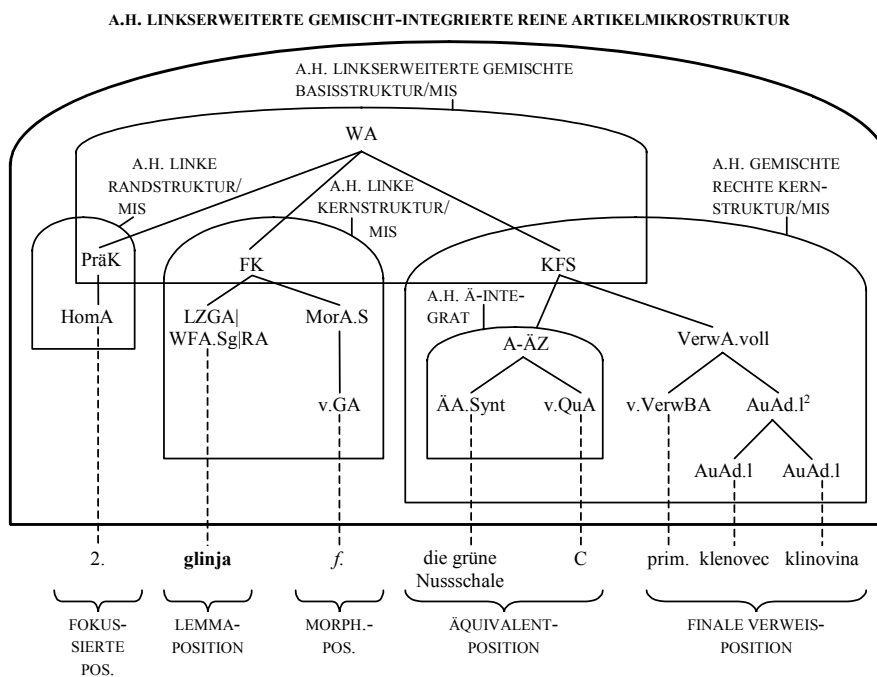


Abb. 8: *Erweitert kommentierter Strukturgraph zur abstrakten (und isomorphen konkreten) linkerweiterten gemischt-integrierten reinen Artikel mikrostruktur, die zu wa₂₃ in Abb. 1 gehören; Abkürzungen: A. = ABSTRAKTE; H = HIERARCHISCHE; POS. = POSITION; Ä-INTEGRAT = ÄQUIVALENT-INTEGRAT; AuAd.l² = homosegmentäre lemmatische Außenadressenangabe, die aus zwei Außenadressenangaben besteht; VerwA.voll = vollständige Verweisangabe; v.VerBA = verdichtete Verweisbeziehungsangabe; v.QuA = verdichtete Quellenangabe*

Die erweiterten Kommentierungen unterhalb des Strukturgraphen in Abb. 8 bestehen aus Namen für die mikrostrukturellen Artikelpositionen; diese sind Ausschnitte aus der präzedentiven Artikel mikrostruktur. Zu einer mikrostrukturellen Artikelposition gehören Angaben (und in partiell kondensierten Artikeln auch Angebotexte), die in sprachtheoretischer und/oder metalexikographischer Perspektive zusammengehören (vgl. Wiegand 2000a: 242f).

Im Folgenden werden einige Bemerkungen zum Wörterbuchgegenstand und damit zum Datenangebot in den Kurzartikeln von Plet gemacht. Ob das Datenangebot eines Wörterbuchs angemessen ist, lässt sich nur relativ zum Wörterbuchtyp und relativ zu den angesetzten Wörterbuchfunktionen (i.S.v. Wiegand 2001) feststellen. Wir wissen bereits, dass Plet kein typisches allge-

meinsprachliches zweisprachiges Wörterbuch ist. Über die Wörterbuchfunktionen findet man in den Metatexten dieses Wörterbuchs nichts Weiterführendes. Deutlich ist nur, dass Plet ein Wörterbuch für potenzielle Benutzer mit Slowenisch als Muttersprache ist. Für solche Benutzer, die Plet als zweisprachiges Wörterbuch in Situationen der Hinproduktion oder in Situationen der Hinübersetzung nutzen möchten, ist dieses Wörterbuch allerdings relativ schlecht geeignet, wie im Folgenden kurz dargelegt werden soll.

Der slowenische Benutzer findet bei vielen deutschen Äquivalenten eine veraltete Schreibung, z.B. tritt durchgängig *th*-Schreibung auf, so in wa₂₀ „thonhältig“, in wa₁₁ „Thon“, in wa₂₉ „Thongräber“ oder s.v. „**bájanica**“ „Wünschelruthé“; oder statt initialem *k* | *K* steht *c* | *C*, z.B. s.v. „**istobítnost**“ „Consubstantialität“ oder s.v. „**istonárodən**“ „connational“; weiterhin finden sich statt der Umlautbuchstaben häufig Schreibungen von Vokalbuchstabe + *e*, z.B. s.v. „**glicerín**“ „Oelsüß“ oder s.v. „**istomôčnica**“ „Äquivalentzahl“. Häufig findet sich statt *z* | *Z* der Buchstabe *c* | *C*, z.B. s.v. „**istoviti**“ „identifizieren“, s.v. „**bakreníca**“ „Cementwasser“. Öfters finden sich Umlautschreibungen, wenn im Standarddeutsch Umlaut nicht üblich ist, z.B. „thonhältig“ s. v. „**glinovàt**“ oder „kupferfärbig“ s.v. „**bákrast**“. Die genannten Fälle sind keine Einzelfälle, sondern es gibt jeweils Hunderte. Offensichtlich wurde hinsichtlich der Orthographie Quellentreue angestrebt. Neben einer vollständig veralteten Orthographie findet sich aber auch die neueste, z.B. in wa₂₃ „Nusschale“ (mit 3 s!). Schlimmer als dieses orthographische Durcheinander ist aber, dass man Hunderte von deutschen Äquivalenten findet, die kein Deutscher, der seine Sprache beherrscht, kennt und die auch in keinem der einschlägigen deutschen Wörterbücher des heutigen Deutsch (wie z.B. Duden-³GW oder Wahrig-⁸DW) belegbar sind. Hier nur einige Beispiele: *Keuscherstand* als Äquivalent zu *bâjtarstvo*, *Fackelschuh* als Äquivalent zu *bákətnik* und *absprecherisch* als Äquivalent zu *sôdč'ən*. Viele Äquivalente sind veraltet und in deutschen Wörterbüchern entsprechend markiert, z.B. *Oelsüß* als Äquivalent zu *glicerín* (vgl. Duden-³GW). Häufig werden bei Substantiven veraltete Formen des Nominativ Singular verwendet, wie z.B. „Mytholog“ oder „Glaubensgenoss“. Oft sind auch die Syntagmaäquivalente im heutigen Deutsch nicht bekannt wie z.B. „mit Brettern und Balken aussteifen“ als Äquivalent zu *bâjtati*. Viele Komposita, die als Äquivalente genannt werden, sind zwar im Deutschen bildbar, aber ungebrauchlich wie z.B. *Wasserschlamm* als Äquivalent zu *glīb*, *Mitmagd* als Äquivalent zu *sodékla*, oder *Mittheilnehmerin* als Äquivalent zu *sodeléžnica*.

Wollte man die Kurzartikel in Plet so bearbeiten, dass aus ihnen Kurzartikel werden, die ein slowenischer Benutzer in Situationen der Hinübersetzung und in Hinproduktionsituation erfolgreich nutzen könnte, wären mindestens die Folgenden Veränderungen nötig:

- einheitliche neueste deutsche Orthographie bei allen mit Äquivalentangaben genannten Äquivalenten

- Ergänzung der morphologischen Angaben bei allen Äquivalentangaben, und zwar: verweisvermittelnde Deklinationsmusterangaben bei Substantiven, Konjugationsklassenangaben (regelmäßig vs unregelmäßig) und andere Verbklassenangaben bei Verben, Angaben zur unregelmäßigen Graduierung bei Adjektiven
- Kollokationsangaben und Synonymangaben in Substantiv-, Verb- und Adjektivartikeln
- Verzicht auf die rudimentären Quellenangaben.

Die vorgeschlagenen Veränderungen sind ein Minimalprogramm für eine modernisierte Neuauflage. Sie lassen sich hinsichtlich der Wörterbuchform unterschiedlich realisieren. Für Substantivartikel ist z.B. eine Angabe zur Äquivalenz (A-Äz) optimal, die obligatorisch aus einer adjazent rechtsadressierten Artikelangabe (ArtA), mindestens einer Äquivalentangabe (entweder ÄA.W od. ÄA.Synt) besteht, sowie aus einer umtextorientierten numerischen verweisvermittelnden Deklinationsmusterangabe (DekMA). Die Wortäquivalentangaben müssen durch Silbentrennungsangaben binnenerweitert und durch Wortakzentkennzeichnungen, die zugleich Vokalquantitätskennzeichnung sind, unten erweitert sein. Fakultativ folgt eine Angabe eines äquivalenten Kollokationspaares (A.Koll.P.ä) oder eine entsprechende homosegmentäre Angabe (A.KollP.ä², mit $n \geq 2$) und/oder eine Angabe eines äquivalenten Kompetenzbeispielpaares (A.KbeiP.ä).

Beispielsweise der Artikel wa₃₄ in Abb. 1 könnte so umgeschrieben werden, das z.B. folgender Artikel wa₃₄ gegeben ist:

wa₃₄: **sočútje**, *n.* das Mit|ge|füh| 11|| die Sym|pa|thie, 19 || ~ **e do bolnih**
Mitgefühl für Kranke

Anhand der als reduzierte Verweisangaben „11“ und „19“ ausgeprägten numerischen umtextorientierten verweisvermittelten Deklinationsmusterangaben, kann der Benutzer-in-actu einen umtextorientierten Verweis erschließen, dessen Befolgung durch die Ausführung einer externen Verweisbefolgungshandlung (i.S.v. Wiegand 1998: 408ff) ihn zu den numerischen Umtexteingängen **11** und **19** bringt, die Elemente einer numerischen äußeren Umtextzugriffstruktur sind, die auf Deklinationsmuster Zugriff eröffnen, die Teile von Deklinationstabellen sind; diese könnten beispielsweise ähnlich gestaltet sein wie die im Wahrig-^sDW. Dies bedeutet zugleich, dass in einer Neuauflage von Plet eine tabellarische Wörterbuchgrammatik mit mehreren äußeren Umtextzugriffstrukturen vorzusehen wäre.

Die abstrakte (und konkrete isomorphe) hierarchische einfache gemischtintegrierte reine Artikelmikrostruktur, die zu wa₃₄ gehören, sind in Abb. 9 dargestellt.

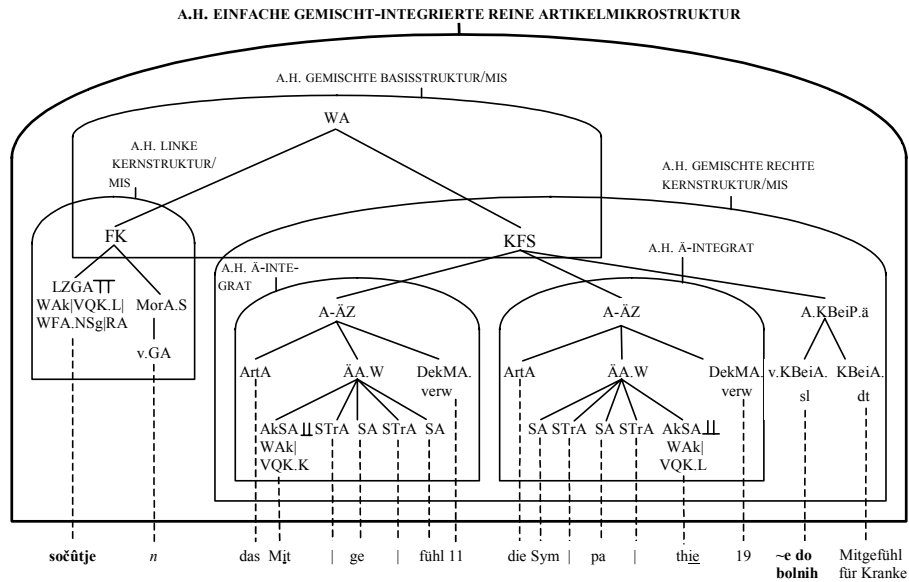


Abb. 9: Einfach kommentierter Strukturgraph zur abstrakten (und konkreten isomorphen) hierarchischen einfachen gemischt-integrierten reinen Artikelmikrostrukturen, die zu wa₃₄ gehören; *Abkürzungen:* AkSA = Akzentsilbenangabe, AkSA || WAK| VQ.K.L = Akzentsilbenangabe, unten erweitert um eine Wortakzent- zugleich Vokalquantitätskennzeichnung zur Länge; SA = Silbenangabe; STra = Silbentrennungsangabe; DekMA.verw = verweisvermittelnde Deklinationsmusterangabe; AkSA || WAK| VQA.L = Akzentsilbenangabe, unten erweitert um eine Wortakzent- zugleich Vokalquantitätskennzeichnung zur Länge; A.KBeiP.ä = Angabe eines äquivalenten Kompetenzbeispielpaares; v.KBeiA.sl = verdichtete slowenische Kompetenzbeispielangabe; KBeiA.dt = deutsche Kompetenzbeispielangabe.

Mit der großen Mehrzahl der Lemmata in Plet werden Lemmazeichen genannt, die als monoseme Wörter interpretiert sind. Als polysem interpretierte Lemmazeichen treten viel seltener auf. Im Folgenden betrachten wir die textuellen Strukturen von wa₁ in Abb.1, dem Verbartikel zum Lemmazeichen *glęštati*, das als zweifach polysem interpretiert ist. Nach dem aus den Verbartikeln von Plet erschließbaren Mikrostrukturprogramm gilt: Formangaben zu den deutschen Äquivalentangaben sind — im Unterschied zu den Substantivartikeln — im Zweitkommentar nicht vorgesehen. Deswegen weisen die Verbartikel als Zweitkommentar auch keinen Kommentar zur Form und Semantik, sondern einen semantischen Kommentar auf.

Bei allen nachfolgenden Strukturanalysen wird der Prozess der Struktur- bildung nicht mehr erläutert, sondern es werden nur dessen Ergebnisse prä- sentiert. Das allgemeine Mikrostrukturbild von wa₁ ist in Abb. 10 zu sehen.

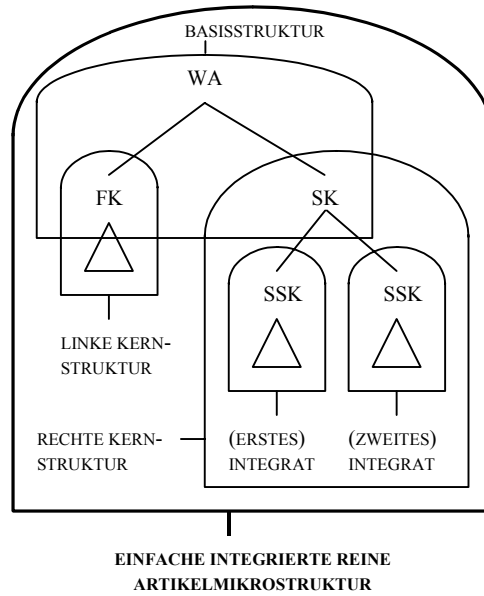


Abb. 10: Allgemeines Mikrostrukturbild zu integrierten reinen Artikelmikrostrukturen, die zu basalen Wörterbuchartikeln zu zweifach polysemen Lemmazeichen gehören; Abkürzungen: SK = semantischer Kommentar; SSK = semantischer Subkommentar

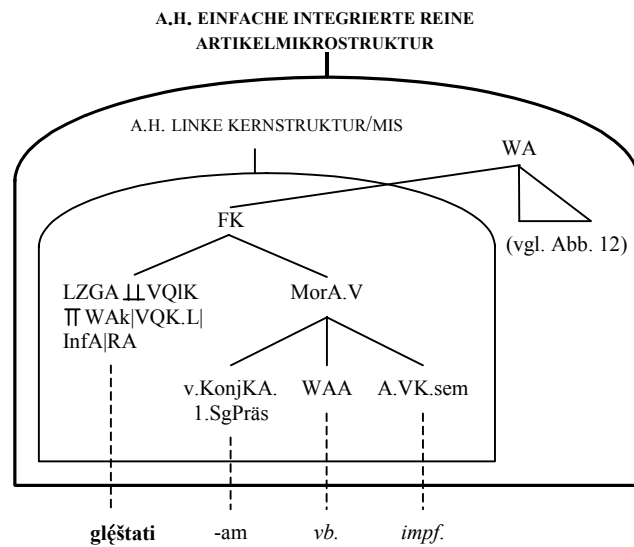


Abb. 11: Einfach kommentierter Strukturgraph zur abstrakten (und isomorphen konkreten) linken Kernstruktur als Teil der abstrakten hierarchischen integrierten reinen Artikelmikrostruktur; Abkürzungen: VQIK = Vokalqualitätskennzeichen

nung; InfA = Infinitivangabe; v.KonjKA = verdichtete Konjugationsklassenangabe; A.VK.sem = Angabe der semantischen Verbklasse

Das erweiterte Artikelstrukturschema, das zu dem allgemein Mikrostrukturbild in Abb. 10 gehört, hat die folgende Form: WA_(Mis) FK<SK[SSk<SSK].

Das Datenangebot im Formkommentar in den Verbartikeln in Plet ist relativ zum Typ des verbalen Lemmazeichens weitgehend einheitlich, so dass die in Abb. 11 dargestellten hierarchischen linken Kernstrukturen (nämlich die abstrakte und die konkrete) zu vielen entsprechenden linken Kernstrukturen isomorph sind. Dies bedeutet auch, dass die Formeigenschaften von Verben in den Verbartikeln meistens einheitlich lexikographisch bearbeitet wurden.

Die Kommentarstrukturen, die zum Formkommentar von wa₁ gehören, also die linken Kernstrukturen, sind als Teilstrukturen der abstrakten hierarchischen und isomorphen konkreten einfachen integrierten reinen Artikelmikrostrukturen in Abb. 11 dargestellt.

Im Folgenden betrachten wir einige Auffälligkeiten im semantischen Kommentar von wa₁. Im ersten semantischen Subkommentar wird das zu *gléštati* gehörige reflexive Verb mit der verdichteten Verbangabe „g. se“ genannt. Im zweiten semantischen Subkommentar finden sich zwei slowenische Kompetenzbeispielangaben, zu denen es keine deutschen Entsprechungen gibt; dies ist kein Einzelfall, und hier zeigt sich, dass Plet kein konsequent bearbeitetes bilinguales Wörterbuch ist, sondern, dass das Interesse in erster Linie der slowenischen Sprache gilt (vgl. 3). Dass das reflexive Verb im ersten semantischen Subkommentar präsentiert wird, ist wohl seiner Bedeutung geschuldet. Es stellt sich hier aber auch eine grundsätzliche Frage, nämlich die nach dem Verhältnis von Makrostruktur und Mikrostruktur in zweisprachigen Wörterbüchern, auf die wir zurückkommen. Die abstrakte (und die konkrete isomorphe) hierarchische rechte Kernstruktur sind als Teilstrukturen der einfachen integrierten reinen Artikelmikrostrukturen, die zu wa₁ in Abb. 1 gehören, in Abb. 12 dargestellt.

Wie angekündigt, kommen wir im Folgenden auf das Problem des Verhältnisses von Makrostruktur und Mikrostruktur in zweisprachigen Wörterbüchern zurück; dies wird am Beispiel von wa₃₆ in Abb. 1 mit dem Lemma **glista** diskutiert. Dabei wird auf den neueren Begriff der Makrostruktur zurückgegriffen, wie er in Wiegand (2008: 230ff) eingeführt wurde. Danach muss zwischen der alphabetischen Makrostruktur und der bzw. den alphabetischen Zugriffsstruktur(en) unterschieden werden. Für die Makrostruktur ist kriterial, dass sie die typspezifische makrostrukturelle Abdeckung eines Wörterbuchs und damit einen genau angebbaren Teil der lexikographischen Abdeckung vollständig anzeigt. Eine Makrostruktur kann mit der äußeren alphabetischen Hauptzugriffsstruktur identisch sein. In diesem Fall liegt eine monoalphabetische Makrostruktur vor. Eine Makrostruktur kann aber auch eine geordnete Menge von solchen äußeren alphabetischen Zugriffsstrukturen sein, deren äußere Zugriffstextelemente zur makrostrukturellen Abdeckung einen Beitrag leisten. Dann liegt eine polyalphabetische Makrostruktur vor. Im Meta-

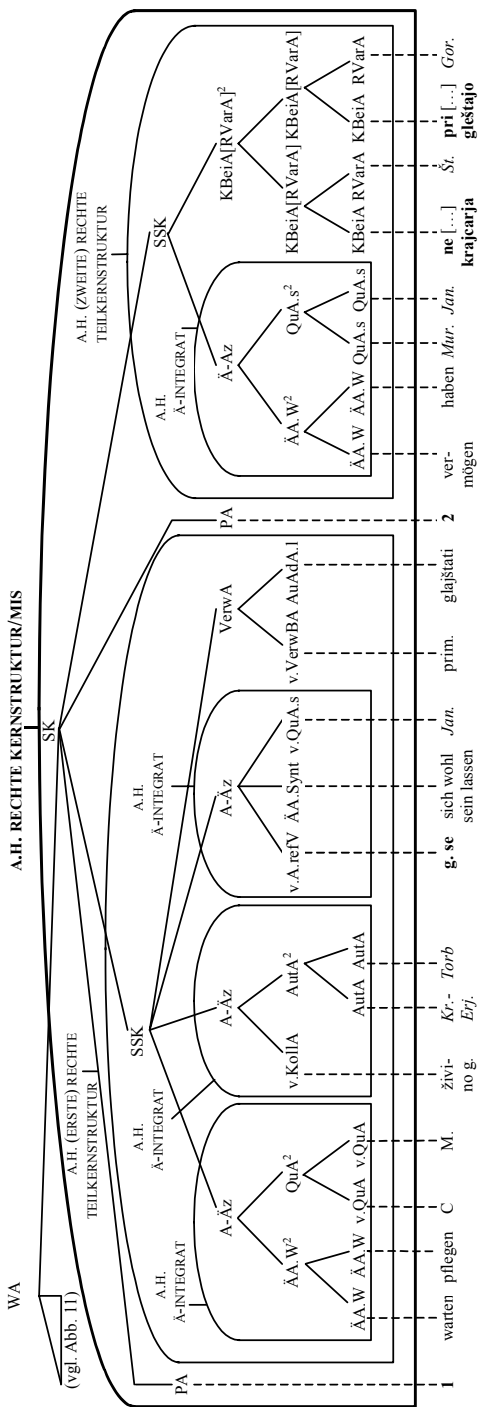


Abb. 12: *Einfach kommentierter Strukturgraphi zur abstrakten (und isomorphen) hierarchischen rechten Kernstruktur als Teilstrukturen der abstrakten (und konkreten isomorphen) hierarchischen einfachen integrierten reinen Artikel mikrostrukturen, die zu wa₁ in Abb. 1 gehören; Abkürzungen: PA = Polysemieangabe; v.Kolla = verdichtete Kollokationsangabe; RVarA = Raumvarietät-angabe; AuAdA.1 = lemmatische Außenadressenangabe; KBeiA[RVarA]² = homosegmentäre um ein Raumvarietätenangabe rechnerweiterte Kompetenzbeispielangabe*

text von Plet erhält man keine Hinweise, wie das im Folgenden näher zu erläuternde Problem zu lösen ist. In wa₃₆ finden sich mehrere verdichtete Mehrwortangaben, anhand derer slowenische Mehrwortbenennungen erschließbar sind, zu denen es im Deutschen äquivalente Komposita gibt, die mit Wortäquivalentangaben genannt werden. Beispielsweise ist „**navadna g.**“ der Spulwurm“ eine lexikographische Bearbeitungseinheit: Die anhand der verdichteten Mehrwortangabe „**navadna g.**“ erschließbare Mehrwortbenennung *navadna glísta* wird mit der linkerweiterten Wortäquivalentangabe „der Spulwurm“ lexikographisch bearbeitet. Das bedeutet, dass „**navadna g.**“ zur lexikographischen Abdeckung einen Beitrag leistet. Entsprechendes gilt für die fünf alphabetisch folgenden verdichteten Mehrwortangaben, nämlich „**otročja g.**“, „**konjska g.**“, „**ovčja g.**“, „**svinjska g.**“ und „**ribja g.**“. Es muss daher die Frage beantwortet werden: Sind die insgesamt sechs verdichteten Mehrwortangaben als Sublemmata zu betrachten? Wenn ja, dann wären sie Elemente der Makrostruktur und leisteten einen Beitrag zur makrostrukturellen Abdeckung; dies hätte weiterhin zur Folge, dass wa₃₆ ein Artikelnest wäre; **navadna g.** wäre dann das erste von sechs Nestlemmata, mit denen die alphabetische Ordnung durchbrochen wird, denn „**glístast**“, das Lemma von wa₃₇, kommt alphabetisch vor „**navadna g.**“. Insgesamt hätte man dann eine nestalphabetische Makrostruktur, die mit einer nestalphabetischen Hauptzugriffsstruktur identisch ist. Gegen diese Lösung spricht vor allem das Angabentupel „1), 2)“ in wa₃₆.

Eine zweite Lösung lässt sich wie folgt erläutern: Plet weist eine glattalphabetische Makrostruktur auf, die mit einer glattalphabetischen Hauptzugriffsstruktur identisch ist. Die alphabetischen geordneten verdichteten Angaben der Mehrwortbenennungen sind nichtlemmatische Angaben und bilden zusammen mit dem kontinuierlichen Lemma **glísta**, zu dem die oben erweiterte Lemmazeichengestaltangabe „**glísta**“ gehört, eine alphabetische innere Schnellzugriffsstruktur im Artikel wa₃₆. Die makrostrukturelle Abdeckung wird dann auf der Basis der glattalphabetisch geordneten Lemmata berechnet. Um die lexikographische Abdeckung insgesamt zu berechnen, müssen dann alle zur Mikrostruktur gehörigen Angaben, mit denen artikelintern lexikalisierte slowenische Ausdrücke genannt werden und die auch an einer alphabetischen Stelle in die Makrostruktur eingeordnet sein könnten, wie z.B. *gléštati se* (vgl. wa₁) in der G-Lemmastrecke und *navadna glísta* (vgl. wa₃₆) in der N-Lemmastrecke, berücksichtigt werden. In zweisprachigen Wörterbüchern heißen Angaben, mit denen lexikalisierte Ausdrücke der lexikographischen Ausgangssprache genannt werden, die bei einem anderen Lemmasatzprogramm als Elemente der Makrostruktur auftreten können und dann einen Beitrag zu makrostrukturellen Abdeckung leisten würden, *makrostrukturelle Inseln* (vgl. Meyer und Wiegand 2000: 140). Diese werden bei der Berechnung der lexikographischen Abdeckung berücksichtigt.

Damit kennen wir die für Plet charakteristischen textuellen Strukturen und wesentliche Teile des Datenangebots. Als zweisprachiges Wörterbuch ist Plet für Slowenen in Produktions- und Übersetzungssituationen und für Deut-

sche in Rezeptionssituationen nur sehr bedingt brauchbar. Es gibt verschiedene Perspektiven für Plet. Eine wurde mit einem Minimalprogramm genannt: Es ist die Weiterentwicklung zu einem aktiven zweisprachigen Wörterbuch für Slowenen.

4.2 PonsN und PonsS

Wir betrachten zunächst PonsN; gegeben seien die Artikel in Abb. 13.

wa₃₈:

größte(r, s) [ˈgrø:stə, -tə, -təs] *adj superl von groß*

wa₃₉:

groß|ziehen *irr vt* ❶ (Kind) vzgajati, vzgoiti *pf* ❷ (Tier) rediti, vzrediti *pf*

wa₄₀:

grub [gru:p] 3. *prät von graben*

wa₄₁:

grün [gry:n] *adj* ❶ (Farbe, a. POL) zelen; G~er Punkt (ÖKOL) zelena pika; jmdn ~ und blau schlagen (*ugs fig*) močno preteptsti koga; sich ~ und blau ärgern (*ugs fig*) pomodreti od jeze; ~es Licht geben (*ugs fig*) dati zeleno luč; auf (k) ~en Zweig kommen (*fig*) (ne) doseči soglasja; die ~e Welle zeleni val ❷ (abw: unreif, unerfahren) nezrel, zelen (*ekspr.*); ein ~er Junge (*fig*) zeleneč

wa₄₂:

groß|schreiben^{RR} *irr vt* pisati z veliko pisavo

wa₄₃:

Grubenarbeiter(in) *m(f)* (BERGB) rudar(ka) *m(f)*, knap *mf* (*pog.*)

wa₄₄:

grummeln [ˈgruməlŋ] *vi* ❶ (Mensch) mrmrati, zamrmrati *pf* ❷ (Donner) grmeti, zagrmeti *pf*

wa₄₅:

ansteckend *adj* nalezljiv

wa₄₆:

anspruchsvoll *adj* zahteven, težaven

Abb. 13: Wörterbuchartikel wa₃₈ bis wa₄₆ aus PonsN

Das Lemmaansatzprogramm in PonsN weist eine ganze Reihe von angemessenen Regelungen auf; dazu gehören u.a. der Ansatz von unregelmäßigen Superlativformen in Adjektivartikeln (vgl. wa₃₈ in Abb. 13), der Ansatz von Präteritumstammformen in Verbartikeln (vgl. wa₄₀ in Abb. 13), die Identifizierung der Partikelverben durch „|“, eine Angabe der Trennungsfuge (vgl. wa₃₉ und wa₄₂ in Abb. 13), der platzsparende Ansatz von verdichteten Lemmata, wenn es eine movierte Form des Lemmazeichens gibt, wie „Grubenarbeiter(in)“ in wa₄₃ in Abb. 13 und die Kenntlichmachung durch eine hochgestellte Neuschreibungsidentifizierungsangabe „^{RR}“ in „groß|schreiben^{RR}“ in wa₄₂ in Abb. 13.

Wie in Plet weisen die Wörterbuchartikel in PonsN einfache integrierte oder einfache gemischt-integrierte Artikelmikrostrukturen auf. Typisch für einen Verbartikel in PonsN mit einem als zweifach polysem interpretierten Verb sind die in Abb. 14 dargestellte abstrakte (und isomorph konkrete) hierarchische einfache gemischt-integrierte Mikrostruktur.

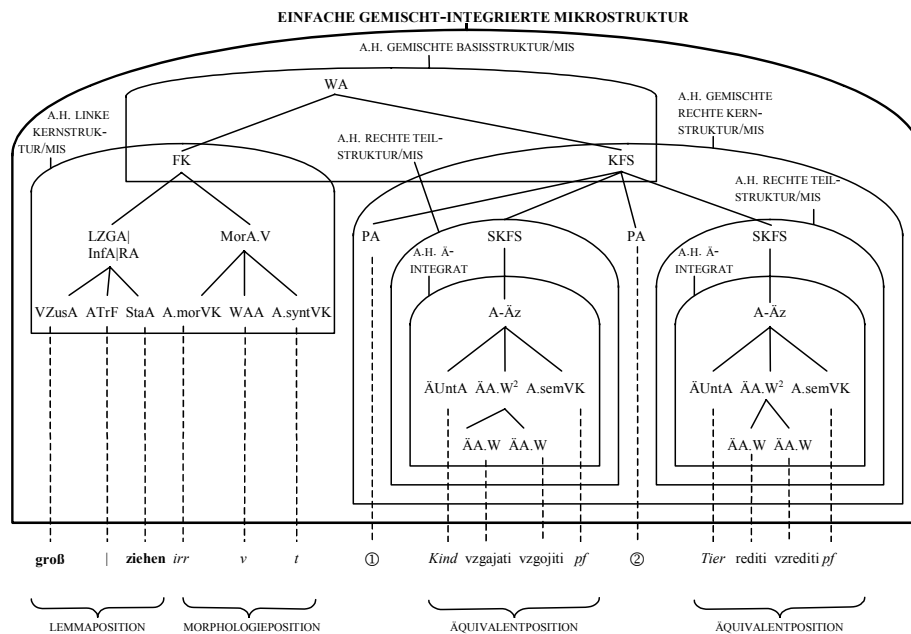


Abb. 14: *Erweitert kommentierter Strukturgraph* zur abstrakten (und isomorphen konkreten) hierarchischen einfachen gemischt-integrierten Artikelmikrostruktur, die wa_{39} in Abb. 13 aufweist; *Abkürzungen:* VZusA = Verbzusatzangabe; ATrF = Angabe der Trennungsfuge (zugleich Partikelverbidentifizierungsangabe); StaA = Stammangabe; A.morVK = Angabe der morphologischen Verbklasse; A.syntVK = Angabe der syntaktischen Verbklasse; ÄUntA = Äquivalentunterscheidungsangabe; SKFS = Subkommentar zur Form und Semantik; A.semVK = Angabe der semantischen Verbklasse; WAA = Wortangabe; ÄÄ.W² = homosegmentäre Wortäquivalentangabe, die aus zwei Wortäquivalentangaben besteht

Es folgen einige Bemerkungen zu wa_{39} in Abb. 13. Das erweiterte Artikelstrukturschema lässt sich aus Abb. 14 ablesen; es hat die Form $WA:FK < KFA[SKFS < SKFS]$. Dieses Strukturschema gehört zu zahlreichen Verbartikeln zu Verben, die als zweifach polysem interpretiert sind, z.B. zu wa_{44} in Abb. 13 und zu den Artikeln zu den Verben *anspannen*, *anspornen*, *abblasen*, *abfragen* und *abrichten* in PonsN. Bei als polysem interpretierten Ausdrücken finden sich meistens kursiv gesetzte Äquivalentunterscheidungsangaben in runden Klammern. In wa_{39} sind dies „Kind“ und „Tier“. Werden Äquivalentunterscheidungsangaben angesetzt, dann werden in dem zugehörigen zweisprachigen Wörterbuchartikel im Wörterbuchgegenstandsbereich gegebene *wenn-dann*-Beziehungen lexikographisch vertextet (vgl. z.B. Wiegand 2009). Beispielsweise gilt für wa_{39} die Formulierung: Wenn *großziehen* (in einer seiner Formen) sich auf Kinder bezieht, dann sind *vzgjati* und *vzgojiti* (in einer entsprechenden Form)

Wortäquivalente zu *großziehen*. In den Benutzungshinweisen wird dies allerdings nicht hinreichend erklärt. Die Angabenadressierungsbeziehungen in den meisten Artikeln sind leicht zu rekonstruieren, so dass die Informationsgewinnung für den Benutzer-in-actu ohne Schwierigkeiten möglich ist. In den meisten Artikeln ist die Adressierungskonstellation der partiell lemmatischen Adressierung gegeben; zu dieser Adressierungskonstellation gehört eine partiell lemmatisch fokussierte Adressierungsstruktur in konkreter und abstrakter Ausprägung (vgl. u.a. Wiegand 2009). Die artikelinterne konkrete partiell lemmatisch fokussierte Adressierungsstruktur, die zu wa₃₉ gehört, ist in Abb. 15 dargestellt.

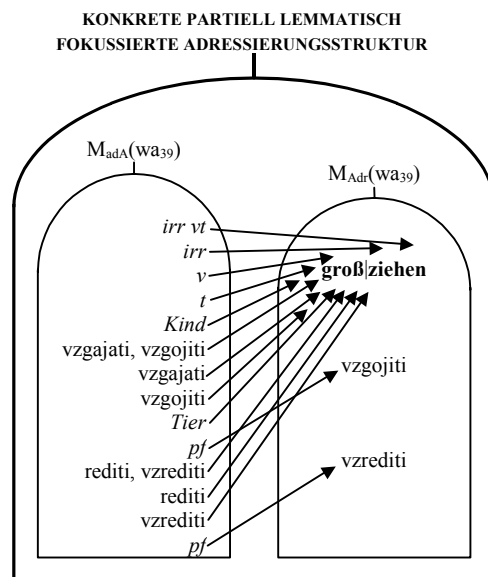


Abb. 15: Kommentiertes Pfeildiagramm zur konkreten partiell lemmatisch fokussierten Artikeladressierungsstruktur, die zu wa₃₉ in Abb. 13 gehört; *Abkürzungen:* $M_{adA}(wa_{39})$ = Menge (M) der adressierten Angaben (adA) von wa₃₉; $M_{Adr}(wa_{39})$ = Menge der Adressen (Adr) von wa₃₉.

Neben der häufig auftretenden Adressierungskonstellation der partiell lemmatischen Adressierung tritt relativ selten die Adressierungskonstellation der vollständig lemmatischen Adressierung auf. Öfters ist dies in Adjektivartikeln der Fall, wenn das deutsche Adjektiv als monosem interpretiert ist, wie z.B. in wa₄₅.

Ausspracheangaben werden nur bei Simplizia gegeben (vgl. z.B. wa₄₁ und wa₄₄). Das führt dazu, dass bei den Partikelverben der Benutzer nicht erfährt, dass (in Unterschied zu den Präfixverben) der Wortakzent auf dem Verbzusatz und damit auf der ersten Silbe liegt. Dies ist ein systematischer lexikographischer Fehler, der in einer nächsten Bearbeitung unbedingt beseitigt werden

sollte und in zahlreichen neueren zweisprachigen Wörterbüchern nicht auftritt. Die einfachste Lösung ist die Hinzufügung eines freistehenden Akzentes vor den Verbzusatz als Wortakzentangabe (WakA) z.B.:

'groß | ziehen, 'groß | schreiben, 'ab | fahren

Die präzendentive Angabestruktur einer linkerweiterten Lemmazeichengestaltungangabe, mit der ein Partikelverb genannt wird, hat dann folgende Form:

WakA < Vzusa < AtrF < StA (mit „<“ für *geht voraus*).

Will man zusätzlich die Vokalquantitätsangabe des Vokals im Verbzusatz berücksichtigen, kann man auch unten erweiternde funktionale Angabezusätze wählen, zum Beispiel einen Unterpunkt für Kürze und einen Unterstrich für Länge (**groß | schreiben, ab | fahren**). Ergänzt sei, dass es für den slowenischen Benutzer auch hilfreich sein dürfte, wenn alle deutschen Lemmazeichengestaltungangaben, mit denen mehrsilbige Lemmazeichen genannt werden, durch Silbentrennungsangaben in Form eines Mittenpunktes binnenerweitert wären. Die optimale Form wäre dann z.B.:

'groß | zie•hen oder 'ab | fah•ren (oder: groß | zie•hen, ab | fah•ren).

Die Angabe der Trennungsfuge wäre dann zugleich eine Silbentrennungsangabe und damit bifunktional.

Im Folgenden betrachten wir wa₄₁, den Artikel zu *grün*, als Beispiel für einen mikrostrukturell reichhaltigeren Artikel. Solche Artikel weisen vor allem dt. Phrasem-, Kollokations- und Kompetenzbeispielangaben auf. Allerdings sind diese nicht als solche ausreichend identifiziert und daher für Slowenen häufig nicht unterscheidbar. Auch dies muss als ein systematischer Fehler angesehen werden. Da keine Formangaben auftreten, weist der Artikel mit dem kontinuierlichen Lemma **grün** keinen Kommentar zur Form und Semantik als Zweitkommentar auf, sondern einen semantischen Kommentar, dessen mikrostrukturelle Kommentarstruktur, die hierarchische rechte Kernstruktur, in abstrakter und konkreter Ausprägung in Abb. 16 und Abb. 17 dargestellt ist.

Auf den ersten Blick scheint es so, als sei die Struktur des Zweitkommentars von wa₄₁ in Abb. 13 hochgradig komplex und ziemlich heterogen; das ist aber nicht der Fall; vielmehr ist die textuelle Struktur relativ übersichtlich. Der Zweitkommentar besteht aus zwei semantischen Subkommentaren (SSK). In beiden SSK kommt zuerst eine Angabe zur Äquivalenz (A-Äz). Mit dieser nichtelementaren Angabe werden die äquivalentrelevante Angaben genannt, die es dem Benutzer ermöglichen die semantisch-pragmatische Äquivalenz (i.S.v. Wiegand 2005b) zu erfassen, die relativ zum deutschen Lemmazeichen *grün* und seinen slowenischen Wortäquivalenten gegeben ist. In beiden SSK folgen auf die Angabe zur Äquivalenz Angaben von Äquivalentpaaren, die meistens binnenerweitert sind; im ersten SSK sind es sechs Angaben dieses Typs, im zweiten SSK ist es eine. Was man in den Strukturdarstellungen erken-

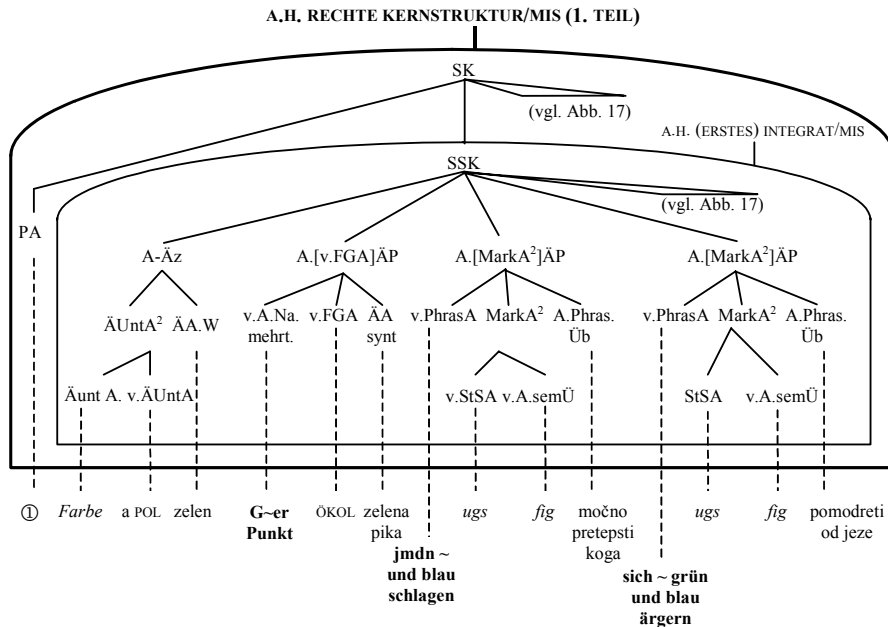


Abb. 16: Nicht vollständig ausgeführter einfach kommentierter Strukturgraph (1. Teil) zur abstrakten (und isomorphen konkreten) hierarchischen rechten Kernstruktur, die der semantische Kommentar von wa₄₁ in Abb. 13 als Teilstruktur der Artikelmikrostrukturen aufweist; *Abkürzungen:* A.[v.FGA]ÄP = um eine verdichtete Fachgebietsangabe binnenerweiterte Angabe eines Äquivalentpaares; A.[MarkA²]ÄP = um eine homosegmentäre Markierungsangabe binnenerweiterte Angabe eines Äquivalentpaares; ÄUntA² = homosegmentäre Äquivalentunterscheidungsangabe, die aus zwei ÄUntA besteht; v.A.Na.mehrt = verdichtete Angabe eines mehrteiligen Namens; v.FGA = verdichtete Fachgebietsangabe; v.PhrasA = verdichtete Phrasenangabe; v.StSA = verdichtete Stilschichtangabe; MarkA = homosegmentäre Markierungsangabe, die aus zwei Markierungsangaben besteht; v.A-semÜ = verdichtete Angabe zum semantischen Übergang; A.PhrasÜb = Angabe der Phrasemübersetzung

nen kann, nämlich von welcher Art die genannten Äquivalentpaare sind, das lässt sich anhand des Artikeltextes gerade nicht erkennen. So weiß ein slowenischer Benutzer keineswegs sicher, ob beispielsweise mit „jmdn ~ und blau schlagen“ ein deutsches Phrasem genannt wird oder ein Kompetenzbeispiel; denn an der zugehörigen slowenischen Angabe „močno pretepsti koga“ kann er das gerade nicht erkennen, da diese eine nichtidiomatische Übersetzung ist. Hinzu kommt, dass zu viele dieser Übersetzungen schlecht oder eindeutig falsch sind, wie z.B. die mit der Angabe der Phrasemübersetzung genannte slowenische Übersetzung „(ne) doseči soglasja“, die auf Deutsch soviel wie *(keine Zustimmung bekommen)* bedeutet. Die mit der verdichteten Phrasemangabe auf (k)~en Zweig kommen genannten Phraseme bedeutet aber soviel wie

(nicht, zu wenig) erfolgreich sein. Im Zweitkommentar von wa₄₁ werden gleich zwei grundsätzliche lexikographische Fehler gemacht. Der erste besteht darin, dass Angaben, die zu unterschiedlichen Angabetypen gehören, ohne dass der jeweilige Angabetyp mit einem der üblichen Identifizierungsmethoden, nämlich typographische, nichttypographische oder positionelle Angabenidentifizierung (vgl. Wiegand 2005a: 298ff) angewandt wurden, einfach beliebig aufeinanderfolgen. Der zweite Fehler besteht darin, dass Phrasemangaben innerhalb der semantischen Subkommentare lexikographisch bearbeitet werden. Dabei spielt es keine Rolle, ob mit den Phrasemangaben teil- oder vollidiomatische Phraseme genannt werden. Denn oft genug werden vollidiomatische Phraseme aufgrund muttersprachlicher Kompetenzen fälschlich remotiviert und dann vom Lexikographen willkürlich der Bedeutung einer Phrasemkomponente zugeordnet. Das Phrasem *jemandem grün und blau schlagen* hat nichts mit der Bedeutung von *grün* (auf Farbe bezogen) zu tun; es bedeutet vielmehr einfach soviel wie *jemanden gehörig verprügeln*. Eine verdichtete Phrasemangabe wie z.B. *jmdn ~ und blau schlagen* gehört daher nicht in den ersten semantischen Subkommentar, weil er dort im Skopus der Angaben „(Farbe [...]) zelen“ steht. Die hier anhand von wa₄₁ geäußerte Kritik gilt für alle Wörterbuchartikel in PonsN, in denen deutsche Phraseme lexikographisch bearbeitet sind.

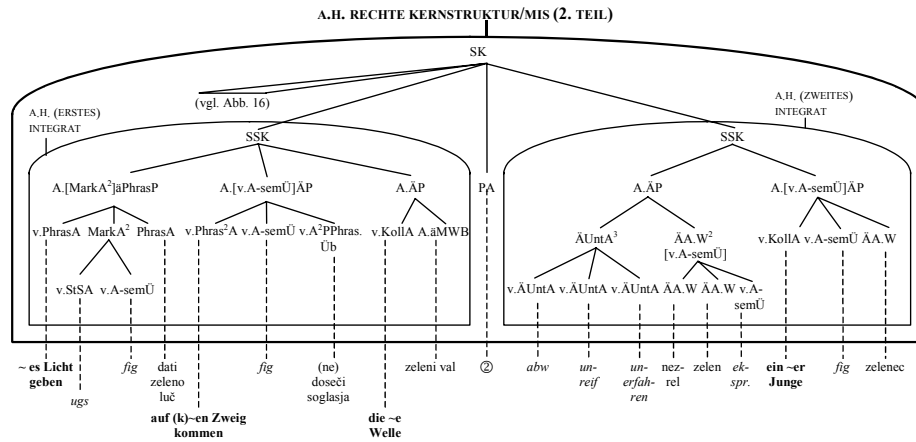


Abb. 17: Nicht vollständig ausgeführter einfach kommentierter Strukturgraph (2. Teil) zur abstrakten (und isomorphen konkreten) hierarchischen rechten Kernstruktur, die der semantische Kommentar von wa₄₁ in Abb. 13 als Teilstruktur der Artikelmikrostruktur aufweist; *Abkürzungen:* A.[MarkA²]äPhrasP = um eine homosegmentäre Markierungsangabe binnenerweiterte Angabe eines äquivalenten Phrasempaars; A.[A-semÜ]ÄP = um eine Angabe zum semantischen Übergang binnenerweiterte Angabe eines Äquivalentpaares; v.A.PhrasÜb = verdichtete Angabe der Phrasemübersetzung; v.KollA = verdichtete Kollokationsangabe; A.ämWB = Angabe der äquivalenten Mehrwortbenennung; ÄA.W²[v.A-semÜ] = um eine verdichtete Angabe zum semantischen Übergang rechtserweiterte homosegmentäre Wortäquivalentangabe

Es gibt verschiedene Artikeltypen (i.S.v. Wiegand 2003), die geeignet sind, um in den zugehörigen Wörterbuchartikeln die Phraseme angemessen lexikographisch zu bearbeiten. Der einfachste Artikeltyp, der dazu geeignet ist, ist der Typ des rechtserweiterten Wörterbuchartikels mit Kommentar zur Form und Semantik. Der Postkommentar ist dabei ein Postkommentar zur Phraseologie (PostK:Phras). Das einfache mikrostrukturelle Artikelstrukturschema für diesen Wörterbuchtyp lautet: WA:(MiS) FK < SK < PostK:Phras. Strukturiert man wa₄₁ so um, dass der umstrukturierte Artikel wa'₄₁ zu diesem Wörterbuchartikeltyp gehört und hält dabei das artikelzugehörige Datenangebot soweit konstant, dass lediglich die falsche Angabe der Phrasemübersetzung geändert und eine Phrasemidentifizierungsangabe „♦“ hinzugefügt wird, dann ergibt sich der folgende Artikel wa'₄₁:

- grün** [gry:n] *adj* ❶ (*Farbe*, a. POL) zelen;
G~er Punkt (ÖKOL) zelena pika; ❷ (*abw:*
unreif, unerfahren) nezrel, zelen (*ekspr.*);
ein ~er Junge (*fig*) zelenec
 ♦ **jmdn ~ und blau schlagen** (*ugs fig*)
 močno pretepsti koga; **sich ~ und blau**
ärgern (*ugs fig*) pomodreti od jeze; **~es**
Licht geben (*ugs fig*) dati zeleno luč; **auf**
(k)~en Zweig kommen (*fig*) (ne) priti na
 zeleno vejo; **die ~e Welle** zeleni val

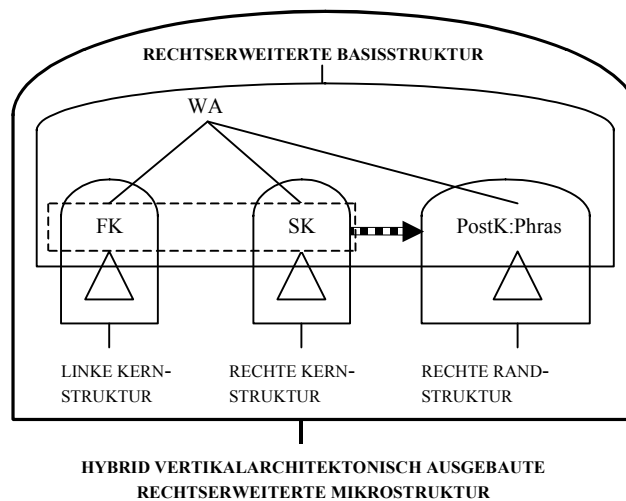


Abb. 18: Allgemeines Mikrostruktur- und Mikroarchitekturmodell für hybrid vertikalarchitektonisch ausgebauter rechtserweiterter integrierter Mikrostrukturen mit als zweifach polysem interpretiertem Lemmzeichen; Abkürzungen: PostK:Phras = Postkommentar zur Phraseologie: „x ♦ y“ bedeutet soviel wie *x ist oberhalb von y*

Durch die Umgruppierung erhält wa₄₁ eine vertikale hybride Mikroarchitektur: Auf einer Zweiermenge, deren Elemente die textuelle Folge des Formkommentars und des semantischen Kommentars sowie der Postkommentar sind, wird eine textarchitektonische *oberhalb*-Relation definiert. Das allgemeine Mikrostruktur- und Mikroarchitekturbild für Artikel mit hybrid vertikalarchitektonisch ausgebauter rechtserweiterter integrierter Mikrostruktur findet sich in Abb. 18.

Damit haben wir zwar keinen Überblick, aber doch einen Einblick in den Bereich der Wörterbuchform und der lexikographischen Bearbeitung des Wörterbuchgegenstands in PonsN, so dass wir im Folgenden noch einen kurzen Blick auf PonsS werfen können. Gegeben seien die Wörterbuchartikel in Abb. 19.

wa₄₇:

zelèn *adj* <-a, -o> ❶ (po barvi) grün; ~i pas (PROM) Grünstreifen *m* ❷ (ekspr: mlad) grün, unreif ❸ (ekspr) **postati** ~ **od zavisti** vor Neid erblassen; **dati komu** ~o luč jmdm grünes Licht geben

wa₄₈:

kráv|a *f* <-e, -i, -e> ❶ (ZOO) Kuh *f*; **morska** ~a Seekuh *f*; **molsti** ~e Kühe melken ❷ (fig): **molzna** ~a Melkkuh *f* (ugs); **saj nisva skupaj** ~ **pasla** wo haben wir denn schon zusammen Schweine gehütet? ❸ (ekspr: slabš) Kuh *f*; **ti** ~a **neumna!** du blöde Kuh!

wa₄₉:

glas|ováti <-ujem; glasoval> *impf vt* (ab-)stimmen; ~ovati **za/proti** dafür-/dagegenstimmen; ~ovati **o uvedbi reform** über die Einführung von Reformen abstimmen

wa₅₀:

episkopál|en *adj* <-na, -no> (REL) bischöflich, episkopal; ~na **cerkev** Episkopalkirche *f*

wa₅₁:

kráv|ji *adj* <-a, -e>: ~e **mleko** Kuhmilch *f*; ~i **zvonec** Kuhglocke *f*

Abb. 19: Wörterbuchartikel wa₄₇–wa₅₁ aus PonsS

Im PonsS finden sich die gleichen Typen von textuellen Artikelstrukturen wie in PonsN, also einfache integrierte und einfache gemischt-integrierte Artikelmikrostrukturen. Besonders bedingt durch den anderen Sprachtyp, zu dem die lexikographische Ausgangssprache gehört, treten jedoch sowohl in den Formals auch in den Zweitkommentaren z.T. Angaben auf, die zu anderen Angabetypen gehören.

Wir betrachten zunächst wa₄₇, und zwar auch vergleichend mit wa₄₁ und mit wa₅₀. In wa₄₁ heißt eine ausgangssprachliche Phrasemangabe „~es **Licht geben**“. In wa₄₇ heißt die zielsprachliche Phrasemangabe für das gleiche Phrasem „jmdm grünes Licht geben“. In wa₄₁ wird auf die Angabe der Objektklasse verzichtet, was zumindest als benutzerunfreundlich zu gelten hat. In wa₄₇ ist die Objektklasse „jmdm“ falsch. Sie muss lauten „jmdm/für etwas“, da man nicht nur einer Person, sondern auch für einen Sachverhalt grünes Licht geben kann (vgl. *Er gab für die Freigabe der Schmuggelware grünes Licht*). In wa₄₁ steht „~es **Licht geben**“, im ersten semantischen Subkommentar, und damit ist das

Phrasem der Bedeutung von *grün* im Sinne der Farbbezeichnung zugeordnet. In wa₄₇ ist das aber nicht der Fall, denn die Phrasemangabe steht hier nicht im ersten semantischen Subkommentar, sondern im dritten. Dieser ist aber kein regulärer semantischer Subkommentar, weil es in ihm keine Äquivalentangabe gibt, die an die Lemmazeichengestaltangabe **zelèn** adressiert ist. Dies bedeutet zugleich, dass „**3**“ in wa₄₇ keine Polysemieangabe ist. Entsprechend werden im zuständigen Metatext von PonsS die arabischen Zahlen auch nur als Gliederungszeichen eingeführt, die den Artikel (so wie es ad hoc gerade schlecht und recht geht) irgendwie gliedern. Entsprechend könnte man den textuellen Teil, der in wa₄₇ auf „**3**“ folgt als Postkommentar auffassen, so dass wa₄₇ ein rechts-erweiterter Artikel wäre.

In anderen Artikeln wird aber wieder anders verfahren: Die Phraseme stehen dann nicht zusammen in einem Postkommentar zur Phraseologie, sondern sind auf die semantischen Subkommentare verteilt, wenn das Lemmazeichen als polysem interpretiert ist. Das bedeutet: In PonsS findet sich bei gleichartigen Eigenschaften des Wörterbuchgegenstandes eine unterschiedliche Artikelform und damit eine heterogene Strukturgebung. Dies erschwert dem Benutzer die artikelinterne Orientierung ungemein und erlaubt auch keine einheitlichen und präzisen Benutzungshinweise; entsprechend vage sind daher auch die entsprechenden Metatexte in PonsS.

Im Folgenden betrachten wird die Formkommentare von wa₄₇ und wa₅₀, also die Formkommentare von zwei Adjektivartikeln. Die Lemmazeichengestaltangabe „**episkopál|en**“ ist durch einen senkrechten Strich binnenerweitert; mit diesem wird eine Identifizierungsangabe für das invariante Wortsegment realisiert, so dass der Benutzer erkennen kann, wo das invariante Wortsegment endet und er die Flexive anhängen muss. Identifizierungsangaben für das invariante Wortsegment dürfen nicht mit Stammidentifizierungsangaben, die sich in deutschen einsprachigen Wörterbüchern finden, verwechselt werden: Nicht jedes invariante Wortsegment ist im Slowenischen ein Stamm. In wa₄₇ ist die Lemmazeichengestaltangabe „**zelèn**“ nicht durch einen senkrechten Strich binnenerweitert. Das bedeutet, dass hier die Flexive an *zelèn* angehängt werden. Für den dt. Benutzer ist diese Angabenerweiterung, die sich auch in anderen slowenisch-deutschen Wörterbüchern findet, sehr nützlich. Hier drängt sich allerdings sofort die Frage auf, ob es für den dt. Benutzer nicht hilfreich wäre, wenn er auch im dt.–slov. Wörterbuch bei den Wortäquivalentangaben Identifizierungsangaben für das invariable Wortsegment anträfe.

Im Folgenden analysieren wir den Substantivartikel wa₄₈ in Abb. 19. Das erweiterte allgemeine Mikrostrukturbild von wa₄₈ findet sich in Abb. 20.

Erweiterte allgemeine Mikrostrukturbilder sind Mikrostrukturbilder, die um Elemente der inneren Zugriffsstruktur sowie um die zugehörigen Klassen erweitert sind. Sie optimieren die Übersicht über die Artikelmikrostruktur; im vorliegenden Fall zeigen sie, dass die inneren Zugriffstextelemente keine Polysemieangaben sind, sondern nichttypographische Mikrostrukturanzeiger vom Typ der Trennzeichen. — In Abb. 21 wird im Folgenden die reine mikrostruk-

turelle Kommentarstruktur des Formkommentars, also die linke Kernstruktur als Teilstruktur der reinen Artikelmikrostruktur, dargestellt.

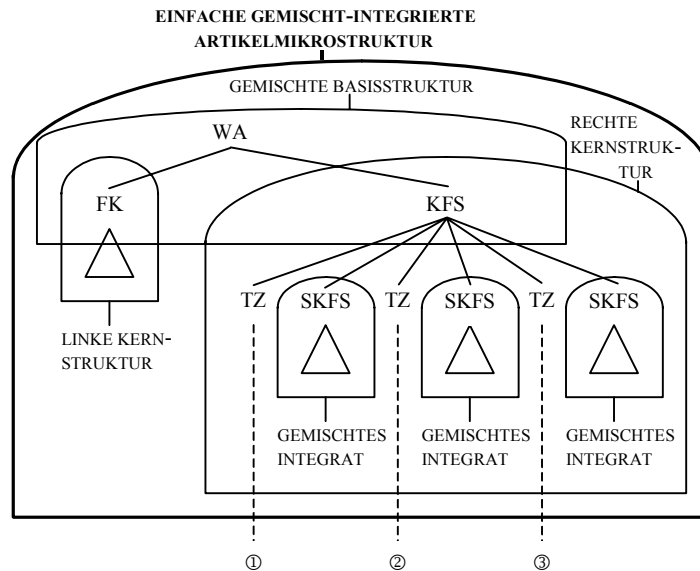


Abb. 20: Erweitertes allgemeines Mikrostrukturbild zu wa₄₈ und allen PonS-Artikeln mit isomorpher Artikelmikrostruktur; Abkürzungen: TZ = Trennzeichen; Darstellungskonvention: „x – – Y“ = x ist ein Element von Y (x ∈ Y) (von unten nach oben gelesen)

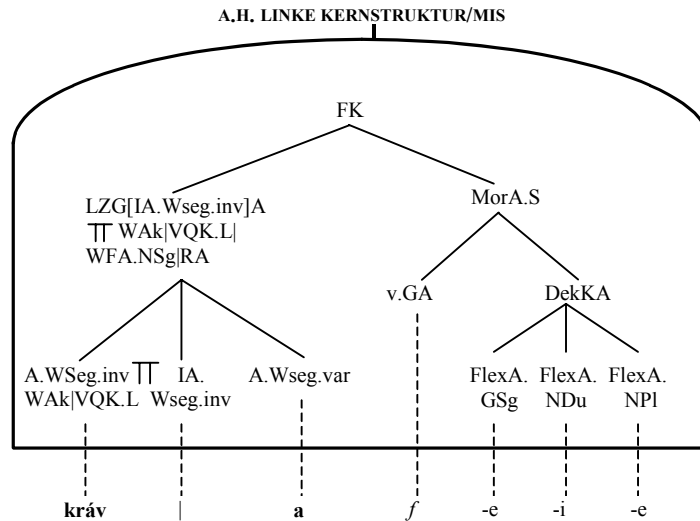


Abb. 21: Einfach kommentierter Strukturgraph zur abstrakten (und isomorphen konkreten) hierarchischen Kernstruktur als Teilstruktur der reinen Artikelmikro-

struktur, die der Formkommentar von wa₄₈ in Abb. 19 aufweist; *Abkürzungen*: IA.W.Seg.inv = Identifizierungsangabe für das invariante Wortsegment; A.Wseg.inv = Angabe des invarianten Wortsegments; A.Wseg.var = Angabe des variablen Wortsegments; LZG[IA.Wseg.inv]A = Lemmazeichengestaltungsangabe, binnenerweitert um eine Identifizierungsangabe für das invariante Wortsegment; FlexA.NDu = Flexivangabe. Nominativ Dual

Die abstrakte hierarchische linke Kernstruktur in Abb. 21 ist zu Hunderten von konkreten linken Kernstrukturen in PonsS isomorph. Die gesamten Formkommentare im PonsS sind bei gleichen Formeigenschaften strukturell gleichwertig gestaltet. Bei gleichem Lemmazeichentyp sind entsprechend die linken Kernstrukturen isomorph. Entsprechendes gilt nicht für die Kommentare zur Form und Semantik. Sie sind strukturell, ohne dass der jeweilige Wörterbuchgegenstand dies erzwingt, ausgesprochen heterogen und die sprachlichen Fehler beim Wörterbuchgegenstand sind zahlreich.

Im Folgenden analysieren wir den Kommentar zur Form und Semantik von wa₄₈ in Abb. 19 hinsichtlich der Artikelform und hinsichtlich des Artikelgegenstands. In Abb. 22 findet sich die Darstellung der abstrakten (und isomorphen konkreten) hierarchischen rechten Kernstruktur, die der Kommentar zur Form und Semantik von wa₄₈ in Abb. 19 aufweist.

Im Folgenden betrachten wir die lexikographische Bearbeitung des Artikelgegenstands in den drei Subkommentaren zur Form und Semantik (SKFS). Im ersten SKFS ist mit „Kuh“ der zoologische Ausdruck für weibliche Tiere gemeint, so wie er z.B. in „Seekuh“ als Zweitkonstituente, aber auch z.B. in *Hirschkuh*, *Elefantenkuh* u.a. vorkommt. Diese Kühe werden aber nicht gemolken! Daher passt die Angabe eines äquivalenten Kollokationspaares „**molsti ~e** Kühe melken“ nicht in den ersten SKFS. Im zweiten SKFS ist „Melkkuh“ veraltet, da es von dem veralteten intransitiven Verb *melken* i.S. von *Milch geben* abgeleitet ist. Das Phrasem: „wo haben wir denn schon zusammen Schweine gehütet?“ kommt im Deutschen selten vor. Die innere Selektion im zweiten SKFS ist daher zumindest äußerst ungeschickt. Das frequenteste Äquivalent, nämlich das alltägliche verwendete Wort *Kuh* (mit dem das weibliche Hausrind nach dem ersten Kalben bezeichnet wird), erscheint erst im dritten SKFS; „**molsti ~e**“ gehört dann in den dritten SKFS. Die innere Selektion ist ein Thema für sich und kann hier nicht angemessen behandelt werden. Es sei nur auf das Folgende aufmerksam gemacht. Zahlreiche dt. Komposita mit innerer Idiomatisierung, wie z.B. *Kuhhandel*, die in PonsN als Lemma angesetzt sind und eine slowenische Mehrwortbenennung als Äquivalent aufweisen, wie z.B. *kravja kupčija*, haben keine Chance im PonsS als Äquivalent zu erscheinen, weil die entsprechende Mehrwortbenennung in dem entsprechenden Adjektivartikel in PonsS nicht gebucht ist. So findet sich in wa₅₁ nur „**~e mleko** Kuhmilch f“ und „**~i zvonec** Kuhglocke f“; es fehlen u.a. *kravja kupčija* und *kravja koža*.

Insgesamt kann festgestellt werden, dass sich positive und negative Aspekte in PonsN und PonsS in etwa die Waage halten. Für gehobene Ansprüche sind beide Wörterbücher ungeeignet. Solange die lexikographische Bearbeitung

innerhalb der Wörterbuchartikel nur im Rahmen von basalen Wörterbuchartikeln mit einfachen integrierten und einfach gemischt-integrierten Mikrostrukturen erfolgt, ist hier kein Fortschritt zu erwarten. Denn falls man solche Wörterbuchartikel mit zusätzlichen Daten anreichert, werden die Artikel des genannten Typs unübersichtlich. Für ein anspruchsvolles zweisprachiges Großwörterbuch sind sie nicht geeignet. Auch das gerade erschienene PonsN/S-KompWb ist keine wirkliche Weiterentwicklung, sondern fasst in den Wörterverzeichnissen im Wesentlichen PonsN und PonsS zu einem Band zusammen.

4.3 DebN und DebS

Wir betrachten zunächst DebN. Nach dem Titel gehört DebN zu den Großwörterbüchern; nach eigenen Angaben weist es ca. 120 500 Lemmata auf. Die monoalphabetische Makrostruktur ist glattalphabetisch und mit der Hauptzugriffsstruktur identisch. Jeder Wörterbuchartikel beginnt mit einer neuen Zeile. DebN ist mithin ausgesprochen zugriffsfreundlich. Ca. 90% aller Wörterbuchartikel sind Kurzartikel: Sie sind nicht länger als zwei Spaltenzeilen. Ca. 90% der Kurzartikel sind Einzeilenartikel. Ein für DebN typischer Artikelstreckenausschnitt findet sich in (1) in Abb. 23.

(1)

Sammelfrucht, *die*, ♣ birni plod
Sammelgebiet, *das*, zbirateljsko področje, področje,
 s katerega kdo zbira
Sammelgefäß, *das*, zbirna/zbiralna posoda
Sammelglas, *das*, zbiralna leča, konkavna leča
Sammelgrab, *das*, skupinski grob
Sammelgut, *das*, zbirno blago
Sammelhefter, *der*, T stroj za vezavo revij
Sammelheizung, *die*, T centralno ogrevanje
Sammelkasse, *die*, osrednja blagajna
Sammelkasten, *der*, T zbiralna posoda
Sammelkatalog, *der*, zbirni katalog
Sammelkonto, *das*, zbirni račun
Sammelladung, *die*, ☞ zbirna pošiljka
Sammellager, *das*, zbirno skladišče
Sammelleidenschaft, *die*, zbiralska strast
Sammelleitung, *die*, zbiralni vod, zbirni vod, kolek-
 tor
Sammellinse, *die*, zbiralna leča, konveksna leča
Sammelliste, *die*, zbiralni seznam, seznam daroval-
 cev
Sammelmappe, *die*, zbiralna mapa
Sammelmulde, *die*, T zbiralna posoda
sammeln (*sich se*) zbirati, zbrati; *Pilze, Holz usw.*:
 nabirati; *Erfahrungen*: pridobivati; *Reichtümer*:
 kopičiti; *all seine Kräfte* ~ zbrati vse moči; *seine*
Gedanken ~ zbrati se

wa73:

Glühbirne, *die*, žarnica

wa74:

vrč *M* (-a ...) *der* Krug (*s pokrovom* Deckelkrug, *za vino* Wein-
 krug, *za vodo* Wasserkrug, *z ročajem* Henkelkrug), *die* Kanne
 (*za vodo* Wasserkanne, *lochat* Bügelkanne, *z ročem* Henkel-
 kanne); *v ~ih/cele* ~e krugweise; *FIG.* ~ **hodi po vodo**,
dokler se ne razbije *der* Krug geht so lange zum
 Brunnen, bis er bricht

wa75:

vrečar [e] *M* (-ja ...) ☞ *das* Beuteltier (*orjaški* Riesen-
 beuteltier); ... ~ Beutel-, -beutler; **jazbec** ~ *der* Beu-
 teldachs; **krt** ~ *der* Wühlbeutler, *der* Beutelmull; **mali**
leteči ~ *der* Zwergflugbeutler; **veliki leteči** ~ *der* Rie-
 senflugbeutler; **medvedek** ~ (*koala*) *der* Koala; **mra-**
vojedi ~ *der* Ameisenbeutler; **obročastorepi** ~ **ple-**
zalec *der* Ringschwanz-Kletterbeutler; **perorepi** ~
der Federschwanzbeutler; ~ **plezalec** *der* Kletter-
 beutler; **polh** ~ *der* Schlafmausbeutler; **progasti**
veveričji ~ *der* Streifenkletterbeutler; **rilčasti** ~
der Rüsselbeutler; ~ **skakač** *der* Springbeutler;
 ~ **svinjenožec** *der* Stutzbeutler; **volk** ~ *der* Beutel-
 wolf; **zverski** ~ *der* Raubbeutler

Abb. 23: Artikelteilstrecke (1) (= wa52–wa72) und Wörterbuchartikel wa73 aus DebN; wa74 und wa75 aus DebS

In DebN sind 65 Komposita mit der Erstkonstituente „Sammel-“ gebucht. Im zehnbändigen Duden-³GW, dem größten Wörterbuch des gegenwärtigen Deutsch, sind es 58; in PonsN sind es neun. Bei anderen überprüften Komposita sind die Zahlen ähnlich. Dies zeigt, dass DebN zu den zweisprachigen Wörterbüchern mit reichhaltiger Makrostruktur gehört. Das Datenangebot in den Wörterbuchartikeln ist dagegen ausgesprochen armselig; weder die slowenischen noch die deutschen Benutzer können damit zufrieden sein. In den Artikeln, in denen es etwas reichhaltiger ist, ist öfters die textuelle Strukturierung wenig klar. Zu den Kurzartikeln ist unter der Voraussetzung, dass es für das Sprachenpaar Deutsch/Slowenisch nur ein Großwörterbuch für dt. und slow. Benutzer geben kann, zunächst festzuhalten, dass das ein solches Wörterbuch polyfunktional sein muss. Unter dieser Voraussetzung sind sämtliche Kurzartikel sowohl in DebN als auch in DebS eine Zumutung für den Benutzer. Am Beispiel von wa₇₃ in Abb. 23 sei dies erläutert. Will ein slow. Benutzer wissen, wie *Glühbirne* dekliniert wird, muss er unter **Birne** nachschlagen, und will er wissen wie *Glühbirne* getrennt und ausgesprochen wird, erfährt er in DebN nichts. Will der dt. Benutzer wissen, wie *Glühbirne* im Slowenischen heißt, bekommt er eine halbe Antwort ohne morphologische Angaben. Ein brauchbarer Kurzartikel, der einem slow. und dt. Benutzer weiteres Nachschlagen erspart, kann z.B. die folgende Form aufweisen:

wa₇₃: **Glüh|bir-ne** ['gly:birnə], die <-,-n> žárnic|a f <-e,-i,-e>

DebN ist erst dann bei den Substantivartikeln ein brauchbares Großwörterbuch von mittlerem Wert, wenn das in wa₇₃ realisierte Minimalprogramm erfüllt ist. Das Wörterbuch kann zu einem guten Großwörterbuch weiterentwickelt werden, wenn die deutsche Deklinationsklassenangabe <-,-n> durch eine numerische verweisvermittelnde Deklinationsmusterangabe angereichert wird (z.B. <-,-n; 26>, denn bekanntlich reicht die Deklinationsklassenangabe für viele Nichtmuttersprachler nicht aus, um alle Substantivformen bilden zu können; „26“ verweist auf einen tabellarischen Umtext mit Deklinationstabellen. Zu einer angemessenen Weiterentwicklung gehört weiter, dass wichtige Kollokationen und entsprechende Äquivalente genannt werden und im Postkommentar die wichtigsten Phraseme, in denen das Lemmazeichen als Phrasemkomponente auftritt. Soll ein wirklich sehr gutes Wörterbuch entstehen, müssen sehr viel mehr Lemmazeichen als polysem interpretiert werden und in den semantischen Subkommentaren müssen textuelle Synonym- und Antonympositionen vorgesehen werden. Nach DebN sind deutlich über 90% der deutschen Lemmazeichen als monosem interpretiert: Von den sprachlichen Gegebenheiten im Wörterbuchgegenstandsbereich, also bei der lexikographischen Ausgangssprache und ihrem Gebrauch, ist eine solche lexikographische Darstellung meilenweit entfernt.

Die Verbartikel sind in der Regel mikrostrukturell reichhaltiger als die Substantivartikel, aber auch sie weisen systematische Mängel auf. Wir nennen nur einige: Die Partikelverben sind von den Präfixverben nicht dadurch unterschieden, dass die Trennungsfuge markiert ist. Das Hilfsverb zur Bildung der

zusammengesetzten Tempusformen (*sein vs hat*) wird nicht genannt. Weder die syntaktische Verbklasse (transitiv vs intransitiv) bei der lexikographischen Ausgangssprache wird angegeben, noch ist eine Angabe der semantischen Verbklasse (perfektiv vs imperfektiv) an die Verbäquivalentangabe adressiert. In diesen Hinsichten sind die Verbartikel in PonsN erheblich angemessener gestaltet. Ein schwieriges Problem bilden für Nichtmuttersprachler die korrekten deutschen Anschlusspräpositionen. Diese sind vom Verb regiert und müssen angegeben werden; nur selten ist das der Fall. — Eine lexikographische Unsitte ist es auch, den Lemmabestand dadurch zu vermehren, dass substantivierte Verben als Lemmata angesetzt werden, z.B. „**Bewegen**, *das*“; „**Tränken**, *das*“; „**Tränen**, *das*“ und „**Tragen**, *das*“. Auch wenn dieser Lemmasatz z.T. dadurch bedingt sein kann, dass besondere slowenische Äquivalente gegeben sind, gehören die substantivierten Infinitive in die Verbartikel.

Bei den Adjektivartikeln fehlen häufig die Angaben der graduierten Formen, wenn der Stammvokal umgelautet wird (vgl. z.B. s.v. **groß**, **kurz**, **lang**¹). Bei anderen Adjektiven sind die Angaben falsch, vgl. z.B. s.v. **hoch**. Hier findet man im Formkommentar „hoher, hohe, hohes“! Korrekt wäre: *höher, höchste*. Selbst die Suppletivformen bei *gut* fehlen. Zwar ist *besser* als Lemma angesetzt, erhält aber als verdichtete Wortartangabe „ADV.“ (= Adverb), und auf diese verdichtete Wortangabe folgt die Beispielangabe „~e Tage“! Neben den genannten systematischen Defiziten in den Substantiv-, Verb- und Adjektivartikeln finden sich in DebN zahlreiche sprachliche Fehler. Wir wollen es bei dieser Feststellung belassen.

Verglichen mit den bisher in Plet, PonsN und PonsS behandelten treten in DebN keine andersartigen Typen von Artikelmikrostrukturen auf, so dass sich Strukturdarstellungen erübrigen.

Abschließend betrachten wir DebS und beginnen mit dem Artikel wa₇₄ in Abb. 23. Die abstrakte (und isomorphe konkrete) hierarchische einfache gemischt-integrierte Artikelmikrostruktur ist in Abb. 24 dargestellt.

Wie Abb. 24 zeigt, weist wa₇₄ keine Subkommentare zur Form und Semantik auf, da die Polysemieangaben fehlen, obwohl das Substantiv *vrč*, wie die beiden nichtsynonymen Äquivalentangaben „Krug“ und „Kanne“ zeigen, als zweifach polysem interpretiert ist. In anderen vergleichbaren Artikeln treten aber Polysemieangaben auf. Weiterhin ist es wenig sinnvoll, an Sprichwortangaben die verdichtete Markierungsangabe „FIG“ zu adressieren. Besser wäre eine Sprichwortidentifizierungsangabe, z.B. SPW.

Sehr oft sind die längeren Wörterbuchartikel in DebS — wie wa₇₄ — nicht übersichtlich gegliedert. Daher gibt es in den Benutzungshinweisen keine brauchbaren Hinweise zu Artikelgliederung. In zahlreichen Substantivartikeln — wie z.B. wa₇₅ in Abb. 23 — mit dem Lemma **vrečar**, die entweder mit dem Symbol für Tierkunde oder dem für Pflanzenkunde versehen sind, dessen Skopus — wie in wa₇₄ — häufig der gesamte Kommentar zur Form und Semantik ist, werden slowenische Mehrwortbenennungen und zugehörige deutsche Komposita genannt, die in einem allgemeinen zweisprachigen Wörterbuch nichts zu suchen haben. Sie gehören vielmehr in entsprechende Fachwörterbücher.

Zusammenfassend muss gesagt werden, dass DebN und DebS den Anforderungen, die an ein modernes allgemeines zweisprachiges Großwörterbuch gestellt werden, nur hinsichtlich des Umfangs des Lemmabestandes genügen. Die textuellen Strukturen der Wörterbuchartikel und das Datenangebot in diesen genügen diesen Anforderungen nicht.

5. Schlussbemerkung

Zum Stand der allgemeinsprachlichen Lexikographie im Sprachenpaar Deutsch und Slowenisch kann abschließend festgestellt werden: Die lexikographische Versorgung ist für relativ anspruchslose Nachschlagebedürfnisse sichergestellt. Für jede Form von Wörterbuchbenutzung, die in Zusammenhang mit der Produktion, Rezeption und Übersetzung von anspruchsvolleren allgemeinsprachlichen Texten steht, ist die lexikographische Versorgung nur sehr eingeschränkt sichergestellt. Wie sich diese unbefriedigende Lage schrittweise verbessern lässt, wurde für die untersuchten Wörterbücher u.a. dadurch gezeigt, dass hierzu jeweils ein Minimalprogramm für angemessener Wörterbuchartikel und ihre Vernetzung mit akzessiven Umtexten genannt wurde, so dass bestimmte realistische Perspektiven zu einer Verbesserung der lexikographischen Versorgung sichtbar werden.

Eine grundlegende Verbesserung der lexikographischen Versorgung im Sprachenpaar Deutsch und Slowenisch ist aber nur möglich, wenn ein neues modernes bidirektionales und polyfunktionales Großwörterbuch mit modernen Methoden erarbeitet wird, in dem das Wissenskorporus der modernen Wörterbuchforschung genutzt wird, so dass die in der Praxis auftretenden grundsätzlichen Fehler nicht stets auf Neue reproduziert werden.

Literatur

- DebN** = Debenjak, Doris, Božidar Debenjak und Primož Debenjak (Hrsg.). 2003. *Veliki nemško-slovenski slovar*. 2. Aufl. Ljubljana: DZS.
- DebS** = Debenjak, Doris, Božidar Debenjak und Primož Debenjak (Hrsg.). 2003. *Veliki slovensko-nemški slovar*. 2. Aufl. Ljubljana: DZS.
- Duden-³GW** = Wissenschaftlicher Rat der Dudenredaktion. 1999. *Duden. Das große Wörterbuch der deutschen Sprache in zehn Bänden*. 3., völlig neu bearb. u. erw. Aufl. Mannheim [usw.]: Dudenverlag.
- Furlan, Metka**. 2006. O Pleteršnikovem slovarju skozi čas. K transliterirani izdaji Pleteršnikovega slovarja. Pleteršnik, Maks und Metka Furlan (Hrsg.). 2006. *Slovensko-nemški slovar (1894–1895)*. Bd. 2, I–XV. Translitierte Ausg. Ljubljana: Založba ZRC SAZU.
- Furlan, Metka**. 2008. Ob izidu nove izdaje Pleteršnikovega Slovensko-nemškega slovarja. Jesenšek, Marko (Hrsg.). *Od Megiserja do elektronske izdaje Pleteršnikovega slovarja*. 2008: 363–367. Maribor: Filozofska fakulteta, Oddelek za slovanske jezike in književnosti.
- Hausmann, Franz Josef, Oskar Reichmann, Herbert Ernst Wiegand und Ladislav Zgusta (Hrsg.)**. 1989–1991. *Wörterbücher. Ein internationales Handbuch zur Lexikographie/Dictionaries. An Inter-*

national Encyclopedia of Lexicography/Dictionnaires. Encyclopédie internationale de lexicographie. Handbücher zur Sprach- und Kommunikationswissenschaft. 3 Teilbände. Berlin/New York: Walter de Gruyter.

- Hausmann, Franz-Josef und Herbert Ernst Wiegand.** 1989. Component Parts and Structures of General Monolingual Dictionaries: A Survey. Hausmann, F.J., O. Reichmann, H.E. Wiegand und L. Zgusta (Hrsg.). 1989–1991: 328-360.
- Jakopin, Franc.** 1994. Maks Pleteršnik — mojster slovenskega slovaropisja. Ob stoletnici izhajanja njegovega slovensko–nemškega slovarja. Orožen, Martina (Hrsg.). 1994. *XXX. seminar slovenskega jezika, literature in kulture*: 23-34. Ljubljana: Seminar slovenskega jezika, literature in kulture pri Oddelku za slovanske jezike in književnosti Filozofske fakultete.
- Jakopin, Franc.** 1997. Pleteršnikovo mesto v sočasnem slovaropisju. Novak, France (Hrsg.). 1997. *Maks Pleteršnik*: 9-15. Ljubljana: Zavod Republike Slovenije za šolstvo.
- Jesenšek, Vida.** 2004. Frazeologija v Pleteršnikovem slovensko–nemškem slovarju: jezikovnosistemska in prevodna ustreznost. Jesenšek, Marko (Hrsg.). 2004. *Besedoslovne lastnosti slovenskega jezika: slovenska zemljepisna imena*: 241-252. Ljubljana: Slavistično društvo Slovenije/Pišce: Društvo Pleteršnikova domačija.
- Jesenšek, Vida.** 2005. Frazeologija v splošnih dvojezičnih slovarjih: razmislek ob nemško–slovenskih slovarjih avtorjev Debenjak. Jesenšek, Marko (Hrsg.). 2005. *Knjižno in narečno besedoslovje slovenskega jezika*: 154-172. Maribor: Slavistično društvo.
- Kammerer, Matthias und Herbert Ernst Wiegand.** 1998/99. Über die textuelle Rahmenstruktur von Printwörterbüchern. Präzisierungen und weiterführende Überlegungen. *Lexicographica* 14: 224-238.
- Konerding, Klaus-Peter und Herbert Ernst Wiegand.** 1994. Framebasierte Wörterbuchartikel. Zur Systematisierung der lexikographischen Präsentation des Bedeutungswissens zu Substantiven. *Lexicographica* 10: 100-170.
- Meyer, Meike und Herbert Ernst Wiegand.** 2000. Gemischt-semiintegrierte Mikrostrukturen für deutsch–spanische Printwörterbücher. Wiegand, Herbert Ernst (Hrsg.). 2000. *Studien zur zweisprachigen Lexikographie mit Deutsch V*: 87-171. Germanistische Linguistik 151–152. Hildesheim/New York: Georg Olms Verlag.
- Pan Zaiping und Herbert Ernst Wiegand.** 1995. Über die Musterartikel für das große Deutsch–Chinesische Wörterbuch. Zugleich ein Beitrag zu einer Theorie zweisprachiger lexikographischer Texte. Wiegand, Herbert Ernst (Hrsg.). 1995. *Studien zur zweisprachigen Lexikographie mit Deutsch II*. Germanistische Linguistik 127–128. Hildesheim/Zürich/New York: Georg Olms Verlag.
- Plet** = Pleteršnik, Maks und Metka Furlan (Hrsg.). 2006. *Slovensko–nemški slovar (1894–1895)*. Transliterierte Ausg. Ljubljana: Založba ZRC SAZU.
- PonsN** = Jemec Škoda, Marjeta et al. (Hrsg.). 2006. *Splošni nemško–slovenski slovar*. Ljubljana: Rokus.
- PonsS** = Martinčič, Polona et al. (Hrsg.). 2008. *Splošni slovensko–nemški slovar*. 1. Aufl. Ljubljana: Rokus.
- PonsN/S-KompWb** = *Kompaktwörterbuch Slowenisch. Slowenisch–Deutsch/Deutsch–Slowenisch*. 2009. Neubearb. Stuttgart: Klett.
- Schierholz, Stefan und Herbert Ernst Wiegand.** 2005. Die Wörterbücher zur Sprach- und Kommunikationswissenschaft. Eine neue Konzeption der linguistischen Fachlexikographie und ihre computergestützte Praxis. *Lexicographica* 20: 164-264.

- Toporišič, Jože.** 1998. Pleteršnikov pravopis v Slovensko-nemškem slovarju. Toporišič, Jože (Hrsg.). *Pleteršnikov slovensko-nemški slovar. Zbornik s simpozija '96 v Pišecah*: 5-15. Novo mesto: Dolenjska založba.
- Wahrig-^sDW = Wahrig-Burfeind, Renate** (Hrsg.). 2006. *Deutsches Wörterbuch. Mit einem Lexikon der Sprachlehre*. 8. vollst. neu bearb. Aufl. Gütersloh/München: Wissen Media Verlag GmbH.
- Wiegand, Herbert Ernst.** 1989. Aspekte der Makrostruktur im allgemeinen einsprachigen Wörterbuch: Alphabetische Anordnungsformen und ihre Probleme. Hausmann, F.J., O. Reichmann, H.E. Wiegand und L. Zgusta (Hrsg.). 1989-1991: 246-280.
- Wiegand, Herbert Ernst.** 1989a. Der Begriff der Mikrostruktur. Geschichte, Probleme, Perspektiven. Hausmann, F.J., O. Reichmann, H.E. Wiegand und L. Zgusta (Hrsg.). 1989-1991: 409-462.
- Wiegand, Herbert Ernst.** 1989b. Arten von Mikrostrukturen im allgemeinen einsprachigen Wörterbuch. Hausmann, F.J., O. Reichmann, H.E. Wiegand und L. Zgusta (Hrsg.). 1989-1991: 462-501.
- Wiegand, Herbert Ernst.** 1990. Printed Dictionaries and their Parts as Texts. An Overview on Recent Research as Introduction. *Lexicographica* 6: 1-126. [Auch in Wiegand 2000: 951-1062.]
- Wiegand, Herbert Ernst.** 1990a. Die deutsche Lexikographie der Gegenwart. Hausmann, F.J., O. Reichmann, H.E. Wiegand und L. Zgusta (Hrsg.). 1989-1991: 2100-2246.
- Wiegand, Herbert Ernst.** 1991. Über die Strukturen der Artikeltexte im *Frühneuhochdeutschen Wörterbuch*. Zugleich ein Versuch zur Weiterentwicklung einer Theorie lexikographischer Texte. Goebel, Ulrich und Oscar Reichmann (Hrsg.). 1991. *Historical Lexicography of the German Language. Volume 2*: 361-673. Studies in Russian and German 3. Lewiston/Queenston/Lampeter: Edwin Mellen Press.
- Wiegand, Herbert Ernst.** 1995. Lexikographische Texte in einsprachigen Lernerwörterbüchern. Kritische Überlegungen anlässlich des Erscheinens von *Langenscheidts Großwörterbuch Deutsch als Fremdsprache*. Popp, Heidrun (Hrsg.). 1995. *Deutsch als Fremdsprache. An den Quellen eines Faches. Festschrift für Gerhard Helbig zum 65. Geburtstag*: 463-499. München: Judicium Verlag.
- Wiegand, Herbert Ernst.** 1995a. *Deutsch-Türkmenisches Wörterbuch*. Einblicke in die Wörterbucharbeit an der Türkmenischen Staatlichen Magtymguly-Universität in Aschghabat. *Lexicographica* 10: 249-300.
- Wiegand, Herbert Ernst.** 1996. Das Konzept der semiintegrierten Mikrostrukturen. Ein Beitrag zur Theorie zweisprachiger Printwörterbücher. Wiegand, Herbert Ernst (Hrsg.). 1996. *Wörterbücher in der Diskussion II. Vorträge aus dem Heidelberger Lexikographischen Kolloquium*: 1-82. *Lexicographica. Series Maior* 70. Tübingen: Max Niemeyer Verlag. [Auch in Wiegand 2000: 1193-1277.]
- Wiegand, Herbert Ernst.** 1996a. *Deutsch-Uzbekisches Wörterbuch*. Einblicke in die Wörterbucharbeit an der Staatlichen Usbekischen Weltsprachen-Universität in Taschkent. *Lexicographica* 12: 190-254. [Auch in Wiegand 2000: 1339-1409.]
- Wiegand, Herbert Ernst.** 1996b. Über die Mediostrukturen bei gedruckten Wörterbüchern. Zettersten, Arne und Viggo Hjørnager Pedersen (Hrsg.). 1996. *Symposium on Lexicography VII. Proceedings of the Seventh Symposium on Lexicography May 5-6, 1994 at the University of Copenhagen*: 11-43. *Lexicographica. Series Maior* 76. Tübingen: Max Niemeyer Verlag. [Auch in Wiegand 2000: 1163-1192.]
- Wiegand, Herbert Ernst.** 1997. Printed Language Dictionaries and their Standardization: Notes on the Progress toward a General Theory of Lexicography. Hock, Hans Heinrich (Hrsg.). 1997. *Historical, Indoeuropean and Lexicographical Studies. A Festschrift for Ladislav Zgusta on the*

- Occasion of his 70th Birthday*: 319-380. Trends in Linguistics. Studies and Monographs 90. Berlin: Mouton de Gruyter.
- Wiegand, Herbert Ernst.** 1998. *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie. 1. Teilbd.* Mit 159 Abbildungen im Text. Berlin/New York: Walter de Gruyter.
- Wiegand, Herbert Ernst.** 1998a. Altes und Neues zur Makrostruktur in alphabetischen Printwörterbüchern. Wiegand, Herbert Ernst (Hrsg.). 1998. *Wörterbücher in der Diskussion III. Vorträge aus dem Heidelberger Lexikographischen Kolloquium*: 348-372. Lexicographica. Series Maior 84. Tübingen: Max Niemeyer Verlag. [Auch in Wiegand 2000: 1428-1453.]
- Wiegand, Herbert Ernst.** 1998b. Lexikographische Textverdichtung. Entwurf einer vollständigen Konzeption. Zettersten, Arne, Viggo Hjørnager Pedersen und Jens Erik Mogensen (Hrsg.). 1998. *Symposium on Lexicography VIII. Proceedings of the Eighth International Symposium on Lexicography May 2-4, 1996, at the University of Copenhagen*: 1-35. Lexicographica. Series Maior 90. Tübingen: Max Niemeyer Verlag. [Auch in Wiegand 2000: 1454-1489.]
- Wiegand, Herbert Ernst.** 1999. Immken, Antje und Werner Wolski (Hrsg.). *Semantics and Lexicography. Selected Studies (1976-1996)*. Lexicographica. Series Maior 97. Tübingen: Max Niemeyer Verlag.
- Wiegand, Herbert Ernst.** 1999a. Artikel einsprachiger Lernerwörterbücher, Textgestaltwahrnehmung und Suchbereichsstrukturen. Plädoyer für übersichtliche Printwörterbücher im Zeitalter der Neuen Medien. Skibitzki, Bernd und Barbara Wotjak (Hrsg.). 1999. *Linguistik und Deutsch als Fremdsprache*: 259-281. Tübingen: Max Niemeyer Verlag.
- Wiegand, Herbert Ernst.** 2000. Kammerer, Matthias und Werner Wolski (Hrsg.). *Kleine Schriften. Eine Auswahl aus den Jahren 1970 bis 1999 in zwei Bänden. Bd. 1: 1970-1988; Bd. 2: 1988-1999*. Berlin/New York: Walter de Gruyter.
- Wiegand, Herbert Ernst.** 2000a. Über Suchbereiche, Suchzonen und ihre textuellen Strukturen in Printwörterbüchern. Ein Beitrag zur Theorie der Wörterbuchform. Wiegand, H.E. (Hrsg.). 2000. *Wörterbücher in der Diskussion IV. Vorträge aus dem Heidelberger Lexikographischen Kolloquium*: 233-301. Lexicographica. Series Maior 100. Tübingen: Max Niemeyer Verlag.
- Wiegand, Herbert Ernst.** 2000b. Über tabellarische Wörterverzeichnisse und deren Wörterbuchartikel. Ein Beitrag zur Theorie der Wörterbuchform. *Lexicographica* 16: 212-234.
- Wiegand, Herbert Ernst.** 2000c. Adressierung in der ein- und zweisprachigen Lexikographie. Eine einführende Übersicht über die Forschungs- und Problemlage. *Lexikos* 10: 32-74.
- Wiegand, Herbert Ernst.** 2000d. Wissen, Wissensrepräsentation und Printwörterbücher. Heid, Ulrich, Stefan Evert, Egbert Lehmann und Christian Rohrer (Hrsg.). 2000. *Proceedings of the Ninth EURALEX International Congress. EURALEX 2000. Stuttgart, Germany, August 8-12, 2000*: 15-38. Stuttgart: Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.
- Wiegand, Herbert Ernst.** 2001. Was eigentlich sind Wörterbuchfunktionen? Kritische Anmerkungen zur neueren und neuesten Wörterbuchforschung. *Lexicographica* 17: 217-248.
- Wiegand, Herbert Ernst.** 2001a. Sprachkontaktwörterbücher, Typen, Funktionen, Strukturen. Igla, Birgit, Pavel Petkov und Herbert Ernst Wiegand (Hrsg.). 2001. *Theoretische und praktische Probleme der Lexikographie. 1. Internationales Kolloquium zur Wörterbuchforschung am Institut Germanicum der St. Kliment Ohridski-Universität, Sofia, 7. bis 8. Juli 2000*: 115-224. Germanistische Linguistik 161-162. Hildesheim/Zürich/New York: Georg Olms Verlag.
- Wiegand, Herbert Ernst.** 2002. Zur Makrostruktur und zu den äußeren Zugriffsstrukturen im *de Gruyter Wörterbuch Deutsch als Fremdsprache*. Wiegand, Herbert Ernst (Hrsg.). 2002. *Perspektiven der pädagogischen Lexikographie des Deutschen II. Untersuchungen anhand des de Gruyter*

- Wörterbuchs Deutsch als Fremdsprache: 413-442. *Lexicographica*. Series Maior 110. Tübingen: Max Niemeyer Verlag.
- Wiegand, Herbert Ernst.** 2002a. Altes und Neues zur Mediostruktur in Printwörterbüchern. *Lexicographica* 18: 168-252.
- Wiegand, Herbert Ernst.** 2002b. Über textuelle Strukturen der Wörterbuchartikel und Artikelnischen im *de Gruyter Wörterbuch Deutsch als Fremdsprache*. Zugleich ein Beitrag zur Weiterentwicklung einer Theorie der Wörterbuchform. Wiegand, Herbert Ernst (Hrsg.). 2002. *Perspektiven der pädagogischen Lexikographie des Deutschen II. Untersuchungen anhand des de Gruyter Wörterbuchs Deutsch als Fremdsprache*: 497-595. *Lexicographica*. Series Maior 110. Tübingen: Max Niemeyer Verlag.
- Wiegand, Herbert Ernst.** 2002c. Adressierung in zweisprachigen Printwörterbüchern. Wiegand, Herbert Ernst (Hrsg.). 2002. *Studien zur zweisprachigen Lexikographie mit Deutsch VIII*: 111-175. Germanistische Linguistik 166. Hildesheim/Zürich/New York: Georg Olms Verlag.
- Wiegand, Herbert Ernst.** 2002d. Equivalence in Bilingual Lexicography: Criticism and Suggestions. *Lexikos* 12: 239-255.
- Wiegand, Herbert Ernst.** 2003. Überlegungen zur Typologie von Wörterbuchartikeln in Printwörterbüchern. Ein Beitrag zur Theorie der Wörterbuchform. *Lexicographica* 19: 169-313.
- Wiegand, Herbert Ernst.** 2003a. Wörterbuch zur Lexikographie und Wörterbuchforschung / Dictionary of Lexicography and Dictionary Research. Städtler, Thomas (Hrsg.). 2003. *Wissenschaftliche Lexikographie im deutschsprachigen Raum. Im Auftrag der Heidelberg Akademie der Wissenschaften*: 417-437. Heidelberg: Winter Verlag.
- Wiegand, Herbert Ernst.** 2004. Überlegungen zur Mediostruktur in Fachwörterbüchern. Auch am Beispiel des „Wörterbuchs zur Lexikographie und Wörterbuchforschung“. Brdar-Szabó, Rita und Elisabeth Knipf-Komlosi (Hrsg.). 2004. *Lexikalische Semantik, Phraseologie und Lexikographie. Abgründe und Brücken. Festgabe für Regina Hessky*: 339-365. Duisburger Arbeiten zur Sprach- und Kommunikationswissenschaft 57. Frankfurt a. M.: Peter Lang Verlag.
- Wiegand, Herbert Ernst.** 2005. Über die Datenakzessivität in Printwörterbüchern. Einblicke in neuere Entwicklungen einer Theorie der Wörterbuchform. *Lexikos* 15: 196-230.
- Wiegand, Herbert Ernst.** 2005a. Angaben, funktionale Angabezusätze, Angabetexte, Angabestrukturen, Strukturanzeiger, Kommentare und mehr. Ein Beitrag zur Theorie der Wörterbuchform. *Lexicographica* 21: 202-379.
- Wiegand, Herbert Ernst.** 2005b. Äquivalenz, Äquivalentdifferenzierung und Äquivalentpräsentation im zweisprachigen Wörterbüchern. Eine neue einheitliche Konzeption. Gottlieb, Henrik, Erik Jens Mogensen und Arne Zettersten (Hrsg.). *Symposium on Lexicography XI. Proceedings of the Eleventh International Symposium on Lexicography May 2-4, 2002 at the University of Copenhagen*: 17-57. *Lexicographica*. Series Maior 115. Tübingen: Max Niemeyer Verlag.
- Wiegand, Herbert Ernst.** 2005c. Über die textuellen Strukturen im GWDS. Wiegand, Herbert Ernst (Hrsg.). 2005. *Untersuchungen zur kommerziellen Lexikographie der deutschen Gegenwartssprache II*. Duden. Das große Wörterbuch der deutschen Sprache in zehn Bänden. Print- und CD-ROM-Version. 2. Bd.: 295-335. *Lexicographica*. Series Maior 121. Tübingen: Max Niemeyer Verlag.
- Wiegand, Herbert Ernst.** 2005d. Äquivalentpräsentation und Wörterbuchfunktionen in zweisprachigen Printwörterbüchern. Mit einem Seitenblick auf die so genannte „moderne lexikographische Funktionslehre“. Iгла, Birgit, Pavel Petkov und Herbert Ernst Wiegand (Hrsg.). 2005. *Kontrastive Lexikologie und zweisprachige Lexikographie. 2. Internationales Kolloquium zur Wörter-*

buchforschung. St. Kliment Ohridski-Universität Sofia, 18. bis 19. Oktober 2002: 1-38. Germanistische Linguistik 179. Hildesheim/Zürich/New York: Georg Olms Verlag.

- Wiegand, Herbert Ernst.** 2006. Das Lern- und Konsultationswörterbuch. Ein neuer Fachwörtertyp am Beispiel der Wörterbücher zur Sprach- und Kommunikationswissenschaft (WSK). *Lexikos* 16: 205-221.
- Wiegand, Herbert Ernst.** 2006a. Die „Wörterbücher zur Sprach- und Kommunikationswissenschaft“ (WSK) und ihre Benutzungsmöglichkeiten im Fach Deutsch als Fremdsprache. Dimova, Ana, Vida Jesenšek und Pavel Petkov (Hrsg.). 2006. *Zweisprachige Lexikographie und Deutsch als Fremdsprache. Drittes Internationales Kolloquium zur Lexikographie und Wörterbuchforschung. Konstantin Preslavski-Universität Schumen, 23.-24. Oktober 2004: 1-35. Germanistische Linguistik 184-185. Hildesheim/Zürich/New York: Georg Olms Verlag.*
- Wiegand, Herbert Ernst.** 2006b. Adressierung in Printwörterbüchern. Präzisierungen und weiterführende Überlegungen. *Lexikographica* 22: 187-261.
- Wiegand, Herbert Ernst.** 2007. Über Zugriffspfade in Printwörterbüchern. Ein Beitrag zur Schnittstelle von Benutzungshandlungen und Wörterbuchform. *Lexikos* 17: 180-211.
- Wiegand, Herbert Ernst.** 2007a. Aspectos escogidos de una teoría de las estructuras textuales de los diccionarios. Souto, Mar Campos, Eugenia Conde Noguero, José Ignacio Pérez Pascual und José Porto Dapena (Hrsg.). 2007. *Reflexiones sobre el diccionario: 35-63. Coruña Anexos de „Revista de Lexicografía“ 6. Coruña: Universidade da Coruña, Servizo de Publicacións.*
- Wiegand, Herbert Ernst.** 2007b. Il riferimento indirizzato nei dizionari bilingui a stampa. *Studi Italiani di Linguistica Teorica e Applicata: 7-87.*
- Wiegand, Herbert Ernst.** 2007c. Sull'accessibilità ai dati nei dizionari a stampa. Sei considerazioni sugli ultimi sviluppi di una teoria sulla forma del dizionario. *Studi Italiani di Linguistica Teorica e Applicata: 89-136.*
- Wiegand, Herbert Ernst.** 2007d. Neuere Aspekte einer Theorie und Typologie von Wörterbuchartikeln und ihre Praxisrelevanz. Gottlieb, Henrik und Jens Erik Mogensen (Hrsg.) 2007. *Dictionary Visions, Research and Practice. Selected Papers from the 12th International Symposium on Lexicography, Copenhagen 2004: 183-200. Amsterdam/Philadelphia: John Benjamins.*
- Wiegand, Herbert Ernst.** 2007e. Zur Adressierung in Printwörterbüchern. Präzisierungen und weiterführende Überlegungen. *Lexikographica* 22: 187-261.
- Wiegand, Herbert Ernst.** 2008. Zugriffsstrukturen in Printwörterbüchern. Ein zusammenfassender Beitrag zu einem zentralen Ausschnitt einer Theorie der Wörterbuchform. *Lexikographica* 24: 209-315.
- Wiegand, Herbert Ernst.** 2008a. Wörterbuchregister. Grundlagen einer Theorie der Register in modernen Printwörterbüchern. *Lexikos* 18: 256-302.
- Wiegand, Herbert Ernst.** 2009. Adressierung in der ein- und zweisprachigen Lexikographie. Eine zusammenfassende Darstellung. Kürschner, Wilfried und Manfred Ringmacher (Hrsg.). *Aus Ost und West. Erträge des 3. und 4. Ost-West-Kolloquiums. [Erscheint].*
- Wiegand, Herbert Ernst.** 2009a. Nichtnatürlich über natürliche Sprache schreiben. Zu einigen formalen Aspekten von Wörterbuchartikeln. *Heidelberger Jahrbücher* 53.2009: 247-285.

Place et nécessité des travaux religieux, ethnohistoriques et linguistiques dans le paysage dictionnaire gabonais¹

P.A. Mavoungou, *Groupe de Recherches en Langues et Cultures Orales (GRELACO), Département des Sciences du Langage, Université Omar Bongo, Libreville, République Gabonaise (moudika2@yahoo.fr)*

Résumé: Les dictionnaires devraient être le reflet d'une grande variété de découvertes enregistrées dans le domaine de la linguistique. Au Gabon, en ce qui concerne les rapports entre la lexicographie et la linguistique, des études doivent être menées. Cet article ne retrace pas seulement l'histoire de la discipline lexicographique au Gabon, mais également comment les travaux pionniers dans les domaines de la linguistique, de la religion et de l'ethnohistoire, ont servi (et devraient continuer de servir) de travaux de référence pour la compilation de nouveaux dictionnaires. Compte tenu du fait que la majorité des dictionnaires existant a été le résultat de la contribution de lexicographes isolés, on peut espérer que les futurs produits lexicographiques ne seront pas seulement des réservoirs de connaissances linguistiques et culturelles mais également le résultat d'un travail d'équipe.

Mots-clés: LEXICOGRAPHIE, UNITÉ(S)-LANGUE(S), LINGUISTIQUE, ÉTUDES RELIGIEUSES, ETHNOHISTOIRE, TRAVAIL D'ÉQUIPE, METALEXICOGRAPHIE

Abstract: **The Place and Necessity of Religious, Ethnohistorical and Linguistic Works in the Gabonese Dictionary Landscape.** Dictionaries should reflect a great diversity of results gained from research in the field of linguistics. In Gabon, as far as relationships between lexicography and linguistics are concerned, studies are still to be started. This article does not only trace the historical development of lexicography in Gabon, but also shows how earlier studies in the fields of linguistics, religion and ethnohistory have served (and should continue to serve) as reference works for the compilation of new dictionaries. Owing to the fact that most existing dictionaries have resulted from the enterprise of single lexicographers, it is hoped that future lexicographic products will not only be containers of linguistic and cultural knowledge but also the result of team efforts.

Keywords: LEXICOGRAPHY, LANGUAGE GROUP(S), LINGUISTICS, RELIGIOUS STUDIES, ETHNOHISTORY, TEAM WORK, METALEXICOGRAPHY

1. Introduction et délimitation du sujet

La liste plus ou moins complète des travaux sur les langues du Gabon a été établie (cf. Hombert et Mortier 1984, 1990). Il en est de même pour les travaux

lexicographiques à proprement parler (cf. Nyangone Assam et Mavoungou 2000). Les rapports entre la lexicographie et la linguistique entre autres n'ont pas encore été abordés de façon systématique par la plupart des publications sur les langues du Gabon. La présente contribution a la prétention de combler un tant soit peu cette lacune.

Dans cet aperçu des travaux sur les langues gabonaises, nous nous sommes imposés les restrictions suivantes:

- (a) Le foisonnement relatif des travaux ethnohistoriques, religieux et linguistiques sur les langues gabonaises, nous contraint dans la plupart des cas à une simple énumération.
- (b) Nous excluons de cet aperçu tous les dictionnaires plurilingues ou multilingues (c'est-à-dire trois langues et plus).
- (c) Parmi les dictionnaires bilingues (puisque c'est d'eux qu'il s'agit dans cet aperçu), nous ne parlerons pas des dictionnaires des langues de spécialités.

L'état des lieux sur les travaux sur les langues du Gabon se fera selon les dix unités-langues de Kwenzi-Mikala (1988, 1998). Le traitement de chaque unité-langue fera apparaître çà et là des disproportions que le lecteur avisé ne manquera pas de noter. En effet, le développement est fonction de la littérature disponible sur chaque groupe.

2. Le Groupe Mazuna

Les parlers de ce groupe se localisent principalement dans trois des neuf provinces du Gabon, à savoir: l'Estuaire, l'Ogooué-Ivindo et le Woleu-Ntem. Le *fañ*-ntumu est parlé à Oyem et Bitam, et le *fañ*-okak à Médouneu, Cocobeach et Mitzié. Le *fañ*-atsi est parlé à Ndjolé et à Lambaréné, le *fañ*-make à Libreville et Kango, le *fañ*-mvaï se retrouve à Minvoul, et le *fañ*-nzaman est parlé dans la région de Makokou–Ovan–Booué. Le *fañ* est une langue transfrontalière parce qu'il est également parlé dans les pays voisins, à savoir en Guinée-Équatoriale, au Cameroun, à São Tomé et Príncipe (Grimes 1996) et en République du Congo.

Tous les travaux disponibles en littérature mazuna touchent les domaines de la religion, l'ethnohistoire, la linguistique et la lexicographie. Dans cette section, nous mettrons beaucoup plus l'accent sur les travaux ethnohistoriques et linguistiques avant de présenter quelques études lexicographiques.

Comme cela apparaît dans la localisation ci-dessus, le *fañ* est divisé en plusieurs dialectes. Le *make* est unanimement reconnu comme le dialecte du *fañ* le plus abondamment documenté en matière religieuse (Afane Otsaga 2004). Ceci tient notamment à une situation ou position géographique stratégique. En effet, disséminées dans la province de l'Estuaire (Libreville, Kango,

Ntoum et Foulenzem), les populations Fan-Make sont très tôt rentrées en contact avec les explorateurs européens et les colons français. Ces derniers seront très tôt confrontés au problème de la communication avec les populations autochtones. Afin de surmonter les barrières linguistiques, les missionnaires, notamment, vont se lancer dans la production de catéchismes, de recueils de cantiques, des traductions d'évangiles et des manuels scolaires pour établir leur oeuvre d'évangélisation et de scolarisation. Cette production religieuse et scolaire intervient sous les auspices de la Mission Catholique de Libreville (Vicariat Apostolique de Libreville). Déjà depuis le Concile de Trente (1545-1563), les prêtres devaient apprendre les langues de leurs futurs convertis, non seulement pour propager le christianisme, mais aussi parce que cela favorisait l'imposition. Pour la langue fan, les missionnaires impliqués dans la traduction de la Bible et des Évangiles la faisaient en fan-make, le dialecte fan le plus usité au siège du clergé catholique du Gabon (la Mission Catholique de Libreville).

L'ouvrage qui fait autorité depuis sa parution en 1962 est la traduction de la Bible en fan par François Ndong. Cette traduction intitulé *Nten Wam* est évidemment rédigée en fan-make, langue maternelle de l'archevêque Ndong. Dans le contexte des dispositions du Concile de Trente qui avait encouragé l'évangélisation dans les langues vernaculaires, monseigneur Ndong va également jouer un rôle instrumental dans la production de nombre de travaux dans le dialecte make (Afone Otsaga 2004).

Pour ce qui est des traductions des Évangiles, citons: *Évangile de Jean: traduction dans la langue des Fan* (1910), *Évangile de Luc: traduction dans la langue des Fan* (1910), et *Évangile de Matthieu: traduction dans la langue des Fan* (1902). À ces Évangiles selon Jean, Luc et Matthieu, il faut ajouter non seulement des livres de prières (*Livre de prières = NTEN WAM minkobe ma nè ge kobe né nzame* 1898), des livres de questions ou catéchismes (*Catéchisme Fan* 1932, *Katesism Nten wa vegele Fang Nsong Katolik = Catéchisme de la Doctrine Catholique* n.d., et *Nteni osi nteni Fañwe* 1889) mais également des ouvrages qui sont beaucoup plus des objets scolaires que des travaux religieux: *Syllabaire Fan* (1925), *Ntèn wa tar'éyé gele fang'ne fala étén nélang = Exercices de lecture et d'écriture en pahouin et en français* (1898).

Tous ces travaux aussi bien liturgiques que bibliques en fan-make constituent une matière lexicographique non négligeable. Une fois intégrés, par exemple au moyen de la méthode de l'OCR (Optical Character Recognition) ou de la saisie, ces textes monolingues pourraient être renforcés par des données orales collectées sur le terrain. On pourrait ainsi obtenir à terme de larges bases de données textuelles électroniques ou corpora en fan-make notamment. Plusieurs types de dictionnaires (monolingues, bilingues ou multilingues) pourraient être dérivés à partir desdits corpora électroniques.

Par ailleurs, les syllabaires et abécédaires disponibles en fan sont, en quelque sorte, des dictionnaires de prononciation associant objets et lettres (voire sons). On dispose ainsi d'extraits textuels à trois colonnes disposées par séries de consonnes en variations sur une voyelle. Le lecteur aura une idée du type de

répertoire dont il s'agit d'après le bref fragment suivant, qui présente une série de consonnes en variations sur la voyelle /a/; dans la première colonne apparaissent une série d'illustrations éminemment didactiques. La deuxième colonne rend compte de l'association des objets et des lettres (et même des sons). En effet, les graphèmes (voire syllabogrammes: signes à valeur "phonétique") apparaissent avec à l'arrière plan les dessins les illustrant. Enfin la troisième colonne correspond aux traductions *fañ* en graphèmes voire syllabogrammes des deux premières colonnes.



Figure 1: Quelques syllabogrammes du *fañ* sur fond d'illustrations (extrait de l'ouvrage *Ma yi lañ nten fañ osua* (1965) publié sous les auspices de l'Église Évangélique du Gabon, tel que cité dans Mayer (1990: 86)

Ces productions religieuses et scolaires ont également le mérite de fixer un premier état de la langue, en particulier une orthographe jusque là très fluctuante. En outre, la plupart de ces travaux comportent un certain nombre de lacunes qui doivent être comblées. Par exemple, ils ne marquent pas les tons ainsi que la prononciation de la langue. La lacune la plus importante et la plus souvent citée est que ces travaux contiennent beaucoup d'inconsistances orthographiques. Par exemple, Alexandre (1961) mentionne que la langue *fañ* a acquis "trois transcriptions [catholiques] et deux [protestantes] pour environ un million de locuteurs, sans parler des différences morphologiques dues à des

calques catholiques sur l'allemand, le français et l'espagnol, et protestants, sur l'américain".

Les premiers travaux à caractère ethnohistorique sur le *Fan* abordent essentiellement la question des migrations. Les récits migratoires sont ponctués d'épisodes guerriers où se côtoient l'imaginaire, l'extraordinaire et le mystique. Par conséquent et d'une manière générale, le récit des migrations des *Fan* contient aussi bien des éléments ou données mythiques qu'historiques. Le récit mythique fait partir le peuple *Fan* d'un lieu difficile à identifier de nos jours appelé "Odzamboga". C'est au cours de leur déplacement que les populations *Fan* seront confrontées à une épreuve physique de taille. En effet, elles vont tomber nez à nez avec un arbre gigantesque appelé "Adzap". Ne pouvant ni contourner ni passer au-dessus de l'"Adzap", les populations n'avaient pas d'autre choix que de pratiquer une ouverture dans l'arbre. Ainsi va naître et se cristalliser dans la mémoire collective la légende du "trou de l'Adzap". Traverser l'"Adzap" de part en part n'était pas chose facile, les *Fan* y arrivèrent au prix de dures et tenaces efforts. C'est seulement après la rencontre et le franchissement de cet obstacle naturel que ces populations vont se retrouver disséminées à des proportions diverses dans les provinces du Gabon: Estuaire, Moyen-Ogooué, Ogooué-Ivindo et Woleu-Ntem.

D'un point de vue historique, les migrations du peuple *Fan* remontent approximativement aux 18^{ème} et 19^{ème} siècles. Les régions du Dja, l'Ivindo et le Ntem sont considérées comme les dernières étapes avant la dispersion des populations *Fan* (Mouguiama-Daouda 1995). Selon un certain nombre d'auteurs (Alexandre 1967 et Balandier 1957, 1963 entre autres), c'est notamment sous la pression des Foulbé que les *Fan* à partir du Cameroun vont atteindre le Moyen-Ogooué au début du 19^{ème} siècle. Leur arrivée dans la province de l'Estuaire pourrait se situer autour de 1860. Les superlatifs ne manquent pas pour qualifier ces nouveaux venus, "une belle race [...] des hommes grands, bien faits" (le Marquis de Compiègne, cité par Pierre Alexandre 1967). Tous ces récits sur les migrations des *Fan* constituent de facto le matériel de base pour la rédaction des notes introductives traitant des questions ethnohistoriques dans les dictionnaires en *fan*.

D'un point de vue purement linguistique, le terme *fan* désigne aussi bien la langue que ses locuteurs. Il s'agit d'une langue bantou de la zone A70, avec plus précisément le sigle A75 selon la classification de Guthrie (1948, 1953 et 1969-1971). En outre, le glossonyme et ethnonyme *Pahouin*, écrit selon une orthographe administrative française tronquée, est souvent utilisé pour désigner les *Fan*. Les travaux linguistiques en *fan* sont exogènes. Les travaux de Marling (1872), Martrou (1924), etc. sont certes des travaux lexicographiques mais ils sont précédés d'une description grammaticale du *fan*. Mis à part les travaux de Raponda-Walker notamment, c'est avec l'ouverture à l'Université Omar Bongo d'une option linguistique au sein du Département des Lettres Modernes que l'on va constater que de plus en plus de travaux linguistiques sur les langues locales seront menés par les Gabonais eux-mêmes. Nombre de travaux linguistiques portant sur le *fan* ont été présentés et soutenus à l'Uni-

versité Omar Bongo ou faits par des étudiants gabonais dans les universités européennes. Il s'agit pour la plupart des Rapports de Licence, Mémoires de Maîtrise et Thèses de Doctorat. Citons: Mbougou Asseko (1979), Ndong Menini (1980), Mba-Nzué (1981), Ondo Essono Aba'a (1981), Mba-Nkoghe (1981), Andeme Allogho (1983), Nzang-Bié (1986), Rekanga (1986), Nguema Nzue (1988), Cinnamon (1990) et Ondo-Mebiame (1992). Les travaux de Ndong Menini, Ondo Essono Aba'a, Nzang-Bié, Rekanga, Nguema Nzue et Ondo-Mebiame décrivent le *fan*-ntumu. Mba-Nzué propose une description du *fan*-mvai, tandis qu'Andeme Allogho et Cinnamon décrivent le *fan*-nzaman. Les données utilisées dans l'étude menée par Mba-Nkoghe sur la phonologie et la morphologie proviennent du *fan*-atsi parlé dans les zones de Ndjolé et Lambaré. Le travail de Mbougou Asseko qui est une étude contrastive entre le *fan* d'Oyem et celui de Libreville est un peu à l'image de ce qui se fait actuellement et cela depuis 1994 au Département des Sciences du Langage de l'Université Omar Bongo. Nombre de travaux linguistiques en *fan* soutenus dans ledit Département abordent la langue *fan* sous l'angle de la comparaison avec notamment le français, la langue officielle au Gabon (sur le statut du français au Gabon ainsi que la place et l'utilité des langues gabonaises dans l'activité nationale, lire Mba-Nkoghe 1991). Citons par exemple: Mekui Missang (1998), Ella Ella (2000), Ella (2000) et Bibang Meyé (2001). Enfin, il convient de signaler que les travaux susmentionnés sur le *fan* concernent la perspective synchronique. Par contre, la contribution de Medjo-Mvé (1997) envisage la perspective historique du *fan*.

Tous ces travaux linguistiques sur le *fan* constituent une matière lexicographique non négligeable. En effet, ils présentent en majorité des corpora plus ou moins étendus qui pourraient être énormément enrichis afin de répondre aux normes dictionnaires.

Sur le plan lexicographique stricto sensu, le premier dictionnaire décrivant le *fan* est celui de Marling, publié en 1872 à New York. La contribution de Marling intitulée *Dictionnaire fang-français* est suivie en 1881 d'un lexique *Panwe Primer and Vocabulary* compilé par Nassau à partir du matériau collecté par H.M. Adam. Ce travail lexicographique est précédé d'une description grammaticale de la langue décrite. La comparaison entre ces deux travaux lexicographiques révèle que dans le travail de Marling, le *fan* est la langue source tandis que le français est la langue cible. Ces langues paires sont inversées dans le travail de Lejeune (1892). Du point de vue de la portée, ces deux dictionnaires ont un dénominateur commun: ce sont des ouvrages monocopals et tous d'eux décrivent le dialecte *fan*-atsi. Un dictionnaire monoscopal est une publication incluant seulement une section traitant la langue source et la langue cible (A > B). En d'autres termes, c'est une publication dans laquelle le lexicographe traite uniquement une langue source et une langue cible. Par exemple: le dictionnaire monoscopal de Marling a le *fan* comme langue source et le français comme langue cible. Ceci implique que seuls les items lexicaux d'un membre du couple de langues (le *fan*) seront lemmatisés. Par contre, dans un ouvrage biscopal, le lexicographe travaillera avec deux langues sources et deux langues

cibles, c'est-à-dire qu'il inclura à la fois des sections A > B et B > A dans le même dictionnaire. À la suite des travaux de Marling et Lejeune, l'autre ouvrage lexicographique d'importance est celui de Largeau (1901). Il s'agit de l'*Encyclopédie pahouine*. D'un point de vue typologique, les ouvrages de référence lexicographique connaissent généralement une présentation tripartite ou trigonométrique: dictionnaire, encyclopédie et dictionnaire encyclopédique. L'*Encyclopédie pahouine* de Largeau est en réalité un dictionnaire encyclopédique. Il contient notamment des principes grammaticaux et une nomenclature français-pahouin. Les lemmes dans une encyclopédie sont classés par ordre thématique. Or, dans la nomenclature français-pahouin de Largeau, les unités de traitement apparaissent dans un ordre strictement alphabétique.

À la suite du travail de Largeau, l'autre ouvrage lexicographique qui mérite d'être mentionné est le bilingue *Lexique fān-français* produit par monseigneur Martrou (1924), un missionnaire de la Congrégation des Pères du Saint-Esprit. Bien que le titre de cet ouvrage soit modestement décrit comme un "lexique", il s'agit en réalité d'un dictionnaire de volume réduit comprenant 3 431 articles et couvrant 137 pages. Sur le plan microstructurel, ce travail lexicographique fournit aux usagers une variété de catégories de données, à savoir: la paraphrase du sens, les traductions, les exemples d'emploi, etc. (Mavoungou 2001: 127-130 et Mavoungou, Afane Otsaga et Mihindou 2003: 136).

Après la publication du dictionnaire de Largeau, l'activité de la lexicographie du fān-atsi va connaître un ralentissement. Cette activité va reprendre à la faveur de l'accession du Gabon à la souveraineté internationale. C'est précisément quatre ans (en 1964) après les Indépendances que le *Dictionnaire fang-français et français-fang* est publié sous les auspices de l'Association Française des Amis d'Albert Schweitzer. Cette publication posthume du pasteur genevois Samuel Galley contient deux parties: le dictionnaire à proprement parler suivi d'une grammaire fān.

Le dernier travail en lexicographie fān qui mérite d'être cité concerne un travail de métalexigraphie (recherche sur les dictionnaires). Il s'agit de l'*Esquisse d'un dictionnaire fang-mekè*. Ce travail mené par Nzang-Obame (2004) constitue une espèce de rupture épistémologique parce qu'il présente un modèle théorique d'un dictionnaire en fān-make. L'étude se donne notamment pour objectif de concevoir l'ensemble des structures dudit dictionnaire complété par quelques échantillons d'articles traités. Pour les ouvrages lexicographiques qui l'ont précédé, on peut arguer qu'à un degré plus ou moindre, ils manquent d'une base métalexigraphique. Toutefois, ces manquements sont excusables si l'on considère que les lexicographes de ces travaux précurseurs n'étaient ni linguistes, ni lexicographes qualifiés. Dans le même esprit que le travail de Nzang-Obame (2001), Afane Otsaga (2004) propose de A à Z la démarche pour la compilation et la publication d'un dictionnaire standard du fān. Concrètement, l'auteur analyse un certain nombre de critères scientifiques susceptibles de permettre qu'un des six dialectes du fān soit érigé ou élevé au rang de variante standard. C'est ce dialecte du fān susceptible de jouer le rôle de langue standard qui bénéficiera de l'ensemble des avantages liés à l'utilisa-

tion de ce dialecte choisi dans l'administration, l'enseignement et les médias, ainsi qu'à la compilation et publication de matériels didactiques (dictionnaires, syllabaires, abécédaires, grammaires, journaux, etc.). Cette étude qui s'intitule *The Standard Translation Dictionary as an Instrument in the Standardization of Fang* n'apporte pas seulement des éléments de réponses à la standardisation du fan, elle fait également entrevoir aux linguistes et lexicographes travaillant sur d'autres langues gabonaises des pistes de recherche dans le domaine de la standardisation des langues.

Enfin, à côté de ces travaux publiés ou encore présentés et soutenus dans le cadre de mémoires ou de thèses, il faut souligner l'existence de recherches non datées. Il s'agit notamment des ouvrages intitulés *Diccionario Pamue-Español, Vocabulary of the Fan Language in Western Africa* (Osorio 1887), "Contribution à l'étude du folklore bantou: fables, devinettes et proverbes fang" (Tardy 1933) et *Gramática pamue* (Ndongo Esono 1956). Le travail d'Osorio, ajouté aux études susmentionnées, pourrait entrer dans la constitution d'une large base de données sur le fan. Le travail de Tardy, considérablement plus intéressant du point de la production de dictionnaires syntagmatiques, pourrait constituer le matériel de base pour la compilation et publication d'un dictionnaire des devinettes et proverbes en fan. De tels ouvrages de référence font encore défaut aux communautés linguistiques Fan.

En conclusion, tous les dictionnaires disponibles dans la langue fan sont des dictionnaires de traduction et centrés principalement sur le français. En outre, ils comportent un certain nombre de lacunes qui doivent être comblées. La plupart de ces travaux ne marquent pas les tons ainsi que la prononciation de la langue.

Grosso modo, ces lacunes s'appliquent également à l'unité-langue qui fait l'objet de la section suivante.

3. Le Groupe Myene

L'omyene est la langue maternelle de 5 pour cent de la population gabonaise. Les parlers de cette unité-langue sont principalement localisés dans trois provinces du Gabon: la province de l'Estuaire, le Moyen-Ogooué et l'Ogooué-Maritime. L'enenga est parlé dans le nord-est de Lambaréné, à proximité du lac Zilé. Le ghalwa se rencontre autour des lacs Onangué, Avanga, et Ezanga et sur l'île de Lambaréné. Le mpongwe est parlé à Libreville ainsi qu'à la Pointe Denis. Enfin, le nkomi est utilisé dans le sud-est de Port Gentil et dans les environs de la lagune du Fernand-Vaz.

Tous les travaux disponibles en littérature myene touchent aux domaines de la religion, l'ethnohistoire, la linguistique et la lexicographie. Les contributions religieuses incluent des traductions de la Bible, ainsi que des catéchismes, des recueils de cantiques et de prières (cf. Jacot 1894, Mission Catholique de Libreville 1894 et 1948). Ces travaux ont permis en outre de consolider l'œuvre d'évangélisation et de scolarisation (notamment grâce à la production de gram-

maires, syllabaires et abécédaires dans les langues locales) des missionnaires catholiques et protestants.

Comme ce fut le cas pour le *fan* ci-dessus, il n'est pas sans intérêt de souligner que les syllabaires et abécédaires en question disponibles en langue *omyeng* sont, en quelque sorte, des dictionnaires de prononciation. En effet, ils sont typiquement organisés de telle manière que les syllabogrammes (signes à valeur "phonétique") sont énumérés dans l'ordre vocalique a, i, u. Comparez les exemples suivants:

A a <i>A a</i>											
a	a	a	a	a	a	a	a	a	a	a	a
a	a	a	a	a	a	a	a	a	a	a	a
I i <i>I i</i>											
i	i	i	i	i	i	i	i	i	i	i	i
i	i	i	i	i	i	i	i	i	i	i	i
U u <i>U u</i>											
u	u	u	u	u	u	u	u	u	u	u	u
i	a	a	u	a	i	u	i	a	u	a	i
u	i	a	i	u	a	u	i	a	u		
a i u <i>A U I</i>											

Figure 2: Quelques syllabogrammes du ghalwa (extrait du manuel *Ezangô elônga zi galwa* 1920, cité dans Mayer 1990: 83)

Sur le plan ethnohistorique, un certain nombre de travaux d'ensemble abordent entre autres les questions liées à l'histoire des migrations, à l'organisation sociale et à la description physique du milieu dans lequel vivent les populations *Ngwe-Myeng* (la mère *Myeng*). Citons par exemple Haug (1903), Hauser (1954) et Raponda-Walker (1961). À côté de ces travaux de la première heure, figurent des travaux un peu plus récents, comme Ambouroué-Avaro (1981) et Ratanga-Atoz (1999). Le travail d'ensemble réalisé par Haug découle d'une observation minutieuse des peuples du Bas-Ogooué et de leurs mœurs. La contribution de Hauser constitue une étude de sociologie sur les *Omyeng* du Bas-Gabon. Enfin le travail de Raponda-Walker publié avec une introduction et des cartes de M. Soret contient des documents indispensables, non seulement sur l'histoire des *Ngwe-Myeng* mais également sur d'autres peuples du Gabon. À côté de ces travaux d'ensemble, il y a des études limitées à certains groupes ethnolinguistiques de l'entité *Myeng*. Citons par exemple les travaux de Gautier (1950, sur les *Mpongwe*), Buléon (1887, 1892, sur les *Nkomi*), Agondjo-Okawe (1967, traité

majeur de droit coutumier chez les Nkomi) et Gaulme (1976 et 1981, sur le royaume Nkomi).

Les travaux linguistiques en omyeng sont exogènes. Les travaux de Delorme (1877) et de Raponda-Walker (1930–1934, 1961, 1998) sont certes des travaux lexicographiques mais ils sont précédés d'une description grammaticale de l'omyeng. Quant aux travaux de description linguistique strictu sensu, on peut citer entre autres les études de Jacquot (1978, 1983), Mouguiama-Daouda (1991) et Kane (1998). Jacquot analyse le système phonologique et morphologique de l'omyeng. Plus tard, il réalise une étude d'ensemble sur les classes nominales dans les langues bantoues des groupes B10 (où figurent les six dialectes de l'omyeng), B20 et B30. Mouguiama-Daouda se focalise pour sa part sur les niveaux d'analyse phonologique et morphologique du dialecte mpongwe. Enfin, Kane aborde également la variante mpongwe mais ici sous l'angle de la comparaison avec le français, langue officielle. Cette étude contrastive sur le mpongwe et le français réalisée par Kane a été précédée par aussi bien des descriptions synchroniques que diachroniques sur l'omyeng. On peut citer Tchen-Damas (1990) et Gregoire (1990) sur la phonologie du mpongwe et du nkomi respectivement. Quant aux études diachroniques, la contribution de Rekanga (1986) sur l'omyeng-nkomi figure en bonne place. Tous ces travaux linguistiques sont une mine d'informations pour tout lexicographe qui s'intéresserait à l'omyeng.

La première publication omyeng qui porte le titre de dictionnaire est l'ouvrage intitulé *Dictionnaire français–mpongwe* par le R.P. Amable Delorme (1877). Ce dictionnaire qui couvre 354 pages, a été publié sous les auspices des Missionnaires de la Congrégation du Saint-Esprit et du Saint Cœur de Marie. Les lemmes sont présentés en français et accompagnés d'une traduction en mpongwe ainsi que de quelques exemples. L'ouvrage de Delorme est suivi quelques années plus tard de la publication du *Dictionnaire mpongwè–français* de Gachon (1881). L'ouvrage couvre 287 pages et se compose de deux sections. La première présente la grammaire du mpongwe, tandis que la deuxième contient le dictionnaire à proprement parler. Les lemmes sont classés par ordre alphabétique et selon la tradition du mot.² Le dictionnaire a été une grande contribution à la traduction par Gachon de la *Bible gnango ine agamba mi re tendo pa gou'ejango j'agnambie ji felio né TESTAMAN NOUNGOU NI TESTAMAN GNONA gou'inongo gni mpongoue* publiée en 1891.

Le premier ouvrage majeur avec l'omyeng et le français ne voit le jour qu'en 1934. Il s'agit du *Dictionnaire mpongwé–français* bilingue de l'Abbé Raponda-Walker (1930–1934) qui paraît chez les Presses de la Libre Lorraine à Metz. Le même auteur achève quelques années plus tard un ouvrage dans le sens inverse portant le titre *Dictionnaire français–mpongwé* (en abrégé DFM dans le reste de l'article) publié par l'Imprimerie Saint Paul à Brazzaville. Cette version du dictionnaire a été rééditée depuis sous les auspices de la Fondation Raponda Walker en 1995. La dernière version du travail comporte environ 8 000 articles classés par ordre alphabétique selon la tradition du mot. La première par-

tie de l'ouvrage est le dictionnaire à proprement parler, tandis que les posttextes ou deuxième partie du dictionnaire contiennent un exposé de la grammaire du mpongwe. Les deux volumes du dictionnaire (celui de 1930–1934 et celui de 1961) constituent indéniablement un dictionnaire complet. Ils se distinguent de leurs prédécesseurs par une élaboration soignée et par une grande richesse des mots-entrées ou lemmes. Dans la macrostructure, les substantifs prédominent, mais les autres classes de mots n'en sont pas pour autant négligées.

Communiquer	V. A. (<i>transmettre</i>) nóza; pa. <i>Communiquer une lettre, une maladie, nóz'ezango; nóze nkani. Communiquer un ordre, bol'okenge. (Correspondre). Communiquer par lettre, tendane yango; une nouvelle, myeze ntsango. (Aboutir) punda, pókósa. Le Fernand-Vaz communique avec l'Océan, Eliwe-Nkómi zi punda go ntsuwa. Ce chemin communique avec la plaine, mpónó yinó yi pókósa g'orove. Se communiquer, nógana. Le pian se communique, abukwè mi nógana.</i>
-------------	---

Exemple textuel 1: Article communiquer (extrait de DFM 1995: 119)

En résumé, dans cet état des lieux sur la littérature myene, il apparaît clairement que le dialecte le plus abondamment documenté est le mpongwe. La raison la plus évidente de cet état de fait est la position géographique. En effet, situées sur le littoral, les populations Mpongwe furent les premières à entrer en contact avec les explorateurs européens. Le domaine que nous allons à présent examiner fournit une littérature moins riche que le précédant ainsi qu'une matière lexicographique quasi inexistante.

4. Le Groupe Mekana-Menaa

Les parlers qui constituent cette unité-langue sont disséminés dans six provinces du Gabon, à savoir: l'Estuaire, le Haut-Ogooué; le Moyen-Ogooué, la Ngounié, l'Ogooué-Lolo et l'Ogooué-Ivindo. L'akele est parlé à Lambaréné et Sindara, l'ungom au nord de Franceville, à Koulamoutou et aux environs de Mékambo. Le lisighu et le metombolo se rencontrent à Lastoursville, le seki à Cocobeach. Le shake est présent à Booué et Lastoursville, le tumbidi à Mbigou et Malinga. Le lendambomo est parlé à Booué, Okondja ainsi qu'au sud de Mékambo. Enfin, le wumpfu se rencontre à Mbigou, Malinga et au nord de Franceville.

L'essentiel des travaux sur le groupe mekana-menaa couvre les domaines de l'ethnohistoire et la linguistique. Parmi les travaux consacrés à l'ethnohistoire mentionnons ceux d'Avelot (1911) et Deschamps (1962). Dans le premier essai, l'auteur décrit avec minutie les pratiques religieuses chez les populations de l'ethnie Akele. Le second essai est un ouvrage de référence en ethnohistoire.

En effet, il contient les traditions orales recueillies sur l'ensemble des ethnies du Gabon. D'une manière générale, les migrations des ethnies de cette unité-langue ne sont pas très bien connues. Toutefois, pour les peuples Ndambomo, Wumpfu et Shake entre autres, les récits oraux mentionnent avec insistance deux faits, à savoir: la guerre de pupu et la querelle des ancêtres à propos d'une antilope prise au filet au cours d'une partie de chasse en groupe. Ces faits sont à l'origine de la dispersion des ethnies précitées.

En proie également à la dispersion suite à la même guerre de pupu, le peuple Akele venu du Nord-Est va essaimer les provinces du Haut-Ogooué (l'ethnie Mbanwe), la Ngounié (les ethnies Tumbidi et Metombolo) et l'Ogooué-Lolo (les ethnies Sighu et Bungom). Dans leur migration, les Bakele croisent le chemin des Beseki. S'engage alors une épreuve de force qui va contraindre les Beseki à se scinder en quatre groupes qui occuperont respectivement la Noya jusqu'à l'Estuaire de la Mondah et ses affluents, Libreville, la zone littorale du Como et l'Ogooué.

L'ensemble des travaux sur les parlers du groupe mekana-menaa s'inscrivent dans la perspective de la linguistique synchronique et inclut la contribution des auteurs suivants: Jacquot (1983), Ondo-Mebiame (1989) et Manfoumbi (1989). L'ouvrage de Jacquot (qui s'inscrit dans la longue lignée des travaux de classification sur les langues Bantu) examine le système des classes nominales dans les langues Bantu des groupes B10, B20, et B30. Pour chacune des langues retenues dans ce groupe, l'auteur propose un tableau des classificateurs dans lequel les classes s'associent éventuellement par paires en opposition singulier-pluriel pour former des genres. Les nombreux commentaires qui sous-tendent ces travaux font de cet ouvrage un outil morphologique indispensable à tout chercheur intéressé par les parlers de l'unité-langue mekana-menaa qui appartiennent majoritairement au groupe B20 de Guthrie. Il n'est pas inutile de signaler que ce travail a fait l'objet d'une critique dans Blanchon (1988).

Les deux dernières études se penchent respectivement sur la phonologie et la morphologie du seki et du shake. La première partie du travail d'Ondo-Mebiame est consacrée à la phonologie paradigmatique (tonèmes, voyelles et consonnes) et à la phonologie syntagmatique (étude des structures syllabiques). Dans la seconde partie de l'essai, le descripteur se propose de prouver l'appartenance du seki au domaine Bantu. Pour y arriver, il sollicite les critères élaborés par Guthrie (1948). Cette partie s'achève sur la présentation du système nominal du seki qui compte douze classes pour onze genres. On observera que la démarche de l'auteur épouse dans ses grandes lignes les principes développés par les linguistes du cercle de Prague. Avec cependant une préférence accordée au système théorique du fonctionnalisme martinétien.

Manfoumbi épouse la même démarche classique: au niveau d'analyse phonologique, l'auteur procède à l'identification des phonèmes du shake en recourant notamment au jeu des paires minimales et à la commutation. La composante descriptive morphologique aborde pour sa part les classes nominales du shake. In fine, mentionnons que ces deux mémoires de maîtrise ont

fait l'objet de publication d'articles, respectivement Ondo-Mebiame et Nzang-Bié (1990) et Manfoumbi et Zambibonda (1990). L'ensemble de ces travaux linguistiques sont indispensables à la rédaction de la partie introductive du dictionnaire, parce que c'est cette dernière qui est censée présenter entre autres les informations phonologiques et morphologiques des parlers qui font l'objet de production dictionnaire. Cependant, force est de constater qu'il n'existe pas dans cette unité-langue de matière lexicographique stricto sensu. Qu'à cela ne tienne, la production d'une telle littérature n'est pas impossible à partir de ce qui existe déjà. À titre d'illustration, les travaux ethnohistoriques dans ce groupe pourraient après synthèse entrer en ligne de compte dans la rédaction des aperçus historiques et les présentations ethnohistoriques des populations utilisant les langues du groupe dit mekana-mena. Le regroupement, notamment, de toutes les dénominations des lieux d'importance historique, les noms des chefs les plus importants dans chaque groupe ethnolinguistique, etc. pourraient ainsi constituer le matériau de base des documents annexes susceptibles d'apparaître dans les posttextes des ouvrages de référence lexicographique. Par rapport à ce point, les dictionnaires et les encyclopédies doivent être considérés comme des véhicules de types textuels, c'est-à-dire que les dictionnaires renferment une pluralité de textes d'origine diverse. Les prétextes (ensemble des textes présentés avant la section A-Z ou nomenclature) abritent généralement le guide pour les utilisateurs ainsi que la mini grammaire de la langue traitée; par contre, les posttextes (ensemble des textes figurant après la section A-Z ou nomenclature) offrent généralement aux lexicographes une plateforme pour la présentation des proverbes, devinettes, anthroponymes, toponymes, etc. Ce sont là des possibilités que ne devraient pas sous-estimer les lexicographes travaillant sur les langues du groupe mekana-mena en particulier. L'unité-langue qui va suivre présente également une matière lexicographique relativement pauvre.

5. Le Groupe Mekona-Mangote

Les parlers qui composent cette unité-langue sont éparpillés dans cinq des neuf provinces que compte le Gabon, à savoir: l'Estuaire, le Haut-Ogooué, la Ngounié, l'Ogooué-Lolo et l'Ogooué-Ivindo. L'ikota est parlé à Booué et Makokou, le benga est signalé au nord de Libreville au Cap Estérias et à la Pointe Santa-Clara. Le shamayi se rencontre à Makokou et Okondja, le mahongwe et le bakola à Makokou. Enfin, le ndasha est parlé à Mbigou, Mandjaye et Okondja (les Bandasha sont également signalés en République du Congo-Brazzaville, plus précisément à Mossendjo et à Dolosie dans la région du Nyari, dans la commune de Moutamba).

L'essentiel de la littérature mekana-mangote couvre les domaines de la religion, l'ethnohistoire et la linguistique. Du point de vue religieux, mentionnons les contributions de Biton (1903) et Lamour (1936). Les ouvrages de ces deux auteurs ont une caractéristique commune: ce sont des livres de questions

ou catéchismes sur l'idéologie catholique en langue ikota. Ce qui distingue ces deux ouvrages est évidemment le fait que le premier est publié tandis que le second nous est parvenu uniquement sous sa forme manuscrite.

Nombre d'ouvrages ethnohistoriques existent dans ce groupe. Mentionnons notamment le travail de Even (1937) sur les ethnies Obamba et Ndasha. Pour l'objet qui nous occupe, nous nous intéresserons à l'ethnie Ndasha. Un des chefs historiques de ce groupe, Oloupi, servit d'informateur à l'auteur, à l'époque où ce dernier occupait les fonctions d'administrateur adjoint aux colonies. Dans cet ouvrage, l'auteur décrit avec diligence les rites de passage dans les sociétés initiatiques suivantes: Mungala, Ngoyi, et Lishimbu. En outre, nous trouvons le récit de la migration des ethnies de ce groupe chez Deschamps (1962). Au chapitre 7 de son ouvrage, l'auteur mentionne deux faits récurrents dans les traditions orales: la guerre de pupu et la querelle des ancêtres à propos d'une antilope prise au filet. Ces faits sont à l'origine de la dispersion des ethnies précitées.

Le pays d'origine des Benga se situerait aux confins du Cameroun, de Centrafrique et du Congo. Sous la pression des Bekwel et des Bafan, les Benga ainsi que les Kota se dirigèrent vers le Gabon. C'est dans le Haut-Ivindo que les deux peuples se séparèrent à la suite de la guerre de pupu. Après cet épisode guerrier, les Benga se dirigèrent en ordre dispersée vers le Cameroun, la Guinée-Équatoriale et le Cap Estérias ainsi que la Pointe Santa-Clara (qui constituent avec Cocobeach les principaux foyers de concentration des Benga au Gabon).

Du point de vue linguistique, les parlers benga et ikota apparaissent comme les plus connus et les plus abondamment documentés. Pour ce qui est du benga, mentionnons les contributions des auteurs suivants: Mackey (1855), Meinhof (1889-1890), Perez et Sorinas (1928) et Essono (1988). Il convient de s'arrêter sur le dernier auteur dont le travail s'inscrit dans le fonctionnalisme. Ce travail comporte deux parties: la phonologie et la morphologie. Dans la première composante descriptive, l'auteur procède à l'identification des phonèmes du benga. Dans le second niveau d'analyse, le descripteur établit d'abord l'appartenance du benga au domaine Bantu en s'appuyant sur les critères élaborés par Guthrie (1948). Ensuite, l'auteur s'attèle à l'examen des classes nominales. L'analyse établit vingt-quatre classes nominales qui s'associent par paires en opposition singulier-pluriel pour former huit genres. Ce travail a fait l'objet de la publication d'un article "Benga" (1990) par le même auteur dans la *Revue Gabonaise des Sciences de l'Homme*. S'agissant de l'ikota, l'article d'Hombert (1990) porte sur la phonologie paradigmatique (systèmes vocalique et consonantique) de ce parler. Cet article s'achève sur un inventaire des symboles retenus pour l'Alphabet Scientifique du Gabon (ASG) ainsi que quelques proverbes. Pour l'ikota, il faut mentionner le travail de Gamille (1997) dont l'intérêt est beaucoup plus sociolinguistique que linguistique stricto sensu. Ce travail est abordé avec une arrière-pensée comparative: l'auteur examine un ensemble de textes kota du Gabon et du Congo sous l'angle de la variation. In fine, men-

tionnons les contributions des auteurs suivants: Sima Mvé (1996) et Mavoungou (1998) sur respectivement le mahongwe et le ndasha. Sur un plan morphologique, l'ouvrage de Jacquot (1978) que nous citons plus haut a valeur de source. Le seul ouvrage à caractère lexicographique dont nous disposons sur ce groupe est du R.P. Perron (1964). Il s'agit d'un lexique bilingue français-ikota publié en deux volumes sur l'initiative de la Mission Catholique de Makokou. Comme informations microstructurelles, ce bilingue est une reprise des notes manuscrites du R.P. Lamour. Comme ce fut le cas pour les autres unités-langues, les données disponibles sur les plans ethnohistorique et linguistique notamment constituent une matière dictionnaire non négligeable. Les travaux lexicographiques en cours au Gabon gagneraient à inclure les langues du groupe dit mekonangote. L'unité-langue qui va suivre présente une matière lexicographique relativement plus étoffée.

6. Le Groupe Membe

Les parlers de ce groupe se retrouvent principalement dans trois des neuf provinces du Gabon, à savoir: la Ngounié, la Nyanga, l'Ogooué-Ivindo et l'Ogooué-Lolo. Le ghetsogho est parlé à Mouila, Moabi et Sindara. Le ghepinzi et le ghevhiya sont localisés à Mouila et le ghevho vhe à Koulamoutou. Le ghehimbaka, l'ebongwe et le kota-kota se rencontrent principalement à Booué et à Mimongo.

Tous les travaux disponibles en littérature membe touchent aux domaines de la linguistique et de la religion. Ces travaux, qui contiennent beaucoup d'informations concernant le lexique ainsi que la grammaire des parlers de ce groupe, peuvent être employés comme point de départ pour la production de nouveaux dictionnaires. Nombre de travaux linguistiques portant sur les parlers de ce groupe ont été présentés et soutenus à l'Université Omar Bongo ou faits par des étudiants (gabonais ou non) dans les universités d'Europe. Mentionnons les contributions de Marchal-Nasse (1979 et 1990, sur le ghetsogho), Idiata (1989, sur le ghepinzi), Blanchon (1988, sur le ghevhiya) et Van der Veen (1991, 1992, 1999, sur le groupe B30 et le ghevhiya).

Les études qui ont spécifiquement traité des questions lexicographiques dans cette unité-langue incluent des travaux des auteurs suivants: Raponda-Walker (n.d.) et Bodinga-bwa-Bodinga et Van der Veen (1990). La contribution de Raponda-Walker est un dictionnaire bilingue intitulé *Dictionnaire getsogo-français*. L'ouvrage contient plus de 5 000 articles couvrant 237 pages. Des lemmes en ghetsogho sont proposés avec des traductions en français ainsi que quelques exemples. L'auteur a également produit une édition du dictionnaire dans le sens inverse avec le français comme langue source et le ghetsogho comme langue cible. L'article de Bodinga-bwa-Bodinga et Van der Veen "Plantes utiles des Evia" traite des noms de plantes ainsi que de leurs propriétés médicinales. Cette étude rejoint dans ses grandes lignes l'ouvrage *Les plantes utiles du Gabon* publié en 1961 par Raponda-Walker et Sillans et contient envi-

ron 608 items lexicaux en ghevhiya. Les entrées du lexique apparaissent en trois colonnes. Les noms scientifiques des plantes sont présentés dans la première colonne dans l'ordre alphabétique. Les noms en ghevhiya apparaissent dans la deuxième colonne. La dernière colonne traite des propriétés thérapeutiques des plantes énumérées. Ceci est illustré par l'exemple suivant:

Abrus precatorius	dì-ndèndè	Propriétés adoucissantes (feuilles), soins de la voix (chanteurs)
-------------------	-----------	---

Exemple textuel 2: Article **Abrus precatorius** (extrait de Bodinga-bwa-Bodinga et Van der Veen 1990: 30)

Enfin, la collaboration de Van der Veen et Bodinga-bwa-Bodinga inclut également la production d'un dictionnaire bilingue: *Gedandedi sa Geviya/Dictionnaire Geviya-Français* (2002, en abrégé DGF dans le reste de l'article) paru aux Éditions Peeters. Ce dictionnaire, fruit d'une longue collaboration scientifique fructueuse entre Van der Veen et Bodinga-bwa-Bodinga, marque le début d'une nouvelle ère dans la compilation des dictionnaires au Gabon (Mavoun-gou 2004). La macrostructure du DGF contient environ 6 200 articles classés par ordre alphabétique sous la lettre initiale du thème de chaque lemme. Les unités de traitement sont précédées de leurs préfixes respectifs entre parenthèses d'une part, et sont d'autre part suivies d'une indication de leurs schèmes tonaux respectifs accompagnée d'une spécification de la catégorie grammaticale (partie du discours) ainsi que du numéro de classe auxquels appartiennent les lemmes. Considérons l'exemple suivant:

(mo-)gadi	H n 1/2 • 1°épouse. <i>Tsikà mógàdì ámé wá mósòngá nà lélé.</i> Laisse ma première épouse tranquille. <i>Mògàdì á ngà.</i> Femme d'autrui. Voir mo-geto. • 2°(fig.) femelle (d'un animal) <i>Mògàdì á ngòdè.</i> La femelle du potamochère. <i>Mògàdì á èd'èmbè.</i> Brébis. <i>Mògàdì á gèvòlàngò.</i> Cane.
-----------	---

Exemple textuel 3: Article (mo-)gadi (extrait de DGF 2002: 136-137)

Après la parution du DGF, le ghevhiya, compte tenu de la sérieuse menace de sa disparition³, pourrait s'enrichir avec la publication d'autres ouvrages de référence lexicographique notamment les dictionnaires syntagmatiques, c'est-à-dire les dictionnaires des proverbes et devinettes. À ces dictionnaires, on pourrait également ajouter des dictionnaires de spécialité dit dictionnaires LSP couvrant notamment le domaine de la pharmacopée traditionnelle (mais ce n'est pas là l'objet de notre discussion). Les publications suivantes pourraient constituer la matière lexicographique de tous ces dictionnaires: Bodinga-bwa-Bodinga (1969) et Bodinga-bwa-Bodinga et Van der Veen (1993, 1995). Après la parution du DGF, l'autre travail de référence lexicographique pour ce groupe est la contribution de Mickala Manfoumbi (2004) qui s'intitule *Lexique pove-français/français-pove* (en abrégé LPFFP dans le reste de l'article). Ce lexique présente une

nomenclature d'environ 8 227 mots répartis sur 725 pages. Cette nomenclature déploie ses données en quatre colonnes. Dans la section *pove-français* en particulier, la première colonne identifie les lemmes (classés par ordre alphabétique sous la lettre initiale du thème de chaque mot) ainsi que les éléments de formation (affixes et formants savants) de même que les variantes des mots-entrées ou vedettes. La deuxième colonne donne les indices ou numéros de classe. La troisième présente la forme entière des lemmes selon l'opposition singulier-pluriel et la quatrième colonne donne la traduction en français des lemmes proposés. L'auteur simplifie ses données dans le souci majeur de fournir un manuel de première initiation. Il y a donc chez lui un souci constant d'adapter l'ouvrage au public. Le lecteur aura une idée du type de répertoire dont il s'agit d'après le bref fragment suivant:

+ sòlà	3, 4	musola misola	"purée, bouillie"
+ sòlè-sòlè	7, 8	gesolesole bisolesole	"instigateur"
+ sòlò	5, 6	esolo masolo	"détection, flair"

Exemple textuel 4: Articles **+ sòlà**, **+ sòlè-sòlè** et **+ sòlò** (extrait du LPFFP 2004: 312)

En revanche dans la section *français-pove*, les items lexicaux en français et en italique apparaissent dans la première colonne. La deuxième colonne présente leurs équivalents pove en gras et sous la forme du thème. Les indices de classe apparaissent dans la troisième colonne et les items lexicaux correspondant aux thèmes en pove apparaissent dans la dernière colonne sous leur forme complète. La typographie est soignée, claire et aérée et elle facilite la lisibilité. Ce lexique qui frappe par sa richesse et sa variété est le premier ouvrage de référence sur le pove. C'est dire tout le symbole qu'il représente pour la communauté Pove toute entière. L'unité-langue qui va suivre présente une matière lexicographique un peu plus importante.

7. Le Groupe Merye

Les parlers qui constituent cette unité-langue se rencontrent dans quatre des neuf provinces que compte le pays, à savoir: la Ngounié, la Nyanga, l'Ogooué-Maritime et l'Ogooué-Lolo. Le ghisira est parlé dans les localités de Fougamou, Mandji et Ndougou, le ghivharama à Setté-Cama et Gamba. Le ghivhungu est signalé à Mandji, Yetso et Moabi. Le yipunu est parlé à Mouila, Tchibanga, Ndendé, Mabanda et Moabi. Le yilumbu est parlé le long de l'axe routier Tchibanga-Mayumba ainsi qu'à Gamba et Setté-Cama. Le yisangu se rencontre dans les localités de Mimongo, Mbigou, Koulamoutou jusqu'à Iboundji. Le ngubi est

parlé dans la lagune Iguéla, le civil à Mayumba et Ndindi et le yirimba à Moabi. Enfin, le yighama est parlé entre Mayumba et Tchibanga. Signalons en outre que quelques parlers de ce groupe se rencontrent également dans quelques pays voisins du Gabon. C'est le cas du yipunu et du yilumbu que l'on rencontre au Congo-Brazzaville et au Congo-Kinshasa d'une part; d'autre part, l'on trouve également le civil au Congo-Brazzaville (Pointe Noire) et dans l'enclave du Cabinda.

De façon générale, la littérature merye couvre les domaines de la religion, l'ethnohistoire, la littérature orale et la linguistique. Parmi les auteurs de productions religieuses, citons Garnier (1897, 1903, 1904; 1900) et Murard (1903, 1903a). Le premier est l'auteur de quatre essais consacrés au yilumbu (variante dialectale de la Nyanga, lagune Banio) et au civil, à savoir: *Katesisa i gheghe nesi ma loghi ma dzambi mu mbembo i-lumbu* (1897), *Mu mbembo kivili ki mayumba* (1903), *M'ambu ma nzambi mo make mu katesisa* (1904) et *Syllabaire i-lumbu keti miganda mio mi teti mi u ranganga mu mbembo i-lumbu* (1900). Les trois premiers ouvrages ressortissent de la liturgie tandis que le dernier constitue un exercice scolaire. Le second, à la différence du premier, est l'auteur d'un ouvrage en deux volumes consacré au yilumbu parlé dans la province de l'Ogooué-Maritime (cantons de Ndugu-Lagune et Basse-Nyanga). Il s'agit de *Katsisu i neni i rendilu mu mbembo bis'Sette-Cama* (1903), et *Katsisu i keki i rendilu mu mbembo bis'Sette-Cama* (1903a). Mentionnons également pour yilumba le travail *Catéchisme de la doctrine chrétienne = Katsisu mu mbembo hi lumba hi bis'sette-cama* (Mission Catholique de Sette-Cama 1933). Les autres publications à caractère biblique et liturgique sont consacrées aux parlers yipunu et ghisira. Nous avons respectivement les ouvrages: *Cantiques = Gnimbu tsi niambi* (1903b) ainsi que *Catechisme gisira* (1924). Enfin, pour ce dernier ouvrage consacré au ghisira mentionnons également les contributions de Buléon (1899) sur ce même parler. L'ensemble de ces travaux qui vont du genre monolingue aux productions interdialectales constitue une matière lexicographique non négligeable.

Nombre de travaux linguistiques existent dans cette unité-langue. Citons par exemple les contributions des auteurs suivants sur le ghisira, le yipunu, le yisangu et le civil respectivement: Koumba (1990), Mouguiama-Daouda (1991) et Dodo-Bounguedza (1992); Nsuka-Nkutsi (1980) et Mickala-Manfoumbi (1986); Ondo-Mebiame (1989), Idiata (1993, 1998) et Mickala-Manfoumbi (1989); Ndamba (1977), Mabika Mbokou (1999) et Ndinga-Koumba-Binza (2000).

Les travaux lexicographiques à proprement parler dans cette unité-langue débutent avec la publication du bilingue *Dictionnaire vili-français* par le R.P. Marichelle (1902). Ce dictionnaire, qui couvre 224 pages, a été publié à Loango. Dans la préface, l'auteur précise que la parution du *Dictionnaire vili-français* vient compléter l'œuvre du R.P. Derouet (n.d.), auteur du *Dictionnaire français-vili*. L'ouvrage de Marichelle comporte un bref aperçu du parler civil. Les articles du dictionnaire fournissent les données lexicographiques suivantes: une unité de traitement en civil, une traduction en français ainsi que quelques exemples. Comparez l'exemple suivant:

<p>Dasuka, vn. P. <i>isi</i>, — Se mettre en colère, se fâcher, s'irriter, se monter. <i>I be n'-kamba, u a dusuka to</i>, je lui ai dit la chose, mais il ne fait que se fâcher. <i>Lemboa ku</i> —, ne te fâche pas.</p>
--

Exemple textuel 5: Article *Dasuka* (extrait de Marichelle 1902: 46)

Mis à part les travaux lexicographiques sur le *civili*, on note également quelques productions à caractère dictionnaire en *yisangu*, *yipunu* et en *ghisira*. Naidailac (1992, 1992a) en collaboration avec l'équipe de l'ALGAB⁴ a produit un travail lexicographique intitulé *Lexique isangu-français*. Idiata (1993) aborde également quelques aspects des lexiques spécialisés en *yisangu*. Quant au *yipunu*, il existe une littérature lexicographique pionnière réalisée par exemple par le R.P. Bonneau (1956) ou encore par le Christian Missionary Alliance (CMA) sous les auspices de l'Église Évangélique du Sud-Gabon (1966). En dehors de ces travaux de la première heure, il y a un ensemble de travaux lexicologiques et lexicographiques relativement plus récents. Citons par exemple Rittaud-Hutinet (1980) et Kwenzi-Mikala (1989, 1990). Pour ce dernier auteur, il faut souligner l'existence d'un ensemble de travaux sous presse sur notamment des questions d'anthroponymie (Kwenzi-Mikala 1990a) et des lexiques spécialisés (qui ne seront pas abordés ici). Enfin, pour le *yipunu* citons un travail lexicographique sous forme électronique, il s'agit *Dictionnaire informatisé du pounou* par Blanchon (cité dans Blanchon 1997a). Ce dictionnaire compte pour le moment 4 272 lemmes, accompagnés de leurs étymons.

Pour le *ghisira*, l'ouvrage lexicographique sur lequel il convient de s'arrêter s'intitule *Petit dictionnaire français-ghisira-ngosi et éléments de verbes* rédigé par une religieuse de la congrégation de l'Immaculé Conception: la sœur Lemoine (1994). Comme son titre l'indique, cet ouvrage est un lexique présentant une nomenclature de 2 619 lemmes. À la suite du lemme ou unité de traitement (en français et en gras), le traitement microstructurel présente les données suivantes: la traduction en *ghisira* du lemme français (il s'agit précisément de l'unité lexicale *ghisira* au singulier suivie du morphème du pluriel) et éventuellement d'autres unités lexicales en français (en italique) appartenant au même champ sémantique que le lemme avec leurs équivalents en *ghisira*. Comparez l'exemple suivant:

<p>mouton: didombi-ma <i>bélier</i>: didombi di mulumi <i>brebis</i>: mwana didombi- dindombi di mugatsi</p>
--

Exemple textuel 6: Article *mouton* (extrait de Lemoine 1994: 74)

Par rapport aux sources primaires du lexique, dans les notes introductives de l'ouvrage il est clairement stipulé que ce travail repose sur des exemples d'oc-

currences orales du ghisira, qui ont été collectées sur plusieurs années (sœur Lemoine a exercée pendant 24 ans à la Mission Sainte-Croix de Mandji-Ndolou) à partir de diverses sources (citons notamment les 600 lectures et évangiles des trois années liturgiques traduits par l'auteur pour le compte de la communauté Ghisira de Mandji-Ndolou). L'orthographe utilisée dans cette production à caractère dictionnaire repose largement sur l'ensemble des graphies du français.

Les items lexicaux de la nomenclature présentent des lacunes à la fois sur le plan segmental et suprasegmental. Pour tous les travaux précurseurs, réserves (voir Nyangone Assam et Mavoungou 2000 et Mavoungou 2001, 2001a, 2002) ont déjà été faites en ce qui concerne l'absence de tons dans les transcriptions écrites des mots (a priori sans indications tonales, on ne peut pas s'attendre à ce qu'un usager reproduise correctement le système phonétique des langues décrites). S'agissant des voyelles en particulier, avec le système orthographique du français comme point de référence, Lemoine essaie de distinguer en ghisira les voyelles qui sont phonologiquement distinctives à l'aide notamment de signes diacritiques (signes graphiques qui permettent de modifier la valeur de base d'un graphème). Par exemple, [ɛ] et [ɔ] sont toutes les deux représentées par è et ô.

Considérablement plus intéressant du point de vue métalexigraphique, le lexique de Lemoine présente un cadre structurel. En effet, la nomenclature est encadrée des prétextes et des posttextes. La grande richesse de l'ouvrage de Lemoine réside dans la présentation d'une variété de types textuels comme posttextes. Le texte central au niveau des posttextes sur lequel il convient de s'arrêter est intitulé **VERBES**. C'est dans ce texte que nous trouvons la conjugaison de certains verbes (entre autres, *être, voir, aimer, manger, boire, entendre* à l'indicatif présent, à l'imparfait, au passé, au futur, à l'impératif, au subjonctif, etc.) en ghisira. À la périphérie de ce texte central gravitent un ensemble d'autres textes et non des moindres. Le premier texte qu'il convient de mentionner s'intitule **MONNAIE**. Comparez l'illustration suivante:

5F	doli mosi
10F	doli bèi
15F	doli irèru
20F	doli ina
25F	doli iranu
30F	doli siamanu
35F	doli ina gu rèru
40F	doli gimana
45F	doli kambumosi
50F	digumi di doli
...	...

Exemple textuel 7: Posttexte **MONNAIE** (extrait de Lemoine 1994: 115)

Le tableau ci-dessus présente une liste des dénominations allant de 5F CFA à 200 000 F CFA. Cette liste pourrait être utile pour la réception et la production des textes aussi bien pour les locuteurs natifs que pour les apprenants du ghisira.

Le second posttexte qu'il convient de souligner est intitulé **DANSES**. Considérons l'exemple suivant:

Bwiti:	Pour trouver le coupable Initiation avec iboga: ama gwingène gu mbwiti: il rentre dans le bwiti avec l'iboga
Mwiri:	Pour un garçon devenu homme, fort, courageux
Nyemba:	Pour une fille devenue femme
Mugulu:	Pour se soigner avec la musique: Disigi
Mabanzi:	Pour avoir des enfants, chasser les esprits: avec l'iboga
Girina:	Pourquoi enfants morts avec l'iboga
Bilumbu:	Avec les esprits avec l'iboga
Mavassa:	Pour avoir des jumeaux
Divanda-mavanda:	Mutu a ne mavanda: homme avec fétiches
Mukwèlengi:	(nom d'un oiseau): garçons et filles
Ntsimba:	Réunion secrète du bwiti

Exemple textuel 8: Posttexte **DANSES** (extrait de Lemoine 1994: 129)

Dans l'exemple textuel ci-dessus, il est intéressant de noter que l'auteur ne se contente pas de présenter des traductions pour chaque unité lexicale ghisira. Le traitement inclut aussi bien les traductions que les données contextuelles (étiquettes) et les données cotextuelles (les exemples). L'accent dans cet extrait textuel est mis sur les items lexicaux ghisira ayant une forte charge sémantico-culturelle notamment les pratiques médico-magiques.

D'un point de vue graphique, il faut noter que le système orthographique du français a des signes pour la gémation ou redoublement consonantique, cependant ce n'est pas le cas des langues gabonaises et le ghisira en particulier. Sous l'influence de sa langue maternelle, Lemoine a écrit le *s* du ghisira comme le français *ss* qui est prononcé [s], par exemple *Mavassa* (au lieu de *Mavasa*). Pour ce qui est de la division des mots, les langues africaines ont deux traditions d'écriture distincte, à savoir: une écriture conjonctive et une écriture disjonctive. Dans une écriture conjonctive, les mots présentent une structure interne complexe (par exemple les préfixes verbaux, les formatifs, les radicaux et leurs extensions formant un tout). Par contre, une écriture disjonctive présente une structure interne plus simple: les affixes (préfixes, infixes et suffixes) des unités lexicales apparaissent séparément. Lemoine a fait le choix d'une écriture disjonctive (cf. *ama gwingène gu mbwiti*). Ce choix est beaucoup plus visible dans le troisième et dernier posttexte de notre étude: **PRIÈRES**. Comparez l'exemple suivant:

Taï étu ndya u gu yulu, dina dyagu di ki du misa, divinda dyagu di ki ruga, tondini agu i ki vagu, gu tsi na gu yulu. Du vègi ne muni bigui byètu bi tsufu tsyotsu, u ku du kébila gu péchés, tyètù; nève dua kébili, bisi ba du va mabi; u ya du sila gu fama gu péchés tsi dua sindigitsa, tu du kani gu mabi. Amen

Exemple textuel 9: Posttexte **PRIÈRES** (extrait de Lemoine 1994: 131)

Cet exemple textuel ajouté notamment aux 600 lectures et évangiles des trois années liturgiques traduits du français au ghisira par Lemoine pourraient entrer dans la constitution d'un corpus électronique ghisira d'où l'on pourrait dériver un ensemble d'ouvrages lexicographiques (dictionnaires monolingues, bilingues ou multilingues). L'unité-langue qui va suivre présente comme le groupe merye un fonds dictionnaire et d'intéressantes perspectives de développement de la discipline lexicographique.

8. Le Groupe Metye

Les parlers de ce groupe sont principalement dispersés dans quatre des neuf provinces du Gabon, à savoir: le Haut-Ogooué, la Ngounié, l'Ogooué-Ivindo et l'Ogooué-Lolo.

Le yinzebi est le parler le plus largement répandu du groupe et il est localisé dans des régions telles que Mbigou, Lebamba, Malinga, Lastoursville, Koulamoutou, Pana-Iboundji et Mounana-Bakoumba. Le yitsengi est parlé à Lastoursville, à Moanda et à Bakoumba, et le yimwele à Mbigou. Le yivhili est parlé à Booué et à Sindara, et liwanzi à Lastoursville et à Moanda. Le liduma est parlé à Lastoursville et le yibongo à Moabi.

La littérature metye couvre essentiellement les domaines de la religion et de la linguistique. Il y a toutefois un certain nombre de travaux qui proviennent de l'ethnohistoire et de la littérature orale mais dont nous ne parlerons pas ici. Les publications à caractère biblique et liturgique sont essentiellement consacrées à liduma. Citons par exemple *Catéchisme en langue adouma* (Dahin 1891) et *Bivovili bya bakretyeni ba baduma* (Reeb 1895). Les travaux linguistiques incluent pour leur part les contributions des auteurs suivants sur liduma, liwanzi et le yibongo respectivement: Mouélé (1997), Dodo-Bounguendza (1986), et Raponda-Walker (1937). À côté des études sur liduma, liwanzi et le yibongo entre autres, il y a un certain nombre de travaux sur le yinzebi présentés et soutenus à l'Université Omar Bongo ou dans les universités européennes. Citons par exemple: Chela (1998), Makoungou (1999), Ndoumou Mba (1999) et Marchal-Nasse (1989). Pour cette dernière contribution, en 1990, le même auteur publie dans la *Revue Gabonaise des Sciences de l'Homme* l'article "Yinzebi". Tous ces travaux aussi bien religieux que linguistiques dans les langues précitées constituent une matière lexicographique non négligeable.

Le premier travail en lexicographie dans ce groupe avec le français et liduma est une contribution du R.P. Dahin intitulé *Vocabulaire français-aduma* (1893). Le même auteur achève quelques années plus tard l'édition dans le sens inverse du même ouvrage: *Vocabulaire aduma-français* (1895). Ces lexiques constituent certainement un ouvrage complet en deux volumes. Les contributions du R.P. Dahin ont été suivies d'une publication modeste en un volume. Il s'agit d'un dictionnaire de poche du yinzebi intitulé *Petit dictionnaire bantou du Gabon: français-ndjabi; ndjabi-français* produit par Muroi (1989). Sur la même langue, Nadaillac a produit un *Lexique yinzebi-français* en collaboration avec l'équipe de l'ALGAB. Le dictionnaire incluant le yinzebi et le français comme langues paires est en préparation. Enfin, l'ouvrage lexicographique le plus important de ce groupe est certainement le *Dictionnaire ndumu-mbede-français* par Biton et Adam (1969). Cet ouvrage de Biton et Adam est intéressant parce qu'il n'implique pas que le français et l'un des parlers du Gabon comme c'est le cas dans la plupart des dictionnaires existants, mais il implique à la fois deux langues du Gabon (liduma et lateghe) et le français. Les sources primaires de ce dictionnaire multilingue sont probablement la *Grammaire composée mbede-ndumu-duma* qu'Adam a publié en 1954. Cependant, ce livre de grammaire doit être considéré comme seulement une des diverses sources de l'auteur. En effet, après 52 ans au Gabon — un pays où il s'est éteint en 1981 — la contribution d'Adam à la littérature orale gabonaise inclut plusieurs articles sur les différents parlers de la province du Haut-Ogooué. Ce dictionnaire de Biton et Adam constitue également une passerelle entre certaines langues du groupe dit metye et celles du groupe membre qui fait l'objet de la section suivante.

9. Le Groupe Membre

Les parlers de ce groupe sont principalement localisés dans la province du Haut-Ogooué. Le lembaama est parlé dans les régions de Franceville-Okondja-Akiéni, le lekanini à Franceville et à Boumango, le lateghe à Akiéni et à Léconi et le latsitseghé à Léconi. Le lindumu est utilisé à Franceville.

La littérature disponible sur ce groupe inclut des travaux religieux, des études de littérature orale ou encore à caractère ethnohistorique ainsi qu'un certain nombre de travaux linguistiques. Nous mettrons ici l'accent sur les travaux linguistiques.

Les travaux linguistiques dans ce groupe apparaissent déjà dans les premières études lexicographiques. Les travaux de Biton (1907) et Biton et Adam (1969) sont certes des travaux lexicographiques mais ils sont précédés d'une description grammaticale des langues faisant l'objet de l'étude. Les travaux réalisés dans le cadre de la linguistique moderne incluent les contributions des auteurs suivants sur le lateghe, le lekanini, le lembaama et le lindumu respectivement: Puech (1990), Kwenzi-Mikala (1990), Alihanga, Teta et Voltz (1990) et enfin Binga et Puech (1990). Les contributions de Binga et Puech sur le lindumu ont été précédées par un certain nombre de travaux: Mouélé (1988) et

Medjo-Mvé (1989). La contribution de Kwenzi-Mikala sur le *lekanini* a été suivie quelques années plus tard par le travail d'Evouna Ntoutoume (2003). Tous ces travaux présentent à un degré plus ou moins les éléments de phonologie et de morphologie des langues décrites. À ce titre, ils gagneraient à être pris en compte par les lexicographes travaillant sur lesdites langues.

La lexicographie *memberé* commence avec la publication du dictionnaire biscopal intitulé *Dictionnaire français-ndumu et ndumu-français* par le R.P. Biton (1907). L'ouvrage couvre 97 pages et est divisé en deux parties. La première partie présente au lecteur la grammaire du *lindumu*, tandis que la deuxième partie contient le dictionnaire à proprement parler. La première section du dictionnaire commence par des lemmes en français. Ensuite, le lexicographe renverse les langues dans la deuxième section du dictionnaire. Le dictionnaire en question a servi comme point de départ pour la production de matériel didactique religieux tels que le *Kateçism Ndumu* (1962) et quelques années plus tard, le *Dictionnaire ndumu-mbede-français et français-ndumu-mbede* produit par Biton en collaboration avec Adam en 1969. Pour rappel, la caractéristique de ce dictionnaire est qu'il implique trois langues (le français, *lindumu* (*ndumu*) et *lateghe* (*mbede*)) et que les langues sont renversées dans le seul et même dictionnaire. En outre, les lemmes sont extraits du fonds lexical des noms de plantes de la région de Franceville. Le prochain travail à examiner est un lexique interdialectal intitulé "Vocabulaire comparé des principaux dialectes ayant cours en Haut-Ogooué" publié par Castex en 1938. L'ouvrage est une contribution intéressante sur le domaine de la dialectologie (également appelé géographie linguistique). Par exemple, il pourrait être utilisé comme un travail de référence valable pour la production d'un dictionnaire multilingue ou *pan-dialectal* impliquant les différents parlers du groupe *memberé*. Par rapport au groupe *memberé*, les unités-langues *mekana* et *baka* quoique présentant une littérature lexicographique quasi inexistante ont de riches perspectives de développement.

10. Le Groupe Mekana

Les parlers de ce groupe sont principalement localisés dans la province l'Ogooué-Ivindo. Le *bekwil* (ou *bekwel*) est parlé à Makokou, le *shiwa* (ou *maki-na*) à Booué et le *mwesa* (ou *yesa*) à Mékambo. En ce qui concerne ce groupe, il reste encore beaucoup à faire. Cependant, les contributions de Puech (1990), Yembi Bouka (1995), Kwenzi-Mikala (1996: 10-16) et Mouloungui (1999) contiennent un aperçu de la grammaire et de la littérature orale du *shiwa*, du *bekwel* et du *mwesa* respectivement. En ce qui concerne les publications lexicographiques, rien n'est pour le moment disponible.

11. Le Groupe Baka

Le *baka* est la seule langue de ce groupe. Il est parlé à Minvoul, à Makokou et à

Belinga. La littérature disponible dans ce groupe inclut le travail de Mayer (1987) et de Mayer et Voltz (1990) dans le domaine de l'anthropologie, entre autres. Les activités lexicographiques restent à entreprendre.

12. En guise de conclusion

L'état des connaissances ci-dessus sur les travaux religieux, ethnohistoriques, linguistiques et lexicographiques sur les langues gabonaises montre clairement et de façon générale l'existence d'une riche littérature. Toutefois, une vue d'ensemble de cette littérature fait également apparaître des déséquilibres entre les langues et les unités-langues. Certaines sont de loin plus avancées que d'autres. Le traitement de chaque groupe varie en importance selon les données religieuses, ethnohistoriques, linguistiques et lexicographiques disponibles.

Ces dernières lignes ne sauraient à proprement parler constituer une conclusion. Si la parution de cet article suscite dans le pays une coopération beaucoup plus poussée entre chercheurs d'origines et de formations diverses (linguistes, sociologues, historiens, lexicographes, etc.), il aura gagné son pari. Ce propos est motivé par le constat suivant de Mavoungou (À paraître):

La majorité des sources de référence lexicographique existant dans les langues gabonaises auxquelles nous avons fait allusion plus haut ont requis des efforts considérables et de l'abnégation si on considère le fait que la plupart de ces travaux était le résultat d'une vie de travail par des auteurs isolés. Comme pour toute œuvre pionnière, ces dictionnaires, lexiques, encyclopédies de la première heure comportent des lacunes tant au niveau linguistique que métalexigraphique. Cependant, ces manquements sont excusables si l'on considère que les lexicographes de ces travaux précurseurs n'étaient ni linguistes, ni lexicographes qualifiés.

Les travaux lexicographiques sur les langues gabonaises publiés en ce début de troisième millénaire devraient être des ouvrages collectifs, c'est-à-dire le fruit d'un travail d'équipe, nouveau défi, dans ce monde qui ne cesse de se mondialiser.

Notes

1. Dans cet article l'orthographe officielle du Gabon est utilisée en lieu et place de la traditionnelle transcription semi-phonétique par Guthrie, Kwenzi-Mikala, et les autres. Les e, o et n souscrits représentent [ɛ], [ɔ] et [ɲ] respectivement, tandis que [ɣ], [ʃ] et [β] sont écrits gh, sh et vh respectivement. Le son phonétique [ə] est représenté par le symbole ə dans l'orthographe gabonaise. Toutefois pour des raisons phonologiques et pratiques, il est représenté par a comme dans Make et Ghisira.
2. Selon la tradition du mot, les items lexicaux sont présentés sous leur forme entière, par exemple leur préfixe plus le thème. Par contre, dans la tradition du thème les lexèmes sont lemmatisés sous le thème sans leurs préfixes.

3. Dans les notes introductives du DGF (2002: 4), les auteurs soulignent qu'"on peut estimer le nombre de bons locuteurs à une quarantaine seulement. Ce chiffre extrêmement bas indique sans ambiguïté aucune que le geviya est une langue dont la disparition est sur le point de s'achever."
4. *Atlas Linguistique du Gabon*. La réalisation de cet atlas échoit aux chercheurs du Laboratoire de Phonétique et de Linguistique Africaine (LAPHOLIA) de l'Université Lumière Lyon 2 sous la direction de Jean-Marie Hombert.

Références bibliographiques

- Adam, J.** 1954. *Grammaire composée mbede-ndumu-duma*. Brazzaville: Institut d'Étude Centrafricaines.
- Afane Otsaga, T.** 2004. *The Standard Translation Dictionary as an Instrument in the Standardization of Fang*. Thèse de Doctorat non publiée. Stellenbosch: Université de Stellenbosch.
- Agondjo-Okawe, P.** 1967. *Structures parentales gabonaises et développement*. Thèse de Doctorat non publiée. Paris: Université de Paris I.
- Alexandre, P.** 1961. Problèmes linguistiques des états négro-africains à l'heure de l'indépendance. *Cahiers d'études africaines* 2(6): 177-195.
- Alexandre, P.** 1967. *Langues et langage en Afrique Noire*. Paris: Payot.
- Alihanga, M., R. Teta et M. Voltz.** 1990. Lembaama. *Revue Gabonaise des Sciences de l'Homme* 2: 235-241.
- Ambouroué-Avaro, J.** 1981. *Un peuple gabonais à l'aube de la colonisation*. Présence Africaine. Paris: Karthala.
- Andeme Allogho, M.F.** 1983. *Esquisse phonologique du Nzaman de Makokou*. Mémoire de Licence non publié. Libreville: Université Omar Bongo.
- Avelot, R.** 1911. Notes sur les pratiques religieuses des Bakalè. *Bulletin et Mémoire de la Société d'Anthropologie de Paris* 6: 282-296.
- Balandier, G.** 1957. *Afrique ambiguë*. Paris: Plon.
- Balandier, G.** 1963. *Sociologie actuelle de l'Afrique Noire*. Paris: PUF.
- Bibang Meyé.** 2001. *Étude des compétences narratives chez les enfants bilingues Fang-Ntumu-Français d'Oyem (psycholinguistique): Étude transversale sur le développement de la temporalité et de la connectivité chez des sujets âgés de 7 à 12 ans*. Mémoire de Maîtrise. Libreville: Université Omar Bongo.
- Binga, H. et G. Puech.** 1990. Lendumu. *Revue Gabonaise des Sciences de l'Homme* 2: 243-246.
- Biton, A. (R.P.).** 1903. *Cantiques endumu-ambété-akota-alimi*. Mission Catholique de Brazzaville-Franceville. Nantes: Imprimerie Bourgeois.
- Biton, A. (R.P.).** 1907 *Dictionnaire français-ndumu et ndumu-français, précédé d'éléments de grammaire*. Nantes: Imprimerie Dupas.
- Biton, A. (R.P.).** 1962. *Kateçism Ndumu*. Rome: Société St Pierre Claver.
- Biton, A. (R.P.) et J. Adam (Mgr.).** 1969. *Dictionnaire ndumu-mbede-français et français-ndumu-mbede. Petite flore de la région de Franceville (Gabon). Grammaire ndumu-mbede*. Archevêché de Libreville. Bar-le-Duc: Imprimerie St Paul.
- Blanchon, A.** 1988. Tonalité des nominaux à thème dissyllabique dans le groupe bantou B20. *Pholia. Revue du Laboratoire de Phonétique et Linguistique Africaine* 3: 37-52.

- Blanchon, J.A.** 1988a. Une langue mixte en voie de disparition: le geviya. *Pholia. Revue du Laboratoire de Phonétique et Linguistique Africaine* 3: 53-69.
- Blanchon, J.A.** 1997. Les formes nominales de citation à préfixe haut en pounou (Bantu B43). *Journal of African Languages and Linguistics* 18(2): 129-138.
- Blanchon, J.A.** 1997a. *Dictionnaire informatisé pounou-français*. Manuscrit. Lyon: Université Lumière Lyon 2.
- Bodinga-bwa-Bodinga, S.** 1969. *Traditions orales de la race eviya*. Paris: T.M.T.
- Bodinga-bwa-Bodinga, S. et L.J. van der Veen.** 1993. Plantes utiles des Evia. *Pholia. Revue du Laboratoire de Phonétique et Linguistique Africaine* 8: 26-65.
- Bodinga-bwa-Bodinga, S. et L.J. van der Veen.** 1995. *Les proverbes evia et le monde animal, la communauté evia à travers ses expressions proverbiales*. Paris: L'Harmattan.
- Bonneau, J. (R.P.)** 1956. *Grammaire pounoue et lexique pounoue-français*. Brazzaville: Institut d'Études Centrafricaines.
- Buléon, J. (R.P.)** 1887. Le lac et la tribu des Nkomi. *Annales de la Propagation de la Foi*: 175-182.
- Buléon, J. (R.P.)** 1892. Le lac et la tribu des Nkomi. *Annales de la Propagation de la Foi*: 437-448.
- Buléon, J. (R.P.)** 1899. *Catéchisme de la Doctrine Catholique en langue Eshira*. Paris: Mission Catholique de Libreville (St Croix).
- Castex (Dr.)** 1938. Vocabulaire comparé des principaux dialectes ayant cours en Haut-Ogooué. Essai de classification. *Bulletin de la Société des Recherches Congolaises* 26: 23-54.
- Chela, P.** 1998. *Étude des emprunts linguistiques du nzébi, langue bantu du Gabon (B52), au français; procédés d'intégration*. Mémoire de Maîtrise. Libreville: Université Omar Bongo.
- Cinnamon, J.** 1990. Fan (nzaman). *Revue Gabonaise des Sciences de l'Homme* 2: 175-183.
- Dahin (R.P.)** 1891. *Catéchisme en langue adouma (Haut-Ogooué) = Mambo ma ndjambi ghu evovili si a tshenge a baduma*. Rixheim: Imprimerie A. Sutter.
- Dahin (R.P.)** 1893. *Vocabulaire français-aduma*. Kempten (Bavière): Kosel.
- Dahin (R.P.)** 1895. *Vocabulaire aduma-français*. Kempten (Bavière): Kosel.
- Delorme, A.** 1877. *Dictionnaire français-mpongwè*. Paris: Missionnaires de la Congrégation du St Esprit et du St Cœur de Marie.
- Derouet (R.P.)** n.d. *Dictionnaire français-vili*. Loango: Imprimerie de la Mission.
- Deschamps, H.** 1962. *Traditions orales et archives au Gabon*. Paris: Berger-Levrault.
- Dodo-Bounguedza, E.** 1986. *Perspectives linguistiques pour l'enseignement du Liwandji (cours préparatoires première et deuxième année)*. Mémoire de Maîtrise. Libreville: Université Omar Bongo.
- Dodo-Bounguedza, E.** 1992. *Esquisse phonologique et morphologique du gisira, langue bantoue (B41) du Gabon*. Thèse de Doctorat non publiée. Bruxelles: Université Libre de Bruxelles.
- Ella, E.M.** 2000. *Étude linguistique d'une oeuvre littéraire: Un moët de Zué Nguema, chant épique Fang recueilli par Herbert Pepper*. Mémoire de Maîtrise. Libreville: Université Omar Bongo.
- Ella Ella, E.** 2000. *Étude contrastive des pronoms relatifs en Français et en Ntumu*. Mémoire de Maîtrise. Libreville: Université Omar Bongo.
- Essono, E.** 1988. *Esquisse phonologique et morphologique du Benga, langue bantu du Gabon*. Mémoire de Maîtrise. Libreville: Université Omar Bongo.
- Essono, E.** 1990. Benga. *Revue Gabonaise des Sciences de l'Homme* 2: 129-139.
- Évangile de Jean: traduction dans la langue des Fan*. 1910. Paris.
- Évangile de Luc: traduction dans la langue des Fan*. 1910. Paris.
- Évangile de Matthieu: traduction dans la langue des Fan*. 1902. Paris.

- Even, A.** 1937. Les confréries secrètes chez les Babamba et les Mindassa d'Okondja. *Bulletin de la Société des Recherches Congolaises* 23: 31-133.
- Evouna Ntoutoume, F.** 2003. *Description phonologique du lekaniji (langue bantoue du Gabon B.66)*. Mémoire de Maîtrise. Libreville: Université Omar Bongo.
- Gachon, J.-B.** 1881. *Dictionnaire mpongwè-français, précédé des principes de la langue mpongwè*. Paris.
- Gachon, J.-B.** 1891. *Bible gnango ine agamba mi re tendo pa gou'ejango j'agnambie ji felio né TESTAMAN NOUNGOU NI TESTAMAN GNONA gou'inongo gni mpongoue*. Sente-Maria yi Gabon. Fribourg-en-Brisgau: Herber B. Libr. Edit. Pontifical.
- Galley, S.** 1889. *Nteni osi nteni Fañwe*. Talagouga: Imprimerie de la Mission.
- Galley, S.** 1964. *Dictionnaires Fang-Français et Français-Fang, suivi d'une grammaire Fang*. Neuchâtel: Henri Messeiller.
- Gamille, G.** 1997. *Le Mungala entre le Congo et le Gabon: Étude sociolinguistique de la variation d'un récit mythique*. Rapport de Licence. Libreville: Université Omar Bongo.
- Garnier, A.** 1897. *Katesisa i gheghe nesi ma loghi ma dzambi mu mbembo i-lumbu*. Loango: Imprimerie de la Mission.
- Garnier, A.** 1900. *Syllabaire i-lumbu keti mi ganda mio mi teti mi ranganga mu mbembo i lumbu*. Loango: Imprimerie de la Mission.
- Garnier, A.** 1903. *Mu mbembo kivili ki mayumbe*. Loango: Imprimerie de la Mission.
- Garnier, A.** 1904. *M'ambu ma nzambi mo make mu katesisa*. Loango: Imprimerie de la Mission.
- Gaulme, F.** 1976. *L'ancien pays de Cama, société et organisation politique du XVIe au XIXe siècle*. Thèse de Doctorat. Paris: Université de Paris V.
- Gaulme, F.** 1981. *Le pays de Cama. Un ancien État-côtier du Gabon et ses origines*. Paris: Karthala.
- Gautier, D. (R.P.)** 1950. *Étude historique sur les Mpongoués et les tribus avoisinantes*. Brazzaville: A.E.F.
- Gregoire, C.** 1990. Omyène (ɲkɔmi). *Revue Gabonaise des Sciences de l'Homme* 2: 273-280.
- Grimes, B.F. (Éd.)** 1996¹³. *Ethnologue: Languages of the World*. Dallas: Summer Institute of Linguistics.
- Guthrie, M.** 1948. *The Classification of the Bantu Languages*. Londres: Oxford University Press.
- Guthrie, M.** 1953. *The Bantu Languages of Western Equatorial Africa*. Londres: Oxford University Press.
- Guthrie, M.** 1969-1971. *Comparative Bantu: An Introduction to the Comparative Linguistics and Prehistory of the Bantu Languages*. 5 tomes. Farnborough: Gregg International.
- Haug, E.** 1903. Le Bas-Ogoué. Notice géographique et ethnographique. *Annales de géographie* 12: 169-171.
- Hauser, A.** 1954. Notes sur les Omyéné du Bas-Gabon. *Bulletin de l'Institut Français d'Afrique Noire (BIFAN)* 16(3-4): 402-415.
- Hombert, J.-M.** 1990. Ikota. *Revue gabonaise des Sciences de l'Homme* 2: 211-212.
- Hombert, J.-M. et A.-M. Mortier.** 1984. Mise à jour de la bibliographie publiée dans *Pholia. Revue du Laboratoire de Phonétique et Linguistique Africaine* 1: 165-187.
- Hombert, J.-M. et A.-M. Mortier.** 1990. Bibliographie des langues du Gabon. *Revue gabonaise des Sciences de l'Homme* 2: 335-355.
- Idiata, D.F.** 1989. *Éléments de phonologie du yepinzipinzi*. Mémoire de Maîtrise. Libreville: Université Omar Bongo.

- Idiata, D.F.** 1993. *Éléments de phonologie diachronique du isangu, correspondances Proto-Bantu-Isangu et lexique spécialisé*. Mémoire de D.E.A. Lyon: Université Lumière Lyon 2.
- Idiata, D.F.** 1998. *Quelques aspects de l'acquisition de la langue isangu par les enfants*. Vol. 1: *Étude des processus d'acquisition*; Vol. 2: *Les protocoles expérimentaux*; Vol. 3: *Esquisse de description grammaticale de la langue isangu*. Thèse de Doctorat (NR). Lyon: Université Lumière Lyon 2.
- Jacot, H.** 1894. *Ezangô elônga zi mpôngwe gw'itangula nl'itenda. Premier livre de lecture en mpongwè*. New York: Société américaine des Traités Religieux.
- Jacquot, A.** 1978. Le Gabon. Barreteau, D. (Éd.). 1978. *Inventaire des études linguistiques sur les d'Afrique noire d'expression française et sur Madagascar*: 493-503. Paris: CILF.
- Jacquot, A.** 1983. *Les classes nominales dans les langues bantoues des groupes B10, B20, B30 (Gabon-Congo)*. Travaux et documents de l'ORSTROM 157. Paris: ORSTROM.
- Kane, F.** 1998. *Étude contrastive du français et du mponwè. Approche phonétique et phonologique*. Mémoire de Maîtrise. Libreville: Université Omar Bongo.
- Koumba, V.** 1990. *Essai de phonologie du gisir (langue bantu du Gabon B41)*. Mémoire de Maîtrise. Libreville: Université Omar Bongo.
- Kwenzi-Mikala, J.T.** 1988. L'identification des unités-langues bantu gabonaises et leur classification internes. *Muntu* 8: 54-64.
- Kwenzi-Mikala, J.T.** 1989. Contribution à l'analyse des emprunts nominaux du yipunu au français. *Pholia. Revue du Laboratoire de Phonétique et Linguistique Africaine* 4: 157-170.
- Kwenzi-Mikala, J.T.** 1990. Lekaniji. *Revue Gabonaise des Sciences de l'Homme* 2: 227-234.
- Kwenzi-Mikala, J.T.** 1990a. L'anthroponymie chez les Bapunu du Sud-Gabon. *Pholia. Revue du Laboratoire de Phonétique et Linguistique Africaine* 5: 113-120.
- Kwenzi-Mikala, J.T.** 1996. *Parémies d'Afrique Centrale (Proverbes et Sentences)*. Libreville: Éditions Raponda-Walker.
- Kwenzi-Mikala, J.T.** 1998. Parlers du Gabon: classification du 11.12.97. Raponda-Walker, A. (Éd.). 1998: 217.
- Lamour (R.P.)**. 1936. *Petit catéchisme ikota*. Manuscrit. Libreville.
- Largeau, V.** 1901. *Encyclopédie pahouine. Eléments de grammaire et dictionnaire français-pahouin*. Paris: E. Leroux.
- Lejeune, L. (R.P.)** 1892. *Dictionnaire français-fang ou pahouin, précédé de quelques principes grammaticaux sur cette même langue*. Paris: A. Faivre/H. Teillard.
- Lemoine, G.-M. (Sœur)**. 1994. *Petit dictionnaire français-gisira-ngosi et éléments de verbes*. Libreville: Mission Sainte-Croix de Mandji-Ndolou.
- Mabika Mbokou, L.** 1999. *Les phénomènes analogiques en civil: Étude phonologique et morphologique*. Mémoire de Maîtrise. Libreville: Université Omar Bongo.
- Mackey, J.L.** 1855. *Grammar of the Benga Language*. New York: Mission House.
- Makoungou, E.** 1999. *L'étude comparée du lexique du français et des parlers de l'unité-langue metye; le yinzébi*. Mémoire de Maîtrise. Libreville: Université Omar Bongo.
- Manfoumbi, M.** 1989. *Éléments de phonologie du Saké í mèmbozù dialecte saké de Booué*. Mémoire de Maîtrise. Libreville: Université Omar Bongo.
- Manfoumbi, M. et M.-E. Zambibonda.** 1990. Jake. *Revue Gabonaise des Sciences de l'Homme* 2: 287-290.
- Marchal-Nasse, C.** 1979. *Esquisse de la langue tsoyo: phonologie et morphologie*. Mémoire de Licence Spéciale. Tervuren: Université Libre de Bruxelles.

- Marchal-Nasse, C.** 1989. *De la phonologie à la morphologie du nzebi, langue bantoue (B52) du Gabon*. Thèse de Doctorat de 3^{ème} cycle. Tervuren: Université Libre de Bruxelles.
- Marchal-Nasse, C.** 1990. γ etsɔɔ. *Revue Gabonaise des Sciences de l'Homme* 2: 185-188.
- Marchal-Nasse, C.** 1990a. Yinzébi. *Revue Gabonaise des Sciences de l'Homme* 2: 295-298.
- Marichelle, C.** 1902. *Dictionnaire français-vili*. Loango: Imprimerie de la Mission.
- Marling (R.P.)**. 1872. *Dictionnaire fang-français*. New York.
- Martrou, L.** 1924. *Lexique fân-français*. Paris: Procure Générale (des Pères du St Esprit)/Abbeville: Imprimerie Paillard.
- Mavoungou, P.A.** 1998. *Esquisse phonologique et morphologique du ndaƿa parlé à Mandjaye*. Mémoire de Maîtrise. Libreville: Université Omar Bongo.
- Mavoungou, P.A.** 2001. Macro- and Microstructural Issues in Mazuna Lexicography. *Lexikos* 11: 122-138.
- Mavoungou, P.A.** 2001a. La mondialisation et la lexicographie trilingue ou plurilingue au Gabon. Emejulu, J.D. (Éd.). 2001. *Éléments de lexicographie gabonaise Tome I*: 160-183. New York: Jims-Hillman.
- Mavoungou, P.A.** 2002. *Metalexical Criteria for the Compilation of a Trilingual Dictionary: Yilumbu-English-French*. Thèse de Doctorat non publiée. Stellenbosch: Université de Stellenbosch.
- Mavoungou, P.A.** 2004. L.J. van der Veen et Bodinga-bwa-Bodinga: *Gedandi sa Geviya/Dictionnaire Geviya-Français*. *Lexikos* 14: 440-448.
- Mavoungou, P.A.** À paraître. *Lexicographie et confection des dictionnaires au Gabon*. Stellenbosch: SUN PreSS.
- Mavoungou, P.A., T. Afane Otsaga et G.-R. Mihindou.** 2003. The Reproduction of Cultural Aspects in Dictionaries in French and the Gabonese Languages. *Lexikos* 13: 133-153.
- Mayer, R.** 1987. Langues des groupes Pygmées du Gabon: un état des lieux. *Pholia. Revue du Laboratoire de Phonétique et Linguistique Africaine* 2: 119-127.
- Mayer, R.** 1990. Histoire de l'écriture des langues du Gabon. *Revue Gabonaise des Sciences de l'Homme* 2: 65-91.
- Mayer, R. et M. Voltz.** 1990. Dénominations ethnoscience des langues et des ethnies du Gabon. *Revue Gabonaise des Sciences de l'Homme* 2: 43-53.
- Mba-Nkoghe, J.** 1981. *Phonologie et classes nominales en fang (Langue bantoue de la zone A)*. Thèse de 3^{ème} cycle. Paris: Université Sorbonne Nouvelle.
- Mba-Nkoghe, J.** 1991. Place et utilité des langues gabonaises dans l'activité nationale. *Bulletin d'information et de liaison du Ministère de l'Éducation Nationale. Message* 5: 18-22.
- Mba-Nzué, N.** 1981. *Esquisse phonologique du mvai (Parler de Minvoul)*. Mémoire de Licence. Libreville: Université Omar Bongo.
- Mboungou Asseko, M.** 1979. *Étude contrastive du Fang d'Oyem et du Fang de Libreville*. Mémoire de Licence. Libreville: Université Omar Bongo.
- Medjo-Mvé, P.** 1989. *Exploration morpho-syntaxique du dialecte le nyani ou contribution à l'étude du verbe ndumu*. Mémoire de Maîtrise. Libreville: Université Omar Bongo.
- Medjo-Mvé, P.** 1997. *Essai sur la phonologie panchronique des parlers fang du Gabon et ses implications historiques*. Thèse de Doctorat (NR). Lyon: Université Lumière-Lyon 2.
- Meinhof, C.** 1889-1890. Das Zeitwort in der Benga-Sprache. *Zeitschrift für afrikanische Sprachen* 3: 265-284.

- Mekui Missang, E.** 1998. *La présence des faits suprasegmentaux du fang ntumu dans l'utilisation du français pour les enfants âgés de 10 à 12 ans.* Mémoire de Maîtrise. Libreville: Université Omar Bongo.
- Mickala-Manfoumbi, R.** 1986. *Une approche linguistique pour l'enseignement du yipumu.* Mémoire de Maîtrise. Libreville: Université Omar Bongo.
- Mickala-Manfoumbi, R.** 1989. *Éléments de description du дума, langue bantu du Gabon (B51).* Mémoire de Licence Spéciale. Bruxelles: Université Libre de Bruxelles.
- Mickala-Manfoumbi, R.** 2004. *Lexique pove-français/français-pove.* Libreville: Éditions Raponda-Walker.
- Mission Catholique de Libreville.** 1948. *Ezango zi chrétien mpongwe-orungu-galoo-nkomi.* Issy-les-Moulineaux: Presses Missionnaires Imprimerie St Paul.
- Mission Catholique de Libreville. Vicariat Apostolique de Libreville.** 1894. *I kambisa si kretien = Livre de prières en mpongwe-français-latin.* Libreville: Vicariat Apostolique du Gabon.
- Mission Catholique de Sette-Cama.** 1933. *Catéchisme de la doctrine chrétienne = Katsisu mu mbembo hi lumbu hi bis' sette-cama.* Paris: Imprimerie des Orphelins Apprentis d'Auteuil.
- Mission Catholique du Gabon. Vicariat Apostolique du Gabon.** 1925. *Syllabaire Fan.* Merville (Nord): Imprimerie Ceugnard Le Sage.
- Mission Catholique du Gabon. Vicariat Apostolique du Gabon.** 1932. *Catéchisme Fan.* Rome: Solidarité St Pierre Claver.
- Mission de la Sainte-Croix.** 1924. *Catéchisme Gisira.* Abbeville: Imprimerie Paillart.
- Mouélé, M.** 1988. *Description d'un parler Ndumu le lenyani (phonétique, phonologie, morphologie nominale et pronominale).* Mémoire de Maîtrise. Libreville: Université Omar Bongo.
- Mouélé, M.** 1997. *Étude synchronique et diachronique du groupe дума.* Thèse de Doctorat (NR). Lyon: Université Lumière Lyon 2.
- Mouguiama-Daouda, L.** 1991. *Étude morpho-syntaxique du constituant nominal en ygòsi, dialecte du gisira.* Mémoire de D.E.A. Lyon: Université Lumière Lyon 2.
- Mouguiama-Daouda, P.** 1995. *Les dénominations ethnoichtyologiques chez les bantous du Gabon.* Thèse de Doctorat (NR). Lyon: Université Lumière Lyon 2.
- Mouloungui, M.P.** 1999. *Esquisse phonologique et morphologique du yesa.* Mémoire de Maîtrise. Libreville: Université Omar Bongo.
- Murard, P. (Trad.).** 1903. *Katsisu i neni i rendilu mu mbembo bis' Sette-Cama (grand catéchisme).* Lyon: Imprimerie Paquet.
- Murard, P. (Trad.).** 1903a. *Katsisu i keki i rendilu mu mbembo bis' Sette-Cama (petit catéchisme).* Lyon: Imprimerie Paquet.
- Murard, P. (Trad.).** 1903b. *Cantiques = Gnimbu tsi niambi.* Lyon: Imprimerie Paquet.
- Muroni, J.-M.** 1989. *Petit dictionnaire bantou du Gabon: français-ndjabi, ndjabi-français.* Paris: L'Harmattan.
- Naidailac, L.** 1992. *Lexique isangu-français.* Inédit. Lyon: Université Lumière Lyon 2.
- Naidailac, L.** 1992a. *Lexique yinzebi-français.* Inédit. Lyon: Université Lumière Lyon 2.
- Nassau, R.H.** 1881. *Panwe Primer and Vocabulary Compiled from Materials Collected by H.M. Adam.* New York: American Tract Society.
- Ndamba, J.** 1977. *Syntagème nominal et groupe nominal en vili: Langue bantu du Congo.* Thèse de Doctorat 3ième cycle. Paris: Université de Paris III Sorbonne Nouvelle.

- Ndinga-Koumba-Binza, H.S.** 2000. *Phonologie du civili de Mayumba: Langue bantu du Gabon (H12a)*. Mémoire de Maîtrise. Libreville: Université Omar Bongo.
- Ndong, F.** 1962. *Nten Wam*. Rome: Société St Pierre Claver.
- Ndong Menini, I.** 1980. *Esquisse phonologique du ntumu*. Mémoire de Licence. Libreville: Université Omar Bongo.
- Ndongo Esono, S. (R.P.)** 1956. *Gramática pamue*. Madrid: Instituto de Estudios Africanos.
- Ndoumou Mba.** 1999. *Étude diachronique des phonèmes du yinzebi et du proto-bantu*. Mémoire de Maîtrise. Libreville: Université Omar Bongo.
- Nguema Nzue, R.** 1988. *Essai de morpho-syntaxe du Fang-Ntumu*. Mémoire de Licence. Libreville: Université Omar Bongo.
- Nsuka-Nkutsi, F. (Éd.)** 1980. *Éléments de description du Punu*. Lyon: Presses Universitaires de Lyon (PUL).
- Nyangone Assam, B. et P.A. Mavoungou.** 2000. Lexicography in Gabon: A Survey. *Lexikos* 10: 252-274.
- Nzang-Obame, J.E.** 2001. *Essai de description phonologique de fang-mekè (A75)*. Mémoire de Maîtrise. Libreville: Université Omar Bongo.
- Nzang-Bié, Y.** 1986. *Quelques aspects de la technique du contage chez les Fang Ntumu*. Mémoire de Licence. Libreville: Université Omar Bongo.
- Ondo Esono Aba'a.** 1981. *Esquisse morphologique du Ntumu*. Mémoire de Licence. Libreville: Université Omar Bongo.
- Ondo-Mebiame, P.** 1989. *Esquisse de description du sangu, langue bantu du Gabon de zone B42*. Mémoire de Licence Spéciale. Bruxelles: Université Libre de Bruxelles.
- Ondo-Mebiame, P.** 1992. *De la phonologie à la morphologie du Fang parlé à Aboumezok (Langue Bantu A.78)*. 2 volumes. Thèse de Doctorat (NR). Tervuren: Université Libre de Bruxelles.
- Ondo-Mebiame, P. et Y. Nzang-Bié.** 1990. Seki. *Revue Gabonaise des Sciences de l'Homme* 2: 281-286.
- Osorio, Z.** 1887. *Diccionario Pamue-Español, Vocabulary of the Fan Language in Western Africa*. London.
- Perez, G. et L. Sorinas.** 1928. *Gramática de la lengua Benga*. Madrid: Editorial del Corazón de María.
- Perron, H.P. (R.P.)** 1964. *Lexique français-ikota*. 2 vol. (reprise des notes manuscrites du R.P. Lamour). Makokou: Mission Catholique.
- Puech, G.** 1990. Lateye. *Revue Gabonaise des Sciences de l'Homme* 2: 213-226.
- Puech, G.** 1990. Bèkwel. *Revue Gabonaise des Sciences de l'Homme* 2: 127-128.
- Raponda-Walker, A.** (n.d.) *Dictionnaire getsogo-français/français-getsogo*. Manuscrit inédit.
- Raponda-Walker, A.** 1930-1934. *Dictionnaire mpongwè-français, suivi d'éléments de grammaire*. Metz: Imprimerie de la Libre Lorraine.
- Raponda-Walker, A.** 1937. Initiation à l'ébongwè, langues des Négrilles. *Bulletin de la Société des Recherches Congolaises (BRC)* 23: 129-156.
- Raponda-Walker, A.** 1961. *Dictionnaire français-mpongwé*. Brazzaville: Imprimerie St Paul. 1995². Classiques Africains. Libreville: Éditions Raponda-Walker.
- Raponda-Walker, A.** 1995. *Dictionnaire mpongwe-français*. Versailles/Libreville: Classiques Africains/Éditions Raponda-Walker.
- Raponda-Walker, A.** 1995. *Dictionnaire français-mpongwe*. Versailles/Libreville: Classiques Africains/Éditions Raponda-Walker.

- Raponda-Walker, A. (Éd.).** 1998. *Les langues du Gabon*. Classiques Africains. Libreville: Éditions Raponda-Walker.
- Raponda-Walker, A. et R. Sillans.** 1961. *Les plantes utiles du Gabon: essai d'inventaire et de concordance des noms vernaculaires et scientifiques des plantes spontanées et introduites du Gabon*. Paris: Lechevalier.
- Ratanga-Atoz, A.F.** 1999. *Les peuples du Gabon Occidental de 1839–1941. Tome 1*. Libreville: Éditions Raponda-Walker.
- Reeb, A.** 1895. *Bivovili bya bakretyeni ba baduma = Manuel de prières à l'usage des chrétiens doumas*. Mission Catholique de Libreville-Lastoursville. Fribourg-en-Brisgau: Imprimerie par Herder.
- Rekanga, J.P.** 1986. *Pour une lecture morpho-syntaxique du chant II du Moet de Zué Nguéma*. Mémoire de Licence. Libreville: Université Omar Bongo.
- Rittaud-Hutinet, C.** 1980. Lexique. Nsuka-Nkutsi, F. (Éd.). 1980. *Éléments de description du Punu*: 193-245. Lyon: Presses Universitaires de Lyon (PUL).
- Sima Mvé, J.B.** 1996. *Esquisse phonologique et morphologique du mahongwe*. Rapport de Licence. Libreville: Université Omar Bongo.
- Tardy, L.** 1933. Contribution à l'étude du folklore bantou: les fables, devinettes et proverbes fang. *Anthropos* 28(3–4): 277-303.
- Tchen-Damas, E.** 1990. Omyène (mpongwe). *Revue Gabonaise des Sciences de l'Homme* 2: 261-271.
- Trilles, H.** 1898. *Livre de prières = NTEN WAM minkobe ma nè ge kobe né nzame*. Mission catholique de Libreville. Bar-le-Duc: Imprimerie de l'œuvre St Paul.
- Trilles, H.** 1898. *Ntèn wa tar'éyé gele fang'ne fala étén nélang = Exercices de lecture et d'écriture en pahouin et en français*. (2e partie). Mission Catholique de Libreville. Tours: Imprimerie Louis Dubois.
- Trilles, H.** n.d. *Katesism Nten wa vegele Fang Nsong Katolik = Catéchisme de la Doctrine Catholique*. Paris: CSSP et Mission Catholique de Libreville.
- Van der Veen, L.J.** 1991. *Étude comparée des parlars du groupe Okani — B30 (Gabon)*. Thèse de Doctorat (NR). Lyon: Université Lumière-Lyon 2.
- Van der Veen, L.J.** 1992. Le système tonal du ge-via (Gabon). *Journal of West African Languages* 22(2): 17-41.
- Van der Veen, L.J.** 1999. *Les Bantous eviya (Gabon-B30): langue et société traditionnelle*. Note de synthèse en vue de l'obtention de l'Habilitation à Diriger des Recherches en Sciences du Langage, soutenue le 28 janvier 1999 à l'Université Lumière-Lyon 2.
- Van der Veen, L.J. et S. Bodinga-bwa-Bodinga.** 2002. *Gedandedi sa Geviya/Dictionnaire Geviya-Français*. Louvain/Paris/Sterling, Virginia: Peeters.
- Yembi Bouka, L.** 1995. *Structures phonologiques et structures prosodiques (le modèle bekwel)*. Thèse de Doctorat (NR). Tervuren: Université Libre de Bruxelles.

Compiling the First Monolingual Lusoga Dictionary

Minah Nabirye, *Institute of Languages, Makerere University, Kampala, Uganda (mnabirye@gmail.com)*

Abstract: In this research article a study is made of the approach followed to compile the first-ever monolingual dictionary for Lusoga. Lusoga is a Bantu language spoken in Uganda by slightly over two million people. Being an under-resourced language, the Lusoga orthography had to be designed, a grammar written, and a corpus built, before embarking on the compilation of the dictionary. This compilation was aimed at attaining an academic degree, hence requiring a rigorous research methodology. Firstly, the prevailing methods for compiling dictionaries were mainly practical and insufficient in explaining the theoretical linguistic basis for dictionary compilation. Since dictionaries are based on meaning, the theory of meaning was used to account for all linguistic data considered in dictionaries. However, meaning is considered at a very abstract level, far removed from the process of compiling dictionaries. Another theory, the theory of modularity, was used to bridge the gap between the theory of meaning and the compilation process. The modular theory explains how the different modules of a language contribute information to the different parts of the dictionary article or dictionary information in general. Secondly, the research also had to contend with the different approaches for analysing Bantu languages for Bantu and European audiences. A description of the Bantu- and European-centred approaches to Bantu studies was undertaken in respect of (a) the classification of Lusoga words, and (b) the specification of their citations. As a result, Lusoga lexicography deviates from the prevailing Bantu classification and citation of nouns, adjectives and verbs in particular. The dictionary was tested on two separate occasions and all the feedback was considered in the compilation process. This article, then, gives an overall summary of all the steps involved in the compilation of the *Eiwanika ly'Olusoga*, i.e. the *Monolingual Lusoga Dictionary*.

Keywords: LEXICOGRAPHY, MONOLINGUAL DICTIONARY, THEORETICAL FRAMEWORK, DICTIONARY TESTING, EURO-CENTRED, BANTU-CENTRED, ORTHOGRAPHY, GRAMMAR, WORD CLASSES, BANTU LANGUAGES, LUSOGA, UGANDA

Obufunze: Okuwandiika Eiwanika ly'Olusoga Elyasookela Ilala. Mu kiwandiiko kino mulimu enambulula y'eisomo elyasinzilwaku okuwandiika eiwanika ly'Olusoga elisookeile ilala. Olusoga lugwa mu nnimi dh'ekika kya Bantu ela lwogelwa abantu abaswika mu bukaile obubili mu Uganda. Engeli Olusoga ye lwali nga ti lugundhiivu bulungi mu buwandiike, kyali kyetaagisa okwekalilisa walifu y'Olusoga, okuwandiika gulaama n'okuzimba eitelekelo ly'ebiwandiiko ebisinzilwaku okuwandiika Olusoga. Okuwandiika eiwanika lino kwali kilala ku bigendelelwa by'omusomo gw'omuwandiisi. Noolwekyo engeli y'okunoonheleza ku byetaagisa yalina okuba nga yeetengelela ku mutindo gw'obwevu. Okusookela ilala empandiika y'amawanika ebailewo ebaile esinga kulaga omulimu ogukoleibwa aye ng'engeli gye gutuukiibwaku teinhonholwa bukalamu. Amawanika gasinga kulondoola ntegeelo dha bigambo ela omusingi gw'entegeelo n'ogwalina okwesigamibwaku okusobola okufuna ebyetaagibwa mu mawanika. Kimanhi-

bwe nti entegeelo dh'ebigambo dheesigama ku mbeela y'olulimi enneekusifu nga tetegeelekeka mangu ku idaala ly'okuwandiika amawanika. Noolwekyo omusingi ogundi ogwesigaima ku mitendela egyekalilisibwaku olulimi n'ogwakozezebwa okugaitania omusingi gw'entegeelo ku mitendela gy'okuwandiika amawanika. Emitendela egyekalilisibwaku olulimi giinhonhola engeli emitendela egy'endhawulo bwe gikolagana okusobola okutuuka ku bitundu by'eiwanika. Ekyokubili, okunoonheleza kuno ela kwalina okwekalilisa engeli ennimi dh'ekika kya Bantu bwe dheeyoleka eli beene nnimi n'abo abagwila abenda okudheega. Okwawula n'okugelaagelania engeli ennimi dha Bantu bwe dheeyoleka eli ebika by'abantu bano ab'endhawulo yakolebwa mu (a) emmindhawaza y'ebigambo ni (b) okusalawo engeli ebigambo by'eiwanika bwe bilina okwingizibwa. Oluvainhuma lwa byonabyona empandiika y'amawanika mu Lusoga yakyusa mu ngeli edhibaillewo ey'okubindhawaza ebigambo okusingila ilala embu dha naliina, nambeela, ni kinantabila. Eiwaniika ly'Olusoga lyagezezebwa emilundi ebili ela ebyava mu kugezezebwa kuno byafulumila mu Eiwaniika. Olupapula luno n'olulaga mu bufunze omusingi ogwasinziilwaku okuwandiika *Eiwaniika ly'Olusoga*.

Ebigambo ebikulu: EMPANDIIKA Y'AMAWANIKA, EIWANIKA ELY'OLULIMI OLU-LALA, OMUSINGI GW'OKUNOONHELEZA, OKUGEZESA EIWANIKA, AMASOMO AGE-EKUUSA KU BULAAAYA, AMASOMO AGEKUUSA KU AFRIKA, WALIFU, GULAAMA, EMBU DH'EBIGAMBO, ENNIMI DHA BANTU, OLUSOGA, UGANDA

1. The status of Lusoga

Lusoga (J16 in Guthrie's (1948) classification) is the third-largest language in Uganda with a population of 2 062 920 people, which corresponds to 8.6% of the total population (Uganda Bureau of Statistics 2005: 12). Lusoga is spoken in an area called Busoga, in the eastern part of Uganda. The sociolinguistic situation in Jinja, which is the main administrative town of Busoga, is multilingual. The historical linguistics of the Busoga region emanate from the 1944 Makerere Conference on Language which deliberated on languages of instruction in schools and decided that Luganda, Acoli, Runyoro, Ateso and Lugbara be the media of instruction in the entire country (Ladefoged et al. 1971: 87-89). Luganda was chosen as the medium of instruction in Busoga. Although this policy was abandoned shortly after Uganda's independence in 1962, Luganda had already been established in the Busoga region as medium of instruction in lower primary school. The 1965 language policy left English as the inevitable lingua franca of Uganda and as the language used as medium of instruction from the seventh school year onwards (Ladefoged et al. 1971: 90). In June 2005, the parliament of Uganda passed the teaching of Lusoga as one of the nine regional indigenous Ugandan languages. Documentation of Lusoga was hence required.

Lusoga was categorized as an undocumented language because available literature in Lusoga was very substandard. Although attempts into the standardization of the Lusoga orthography were made by the Cultural Research Centre (CRC and LULANDA 2001) and the Lusoga Ecumenical Board (LEB 2000), both of the orthographies were inconsistent in their description of the

Lusoga orthography and their coverage was very shallow. On the other hand, the only available grammars of Lusoga (CRC 2000 and Babyale 1999) had a pedestrian consideration of grammar with English translations for tourists. There were no operational Lusoga language boards to regulate the documentation of Lusoga. Interested Lusoga speakers embarked on writing what they thought fit with regard to the language, without addressing the structural composition of Lusoga. As a result, an overhaul analysis of the structure of Lusoga was necessary to provide a foundation for the compilation of the first monolingual dictionary of Lusoga.

A review of literature on cognate languages was used to discuss the structure of Lusoga, on which the decisions in the compilation process were based. The cognate Bantu languages selected were Zulu (S42), Shona (S10), Swahili (G41, G42, G43), and Luganda (J15). A comparative analysis of the structure of the Bantu languages was used as the basis for specifying the Lusoga orthography, grammar and lexicography. The resulting dissertation, Nabirye (2008), addressed Bantu lexicography in general and Lusoga lexicography in particular. The different steps taken in the compilation of the monolingual Lusoga dictionary itself — *Eiwanika ly'Olusoga*, Nabirye (2009), henceforth WSG — were arranged in chapters to explain the process of the compilation of the dictionary as an academic study.

2. A review of some proposed frameworks for compiling dictionaries

Lexicography has been alienated from the study of linguistics, so much so that scholars such as Hartmann (2001: 111-112) doubt whether a lexicographical process qualifies as a research study. Piotrowski (1994: 5-8) thinks that perhaps lexicography is not a branch of linguistics but a discipline of its own. Pawley (1985: 99) believes lexicography is not conducted according to stipulated theoretical principles, and that linguists turn into lexicographers at different points in lexicographical research. As a result, Piotrowski (1994: 8) concludes that lexicography is a complex field, and that "a proper approach to its theory is to evolve a flexible framework which could include as many different approaches as possible". The same or similar opinions on lexicography are shared in studies like Wiegand (1984), Hausmann (1986) and Zgusta (1986).

The compilation of a dictionary for academic purposes, from this background, was therefore very challenging because a theoretical framework to explain the compilation of dictionaries was required. Practical considerations, which provided the only existing framework for compiling dictionaries, were insufficient in accounting for all the linguistic data considered in dictionaries. The existing methodological frameworks for compiling dictionaries lacked a continuum to the intrinsic theoretical explanations of the foundations for the compilation of dictionaries.

For example, a methodology-based lexicographical study like Van Keymeulen (2003) concentrates on the practical aspects of data collection and users are referred to the relevant handbooks for the theoretical background. Since

language references exist and a minimum level of language proficiency is assumed from the users of the target language, this methodology starts at a higher level in the compilation process than that required for the compilation of the WSG.

As a second example, the methodology introduced in De Schryver (1999) does not specify the theoretical basis of the different activities in the compilation process. It focuses more on the practical considerations for compiling a dictionary based on corpus analysis. This methodology is realized at an even higher level of dictionary compilation than that of Van Keymeulen (2003) and is thus even further removed from the methodology used in the compilation of the WSG.

For instance, the De Schryver (1999) framework — a corpus-based approach to the compilation of paper dictionaries within the framework of Simultaneous Feedback (SF) — was inapplicable to the Lusoga dictionary compilation process since corpus analysis of Lusoga was not considered beyond the compilation of a Lusoga word list. Applying De Schryver's framework to the Lusoga compilation context would have meant postponing the application of steps 2, 4, 6 and 8 of the SF framework until such a time when some of the required parameters involved in the compilation process had been synchronized. Though simultaneous feedback can be applied at the testing stage, dictionary testing is also considered at an advanced stage in the dictionary compilation process, the reason being that prior to testing, a theory and the hypothesis on which the testing feedback is based should be specified and questions eliciting feedback should be geared towards testing this hypothesis. However, since the theoretical foundation of the simultaneous framework is not specified, the compilation of the WSG had no justifiable foundation to support the selection of questions for a questionnaire or even to critique the nature of feedback.

Notably though, both Van Keymeulen (2003) and De Schryver (1999) have made a big contribution to the methodology of compiling dictionaries. What was missing was the specification of how comprehensive the application of the proposed methodologies were to different dictionary compilation contexts. For instance, Van Keymeulen is geared towards the compilation of undocumented languages. Lusoga was categorized as an undocumented language but the contexts in which the dictionaries in the two contexts were compiled is not the same. Lusoga lacks the required language references available in the case of Van Keymeulen, and none of the speakers has ever been taught the language — in most cases Lusoga speakers have only used the language orally, but have never read or written it. Everything about the language and its analysis was new to the target user of the WSG. De Schryver, on the other hand, was geared towards the compilation of a bilingual Cilubà–Dutch dictionary. A bilingual audience has different characteristics from a monolingual audience. Though Cilubà may be a Bantu language like Lusoga, the way data is structured will vary depending on the target user. The context in which the Cilubà–Dutch dic-

tionary was compiled thus does not address the same context as the one in which the WSG was compiled. If the bilingual Cilubà–Dutch dictionary had instead been a monolingual dictionary compiled for Cilubà speakers, then De Schryver would likely have proposed a more comparable methodology to that of the WSG.

The main misgiving noted is that the methodologies above mainly looked at the framework from specific contexts whose conditions ended up being restrictive. A context-free and generally applicable framework for compiling dictionaries was found missing. There was a need for some sort of theory to account for language data to be put in dictionaries regardless of the context or type of dictionary compiled or the ultimate dictionary user.

The methodologies provided in the discussion so far help to show two processes in the compilation process. Firstly, that the compilation process can start from theoretical considerations, bypass corpus considerations, and go on to compile a dictionary especially for undocumented languages. Secondly, that a dictionary can be compiled starting from a theoretical framework, move on to corpus analysis, and end in a final product especially for relatively well-documented languages. The gap that remains is the specification of the foundation to the entire dictionary compilation process. This article therefore aims to explain this foundation by showing how theory and practice were merged in the compilation of the WSG.

3. Presentation of Nabirye's (2008) framework for compiling dictionaries

In support of Piotrowski (1994: 8) no single theory is able to account for the entire dictionary compilation exercise. What a theory for compiling dictionaries entails, therefore, is a series of theoretical and conceptual road maps to guide lexicographers from the uncovering of the smallest bit of meaningful language data to its rendering into a dictionary. The compilation framework helps in the decisions on how such information can be interpreted and placed in a dictionary so that it serves both the purpose of the dictionary and the intended audience adequately.

At any point in the process, the lexicographer should be enabled to examine the language from the most abstract meaning formations that string together the grammar or the lexicon of the language. A swift guide from theory to practice and back is necessary, particularly for a lexicographer compiling a dictionary of a less- or undocumented language, who may also have to establish the description of a language for the first time.

Lexicographers in this context are not only compiling the first dictionaries of a language, they are also specifying the structure of the language on which the dictionary is based. Hence, the main question to ask is: What is the most logical place to start the investigation of an undocumented language? The answer lies in the study of meaning and therefore the theory for compiling dictionaries should also start with the theory of meaning.

3.1 The theory of meaning

The hypothesis drawn from the discussion on the study of meaning in Ogden and Richards (1923: 110-112), Lyons (1977: 27-29), Hurford and Hearsley (1983: 91), and O'Grady, Dobrovolsky and Katamba (1996: 276) states that the study of meaning can be described based on four premises, namely:

- (a) That the properties of a language are specified and defined.
- (b) That the nature of words and their relations are established to provide a foundation for the interpretation of their senses.
- (c) That all observable characteristics in the speech acts are analysed to contextualize the observable utterances that users of a language are likely to make and what they could mean in each case.
- (d) That language in specific contexts should also be analysed to contextualize the different forms of usage that good use of words depends on and also explain how the same words in different contexts can produce different meanings.

If all these premises are analysed then the meaning of the different parts of the language and its properties can be established. The composition of the above-named hypothetical variables enables the understanding of the different meaning categories of an entire language. A lexicographer should thus be enabled to tap into the foundations of the meaning existing in each of the variables in order to get data for a dictionary. The lexicographer then endeavours to figure out how all the meaningful parts of a language ought to be represented in a dictionary.

At this level, the analysis of meaning is however relatively abstract and lexicographers need guidance on how to retrieve information of a language from the theory and to render it into a dictionary. Another theory that classifies meaningful language properties into broader linguistic categories that are the basis for studying linguistics is thus required to bridge the gap between the theory of meaning and the activities of compiling a dictionary. This theory, the modular theory, looks at the different levels of linguistic analysis and the specific information at each level.

3.2 The modular theory

This theory considers interdependent modules or levels of language that have access to information in other modules in order to account for a full analysis of a language. The interconnectivity of the different modules of language analysis helps to specify dictionary information on which lexicographical decisions can be applied. For example, the phonological module provides sounds and their meaning, the morphological module provides the word structure, the semantic

module provides the meaning of words to be defined, the syntactical module provides sentence patterns, the graphology module provides the graphical representation of all linguistic data considered, which in total display the structure of a language.

The modularity of language is analysed to specify the type of information each module or level of linguistic analysis can contribute to a dictionary. The selection of information to consider in a dictionary depends on the type of dictionary a lexicographer would like to compile. A lexicographer compiling a bilingual dictionary will choose different data from the modules when compared to someone compiling a monolingual dictionary, for example. However, the general language repository for all types of dictionaries is the same regardless of the context of compiling.

Svensén (1993: 4-5) divides dictionary information into five categories, namely: formal categories concerned with spelling, pronunciation and inflection, combinational categories mainly based on the morphosyntactic nature of a language, a semantic category dealing with the nature of words and their relationships, encyclopaedic and pragmatic categories providing non-linguistic information such as verbal encyclopaedic information and pictorial illustrations, and lastly the historical perspective category catering for information such as etymology (i.e. word origin in relation to the time axis, which cuts across most of the categories, incorporating elements from several of them).

If the data categories above give a comprehensive representation of dictionary data, then at this level in the compilation process the lexicographer is able to sort this data to match the type of dictionary to be compiled. However, in order for the information to be appropriately selected, entered and organized in a dictionary, principles of compiling also have to be introduced into the framework to guide the lexicographers' decisions. The third level of the framework is therefore the consideration of the principles of lexicography which guide the general formatting of the dictionary.

3.3 Principles of lexicography

Principles of lexicography are practical in nature because they guide the actual compilation activities by reminding the lexicographer on what to look out for in the process of compiling a dictionary. The practical methodology for compiling dictionaries therefore considers the following variables: the type of dictionary to be compiled, optimization methods used to select dictionary information, citation forms of word classes, defining methods, and the organization of the dictionary (with regard to parts of the dictionary and each dictionary article). When all of the above steps are completed, then a study of how the dictionary is received by target users may be undertaken, reflecting how each of the variables in the compilation process was carried out.

In the process of determining the type of dictionary, data to be entered is selected as well. Optimization methods are applied to data already available

and formatted by the optimization techniques a lexicographer decides to employ while choosing data for the dictionary and its representation.

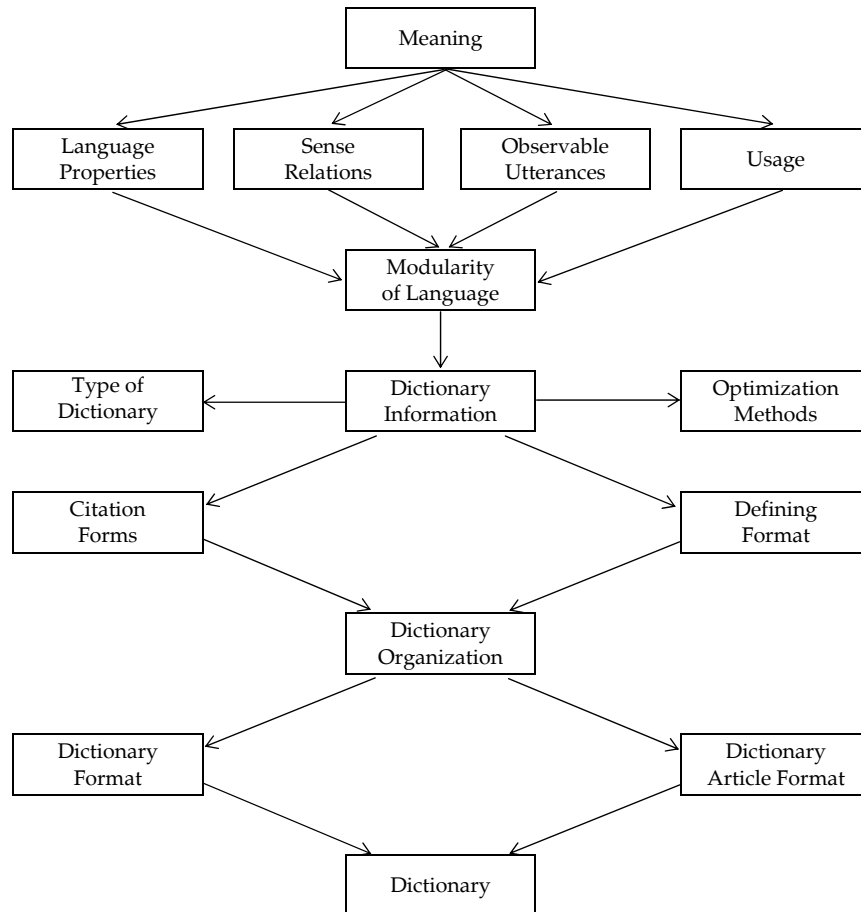


Figure 1: A summary of Nabirye's (2008) framework for compiling dictionaries

Data is then grouped into the different grammatical and/or lexical categories and entered in a unified format throughout the dictionary. Words are defined after deciding on their citation forms. These definitions reflect the linguistic characteristics in each grammatical category throughout the dictionary.

The organization of dictionary data is also essential to the compilation process. The determination of the structure of each dictionary article and the general format of the dictionary is needed.

If all the above steps are systematically followed, a dictionary should be deemed well compiled. It is thereafter left to the users and reviewers to determine how useful that dictionary is and how easy it is to access. This may also

be determined through testing the compiled dictionary with the target readership.

A dictionary that undertakes adequate coverage for the intended user and is easy to access should be well received. However if it fails to meet these objectives, then it is concluded that it has not adequately satisfied its intended objectives. A summary of how the entire dictionary process is conducted using Nabirye's (2008) theoretical framework is shown in Figure 1.

4. Application of Nabirye's (2008) framework in the compilation of the WSG

4.1 Introduction

The summary above is a presentation of the application of the entire framework in the different compilation stages. Other frameworks (Van Keymeulen 2003 and De Schryver 1999) are also considered in this presentation. The compilation process of the WSG was divided into four stages, namely: preparation, word collection, compilation, and testing. The present section focuses on how the framework is used in the compilation of dictionaries in general.

The first stage is the preparation for the compilation, which is mainly library research aimed at equipping the lexicographer with a good understanding of the language of compilation. The formulation and interpretation of all the theoretical considerations is undertaken at this stage to form hypotheses to test later on in the compilation process. This would be the best time to apply Nabirye's (2008) framework.

The second stage is aimed at collecting words for the dictionary. Compilation of dictionaries in undocumented languages utilize this stage looking for words, recordings, speeches and any other available informal sources of words for the dictionary. This would be the best time to consider Van Keymeulen's (2003) framework for compiling dictionaries in undocumented languages. Lexicographers compiling dictionaries in documented languages, however, mainly utilize this stage to explore corpus analytical methods to collect words and their meanings. For documented languages, then, this is the most appropriate time to consider De Schryver's (1999) framework for compiling dictionaries.

The third stage, which is the compilation itself, merges all the steps of the framework depending on the particular data to be considered in the dictionary. If the meaning of a particular language element is not clear then a semantic analysis can be used and a testing measure could be improvised or a lexicographical decision could be employed to assist in the final rendering of such information in the dictionary.

The last stage is the testing of the dictionary. As stated earlier, the testing instrument has also to be grounded in the theory. Diverging views and any new descriptions of a language or changes in the compilation style are tested, based on the theoretical considerations stipulated in the preparation stage and

applied during compilation to generate the simultaneous feedback advocated by De Schryver (1999) while the dictionary is being compiled.

4.2 Preparing the compilation of the WSG

Since Lusoga was not taught or examined in the education system in Uganda, there was no authority on the language. Documentation of Lusoga did not start until a few individuals and the Catholic Diocese in Jinja took the initiative to write what they thought was a description of Lusoga mainly to fill the void. The documentation conducted in this context was not regulated and the literature accrued was mostly based on personal convictions. At the inception of this research, visits were arranged to these individuals and centres hoping to find mentors, literature informants and editors. On approaching the writers of the only existing Lusoga books, however, it was soon discovered that since they were not linguists they could not account for the linguistic decisions taken in the writing of their books. They were very reluctant to respond to questions about the linguistic descriptions of Lusoga because they had no prior formal training on the structure of the language.

Meetings with Basoga personalities were a little more helpful and were mainly constituted of elders, members of the Busoga royal government, the clergy and journalists. Though they showed interest in the research, they had only used Lusoga orally and had not written or read it in their entire lives. Most importantly, these personalities were also not linguists and detailed academic research either interfered with their daily duties or was burdening their normal lives. Notably, these personalities were also intimidated by their ignorance of their own language, this on the most basic levels of the analysis of the nature of words, the spelling of words, and the rules of writing. There was no reference for this information in or on Lusoga, which had to be specified for the dictionary. Unsurprisingly, it was almost impossible to engage these personalities to complete some tasks on time, if at all. Evidently, their anticipated role was overrated, the realization of which led to the reduction of their contribution to the research process. They were only maintained as informants, especially during the testing process. The Cultural Research Centre (CRC) was adopted as the outlet operation centre for meeting informants in order to sustain their involvement in the compilation of the WSG.

The problem of devising the writing system of Lusoga on which to base the compilation was therefore the sole responsibility of the researcher. In order to address the Lusoga literature gap, the study was broadened and instead of focusing on Lusoga lexicography, a wider scope was adopted to focus on Bantu lexicography in the eastern and southern regions of Africa. The major Bantu languages in these regions, namely Luganda (Uganda), Swahili (Kenya and Tanzania), Zulu (South Africa), and Shona (Zimbabwe), were chosen as languages of reference. Fortunately, all the selected cognate languages were relatively well-documented. Therefore the theory of meaning was employed to

assess the descriptions of the different aspects in the structure of Bantu languages, and comparative findings would then be tested for adoption, or be redefined or relegated. Application of the findings depended on the nature of the information and its relation to the nature of Lusoga as it is used by the native speakers.

4.3 Collecting words

The search for written material in or on Lusoga uncovered a bilingual word list¹ entitled *Dictionary Lusoga–English/English–Lusoga* (CRC 2000a); and two attempts at the description of the Lusoga orthography, namely one by the Lusoga Ecumenical Board (LEB 2000),² and one by CRC and LULANDA (2001). Two versions of a Lusoga grammar also exist, one written by Babyale³ (1999) and another by CRC (2000). The CRC had some story books which were procured (CRC 1999, 1999a, 1999b, 1999c, 2000b). The literature was further enriched with the Lusoga version of the New Testament (The Bible Society 1998) and Lusoga newspaper clippings of now defunct publishers, circulation of which ran for two years only (*Kodh'eyo* 1998–1999, and *Ndiwulira* 1998–1999). This was the literature from which words were collected for the compilation of the WSG.

Computational linguistic studies were not available at the host university. Literature on corpora was found at other Bantu lexicography research centres. A visit to the Centre for Kiswahili Research (*Taasisi ya Uchunguzi wa Kiswahili*, TUKI) at the University of Dar es Salaam in June 2005 provided an insight into Swahili lexicography and corpus studies. However, none of the Swahili dictionary compilers employed corpus analysis and there was no corpus bank for Swahili at TUKI. Therefore, TUKI was not able to guide the study on the corpus analysis of Lusoga. The next research visit was to the Department of African Languages at the University of Pretoria in August 2005. This visit mainly served the purpose of providing literature on Shona and Zulu lexicography but none at all on corpus analysis. The research was consequently narrowed down and instead of heading into a full corpus analysis of Lusoga, only methods of collecting words were sought.

All the mentioned texts were scanned and used to generate word lists using Shoebox.⁴ This is as far as the corpus analysis for the WSG could be taken. For this reason the WSG compilation was categorized as belonging to the non corpus-based category.

In order to fill the void left by the limited references on Lusoga and the lack of an analysis of the Lusoga corpus, application of the theory of meaning was utilized to specify the Lusoga lexicon and grammatical categories which guided the specification of the lemmas considered in the WSG. The theory of meaning also helped to provide new evidence for the specification of Lusoga word categories and their respective citation forms, a detailed discussion of which is given in Section 4.4.2.

4.4 Compilation stage

This section will draw only one example from the compilation process which was considered to be the most controversial, forming the basis for the WSG's deviation from the existing citations of Bantu words in dictionaries.⁵ The discussion in this section is therefore centred on a reconstruction of the nature of Bantu words from a Bantu language speaker's perspective (Bantu-centred approach) as opposed to the way the same words are viewed by European Bantu linguists (Euro-centred approach). A distinction between the two perspectives on Bantu lexicography is discussed as basis for the classification and citation of verbs, nouns and adjectives introduced in the WSG. The following discussion presents a relatively new approach to the perceptions of the description of Bantu studies for Bantu users in particular.

4.4.1 Background to the citation of Bantu words in dictionaries

The compilation of dictionaries in Africa was initiated by missionaries for the benefit of missionary governments. The description of the Bantu languages was primarily meant to give European audiences a good understanding of the Bantu languages. English in this context was used as the language of reference and the standard for describing Bantu languages. Studies of this nature are categorized as Euro-centred. Unfortunately, the need to satisfy the European audience compromises the full understanding of Bantu studies from a Bantu perspective and these studies are therefore not Bantu-centred. This was hypothesized as the reason why most monolingual Bantu dictionaries reviewed in Nabirye (2008) were not suited for Bantu audiences, being difficult to access.

For instance, there was no uniform citation of Bantu words in dictionaries. Bantu reference works such as TUKI (1981), Murphy (1972), Snoxall (1967), Hannan (1959), Blackledge and Kitching (1925), and Steere (1870) used varying citation forms for verbs as shown in the selected examples below.

Steere (1870)	sema <i>ku-</i> , to say, to talk, to speak ... zika <i>ku-</i> , to bury ...
Blackledge and Kitching (1925)	-agula , <i>v. tr.</i> , scratch -sesa <i>v.c.</i> , cause to laugh
Snoxall (1967)	<i>ku-menya</i> <i>v. tr.</i> ... break ... <i>ku-duma</i> <i>v. i.</i> ... thunder ...
TUKI (1981)	l.a <i>kt</i> ... 'to eat' lal.a <i>kt</i> ... 'to sleep'

It was not clear which citation method best served Bantu dictionary audiences. Apparently the different citation forms were based on different interpretations of Bantu word classes; therefore the problem lay in the interpretation of Bantu morphology. Different morphological interpretations gave rise to varying cita-

tions because the Euro-centred descriptions of Bantu word classes were inadequate to give a full analysis of Bantu morphology. Compare in this regard Zgusta (1971: 224):

If [...] the posterior excerption of information frequently compels the lexicographer to change the whole construction of some entries, or shows him that there are some direct senses of some lexical units in the sphere of general language [...] which he did not know, then probably the material on which the lexicographer based his construction of entries was not ripe enough, was not sufficient.

Consequently, a new approach to the understanding of Bantu morphology was required to rectify the problem of varying citations in Bantu lexicography.

Citing Bantu verbs and adjectives by their stems is quite contrary to a native speaker's natural usage of words because words are used in full word forms.⁶ Some scholars, for example Kiango (2000: 29) and Hannan (1959: viii), cite as reason for the development of the stem tradition, an attempt to avoid an imbalance in the alphabet. Van Wyk (1995: 94) argues that no language seems to have a balanced distribution of entries across the alphabet, in the sense that each letter in the 26-letter-alphabet has approximately 4% of all entries.⁷ Van Wyk raises the question whether the alphabet imbalance should actually be felt as an obstacle since there are options that could be considered to assist users in overcoming the problem. Van Wyk's statistical interpretation of the treatment of the alphabet balancing helped to dispel the constrained concern for the alphabet because it not only compromises the tone of the lexicographer to being prescriptive, it also does great harm to the language because linguistic information that could result from the natural observation of the language is blocked.

4.4.2 Application of Nabirye's (2008) framework in addressing Bantu citations

In order to understand the composition of verbs and the reasons for their citation as stems, the theory of meaning was used to specify the meaning and function of the noun class prefix in forming Bantu verb infinitives. According to Kiingi (2007, 2007a) Bantu verbs are built around roots and are bound forms with the exception of the verb *-li* which can stand on its own. Because verbal prefixes and suffixes are not primarily part of the verb, an investigation into the meaning and role of these verbal affixes had to be undertaken. This investigation was carried out by analysing the marking of mood and infinitives in languages.

English translations of Bantu verbs in the infinitive always take the preposition "to". The understanding of Bantu infinitives therefore started with an investigation of the meaning and function of "to" in the above-mentioned contexts. According to Duffley (1992: 141) "to" is not, strictly speaking, part of the

infinitive, but it is a dematerialized preposition whose use is called for in certain contexts because of the meaning it expresses. Duffley (1992: 142) describes infinitive as a representation of an event as a whole and says that "to" serves a purpose of contextualizing the element of time situated outside its event. "To" is used with infinitives both for the lexical and grammatical meaning it brings into context. Therefore "to" is not a verb; it only functions with verbs in specific contexts for specific purposes.

English dictionaries do not include "to" in their verb entries. Bantu lexicographers have over a period of time been trying to import this aspect into Bantu lexicography by omitting or not considering the prefix *oku* in verb entries with varying justifications. The problem is that English and Bantu languages have different structures, and what applies in one may not necessarily work in another. The starting point of the analysis of the Bantu languages was from the English perspective, and this perspective was imported into Bantu languages because it is a logical system that can be understood by the European audience of Bantu studies. The ignored disparity in language systems has consequently obscured specification of the most appropriate lemmatization of Bantu words in dictionaries meant for Bantu audiences.

In the analysis of the verb ending vowel, Palmer (1986) was used to analyse the marking of mood and modality in languages. Palmer (1986: 1) introduces two assumptions regarding the study of modality: (a) it is possible to recognize a grammatical category of modality which is similar to aspect, tense, number and gender; and (b) this category can be identified, described and compared across a number of different and unrelated languages. What is less obvious, however, is the characterization of the semantic function of modality. Palmer (1986: 2) notes that modality does not only or primarily relate semantically to the verb but to the whole sentence. For this reason, there are languages in which modality is marked elsewhere than on the verb or within a verbal complex.

According to Thrane (1983: 155) identification of grammatical categories across languages rests upon shared semantic characteristics called cross-linguistic equivalent classes. The identification of a typological category is consequently in terms of meaning. There may not be a precise definition but there appears to be some very basic or "prototypical" feature similar for all languages.

Palmer (1986: 4) also notes that the real problem with the study of modality is not just that there is great variation in meaning across languages, but that there is no clear basic feature. There is a large degree of arbitrariness in the choice of grammatical form in the sense that it is not directly determined by meaning. Even with more easily definable categories such as tense and number, there is a very considerable difference in the extent of grammaticalization in different languages. There are even some languages that do not grammaticalize these familiar categories at all (Palmer 1986: 5).

Palmer (1986: 6) concludes that it is difficult to decide what to include and what to exclude from a grammatical study of modality. He insists on grammatical relates because they enable the investigator to look at the languages themselves and see what is systematized and organized within their grammatical systems. Palmer (1986: 21) gives a clue when he claims that the term "'mood' is traditionally restricted to a category expressed in verbal morphology". He quotes Jespersen (1924: 373) who also insists that "it is a syntactic not a notational category which is in the form of the verb, and dictionary definitions usually refer to verbal inflections".

Nabirye (2008) therefore had to re-examine the full word forms of Lusoga, finding that the infinitive full word forms of Bantu verbs consist of the noun class prefix 15 (*oku*), the root, and mood marked on the verb final; in this case the verb ending vowel as shown below:

(o)ku	+	f	+	a	'to die'
(o)ku	+	gw	+	a	'to fall'
<i>Noun Prefix</i>		<i>Verb Root</i>		<i>Mood</i>	

Both the prefix and suffix have to be present for Bantu verbs to be rendered as infinitives. Omission of either of the two would not create infinitive environments, this being why, without the prefix, stem entries are in the imperative mood. The prefix is a grammatical feature of gender and the verb final is that of mood. The noun prefix functions with the mood in the verb suffix to provide a timeless context for the verb root. This is the grammaticalized function for both the prefix and suffix in forming Bantu verb infinitives. The revised morphology of Lusoga words in the WSG was therefore treated as explained below.

The verb pre-prefix which is used occasionally but is required for the citation of full word forms was placed in parenthesis. The different morphological parts of words which users failed to demarcate during the testing process⁸ were segmented to guide users on Lusoga morphology. A new classification of Lusoga verbs was thus used to represent a word category with both noun and verb qualities. These verbs were moreover lemmatized under the first letter of the pre-prefix, namely "o".

Furthermore, full-form adjectives, which were cited with the noun class prefix 14 (*obu*), were classified as nouns, while words functioning as adjectives in their full form without the addition of noun prefixes were classified as the true adjectives of Lusoga. Dictionary users were guided on all these features in the front matter of the dictionary.

4.5 Testing phase

Testing of each of the theoretical considerations in the compilation process was conducted during the compilation exercise. Questions for testing were based on the practical considerations in the third section of the framework. Each of

the activities was allocated a minimum of four questions and an overall evaluation was based on the average tally of percentages received from each section. The WSG was physically tested twice and the dictionary was very well received in the Lusoga-speaking community.

This testing validated the finding of the research process and resulted in a deviation in the citation of Lusoga words from those in the prevailing Bantu dictionaries. Examples of how Lusoga words were entered are given in the examples in the Addendum.

5. Conclusions from the application of Nabirye's (2008) framework

In the absence of a theory of meaning and modularity, the specification of the composition of the parts of full verbs and adjective forms would have lacked a valid foundation and a relation to the compilation process. The available practical methodologies had not specified how the theoretical lexicographical decisions could be incorporated in the compilation exercise. A combination of both the theoretical and practical considerations in the dictionary compilation process is therefore necessary.

The framework presented in this article can go a long way in furthering the re-examination of more descriptions of African languages, and it could be that new perspectives on Bantu language descriptions could be discovered, based on the Bantu-centred approach for the benefit of Bantu audiences especially.

Endnotes

1. The *Dictionary Lusoga-English/English-Lusoga* (CRC 2000a) was found befitting the category of a 'word list' and is referred to as such in this article in spite of having 'dictionary' in its title.
2. The LEB (2000) was received with missing pages, with no details of authors or publisher.
3. Though Babyale (1999) is mentioned here, it was not available for review during the research process and was not used in the collection of words either.
4. Shoebox is a dictionary formatting program used by Nabirye (2008, 2009) to generate word lists for the WSG.
5. For a recent deviation in (bilingual) Bantu lexicography, however, see De Schryver (2008).
6. The usage of full word forms was tested during the compilation process and was validated by the findings.
7. For an in-depth analysis of the distribution of the alphabetical categories in dictionaries, see De Schryver (2005).
8. Testing on the demarcation of word parts produced 100% failure, as all the respondents were unable to demarcate parts of words correctly, thus prompting the segmentation of entries in the WSG.

References

- Babyale, S.C.** 1999. *Gulama w'Olusoga Omukalamu*. Kampala: Makerere University Press.
- Blackledge, G.R. and A.L. Kitching.** 1925. *A Luganda-English and English-Luganda Dictionary*. Kampala: Uganda Bookshop.
- CRC (Cultural Research Centre). 1999. *Ababita Ababiri*. Jinja: Diocese of Jinja.
- CRC (Cultural Research Centre). 1999a. *Akatabo Akasooka ak'Enfumo Edh'Abasoga*. Jinja: Diocese of Jinja.
- CRC (Cultural Research Centre). 1999b. *Ensambo edh'Abasoga*. Jinja: Diocese of Jinja.
- CRC (Cultural Research Centre). 1999c. *Kintu*. Jinja: Diocese of Jinja.
- CRC (Cultural Research Centre). 2000. *A Lusoga Grammar*. Jinja: Diocese of Jinja.
- CRC (Cultural Research Centre). 2000a. *Dictionary Lusoga-English/English-Lusoga*. Jinja: Diocese of Jinja.
- CRC (Cultural Research Centre). 2000b. *Enkabi Ekifiini mu Busoga*. Jinja: Diocese of Jinja.
- CRC (Cultural Research Centre) and LULANDA. 2001. *Empandiika ey'Olulimi Olusoga Enkalamu*. Jinja: Diocese of Jinja.
- 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.** 2005. Concurrent Over- and Under-Treatment in Dictionaries — The *Woordeboek van die Afrikaanse Taal* as a Case in Point. *International Journal of Lexicography* 18(1): 47-75.
- De Schryver, G.-M.** 2008. A New Way to Lemmatize Adjectives in a User-friendly Zulu-English Dictionary. *Lexikos* 18: 63-91.
- Duffley, P.J.** 1992. *The English Infinitive*. London: Longman.
- Guthrie, M.** 1948. *The Classification of the Bantu Languages*. London: Oxford University Press for the International African Institute.
- Hannan, S.J.** 1959. *Standard Shona Dictionary*. New York: Macmillan.
- Hartmann, R.R.K.** 2001. *Teaching and Researching Lexicography*. Applied Linguistics in Action Series. Harlow: Longman/Pearson Education.
- Hausmann, F.J.** 1986. The Training and Professional Development of Lexicographers in Germany. Ilson, R. (Ed.). 1986: 101-110.
- Hurford, J.R. and B. Hearsley.** 1983. *Semantics: A Coursebook*. Cambridge: Cambridge University Press.
- Ilson, R. (Ed.).** 1986. *Lexicography. An Emerging International Profession*. The Fulbright Papers I. Manchester: Manchester University Press.
- Jespersen, O.** 1924. *The Philosophy of Grammar*. London: George Allen & Unwin.
- Kiango, J.G.** 2000. *Bantu Lexicography: A Critical Survey of the Principles and Process of Constructing Dictionary Entries*. Tokyo: Tokyo University of Foreign Studies.
- Kiingi, K.B.** 2007. *Enkuluze y'Oluganda ey'e Makeerere*. Kampala: Fountain Publishers.
- Kiingi, K.B.** 2007a. *Style Manual for Luganda Dictionaries*. Unpublished manuscript. Kampala: Makerere University.
- Kodh'eyo*. 1998--1999. Lusoga newspaper. Kampala.
- Ladefoged, G. et al.** 1971. *Language in Uganda*. Nairobi: Oxford University Press.

- LEB (Lusoga Ecumenical Board). 2000. *The Lusoga Orthography*. Jinja: Lusoga Ecumenical Board.
- Lyons, J. 1977. *Semantics* (2 volumes). Cambridge: Cambridge University Press.
- Murphy, J.D. 1972. *Luganda-English Dictionary*. Washington: The Catholic University of America Press.
- Nabirye, M. 2008. *Compilation of the Monolingual Lusoga Dictionary*. Unpublished M.A. Dissertation. Kampala: Makerere University.
- Nabirye, M. 2009. *Eiwanika ly'Olusoga*. Kampala: Menha Publishers.
- Ndiwulira. 1998-1999. Lusoga newspaper. Kampala.
- Ogden, C.K. and I.A. Richards. 1923. *The Meaning of Meaning*. London: Routledge and Kegan Paul.
- O'Grady, W., M. Dobrovolsky and F. Katamba. 1996³. *Contemporary Linguistics: An Introduction*. Harlow: Pearson Education.
- Palmer, F.R. 1986. *Mood and Modality*. Cambridge: Cambridge University Press.
- Pawley, A. 1985. Lexicalization. Tannen, D. (Ed.). 1986. *Georgetown University Roundtable on Language and Linguistics*: 98-120. Washington: Georgetown University Press.
- Piotrowski, T. 1994. *Problems in Bilingual Lexicography*. Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego.
- Snoxall, R.A. 1967. *Luganda-English Dictionary*. London: Oxford University Press.
- Steere, E. 1870. *A Handbook of the Swahili Language as Spoken at Zanzibar*. London: Bell and Daldy.
- Svensén, B. 1993. *Practical Lexicography*. New York: Oxford University Press.
- The Bible Society. 1998. *Endagaano Empyaka*. Kampala: The Bible Society.
- Thrane, T. 1983. On the Universality of AUX. *Acta Linguistica Hafniensia* 18: 154-200.
- TUKI (*Taasisi ya Uchunguzi wa Kiswahili*). 1981. *Kamusi ya Kiswahili Sanifu*. Dar es Salaam: Oxford University Press and Institute of Kiswahili Research.
- Uganda Bureau of Statistics. 2005. *2002 Uganda Population and Housing Census: Main Report*. Entebbe: Uganda Bureau of Statistics.
- Van Keymeulen, J. 2003. Compiling a Dictionary of an Unwritten Language: A Non Corpus-based Approach. *Lexikos* 13: 183-205.
- Van Wyk, E.B. 1995. Linguistic Assumptions and Lexicographical Traditions in the African Languages. *Lexikos* 5: 82-96.
- Wiegand, H.E. 1984. On the Structure and Contents of a General Theory of Lexicography. Hartmann, R.R.K. (Ed.). 1984. *LEXeter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9-12 September 1983*: 13-30. *Lexicographica. Series Maior* 1. Tübingen: Max Niemeyer.
- Zgusta, L. 1971. *Manual of Lexicography*. *Janua Linguarum. Series Maior* 39. Prague: Academia.
- Zgusta, L. 1986. Summation. Ilson, R. (Ed.). 1986: 138-146.

Addendum: Dictionary article extracts from *Eitwanika y'Olusoga*

(a)mazaalibwa [(a)mázaalíbwá] *l.* [-/6] **bl:** [Lg: amazaalibwa] 1) Olunaku omuntu lw'azaalibwa. *gez:* *Leelo ndi kukuza amazaalibwa gange ag'emyaka asatu.* 2) **(A)mazaalibwa:** Olunaku olwidhuukililwaku okuzaalibwa kwa Yesu. *gez:* *Amazaalibwa ga Yesu gabaawo nga 25 Desemba.*

duulu [dúúlú] *l.* [1/ -] **bl:** [Lg: duulu] Kika kya muzaanho gwa mpiki. Obala enta ku itaka me waakozesa empiki waaginasula ekube empiki edhindi dhigwe mu kaina.

(e)bbaatule [(e)bbáátulé] *l.* [9/6, 10] [Lng<battery>] **bl:** [Lg: ebbaatule] Ekintu ekili oti kasanduuku nga n'ekivaamu amasaanhalaze agatambuza emmotoka n'ebintu ebindi ebikolela ku masaanhalaze. *gez:* *Bbaatule eweileku emmotoka yaazima. / Emmotoka ezilamu bbaatule tesobola kusimbula.*

goonia [góóniá] **tbk:** gógoniá. *l.* [1/8 bigonia] **bl:** [Lg: ggonya] *kikg:* (*Crocodylus niloticus*) Kika kya kisolo ekili oti mudolome aye nga kyo kibba inho. Kiba mu maadhi ni ku lukalu. Kilya mmamba ela nga kilina amainho amasongovu nga mangi n'ebigalagamba ku mubili gwakyo. Kilina ebigele bina. **Bona:** KFNN (Ebisolo)

Isebataka [Isébatáká] *l.* [1/2 Baisebataka] **bl:** [Ssaabattaka] Omukulu w'ekika.

kaale [káále] *ew.* **bl:** [Lg: kaale] 1) Kagambo akakozesebwa okwikiliza ekintu. *gez:* *A: Mpaamu edaakiika ndala...*

B: kaale. 2) Kagambo akozesebwa okwilamu embooji gy'oikilizigania nayo. *gez:* *A: Tusangaile... B: kaale!* 3) Kakozesebwa okwilamu okwebaza. *gez:* *A: Weebale... B: kaale!* 4) Kakakozesebwa okulaga okusaasila. *gez:* *Kaale nze ndabye.* 5) Kakakozesebwa okuwa olukusa. *gez:* *A: Ka ntwaleku ku kitabo kyo... B: kaale.* 6) Kagambo akakozesebwa okudhulila ekintu ekitagiile bukalamu. *gez:* *Kaale obona by'okoze bwe bituswaziiza. / Kaale nguli maama aliwo.*

manvuuli [manvúulí] **tbk:** mínvuulí. *l.* [9/10 dhimanvuuli] **bb:** [Lsw: mwavuli, Lg: manvuuli] Ekyebwikwa ekiziyiza omusana oba amaadhi okukukuba nga kigemebwa mu ngalo kyabwika omutwe.

(o)bu.bisi [(o)búbisí] *l.* [-/14] **bl:** [Lg: obubisi] 1) Embeela y'obubolu. *gez:* *Enku embisi tidhaaka.* 2) Embeela etayiile (mu kufumba). 3) Embeela y'ekintu ekitali kyenvu. *gez:* *Amenvu gano mabisi tulinde genge.* 4) (*stl*) Obutali bulabufu. *gez:* *Yaakava mu kyalo akaali mubisi.*

(o)bu.kakafu [(o)búkakáfú] *l.* [-/14] **bl:** [Lg: obukakafu] 1) Amazima agalaga obutuufu bw'ekintu. *gez:* *Ky'oyogela okilinaku obukakafu?* 2) Obudhulizi obulaga amazima. *gez:* *Bw'owawabila omuntu olina okuba n'obukakafu obumululiza.*

(o)bwetoowaze [(o)bwétóowáze] *l.* [-/14] [obu- + yetoowaze] **bl:** [Lg: obwetowaze] Empisa y'okwetoowaza.

(o)ku.bak.a [(o)kúbaká] *kt.* [L] [-base] [mbase] **bl:** [Lg: okubaka] 1) Okuwunga ekintu. *gez:* Suula ekisumuluzo ndha kukibaka. 2) (*tgll, yaaye*) Okutegeela by'osomye. *gez:* Bye mmukobye aboneka tabibase bukalamu. 3) (*tgll*) Okugema oba okugemenkeleza. *gez:* Abaile alumuka n'amubaka. 4) Omuzaanho gw'abakazi og'wokusuula n'okuwunga omupiila. *gez:* Eismelo lyaife lyasinze ekikopo ky'okubaka.

(o)ku.beyuk.a [(o)kúbeyúká] *kt.* [T] [-beywike] [mbeywike] **bl:** [Lg: okubeyuka] (*ygl, tgll*) Okwetala oba okutambulatambula nga totyamisaaku wantu walala. *gez:* Ono atubeyukilaku ayenda ki?

bbgz: Okubeyukila, Okubeyusa.

(o)ku.faanana.a [(o)kúfaaná] **tbk:** **(o)kúfaaná** *kt.* [L] [-faananine] [nfaanaine] **bb:** [Lsw: faana, Lg: okufaana] 1) Okuboneka ng'omuntu oba ekintu ekindi. *gez:* Omwana afaanana lata we. 2) Okuba mu mbeela oba n'empisa endala. *gez:* Abantu batyo ni bwe bafaanana. *ggl:* Kizi kibi kifaanana eyakinia; Mwana mubi afaanana inhina; Okufaana ti luganda entondolo efaanana eisooma; Tibifaanana abifuna muk'omwibi aliila ku lugyo; Tufaana ti kyalo.

bbgz: Okufaananibwa, Okufaananika, Okufaananila Okufaananika.

(o)ku.mw.a [(o)kúmwá] *kt.* [L] [-mweile] [mmweile] **bl:** [Lg: okumwa] Okusala enviili ku mubili. *gez:* Idha nkumwe ekyenhi. / Buli wiiki ibili mmwa enviili. *gl:* Kitwe kimwebwa ye bakiliza ye kilaga.

bbgz: Okumwebwa, Okumweka, Okumwela, Okumwesa, Okwemwa.

(o)lu.talo [(o)lútaló] *l.* [11/10] **bl:** [Lg: olutalo] 1) Okulwanagana. *gez:* Bbaabba yafiila mu lutalo. 2) Okuyombagana okw'amaanhi. *gez:* Nkooye okuwuliliza entalo dhaimwe. 3) Ekintu eky'etaagisa aamaanhi oba amagezi amangi okukola. 4) Ekiseela ky'okuzaala. *gez:* Abakazi abafiila mu lutalo beeyongeile.

(o)mudalizo [(o)múdálizó] *l.* [3/4] **bl:** [Lg: omudalizo] Kagoye akadhilige mu langi edh'endhawulo akakookebwa ku mbali y'egomasi.

(o)mu.zila¹ [(o)múzilá] *l.* [3/-] Omulali.

(o)mu.zila² [(o)múzílá] *l.* [1/2] **bl:** [Lg: omuzila] 1) Omuntu alwana einho entalo yaawangula. *gez:* Omusaadha oyo muzila. 2) Omuntu akola ebikolwa ebikuuma obutebenkevu ela nga byegombesa. *gez:* Leelo tukuza olunaku lw'abazila. 3) (*stl, bkl*) Omusaadha alina obulwaile bw'obukaba. *gez:* Omulwaile ali mu waadi y'abazila.

(o)mu.zila³ [(o)múzilá] *l.* [3/-] Omugozi.

(o)mwela [(o)mwélá] *l.* [3] **(o)muyeela / 4 (e)myela, 6 (a)meela** Omunwe gw'eitooke.

(o)mwenge [(o)mwéngé] *l.* [3] **(o)muyenge / 4] bl:** [Lg: omwenge] Omubisi ogutamiiza nga guli mu bika kamaala. **gt:** Omwenge bigele: Kika kya mwenge kye basogozo ebigele.

-onka [-onká] **tbk:** -enká. *kgz.* Kawango akazimba ebigambo ebikugila omuwendo. Kagaitibwa ku nabusigile okulaga nti ezila kindi. *gez:* Bonka, Dhonka, Gonka, Kyonka.

The Representation of Vowel Duration in Civili Dictionaries*

Hugues Steve Ndinga-Koumba-Binza, *Stellenbosch University Centre for Language and Speech Technology (SU-CLaST), Stellenbosch, Republic of South Africa (13751719@sun.ac.za)*

and

Justus C. Roux, *Centre for Text Technology (CTexT), North-West University, Potchefstroom, Republic of South Africa (justusroux@nwu.ac.za)*

Abstract: Civili is a developing language spoken in Gabon and in a few neighbouring countries. This article focuses on the representation of vowel duration in Civili dictionaries. The representation in these dictionaries is inconsistent. In the article, it is argued that this inconsistency stems from a twofold phonetic-phonological issue, which has implications for the word writing system of the language. The article provides an assessment of the existing orthography proposals for Civili and offers materials for a new proposal that takes the vowel duration issue into account. Subsequently, it is recommended that vowel duration be represented by a diaeresis above the vowel for both the orthography and the lemmatization in reference works such as dictionaries and school and religious textbooks.

Keywords: VOWEL DURATION, CIVILI, DICTIONARIES, ORTHOGRAPHY, PHONETIC-PHONOLOGICAL ISSUE, WORD WRITING SYSTEM

Résumé: La représentation de la durée vocalique dans les dictionnaires du *civili*. Le *civili* est une langue en développement parlée au Gabon et dans quelques pays voisins. Cet article traite de la représentation de la durée vocalique dans les dictionnaires du *civili*. La représentation est inconsistante dans ces dictionnaires. Dans l'article, il est soutenu que cette inconsistance trouve son origine dans le double problème phonético-phonologique qui a une implication dans le système d'écriture des mots de la langue. L'article fournit une évaluation des propositions d'orthographe existantes pour le *civili* et présente des éléments pour une nouvelle proposition qui prend en compte la question de la durée vocalique. Ensuite, il y est recommandé que la durée vocalique soit représentée par un tréma au dessus de la voyelle aussi bien pour l'orthographe que pour la lemmatisation dans les travaux de référence tels que les dictionnaires et les manuels scolaires et religieux.

Mots-clés: DURÉE VOCALIQUE, CIVILI, DICTIONNAIRES, ORTHOGRAPHE, PROBLÈME PHONÉTICO-PHONOLOGIQUE, SYSTÈME D'ÉCRITURE DES MOTS

* This article is a revised version of a paper 'Vowel-length Representation in Civili Orthography: Implications for Dictionaries' presented at the Thirteenth International Conference of the African Association for Lexicography, organized by the Bureau of the *Woordeboek van die Afrikaanse Taal*, Stellenbosch, Republic of South Africa, 30 June–3 July 2008.

1. Introduction

Civili, a developing language spoken in Gabon and a few neighbouring countries, has at present three dictionaries published, i.e. Marichelle (1902, 1912) and ILALOK (2008). Ndinga-Koumba-Binza (2005: 138) who identifies different periods of dictionary production in Gabon, classifies the first two dictionaries as belonging to the 'early era' and the last one as belonging to the 'modern era'. Although these three dictionaries deal with language varieties spoken in the Republic of the Congo, they are fully accepted in the Civili community of Gabon.

This article focuses on the representation of vowel duration in these dictionaries. Vowel duration is here understood as the natural time interval of a vowel in the speech production process. The term 'vowel duration' is used as a neutral designation in this article to avoid the labels 'vowel length' and 'vowel lengthening' which have proved to be problematic in Civili studies (cf. Ndinga-Koumba-Binza 2004).

In some languages, this feature is explicitly represented in the spelling of words, and as such lemmatized in dictionaries of these languages. Note the Afrikaans examples from Kromhout and Kritzinger (1998) under (1).

- (1) (a) mak (short) *tame* vs maak (long) *make*
 (b) makker (short) *pal* vs maker (long) *maker*

The representation of Civili vowel duration in existing dictionaries is inconsistent, as can be seen from the examples under (2).

- (2) **Marichelle** **ILALOK**
 (a) bwatu bwätu [b^wà:tù] *boat*
 (b) mbeeli mbëli [mbè:l̥] *knife*

The following can be perceived regarding the vowel duration:

- (i) Marichelle (1902 and 1912) is inconsistent with respect to the indication of vowel duration in the spelling of words. In cases where duration is represented, it occurs by doubling the vowel.
 (ii) In the ILALOK dictionary, vowel duration is systematically indicated and represented by a diaeresis above the vowel.

As this article aims to contribute towards the standardization of the Civili orthography, it seeks to answer the following question: How should vowel duration be represented in dictionaries?

The representation of vowel duration in Civili dictionaries is intimately linked to its representation in the Civili orthography, i.e. the word writing system. For this reason, this article commences with an assessment of existing proposals concerning the Civili writing system, with a focus on the representa-

tion of vowel duration. Findings of this assessment suggest that there are various issues to consider in establishing an orthographic alphabet for smaller languages such as Civili.

It is here hypothesized that the inconsistency of the representation of vowel duration in Civili stems from the fact that vowel duration is essentially a phonetic-phonological issue in the language. The views presented here are the results of systematic studies conducted by Ndinga-Koumba-Binza (2000, 2004 and 2008) at various levels (i.e. acoustic, perceptual and phonological). Results emanating from these studies have definite implications for spelling rules in Civili. It is argued that experimental phonetic and phonological considerations should apply when taking these decisions.

2. Existing Proposals for a Civili Orthography

Orthographic systems for various languages are often founded on the alphabets of these languages, and these alphabets are based on the phonetic and/or phonological systems of the languages (cf. Coulmas 1996; Hombert 1990).

The first record of an orthographic alphabet for Civili appears in Marichelle's dictionaries (1902 and 1912). The representation of vowel duration is seemingly based on the personal perceptions of the compiler.

Proceeding from his description of the language, Ndamba (1977) suggests a phonologically-based alphabet. It should be noted that the aim of Ndamba's work was not to establish an orthographic system for Civili, but to provide a linguistic analysis of Civili. However, the linguistic analysis commenced with a preliminary phonological account which in turn allowed for the transcription of various recorded oral stories into text. The writing system used by Ndamba (1977) is also observed in many of his illustrative texts and examples.

At a workshop held for establishing a scientific alphabet for the Gabonese languages ('Alphabet Scientifique des langues Gabonaises', cf. Hombert 1990a), Blanchon (1990) not only presented an outline of Civili phonology but also proposed an alphabet and orthography for Civili. His work consists of a brief overview of the phonological system of Civili in which he lists consonants, vowels and tones, and then illustrates his proposed alphabet and orthography with a short text written in Civili. Blanchon (1990) does, however, not explain any principles applied in his treatment of vowel length. He uses a succession of two identical segments in each context where it is seemingly possible to perceive long vowels.

Unlike Blanchon (1990), the works of Ndamba (1977), Mabika Mbokou (1999), and Ndinga-Koumba-Binza (2000) were not intended to contribute directly to the establishment of a standardized alphabet and orthography for Civili (cf. Ndinga-Koumba-Binza 2004: 190). These authors do not agree on the representation of vowel duration. For instance, from the Civili texts in Ndamba's (1977) appendices, it appears that he did not intend to indicate vowel length orthographically: the Civili words under (3) have phonetically

audible vowel length; however, orthographically Ndamba (1977), Blanchon (1990) and the dictionary compilers (Marichelle 1902 and 1912; ILALOK 2008) would render them differently.

(3)	Marichelle	Ndamba	Blanchon	ILALOK	
(a)	maama	mama	maama	mäm'	<i>mum, mother</i>
(b)	taata	tata	taata	tät'	<i>dad, father</i>
(c)	saku	saku	saaku	säku	<i>bag</i>
(d)	bana	bana	baana	bän'	<i>children</i>
(e)	mwana	mwaana	mwaana	mwän'	<i>child</i>

It should be noted that in places where the vowel is doubled, as Creissels (1994: 37) puts it, it might lead to a long vowel being regarded as a sequence of two distinct syllabic nuclei, i.e. as two syllables. This could pose some theoretical problems as well as practical difficulties in language description.

To the user, the current orthographic practice seems facultative despite the requirements of the Orthography of Gabonese Languages of 1999 (cf. Idiata 2002 and 2003 for details). In fact, the revised orthographic system for all Gabonese languages requires doubling of the vowel for any vowel duration (Idiata 2002: 55; 2003: 40). However, the writing system of Civili, as for many Gabonese languages, has experienced a number of problems with the implementation of this requirement. Two problems need to be mentioned here:

- (i) Since 1999, there has not been any campaign to inform the public of the proposed Civili writing system. In fact, the Government has been slow to endorse and implement the 'new' orthography.[†]
- (ii) The practice of different users may be based on that of different views existing among researchers, as demonstrated under (3) above. A number of textbooks and publications reflect the non-standardization or non-harmonization of the orthography with regard to vowel duration in Civili. For instance, Garnier (1903 and 1904) does not consider vowel duration in his textbooks, whereas Marichelle (1902 and 1912) consistently doubles vowels when he has to indicate duration in his dictionaries. For their part, the compilers of the ILALOK dictionary chose to represent vowel duration (wherever it can be perceived) with a diaeresis above the vowel.

Because different researchers model different writing systems for users, it is advisable that the practices of these researchers are reconciled.

In order to bring about such a reconciliation, Ndinga-Koumba-Binza (2009) suggested that vowels should be doubled even in predictable contexts, in order to make Civili learners aware of vowel duration. However, this proposal is not based on evident phonetic facts.

Answers to the questions of the physical nature of vowel duration and of the difference between the length and sequence of two adjacent identical vow-

els might help the process of standardizing the orthographic system of Civili.

The next section succinctly presents phonetic-phonological issues regarding Civili vowel duration. The orthographic issue expounded above is a thorough implication of this phonetic-phonological issue.

3. Phonetic-Phonological Issues

The orthographic problem in representing vowel duration in Civili actually stems from phonological studies of the language. In fact, like dictionary compilers, Civili phonologists disagree on vowel duration. The outcome of this disagreement is reflected in the inconsistency regarding the orthography and the ensuing standardization problem as indicated earlier under (3).

For instance, on the basis of minimal pairs, Ndamba (1977), Blanchon (1984 and 1990) and Mabika Mbokou (1999) admit significant contrasts between long and short vowels in Civili as illustrated under (4).

(4) (a)	losu	<i>dirtiness</i>	vs	loosu/lösu	<i>rice</i>
(b)	mbasi	<i>tomorrow</i>	vs	mbaasi/mbäsi	<i>friend</i>
(c)	n'tela	<i>height</i>	vs	n'teela/n'tëla	<i>hunter</i>

However, there are also predictable contexts that allow a vowel lengthening process under certain phonetic and syntactic conditions. This is shown by Ndinga-Koumba-Binza (2000 and 2004) who conducted an autosegmental analysis to describe the vowel duration phenomenon. He postulates that long and/or lengthened vowels are just an outcome of various processes within the underlying system of the language.

The orthographic implication of this claim by Ndinga-Koumba-Binza (2000 and 2004) is that vowel duration should be marked orthographically since long and/or lengthened vowels are predictable. However, this application would have been most appropriate if it were only based on evident phonetic facts.

In view of solving the vowel duration issue in Civili, Ndinga-Koumba-Binza (2008) conducted an extensive experimental study. The study consisted of an acoustic analysis and a perception experiment. The acoustic analysis included vowel duration measurements in various environments: minimal pairs, predictable contexts and syntactic positions. The acoustic analysis was complemented by a statistical analysis.

The perception experiment aimed to determine the perceived duration by means of three perception tests administered to 68 mother-tongue listeners. The 100 prepared electronic stimuli took into account the problematic contexts and environments (minimal pairs and predictable contexts). The perceptual data analysis was also supported by a statistical analysis.

The outcome of this experimental study showed a definite consistency in minimal pairs with respect to the existence of long and short vowels (Ndinga-

Koumba-Binza 2008). This implies that it is a phonetic fact (acoustically and perceptually) that two Civili words can be in significant opposition on the basis of the duration of an identical vowel in the same phonetic environments.

On the other hand, the experimental study showed a certain inconsistency in predictable contexts. This implies that vowels are perceived short or long owing to the phonetic and/or syntactic environment. For instance, a vowel that is perceived short in the word in isolation can be perceived long when the word is in the object position in a sentence.

This is in line with a general phonological rule (in Bantu languages) which assigns duration to the penultimate syllable in a sentence (cf. Watkins 1937: 10; Nurse 1996: 279; Childs 2003: 205). At the same time, a vowel that is perceived long owing to the phonetic environment, e.g. in the environment of nasal or liquid consonants or semivowels, can be perceived short when a word is in the subject position in a sentence (Batibo 1985: 23; Clements 1986: 45; Odden and Odden 1999: 2; Myers and Hansen 2005: 317; Ndinga-Koumba-Binza 2008: 164-168).

4. Implications for Orthography

The assessment previously made of existing proposals for the Civili spelling system has revealed some inconsistencies. The situation with regard to an acceptable and standardized orthography for Civili is indeed extremely confusing. Ndinga-Koumba-Binza (2008: 170-171) indicates that this is due to the fact that most of these proposals did not comply with certain methodological and social-acceptability principles for orthographic conventions for developing languages as suggested by various scholars (cf. Coulmas 1996; Touré 1990; Capo 2002).

With regard to the specific case of vowel duration in Civili, the following that have already been touched on in the previous section should be recalled here:

- (i) It has been experimentally found that Civili distinguishes between short and long vowels through minimal pairs (cf. Ndinga-Koumba-Binza 2008).
- (ii) In many cases, long vowels are predictable owing to either phonetic environments or syntactical positions.
- (iii) Experimental results in Ndinga-Koumba-Binza (2008) have also shown that the phonetic realization in these contexts may vary from speaker to speaker, however less so in minimal pairs.

Given the findings of Ndinga-Koumba-Binza (2008) as discussed in the previous section, the following two recommendations are made with respect to the orthographic representation of vowel duration in Civili.

(i) Vowel duration should only be orthographically represented in case of minimal pairs to avoid semantic confusion. In fact, long vowels in minimal pairs are basically long and should also be written as long. Note the examples under (5) below.

- (5) (a) mbila *call* vs mbiila/mbila *type, kind*
 (b) mbela *wrongness* vs mbeela/mbëla *knife*
 (c) n'totu *earth* vs n'tootu/n'tötu *tube, pipe*
 (d) libaku *clash, stumble* vs libaaku/libäku *fever*
 (e) mbusa *back* vs mbuusa/mbüsa *fish-net*

(ii) Predictable vowel length should not be marked orthographically, i.e. in all other cases a single vowel should be used orthographically. Note the following examples under (6).

- (6) (a) lumbotawu [lùmbótá:wù] *button*
 (b) mabena [mǎbé:nə] *breast*
 (c) simpinda [sǐp^hí:ndə] *peanuts*
 (d) mpokongu [mp^hókò:ŋgù] *problems, harassment*
 (e) sintumbu [sǐt^hú:mbù] *needles, syringes*

On the basis of the above recommendations, an alphabetic list of Civili vowels may be presented by means of the following table. Vowel duration is not represented in this list.

	Small	Capital	Examples
1.	a	A	sabi <i>key</i>
2.	e	E	ndelu <i>intestines</i>
3.	i	I	mili <i>intestines</i>
4.	o	O	bododo <i>completely empty</i>
5.	u	U	ndumu <i>reputation</i>

Alphabetic list of Civili vowels

5. Conclusion: Vowel Doubling or Diaeresis

On the basis of experimental phonetic-phonological considerations, it has been recommended in the previous section that vowel duration should be represented in the case of minimal pairs. However, the main question to deal with in this section is whether the representation should be to double the vowel or to add diaeresis to indicate a long vowel.

It should be recalled that doubling the vowel is the practice adopted by Marichelle (1902 and 1912) and Blanchon (1990) for the orthography of Civili

words, whereas ILALOK (2008) utilizes a diaeresis above each long vowel (despite the predictable phonetic environments).

The practice of adding a diaeresis on a single vowel symbol to indicate a long vowel as one member of a minimal pair should be supported. An application of this view helps to avoid the confusion of two adjacent syllables. Doubling the vowel could mislead language learners in word syllabification especially when the long vowel is at the initial or final position in the word. Note the following Civili examples under (7).

(7) (a)	uuluu	vs	üü	[ú:lú:]	<i>bravo</i>
(b)	voo	vs	vö	[vó:]	<i>nothing</i>
(c)	afanaa	vs	afanä	[áfəná:]	<i>over there</i>
(d)	awunaa	vs	awunä	[áwüná:]	<i>that one</i>

In these examples, it is clear that the use of diaeresis avoids overloading the word. It should be mentioned that tones are not marked in Civili orthography, thus allowing the use of a diacritic with the vowel. In addition, it should be set as writing and reading rule that the diaeresis represents vowel duration (a long vowel).

Finally, the full recommendation is that Civili vowel duration should be represented in the spelling system with a diaeresis above the vowel only for minimal pairs. If this recommendation is considered within dictionary compilation, e.g. in lemmatization and text examples, it may standardize the orthography of Civili, which already suffers from various inconsistent proposals.

Notes

- † Since Gabon gained independence in 1960, the Gabonese Government has shown little will-power with regard to the development of local languages. Most examples of developmental evidence (writing system, learning and teaching, publications, etc.) are still an inheritance from the colonial era accomplished by the work of missionaries and colonial administrators (Mayer 1990; Raponda-Walker 1998; Idiata and Leitch 2000; Mihindou 2001). In contrast to the Government, for whom these matters seem to bear low importance, Gabonese linguists have taken a number of steps towards the development of the Gabonese languages and the implementation of a proper language policy, as evidenced by a number of publications and workshops during the last two decades. (For details, see Ndinga-Koumba-Binza 2005, 2005a and 2007).

References

Dictionaries

- ILALOK (Institut des Langues Locales au Kouilou). 2008. *Dictionnaire vili-français/Mpisukulu bi kum' bi tshi vili ku tshi mputu*. Paris: L'Harmattan.

- Kromhout, J. and M.S.B. Kritzinger.** 1998. *Mini Afrikaans–Engels woordeboek/Mini English–Afrikaans Dictionary*. Cape Town: Pharos.
- Marichelle, C.** 1902. *Dictionnaire vili–français*. Loango: Imprimerie de la Mission.
- Marichelle, C.** 1912. *Dictionnaire français–vili*. Loango: Imprimerie de la Mission.

Other Literature

- Batibo, H.** 1985. *Le kesukuma (langue Bantu de Tanzanie): phonologie, morphologie*. Paris: Éditions Recherche sur les Civilisations.
- Blanchon, J.A.** 1984. Présentation du yi-lumbu dans ses rapports avec le yi-punu et le ci-vili à travers un conte traditionnel. *Pholia* 1: 7-35. [Reprinted in Blanchon, J.A. 1999. *Douze études sur les langues du Gabon et du Congo-Brazzaville*: 5-31. München: Lincom Europa.]
- Blanchon, J.A.** 1990. Civili. *Revue Gabonaise des Sciences de l'Homme* 2: 141-142.
- Capo, H.B.C.** 2002. The Pan-dialectal Approach of Orthographic Conventions: The Case of the Gbe Languages of West Africa. Prah, K.K. (Ed.). 2002. *Writing African. The Harmonization of Orthographic Conventions in African Languages*: 5-25. Cape Town: The Centre for Advanced Studies in African Society (CASAS).
- Childs, G.T.** 2003. *An Introduction to African Languages*. Amsterdam: John Benjamins.
- Clements, G.N.** 1986. Compensatory Lengthening and Consonant Gemination in LuGanda. Wet-zels, J. and E. Sezer (Eds.). 1986. *Studies in Compensatory Lengthening*: 37-77. Dordrecht: Foris.
- Coulmas, F.** 1996. *The Blackwell Encyclopedia of Writing Systems*. Oxford/Cambridge: Blackwell.
- Creissels, D.** 1994². *Aperçu sur les structures phonologiques des langues négro-africaines*. Grenoble: ELLUG.
- Garnier, A.** 1903. *Mumbembo kivili ki Mayumbe*. Loango: Imprimerie de la Mission.
- Garnier, A.** 1904. *M'ambu ma nzambi mo make mu katesisa*. Loango: Imprimerie de la Mission.
- Hombert, J.-M.** 1990. Problèmes phonétiques et phonologiques rencontrés dans les langues du Gabon. *Revue Gabonaise des Sciences de l'Homme* 2: 97-103.
- Hombert, J.-M.** 1990a. Présentation de l'alphabet scientifique des langues du gabonaises. *Revue Gabonaise des Sciences de l'Homme* 2: 105-111.
- Idiata, D.F.** 2002. *Il était une fois les langues gabonaises*. Libreville: Éditions Raponda-Walker.
- Idiata, D.F.** 2003. *Pourquoi le Gabon doit investir sur ses langues vernaculaires*. Cape Town: The Centre for Advanced Studies in African Society (CASAS).
- Idiata, D.F. and M.F. Leitch.** 2000. *Histoire du développement des langues gabonaises*. Paper presented at the UNESCO/ANACLAC International Conference on Adult Education, Yaoundé, Cameroon. December 6–8, 2000.
- Mabika Mbokou, L.** 1999. *Les phénomènes analogiques en civili: étude phonologique et morphologique*. Unpublished Master's Thesis. Libreville: Omar Bongo University.
- Mayer, R.** 1990. Histoire de l'écriture des langues du Gabon. *Revue Gabonaise des Sciences de l'Homme* 2: 65-92.
- Mihindou, G.-R.** 2001. Apports des missionnaires à la lexicographie gabonaise: Dictionnaires bilingues fang–français/français–fang; français–yipounou/yipounou–français; français–mpongwé. Emejulu, J.D. (Ed.). 2001. *Éléments de lexicographie gabonaise. Tome I*: 7-37. New York: Jimacs-Hillman.

- Myers, S. and B.B. Hansen.** 2005. The Origin of Vowel-length Neutralisation in Vowoid Sequences: Evidence from Finnish Speakers. *Phonology* 22: 317-344.
- Ndamba, J.** 1977. *Syntagme nominal et groupe nominal en vili: langue bantu du Congo*. Unpublished Doctoral Thesis. Paris: Sorbonne Nouvelle University.
- Ndinga-Koumba-Binza, H.S.** 2000. *Phonologie du civili de Mayumba: langue bantu du Gabon (H12a)*. Unpublished Master's Thesis. Libreville: Omar Bongo University.
- Ndinga-Koumba-Binza, H.S.** 2004. Vowel Duration Issue in Civili. *South African Journal of African Languages* 24(3): 189-201.
- Ndinga-Koumba-Binza, H.S.** 2005. Considering a Lexicographic Plan for Gabon within the Gabonese Language Landscape. *Lexikos* 15: 132-150.
- Ndinga-Koumba-Binza, H.S.** 2005a. Politique linguistique et éducation au Gabon: un état des lieux. *Journal of Education* 4(1): 65-78. Réduit: Mauritius Institute of Education.
- Ndinga-Koumba-Binza, H.S.** 2007. Gabonese Language Landscape: Survey and Perspectives. *South African Journal of African Languages* 27(3): 97-116.
- Ndinga-Koumba-Binza, H.S.** 2008. *Phonetic and Phonological Aspects of Civili Vowel Duration: An Experimental Approach*. Unpublished D.Litt. Dissertation. Stellenbosch: University of Stellenbosch.
- Ndinga-Koumba-Binza, H.S.** 2009. Annexes: Propositions pour l'orthographe du Civili. Hubert, J. and P.A. Mavoungou (Eds.). 2009. *Écriture et standardisation des langues gabonaises*: 191-199. Stellenbosch: SUN PReSS.
- Nurse, D.** 1996. Prior Pidginization and Creolization in Swahili. Thomason, S.G. (Ed.). 1996. *Contact Languages*: 271-294. Amsterdam/Philadelphia: John Benjamins.
- Odden, D. and M. Odden.** 1999. Kihehe Syllable Structure. Van der Hulst, H. and N.A. Ritter (Eds.). 1999. *The Syllable: Views and Facts*: 417-445. Berlin/New York: Mouton de Gruyter.
- Raponda-Walker, A.** 1998. *Les langues du Gabon*. Libreville: Éditions Raponda-Walker.
- Touré, A.** 1990. Écriture des langues africaines: évolution et principes méthodologiques. *Revue Gabonaise des Sciences de l'Homme* 2: 55-63.
- Watkins, M.H.** 1937. A Grammar of Chichewa: A Bantu Language of British Central Africa. *Language* 24: 5-158.

The Evaluation of the Outside Matter in Dictionary Reviews

Sandro Nielsen, *Centre for Lexicography, Aarhus School of Business, University of Aarhus, Aarhus, Denmark (sn@asb.dk)*

Abstract: Reviewing dictionaries is part of the ongoing work in lexicography, and several lexicographers have discussed the process and guidelines for reviews published in academic journals. However, few have addressed the evaluation of the outside matter and, if so, only in a cursory way. This article examines the evaluation of the outside matter in reviews published in *Lexikos* with a view to proposing some general principles for reviewing outer texts in printed and electronic dictionaries. The study shows that reviewers define the review object differently, some excluding the outside matter altogether, and that the way in which the outside matter is assessed differs within and between reviews. It is proposed that the separate sections of dictionaries should not only be examined independently but that their relationship to each other should also be evaluated so as to represent faithfully the lexicographic elements, i.e. wordlist, front, middle and back matter, their organisation and presentation, as well as three underlying elements: the function(s), data types and structures of the dictionary. Focus on all these elements may result in dictionary reviews that are academically sound because they treat the dictionary as a true research object.

Keywords: BACK MATTER, DICTIONARY REVIEWS, EXTRA-LEXICOGRAPHIC SECTIONS, FRONT MATTER, LEXICOGRAPHIC INFORMATION COSTS, LEXICOGRAPHIC SECTIONS, MIDDLE MATTER, OUTSIDE MATTER, OVERRIDING OBJECTIVE, PREFACES, SCHOLARLY WRITINGS, SUBJECT-FIELD SECTIONS, SUBSTANCE OVER FORM, USER GUIDES, WORDLISTS

Opsomming: Die evaluering van die buitewerk in woordeboekresensies.

Die resensering van woordeboeke is deel van die deurlopende werk in die leksikografie, en verskeie leksikograwe het die metode en riglyne vir resensies wat in akademiese tydskrifte gepubliseer is, bespreek. Min het egter die evaluering van die buitewerk in resensies aangeroer, en, indien wel, slegs op 'n terloopse manier. Hierdie artikel ondersoek die evaluering van die buitewerk in resensies wat in *Lexikos* gepubliseer is, met die doel om 'n aantal algemene beginsels voor te stel vir die resensering van buitetekste in gedrukte en elektroniese woordeboeke. Die studie toon dat resensente die resensieobjek verskillend definieer, met sommige wat die buitewerk heeltemal uitsluit, en dat die manier waarop buitewerk beoordeel word, verskil binne en tussen resensies. Daar word voorgestel dat die aparte afdelings van woordeboeke nie alleen onafhanklik ondersoek behoort te word nie, maar dat hulle verhouding tot mekaar ook beoordeel behoort te word om die leksikografiese elemente getrou te verteenwoordig, d.w.s. woordelys, voor-, middel- en agterwerk, hul rangskikking en aanbieding, sowel as drie onderliggende elemente: die funksie(s), datatipes en strukture van die woordeboek. Fokus op al hierdie elemente kan lei tot woordeboekresensies wat akademies grondig is omdat hulle die woordeboek as 'n werklike navorsingsobjek behandel.

Sleutelwoorde: AGTERWERK, WOORDEBOEKRESENSIES, BUIELEKSIKOGRAFIESE AFDELINGS, VOORWERK, LEKSIKOGRAFIESE INLIGTINGSKOSTE, LEKSIKOGRAFIESE AFDELINGS, MIDDELWERK, BUITEWERK, OORHEERSENDE DOELSTELLING, VOORWOORDE, WETENSKAPLIKE SKRYFWERK, ONDERWERPSVELDAFDELINGS, INHOUD BO VORM, GEBRUIKERSGIDSE, WOORDELYSTE

1. Introduction

Dictionary reviews published in academic journals play an important role in the development of lexicography. This is particularly true of international journals such as *Lexikos*, which contribute to lexicography by publishing a substantial number of reviews in each issue. A natural basis for such reviews is an understanding of the findings reported in scholarly writings on lexicography, including books and articles specifically dealing with dictionary criticism. A study of the literature on dictionary reviewing shows a tendency to focus on linguistic aspects in dictionary articles and to deal mainly with printed language dictionaries for learners, for example Steiner (1979), Tomaszczyk (1986), Osselton (1989), Ripfel (1989), Hartmann (1999), Chan and Taylor (2001) and Jackson (2002). These lexicographers discuss the process involved in reviewing dictionaries and suggest guidelines for reviews published in academic journals. However, few have addressed the evaluation of the outside matter and only in a cursory way. One consequence of this is that dictionaries are often evaluated on the basis of the data in the wordlists instead of being treated as complex units with many types of data found in the often numerous texts surrounding the wordlists. This should be compared with the findings reported in Wolf (1992: 335-338) showing that more than 50% of users read user guides, findings that indicate that outer texts are important to theoretical and practical lexicographers, reviewers and users.

Dictionary reviews can be either scholarly or non-scholarly depending on the actors and media involved. This article will examine reviews of dictionaries in *Lexikos*, which is an academic journal devoted to lexicographic research areas reflecting different theoretical schools of thought for the benefit of lexicographic specialists. So dictionary reviewing will be regarded as essentially a lexicographic exercise which may relate to cultural, factual and linguistic aspects as the case may be. On the basis of selected reviews included in *Lexikos* from its first issue up to issue 18 (published in 2008), the way in which reviewers have treated the outside matter will be examined and discussed. Furthermore, suggestions will be made as to how reviewers may analyse and critically evaluate the outside matter in dictionaries in an attempt to produce better reviews that increase the academic quality of dictionary criticism. This will involve three topics: the identification of the research object for dictionary reviews and its placing in a lexicographic framework; the discussion of the objective of dictionary reviewing in academic settings; and the actual approach to reviewing the outside matter.

2. Dictionary reviews and *Lexikos*

Several researchers have discussed the practice of criticising dictionaries, suggesting various ways to increase the quality of reviews. Some of the first scholarly contributions were Ripfel (1989), the eight contributions in the thematic part of *Lexicographica* 9-1993 as well as the seven contributions in the thematic part of *LexicoNordica* 10-2003, these being more recently followed by a number of individual contributions, including Chan and Loong (1999), Chan and Taylor (2001), Nielsen (2009) and Swanepoel (2008). *Lexikos* is an international journal based on high-quality research as pointed out in the 'Editorial Objectives' (*Lexikos* 2008: xv), and this is reinforced by the statement in the 'Instructions to Authors' that, among others, reviews 'must [...] meet the requirements of academic writing and are evaluated by the editors with this in mind' (*Lexikos* 2008: 466). This means that reviews are evaluated before publication to conform to a certain standard to ensure international research quality, i.e. to contribute to the development of the discipline concerned.

There is no doubt that reviews of dictionaries can contribute to the development of lexicography as a research discipline. The criticism of lexicographic publications, including dictionaries, is one of the fundamental practices of metalexicography. This does not mean, however, that all dictionary reviews can be characterised as 'scientific', 'scholarly' or 'academic', since many reviews are not written with the intention of contributing to the development of lexicography, for example most reviews published in newspapers and popular magazines. Their primary objective is to draw attention to a new publication and to provide information to consumers in general. This does not mean that these reviews cannot contribute to the development of lexicography but they are generally written and published by and for non-academic actors in non-academic media. In contrast, scholarly dictionary reviews are potential contributors to the development of lexicography as a research discipline, they are written and read by scholars in the natural course of their research work, and they are published in academic journals.

Research is often regarded as something special, different from one's usual ordinary activities. The question then is: What makes scholarly reviews special? A possible answer is given by Phillips and Pugh (2005: 48) who argue that 'research goes beyond description and requires analysis'. If this is extended to lexicography, it implies that dictionary reviews should not merely be descriptive but should contain more or less thorough analyses, evaluations and reflections of their objects. This is supported by the following description in the 'Instructions to Authors': '**Reviews:** An analysis and critical evaluation of published scientific sources and products, such as books and computer software' (*Lexikos* 2008: 466). In other words, scholarly reviews that contribute to lexicographic research are at a level above everyday reflection and contain analyses, evaluations and findings that can influence research conducted by others.

This has various implications for the actual work of reviewing. Firstly, reviewers must possess knowledge about the object of review and the reality in which it is rooted that is more extensive than the knowledge of ordinary people. Secondly, reviewers must explicitly use that knowledge in their reviews. Finally, it may be said that scholarly reviewers seek to achieve some sort of intersubjectivity by using generally accepted (lexicographic) methods that enable lexicographers and researchers from other disciplines to learn from the statements made in reviews and to let these statements affect their own research. In order to produce findings that support or contradict this position, it is first necessary to delimit the scope of the study.

3. What is examined?

Throughout its history, *Lexikos* has contained reviews of dictionaries. For practical reasons, however, it has been necessary to limit in the following ways the number of reviews that form part of the discussion. First, only reviews of dictionaries (as defined by their name, e.g. 'dictionary', 'woordeboek', 'lexicon', 'encyclopedia' and 'glossary') found under the heading 'Resensies/Reviews' are considered. In accordance with the 'Instructions to Authors', review articles are regarded as research articles, not as reviews (*Lexikos* 2008: 466). Second, the number of reviews is further limited to those written in English, as they contribute to the development of lexicography in the international academic community. This results in 40 dictionary reviews published in 13 issues. Even though some issues have no dictionary reviews or no reviews that fall within the definition above, the texts examined can be said to span the entire life of the journal. However, not every possible topic will be addressed in the discussion below.

Dictionaries contain many individual textual components and data types. As pointed out by Bergenholtz, Tarp and Wiegand (1999: 1763), the printed dictionary is a collection of text types or genres, and the most general distinction is that between the wordlist and the outside matter. This article will focus on the outside matter and how this has been evaluated in the 40 reviews, though reference will be made to the wordlist when it is relevant for the understanding of arguments put forward. The outside matter, alternatively called outer texts, comprises the front matter, the middle matter and the back matter, this distinction being relevant for any printed dictionary. A similar distinction can be applied to online dictionaries, as they may and often do consist of a number of text files linked to the database in which searches are made, for instance the *Oxford English Dictionary Online* (2009) and the *Merriam-Webster Online Dictionary* (2009), though it is difficult to allege that a particular section is placed before or after another section, while features such as pop-up windows with written or audiovisual data may be regarded as middle matter. How, then, do reviewers perceive their object of review with specific reference to the outside matter?

4. A dictionary is more than just a dictionary

An important aspect of scholarly reviewing is the ontological position of dictionary reviewers which is basically the answer to the following question: What is the nature of the object that is being analysed, examined and evaluated, i.e. the research object? The answer to this question is significant, for research is not value-free but value-laden. No matter how objectively they attempt to conduct their analyses and present their findings, reviewers will do so within a theoretical framework. This does not pose any problem as long as they do so in an academically sound way.

The ontological position directly affects the way in which reviewers treat the dictionaries under review. The *object of dictionary reviewing* is obviously the dictionary, and the theoretical framework within which reviewers place the dictionary is directly or indirectly reflected in their reviews. A small quantitative study may illustrate this point. Out of the 40 reviews examined 25 evaluated outer texts, whereas in 15 cases there was no evaluation of outer texts but the focus was on what the wordlists contain. It is difficult to imagine that these 15 dictionaries had no outer texts, but apparently the outside matter had no place within the ontological positions of the reviewers. This leads to the conclusion that these 15 reviews do not evaluate dictionaries as collections of text types in their entirety but reflect a restricted view of the research object. The question is whether this complies with the *objective of dictionary reviewing*.

It is possible to identify more than one objective, or purpose, of dictionary reviewing in the existing literature. Lexicographers generally agree that reviews should describe, analyse and examine dictionaries (e.g. Ripfel 1989: 31; Bergenholtz and Mogensen 1993: 10), Swanepoel (2008: 209) mentioning that reviews have two major goals:

- (1) to assist readers in their decision-making in acquiring the best dictionaries for their usage needs by presenting them with a well-founded analysis of the positive and negative qualities of a dictionary/dictionaries under review, and (2) to assist lexicographers in optimizing the functionality of their dictionaries.

The above purposes are rather specific, because they deal with individual aspects of a wider picture. Moreover, it is necessary to list several purposes to give a meaningful explanation of the objective of dictionary reviewing, and the purposes selected depend on the individual giving the explanation in each case. One solution to this problem is to introduce what may be called an *overriding objective of dictionary reviewing*. It is provisionally proposed that the overriding objective is to give a faithful representation of the lexicographic elements presented by the dictionary under review. It is further suggested that it is only possible to attain a faithful representation if the depiction of a dictionary is complete when it takes the whole dictionary into account. This approach complies with generally accepted academic requirements, because the dictionary is no longer just a product to be used, but a true research object to be analysed, described and examined.

The introduction of the overriding objective has at least two consequences. First, it impacts directly on the ontological position of reviewers, as it requires them to treat the dictionary as a complex whole. A dictionary is a lexicographic tool that consists of a number of distinct sections whose common function is to provide answers to users so that they can solve particular types of problem. Second, reviewers who do not take the various dictionary sections into account risk that their reviews lack too much information which is useful to readers. For instance, after having read the review of a dictionary of public administration by Marais (1998), readers still do not know what the dictionary looks like, nor do they know which sections it contains. The text of a review of a bilingual dictionary by Nkabinde (1993) is better structured than that of Marais (1998) as it is divided into numbered paragraphs with appropriate headings. However, the review gives no indication of what the actual dictionary looks like and which sections it contains. Even if dictionaries contain a minimum of sections, this would be relevant information for readers, and reviewers should therefore comment on this minimalistic approach to dictionary making.

The 25 reviewers who address the outside matter indicate their ontological position in various ways. One way in which authors of scholarly articles reveal their theoretical framework is through the bibliography or reference section. A closer study shows that only 11 out of 25 reviews contain reference sections. One interpretation of this finding could be that 14 of these reviews do not comply with the requirement that academic texts should be in a different league than reviews published in non-academic media, which are often characterised by having no bibliographies or reference sections. Table 1 summarises the findings of a more detailed examination of the 11 reviews that contain bibliographies and divides the works listed into four broad categories.¹

Reviews with references to dictionaries	Reviews with references to other reviews	Reviews with references to lexicographic literature	Reviews with references to non-lexicographic literature
10	2	6	7

Table 1: Number of reviews dealing with the outside matter with references divided by category.

The numbers in Table 1 give rise to the following comments. First, 10 reviews list dictionaries, both the dictionaries under review and other dictionaries, but this reveals nothing about the ontological position of the reviewers. Second, it is interesting to see that 2 reviewers refer to other dictionary reviews; this seems to underline the importance of reviews in the academic community. Third, 6 reviewers refer to lexicographic literature (books and articles). In one respect this is somewhat disappointing, because dictionary reviewing falls within the discipline of lexicography. As dictionary reviewing is a lexicographic exercise it is not unreasonable to expect that reviewers explicitly state

their theoretical point of view through literature belonging to the discipline, and therefore that all reviewers refer to lexicographic literature. Finally, 7 reviews refer to non-lexicographic literature (books and articles dealing with topics such as grammar, language teaching, musical works, plants and language history). Prima facie, this indicates that most reviewers have an approach to dictionary reviewing that is not based on lexicography.

In order to corroborate the above findings, the actual number of references by category is relevant for a quantitative analysis. Table 2 shows how many individual references divided by category the 11 reviews contain.²

Number of references to dictionaries	Number of references to other reviews	Number of references to lexicographic literature	Number of references to non-lexicographic literature
29	3	9	37

Table 2: Total number of individual references by category

Two of the numbers in Table 2 warrant consideration. It is striking that the number of non-lexicographic literature is more than four times higher than the number of lexicographic references. This reinforces the impression from the numbers in Table 1 that the reviewers do not have a lexicographically based theoretical framework but rather one influenced by other disciplines. References to non-lexicographic literature are natural in dictionary reviews but it does not seem to fit in with the fact that reviewing dictionaries is an exercise in lexicography. From a research point of view, it is important that the texts produced within a discipline are explicitly based on the fundamental theories and principles of that discipline, which also applies to academic dictionary reviewing. It is proposed that an appropriate approach is to say that dictionaries are first and foremost lexicographic products based on lexicographic principles and in the second place relate to other disciplines through their data contents and functions. Consequently, reviews in general and reviews of the outside matter in particular should have a lexicographic core basis with a supplementary basis related directly to the other discipline(s) involved, for instance economics, law, linguistics and translation. This means that the ideal reviewer should be an expert in lexicography plus something else, or that teams of reviewers with different competences are needed. However, bibliographic references are not the only indications of ontological positions of reviewers.

If one looks behind the numbers in Tables 1 and 2, it turns out that a quantitative analysis like this has drawbacks. The main disadvantage is that even though there are references to lexicographic literature at the end of 6 reviews, it is difficult to pinpoint which references have been instrumental in forming the reviewers' perception of the concept of dictionary. Furthermore, the references are mostly so general that it is unclear which concern the issue of the outside matter. This indirect way of indicating references is inferior to

placing references to lexicographic literature directly in the text. Ridge (1993: 293) contains a direct reference to lexicographic literature dealing with the outside matter: 'The introductory section is also kept as short as possible. As Lemmens and Wekker (1986) have shown, users of a dictionary are unlikely to consult detailed discussion in an introduction.' This is the most explicit reference to a theoretical position on the outside matter found in the material studied, but the reviews contain other clues.

Information about the treatment of outer texts can be found outside the reference sections. The use of different linguistic means in their texts directly or indirectly reveals the reviewers' perception of the notion of dictionary. For instance, after having said that 'the dictionary begins with initial information for the users', Batibo (2004: 451) continues by saying that 'the dictionary itself is divided into two main parts', by which he means that the dictionary has two bilingual wordlists. These utterances show that the reviewer has a narrow understanding of what a dictionary is. The reference work reviewed by Batibo (2004) is called *a dictionary* and therefore the entire book is *the dictionary*, i.e. the research object, and the dictionary does not start on, say, page 25. An equally awkward use of terminology is found in Chebanne (2007a: 447): 'The dictionary consists of preliminary remarks and acknowledgements, the table of contents, abbreviations and symbols, an introduction, the Khwe–English dictionary [...]'. Does it really make sense to say that a dictionary contains front-matter texts and a dictionary? At least this does not seem logical. In a similar vein, Mashamaite (2000: 329) claims that 'the dictionary has a simple structure', by which he means that 'the wordlist has a simple structure'. However, other reviewers have a broader view of the object of dictionary reviewing, as explained by Gold (1993: 305):

Reviewing a dictionary means not only scrutinizing the lexical material which it provides, but also examining its front matter, back matter, covers, spine, jacket, and title, as well as the publisher's advertizing for it.

In order to meet the overriding objective of giving a faithful representation, reviewers should have a broad understanding of the object of their reviews. Of course, the publishers' advertising material is not part of the dictionary, but may help reviewers in their work. However, for the purpose of dictionary reviewing, reviewers should regard dictionaries as complex objects that contain a number of independent sections constituting a whole: the dictionary. By comparison, those who review the performance of symphony orchestras evaluate the entire orchestra, not just one or two sections. This is not to say that reviewers who have a narrow understanding of the concept of dictionary write non-academic reviews, but it may indicate that they are perhaps (slightly) off track.

In this context, it may be relevant to make a distinction between two general types of outside matter: *outside matter with a simple composition* containing a maximum of two outside matter sections, and *outside matter with a complex com-*

position containing more than two outside matter sections. A typical example of simple outside matter contains a preface and a user guide, and it will be relevant to discuss whether such a simple set of outer texts is sufficient for fulfilling the function(s) of the dictionary, taking into account the linguistic, factual and cultural competences of intended users. Typically, complex outside matter contains a preface, a user guide, a subject-field introduction, a style guide, and several appendices, and reviewers should examine whether these outer texts individually and collectively live up to expectations and report their findings. In order to give useful information about the complete dictionary and represent it faithfully, these individual sections should be addressed by reviewers. So far, the discussion has had a quantitative focus, but the evaluation of the outside matter also involves a qualitative aspect.

5. What the reviews say and could say

In order to obtain a detailed picture of the evaluation of the outside matter in dictionary reviewing, further examination of the reviews is necessary. This will involve a content analysis of the 25 reviews that address the issue of the outside matter. In addition to the quantitative findings mentioned above, such a content analysis will reveal the degree of scholarliness of the reviews, indicate to what extent the reviews give a faithful representation of the dictionaries, and suggest whether the reviews contribute to the development of lexicography.

When one moves on to a content analysis, the definition of a dictionary as a complex whole in the two-dimensional space with wordlist and outer texts cannot stand alone. In addition to its surface features, a dictionary — whether printed or electronic — has three underlying features, namely:

- (a) It has been designed to fulfil one or more functions.
- (b) It contains data that have been selected because they help to fulfil its function(s).
- (c) It has structures that marshal its data into the task of fulfilling its function(s).

Only when it is defined in terms of both surface and underlying features, does one have a proper account of a dictionary as an object of analysis, description, evaluation and research. And this is what the dictionary consists of for reviewing purposes. However, as pointed out by Nielsen (1999: 301) dictionaries often lack clear statements about their functions, so reviewers have to make educated guesses to assess whether the data support the functions for which the dictionaries are intended. Despite this lack of indications as to the functions of dictionaries, reviewers should be aware of the role of outer texts for, as Gouws (2007: 83) accentuates, the wordlist 'is not the only venue for the realisation of functions'.

The extension of the definition of the review object entails a further specification of the overriding objective of dictionary reviewing. It is proposed that

this objective is to give a faithful representation of the lexicographic elements of the surface features, i.e. wordlist and outer texts, their organisation and presentation, as well as the underlying features, i.e. the function(s), data types and structures of the dictionary. Any evaluation of the outside matter should be conducted with this objective in mind, and it will also be used as a benchmark for assessing the qualitative features of the reviews examined.

The analysis carried out reveals that there are considerable differences in the way in which reviewers comment on outer texts. Some reviewers merely acknowledge the existence of outer texts, such as Schuring (1994: 298) who identifies one outer text and says: 'The bibliography at the end of the dictionary covers eleven pages.' This is a minimalist approach to reviewing, which does not meet the requirements of critical evaluation, though it could be argued that it meets the requirement of quantitative analysis. However, the utterance is merely a description and statement of fact and it does not contribute to the qualitative development of lexicography. On the whole, this reviewer does not give any significant evaluation of the outside matter (e.g. by not relating the bibliography to the function(s) of the dictionary), which may also be an indication of a narrow understanding of the reference work under review. In a different type of review, Snyman (2000: 332) refers several times to pages in the dictionary without explaining which section he is speaking about, for instance 'an orthographic list of symbols' and 'the list of abbreviations'. The forms of the references, which mention p. i and p. iv respectively, indicate that the reviewer refers to text in the front matter, but does not expand on it. An explicit indication of the section to which he refers would have resulted in a more faithful representation, enabled readers to better judge the quality of the dictionary reviewed, and avoided unnecessary speculation and doubt.

Not all the reviewers have taken a minimalist approach to their evaluation of outer texts. In their review of a music dictionary, King and Steyn (2003: 343-344) discuss the contents of the 'Introduction' in 75 lines. Impressive though this is, readers only get a glimpse of what this introductory text contains and what its purpose is, because the reviewers focus on the musical contents of the text and ignore lexicographic aspects. The criticism is no doubt correct, but the discussion does not contribute to the development of lexicography, though it gives a detailed evaluation of the quality of the data presented in the 'Introduction'. Apart from one reference to the publishers' blurb on the back cover of the dictionary, no other outside matter is mentioned in the review. A similar treatment of the outside matter is found in Chebanne (2007b: 481-483), though in this case the two-page discussion of the introduction concerns the status and history of the language covered by the dictionary. Again, this is the only outer text the reviewer considers and the points he raises are no doubt correct, but the evaluation does not contribute to the development of lexicography, and it is therefore doubtful whether the review faithfully represents the dictionary as defined above.

Several reviewers explicitly mention all the outer texts found in the dic-

tionaries. At first glance, this meets the requirement of giving a faithful representation of the entire dictionary, though some reviewers treat outer texts differently in the same review. Feinauer (1996: 280) explicitly mentions that the dictionary under review contains ten appendices, but only two are singled out for comment with the result that the informative value of the review is lower than it could have been. Benjamin (2001: 312) describes the organisation of the dictionary, saying about the user guide: 'The explanation of the entries is a valuable source of learning material and even teachers of grades 8–9 (the first two years of high school) should find the dictionary useful.' Not being an academic analysis of the purpose of user guides, this is merely a description which contributes nothing to the development of lexicography. However, Gold (1993) and Mojela (2003) are good examples of reviews that attempt to give a faithful representation of the dictionary in its entirety, addressing the individual sections in a way that is informative and useful to readers. One minor flaw can be found, though, in that Mojela (2003: 334) states that 'the dictionary has important features such as front and back matter' but does not comment on the back-matter texts.

One area in which the terminology used by reviewers — who no doubt use the terminology found in the dictionaries — is ambiguous concerns the 'introduction', also referred to in the reviews as introductory remarks, introductory pages, introductory guide, initial information for users, and preliminaries. First, not all reviewers make it clear which types of data this 'introduction' contains and what its function is (e.g. Hughes 1996). Second, some reviews show that the introduction is a section giving a preliminary to a technical subject field or the language treated by the dictionary in question (e.g. Sitaram 1996). Third, a number of reviews indicate that the introduction is a collection of text types, in particular the preface and the user guide (e.g. Mojela 2003). For readers to acquire a faithful representation of the object analysed and examined, reviewers should comment on the fact that, by lumping several and distinct types of data together in one outer text, the dictionary authors have not helped users sufficiently. Reviewers should point out that hiding important information from users is not the best option. Finally, ambiguous statements fall short of the requirements of academic standards. In order to contribute to the development of lexicography, reviewers should make precise and unambiguous statements so that they do not leave readers in doubt of what is actually meant, of what was left unsaid, and why. For instance, does the dictionary reviewed by Schuring (1994) really only contain a wordlist and a bibliography? And why did Feinauer (1996) single out only two out of ten appendices?

Directly related to the ambiguous term 'introduction' and its variants, is the effect it has. Most reviewers seem to use an approach that places *form over substance*. This means that their focus is on the formal names of the sections, i.e. it seems to be more important to mention the sections by their formal designations than to focus on the actual data contents. The main drawback of this approach is that it often results in description rather than analysis and critical

evaluation. It is proposed that a better approach is to place *substance over form*, i.e. to focus on what types of data are actually in the sections, how these are accessed, how these are presented, and how these support the functions of the dictionary. One way that can help change the focus from form to substance is to study the lexicographic literature where lexicographers discuss the various sections that can be found in the outside matter, some arguing that prefaces and user guides should be presented separately in dictionaries. Herberg (1989) and Nielsen (1994: 88-93; 1995: 168-169) discuss prefaces in dictionaries, their functions and data types. On the basis of these discussions, it is proposed that reviewers should evaluate prefaces by indicating to what extent they (at least) answer the following questions:

- (a) What is (are) the function(s) of the dictionary (e.g. text reception, text production or translation)?
- (b) What is the subject field of the dictionary (e.g. linguistics, law, mechanical engineering, or a multiplicity of subject fields)?
- (c) What is (are) the target group(s) of the dictionary?
- (d) What is the empirical basis of the dictionary (e.g. sources of lemma or equivalent selection)?

The existing literature also discusses user guides in dictionaries. Reviewers' lack of attention to user guides is noted by Chan and Taylor (2001: 168), who claim that most reviewers do not 'appreciate the potential contribution made by information contained in the user's guide to using a dictionary to its fullest extent'. Theoretical articles such as Kirkpatrick (1989) and Nielsen (2005; 2006) discuss the function and data types in dictionary user guides. These contributions lead to the suggestion that reviewers should evaluate user guides by indicating to what extent they (at least) answer the following questions:

- (a) How do users best access the data when using the dictionary as an aid in, for example, reading, writing and translating texts?
- (b) What types of data are in the dictionary?
- (c) Where can those data types be found (e.g. in separate sections or in the wordlist)?
- (d) How do the data in the individual sections in the outside matter relate to, complement or supplement the data in the wordlist and the other outer texts?
- (e) In what types of user situation are the data in the sections likely to help users?

One benefit of using the substance-over-form approach in discussing the above questions is that it may alert reviewers to aspects relevant for the development of lexicography. For example, prefaces and user guides that answer few questions may be used as examples of how not to write these sections. Conversely, prefaces and user guides that answer all questions or more may be used as

examples to be followed by both theoretical and practical lexicographers. Finally, critical evaluations of these two sections may lead dictionary authors to write really good prefaces and user guides, so that they are read because users find them helpful and relevant.

Reviewers may also examine to what extent the factual contents of outer texts relate to the functions of dictionaries. As discussed in Bergenholtz, Tarp and Wiegand (1999: 1778-1791) and Bergenholtz and Nielsen (2006), lexicographic data are mainly found in the wordlist but may be placed anywhere in the dictionary depending on its distribution structure. The relevant data are found in the *lexicographic sections* of the dictionary, i.e. those textual components that contain data about its use and function(s), for example in sections that inform users of the subject field treated by the dictionary (e.g. geology (Zawada 1997)). Other textual components such as advertising texts, picture credits and single user licence agreements are *extra-lexicographic sections* as described in Nielsen (2006: 7-8), irrelevant for academic dictionary reviewing as they do not contain lexicographic elements. In this context, it is suggested that reviewers look at the number of subject fields treated, and, if only one, whether it is the entire field, or one or more subfields, and whether the section is integrated into the wordlist. Sections to which there are cross-references from the wordlist or vice versa are integrated sections. It would, for instance, have been relevant for readers to know whether there are cross-references from individual articles to the style guide in the dictionary reviewed by Heliel (1998) so that the guide helps users by actually supporting the function of the dictionary. This type of integration can, for instance, be found in *Longman Exams Dictionary* (2006) where articles contain cross-references to the 'Writing Handbook' in its middle matter. An analysis and critical evaluation of such an integration may provide valuable contributions to the development of lexicography.

An aspect that extends to surface as well as underlying features of dictionaries is the concept of lexicographic information costs. According to Nielsen (2008: 173-174), *lexicographic information costs* are defined as the effort that users believe or feel is associated with consulting a dictionary, an article or any other text part of a dictionary such as outer texts. The focus is on the relationship between the expected and actual information costs and the expected and actual informative value, i.e. what users gain from consulting the dictionary. It is important to note, however, that high information costs are not always a disqualification. Even if the lexicographic information costs are high in a particular case, they may be offset by a search result with a large informative value to the user. In accounting terms, the information costs represent expenses and the informative value represents income and if the income exceeds the expenses, the bottom line will post a profit, i.e. show a positive gain.

Two general types of lexicographic information costs are relevant in connection with outer texts. *Search-related information costs* are the efforts related to the look-up activities users have to perform when consulting outer texts to gain access to the data they are searching for; this relates to the search activities nec-

essary to use both printed and electronic dictionaries. Examples of situations involving search-related information costs include cross-references to and within outer texts; the need to scroll up and down the screen to find something in outer texts; the possibility of searching for specific words in outer texts; and the time it takes to open and interact with multimedia elements (see e.g. Louw 2000: 326-327). *Comprehension-related information costs* are the efforts related to the user's ability to understand and interpret the data presented in outer texts. A typical example is whether the text in the subject-field section is in plain intelligible language adapted to the linguistic, factual and cultural competences of intended users. The design and organisation of outer texts may contribute to keeping lexicographic information costs down, as an inappropriate design, arrangement and layout may lead to high information costs. The actual wording and presentation of the data in outer texts, for instance the use of long and complex sentences, may increase information costs, whereas clear search routes may reduce lexicographic information costs. The ease or difficulty with which users can decode messages on sound files with, for example, explanations and pronunciations may reduce or increase lexicographic information costs.

The above discussion mainly deals with outer texts usually found in the front matter of printed dictionaries. This does not mean that some outer texts are more important than others for the purpose of dictionary reviewing. On the contrary, all outer texts that are lexicographic sections are equally important. Furthermore, prefaces, user guides and subject-field sections can be placed in the front, middle and back matter, as shown in Gold (1993: 305), though there is a tendency to place them in the front matter just as appendices are normally placed in the back matter. Some outer texts are placed on CD-ROMs that accompany printed dictionaries, while the distinction front versus back is irrelevant in online dictionaries. The important point is that each individual type of lexicographic section in the outside matter should be analysed, described, examined and critically evaluated on the basis of their substance related directly to the surface and underlying features of the dictionary, no matter where the sections are placed.

6. Concluding remarks

By its focus on linguistic categories in dictionary articles, existing literature on dictionary criticism gives little attention to the evaluation of outer texts. This results in a lacuna, leaving reviewers of dictionaries largely to their own devices. The result is that many reviewers exclude outer texts from their evaluation. Those who do include outer texts treat these sections very differently, ranging from a mere acknowledgement of their existence over a description to a critical evaluation. Consequently, the academic standard of treating outer texts varies much, in each case depending on the reviewer. This indicates that there is a need for general principles that can help academic reviewers analyse and critically evaluate the outside matter in printed and electronic dictionaries.

In order to encourage more reviewers to include outer texts in their reviews and thereby write texts that contribute to the development of lexicography, attention should be given to several factors. It is important to appreciate that dictionary reviewing is an exercise in lexicography. This means that a lexicographically based framework broadens the scope so that reviewers are likely to see the dictionary as a complex whole with surface and underlying features that all form the core of reviewing. This should be linked to the overriding objective of dictionary reviewing, which is to give a faithful representation of the lexicographic elements that make up the dictionary. Reviews that meet the overriding objective will not only examine the separate sections of the dictionary independently but also examine their relationship to each other and the wordlist. Finally, in their evaluation of outer texts, reviewers should place substance over form and focus on data contents and presentation. The application of all these elements may result in dictionary reviews that are academically sound because they treat the dictionary as an object of analysis, examination and evaluation, i.e. as a true research object.

Notes

1. Table 1 only shows numbers for reviews dealing with the outside matter, but these numbers may be compared with the corresponding numbers for all reviews examined. Out of 40 reviews, 18 have bibliographies or reference sections, the following table showing how many of these contain references to the four categories of literature:

Reviews with references to dictionaries	Reviews with references to other reviews	Reviews with references to lexicographic literature	Reviews with references to non-lexicographic literature
14	3	9	13

2. Table 2 only shows numbers for reviews dealing with the outside matter, but these numbers may be compared with the corresponding numbers for all reviews with bibliographies or reference sections. The table below shows how many individual references the 18 reviews with reference sections divided by category contain:

Number of references to dictionaries	Number of references to other reviews	Number of references to lexicographic literature	Number of references to non-lexicographic literature
42	4	26	54

References

Dictionaries

Longman Exams Dictionary 2006 = Bullon, S. et al. (Eds.). 2006. Pearson/Longman.

Merriam-Webster Online Dictionary 2009 = **Mish, F.C. et al. (Eds.)**. 2009. Merriam-Webster. <http://www.merriam-webster.com> [27 March 2009].

Oxford English Dictionary Online 2009 = **Simpson, J. (Ed.)**. 2009. Oxford University Press. <http://dictionary.oed.com> [27 March 2009].

Other literature

Batibo, H.M. 2004. [Review of] Hessel Visser. *Naro Dictionary: Naro-English, English-Naro*. Fourth edition, 2001. *Lexikos* 14: 451-454.

Benjamin, M. 2001. [Review of] Dorothea Mantzel and Bernd Schultz. *Francolin Illustrated School Dictionary for Southern Africa*. 1997. *Lexikos* 11: 311-313.

Bergenholtz, H. and J.E. Mogensen. 1993. Wörterbuchkritik in Dänemark. *Lexicographica* 9: 8-35.

Bergenholtz, H. and S. Nielsen. 2006. Subject-field Components as Integrated Parts of LSP Dictionaries. *Terminology* 12(2): 281-303.

Bergenholtz, H., S. Tarp and H.E. Wiegand. 1999. Datendistributionsstrukturen. Makro- und Mikrostrukturen in neueren Fachwörterbüchern. Hoffmann, L., H. Kalverkämper, H.E. Wiegand together with C. Galinski and W. Hüllen (Eds.). 1999. *Fachsprachen. Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft/Languages for Special Purposes. An International Handbook of Special-Language and Terminology Research: 1762-1832*. Berlin/New York: Walter de Gruyter.

Chan, A.Y.W. and Y. Loong. 1999. Establishing Criteria for Evaluating a Learner's Dictionary. Berry, R., B. Asker, H. Hyland and M. Lam (Eds.). 1999. *Language Analysis: Description and Pedagogy*: 298-307. Hong Kong: Hong Kong University of Science and Technology.

Chan, A.Y.W. and A. Taylor. 2001. Evaluating Learner Dictionaries: What the Reviews Say. *International Journal of Lexicography* 14 (3): 163-180.

Chebanne, A. 2007a. [Review of] Christa Kilian-Hatz. *Khwe Dictionary. With a Supplement on Khwe Place-names of West Caprivi* by Matthias Brenzinger. 2003. *Lexikos* 17: 446-452.

Chebanne, A. 2007b. [Review of] Ferdie Weich. *San Dictionary/San-woordeboek. San-Afrikaans-English/English-San-Afrikaans/Afrikaans-San-English*. First edition/Eerste uitgawe, 2004. *Lexikos* 17: 481-486.

Feinauer, I. 1996. [Review of] Jonathan Crowther (Editor), Kathryn Kavanagh (Assistant Editor) and Michael Ashby (Phonetics Editor). *Oxford Advanced Learner's Dictionary of Current English*. Fifth edition (New International Students' Edition), 1995. *Lexikos* 6: 280-282.

Gold, D.L. 1993. [Review of] P.A. Joubert. *Bilingual Phrase Dictionary/Tweetalige frasewoordeboek (E-A)*. First edition, 1992. *Lexikos* 3: 303-313.

Gouws, R.H. 2007. A Transtextual Approach to Lexicographic Functions. *Lexikos* 17: 77-87.

Hartmann, R.R.K. 1999. *Teaching and Researching Lexicography*. Harlow: Longman.

Hausmann, F.J., O. Reichmann, H.E. Wiegand and L. Zgusta (Eds.). 1989-1991. Wörterbücher. Ein internationales Handbuch zur Lexikographie/Dictionaries. An International Encyclopedia of Lexicography/Dictionnaires. Encyclopédie internationale de lexicographie. Berlin/New York: Walter de Gruyter.

Heliel, M.H. 1998. [Review of] Morton Benson, Evelyn Benson and Robert Ilson. *The BBI Dictionary of English Word Combinations*. Revised edition, 1997. *Lexikos* 8: 316-321.

- Herberg, D.** 1989. Wörterbuchvorwörter. Hausmann, F.J., O. Reichmann, H.E. Wiegand and L. Zgusta (Eds.). 1989–1991: 749-754.
- Hughes, G.** 1996. [Review of] *Longman Dictionary of Contemporary English*. Third edition, 1995. *Lexikos* 6: 300-302.
- Jackson, H.** 2002. *Lexicography. An Introduction*. London/New York: Routledge.
- King, G. and C. Steyn.** 2003. [Review of] Reino Ottermann en/and Maria Smit, Hoofredakteurs/Chief editors, bygestaan deur/assisted by Izak Grové, Winfried Lüdemann, Heinrich van der Mescht en/and Caroline van Niekerk. *Suid-Afrikaanse Musiekwoordeboek/South African Music Dictionary*. Tweede, hersiene en vermeerderde uitgawe/Second, revised and enlarged edition, 2000. *Lexikos* 13: 342-346.
- Kirkpatrick, B.** 1989. User's Guides in Dictionaries. Hausmann, F.J., O. Reichmann, H.E. Wiegand and L. Zgusta (Eds.). 1989–1991: 754-761.
- Lemmens, M. and H. Wekker.** 1986. *Grammar in English Learners' Dictionaries*. Tübingen: Max Niemeyer.
- Louw, P.** 2000. [Review of] Latha Menon (Executive Editor). *Microsoft Encarta Encyclopedia Deluxe 2000*. *Lexikos* 10: 326-328.
- Marais, D.** 1998. [Review of] William Fox and Ivan H. Meyer. *Public Administration Dictionary*. 1995. *Lexikos* 8: 326-329.
- Mashamaite, K.J.** 2000. [Review of] T.J. Kriel (Original Compiler), D.J. Prinsloo and B.P. Sathège (Compilers — Revised edition). *Popular Northern Sotho Dictionary*. 1997. *Lexikos* 10: 329-330.
- Mojela, V.M.** 2003. [Review of] Louis Molamu. *Tsotsitaal: A Dictionary of the Language of Sophiatown*. 2003. *Lexikos* 13: 334-337.
- Nielsen, S.** 1994. *The Bilingual LSP Dictionary. Principles and Practice for Legal Language*. Tübingen: Gunter Narr.
- Nielsen, S.** 1995. Dictionary Components. Bergenholtz, H. and S. Tarp (Eds.). 1995. *Manual of Specialised Lexicography. The Preparation of Specialised Dictionaries*: 167-187. Amsterdam/Philadelphia: John Benjamins.
- Nielsen, S.** 1999. [Review of] R.R.K. Hartmann and Gregory James. *Dictionary of Lexicography*. 1998. *Lexicographica* 15: 296-302.
- Nielsen, S.** 2003. Anmeldelse af ordbøger. *LexicoNordica* 10: 27-44.
- Nielsen, S.** 2005. User's Guides. Barz, I., H. Bergenholtz and J. Korhonen (Eds.). 2005. *Schreiben, Verstehen, Übersetzen, Lernen. Zu ein- und zweisprachigen Wörterbüchern mit Deutsch*: 137-146. Frankfurt am Main: Peter Lang.
- Nielsen, S.** 2006. A Functional Approach to User Guides. *Dictionaries* 27: 1-20.
- Nielsen, S.** 2008. The Effect of Lexicographical Information Costs on Dictionary Making and Use. *Lexikos* 18: 170-189.
- Nielsen, S.** 2009. Reviewing Printed and Electronic Dictionaries: A Theoretical and Practical Framework. Nielsen, S. and S. Tarp (Eds.). 2009. *Lexicography in the 21st Century. In Honour of Henning Bergenholtz*: 23-41. Amsterdam/Philadelphia: John Benjamins.
- Nkabinde, A.C.** 1993. [Review of] G.R. Dent (Compiler) and C.L.S. Nyembezi (Editor). *Compact Zulu Dictionary. English-Zulu; Zulu-English*. Fifth edition, 1990, Second impression, 1992. *Lexikos* 3: 298-280.
- Osselton, N.E.** 1989. The History of Academic Dictionary Criticism with Reference to Major Dictionaries. Hausmann, F.J., O. Reichmann, H.E. Wiegand and L. Zgusta (Eds.). 1989–1991: 225-230.

- Phillips, E.M. and D.S. Pugh.** 2005. *How to Get a PhD. A Handbook for Students and their Supervisors.* Maidenhead: Open University Press.
- Ridge, E.** 1993. [Review of] *Chambers Schoolbag Dictionary including Southern African Usage and Pronunciation.* First edition, 1986, Second impression, 1990. *Lexikos* 3: 295-298.
- Ripfel, M.** 1989: *Wörterbuchkritik. Ein empirische Analyse von Wörterbuchrezensionen.* Tübingen: Max Niemeyer.
- Schuring, G.** 1994. [Review of] A.D. de V. Cluver. *A Dictionary of Language Planning Terms.* First edition, 1993. *Lexikos* 4: 298-300.
- Sitaram, R.** 1996. [Review of] *A Lexicon of South African Indian English.* First edition, 1992. *Lexikos* 6: 307-310.
- Snyman, J.W.** 2000. [Review of] Wilfred H.G. Haacke and Eliphas Eiseb. *Khoekhoegowab-English/English-Khoekhoegowab Glossary/Midi Saogub.* 1999. *Lexikos* 10: 331-333.
- Steiner, R.J.** 1979. Guidelines for Reviewers of Bilingual Dictionaries. *Dictionaries* 1: 166-181.
- Swanepoel, P.** 2008. Towards a Framework for the Description and Evaluation of Dictionary Evaluation Criteria. *Lexikos* 18: 207-231.
- Tomaszczyk, J.** 1986. The Bilingual Dictionary under Review. Snell-Hornby, M. (Ed.). 1986. *Zürich-LEX '86 Proceedings. Papers Read at the EURALEX International Congress, University of Zürich, 9-14 September 1986:* 289-297. Tübingen: Francke.
- Wolf, B.** 1992. Wörterbuch und Benutzer — Versuch einer empirischen Untersuchung. Brauße, U. and D. Viehweger (Eds.). 1992. *Lexikontheorie und Wörterbuch. Wege der Verbindung von lexikologischer Forschung und lexikographischer Praxis:* 295-389. Tübingen: Max Niemeyer.
- Zawada, B.** 1997. [Review of] National Terminology Services of the Department of Arts, Culture, Science and Technology. *Dictionary of Geology/Geologiewoordeboek, English-Afrikaans/Afrikaans-Engels.* 1996. *Lexikos* 7: 315-319.

A Perspective on the Lexicographic Value of Mega Newspaper Corpora — The Case of Afrikaans in South Africa

D.J. Prinsloo, *Department of African Languages, University of Pretoria, Pretoria, Republic of South Africa (danie.prinsloo@up.ac.za)*

Abstract: The aim of this article is to assess the potential use of a mega newspaper corpus, the Media24 archive, in the absence of large balanced and representative corpora, for the compilation of major general dictionaries for Afrikaans. Firstly, an evaluation of Media24 against the lemmalists of both a major single-volume and a multi-volume monolingual dictionary for Afrikaans is undertaken to determine to what extent Media24 correlates with the lemmalists of major dictionaries. Secondly, the strength/suitability of Media24 for lemma selection in categories other than newspapers is evaluated. Finally, it is determined what the contribution could be of Media24 to lexical sense distinction, selection of examples of usage, and typical collocations.

Keywords: AFRIKAANS, MEDIA24 ARCHIVE, NEWSPAPER CORPORA, BALANCED CORPORA, REPRESENTATIVE CORPORA, WORD/LEXICOGRAPHIC/LEXICAL SENSE DISTINCTION, LEMMALIST, CORPUS DESIGN

Opsomming: 'n Perspektief op die leksikografiese waarde van megakoerantkorpuse — Die geval van Afrikaans in Suid-Afrika. Die doel van hierdie artikel is om die bruikbaarheid van 'n megakoerantkorpus, die Media24-argief, te bepaal in die afwesigheid van groot, gebalanseerde en verteenwoordigende korpuse vir die samestelling van omvattende algemene woordeboeke vir Afrikaans. Eerstens word 'n evaluering van Media24 gedoen deur dit met die lemmalyste van 'n groot omvattende enkelvolume-, en 'n multivolumewoordeboek van Afrikaans te vergelyk, ten einde te bepaal tot watter mate Media24 met die lemmalyste van groot woordeboeke korreleer. Tweedens word die gewig/toepaslikheid van Media24 vir lem-maseleksie uit kategorieë wat koerante uitsluit, geëvalueer. Ten slotte word bepaal wat die bydrae van Media24 kan wees tot leksikale betekenisonderskeiding, keuse van gebruiksvoorbeelde en tipiese kollokasies.

Sleutelwoorde: AFRIKAANS, MEDIA24-ARGIEF, KOERANTKORPUSSE, GEBALANSEERDE KORPUSSE, VERTEENWOORDIGENDE KORPUSSE, WOORD/LEKSIKOGRAFIESE/LEKSIKALE BETEKENISONDERSKEIDING, LEMMALYS, KORPUSONTWERP

Introduction

The aim of this article is to assess the contribution that a mega newspaper corpus, in the absence of large balanced and representative corpora, can make to dictionary compilation of major general dictionaries for Afrikaans.

Afrikaans lexicography finds itself in a situation where (a) a number of excellent major dictionaries are available (not traditionally based on corpus material), (b) no large balanced and representative corpora exist, but (c) a mega newspaper archive estimated at 1 000 000 000 (a thousand million tokens) can be consulted.

In this article, the Afrikaans Media24 archive is subjected to three tests in order to determine its effectiveness for the compilation or review of major Afrikaans general dictionaries.

The first test is an evaluation of the newspaper corpus on a macrostructural level against the lemmalists of a modern, major monolingual dictionary for Afrikaans, i.e. the 5th edition of the *Handwoordeboek van die Afrikaanse Taal* (HAT) and the 4th volume of the multi-volume *Woordeboek van die Afrikaanse Taal* (WAT). The intention is to establish to what extent this media archive can be used as a source for inclusion versus omission of lemmas in the revision (or compilation) of major Afrikaans dictionaries. Firstly, lemmas in HAT and WAT are compared to Media24 in order to determine to what extent a word list culled from Media24 matches the existing lemmalists of a single volume of a major dictionary such as HAT and a multi-volume comprehensive dictionary equal to the WAT. Secondly, an attempt is made to determine to what extent a word list culled from Media24 is suitable as an aid to inclusion or omission in future versions of these dictionaries. The question is, therefore, whether such a word list indicates what lemmas could be added to current lemmalists and whether non-occurrence could suggest the need for the omission of certain lemmas from existing dictionaries. Finally, it is suggested that frequency counts over three decades can assist the lexicographer to decide on inclusion or omission.

The second test evaluates the suitability of Media24 for lemmas most likely to be looked for by the target users of a general dictionary in categories not intensively covered by newspapers, for instance, religion, skills, hobbies, government, house organs and fiction which is covered in the BROWN/LOB corpora but collected as separate categories (cf. Table 1 below). Newspapers report on these fields, but the question is whether the coverage of such items is sufficient for lexicographic purposes. The purpose is therefore to determine to what extent terms from these fields, which can be presumed not to be generally associated with newspaper reporting, are covered by the Media24 newspaper archive. The randomly selected categories are gardening, quilting and embroidery. The last two contain precise subject specific terminology and therefore pose an implicit challenge in terms of coverage by a general newspaper corpus.

The third test, on the level of the microstructure of dictionaries, aims to

determine the value of Media24 as an aid in sense distinction, selection of examples of usage, and typical collocations. The question is whether a presumed bias towards typical 'newspaper senses' versus more 'general senses' impedes the value of Media24 in comparison to general corpora.

A brief description of WAT, HAT and Media24 will be given, followed by a calculation of the size of the Media24 archive.

Balance and representativeness as essential but problematic aspects in corpus creation

The debate as to what entails valid/ideal/balanced/representative corpora and whether it will ever be possible to compile such corpora is ongoing (cf. Biber 1993, Summers 1993, Kilgarriff 1997, Kennedy 1998, Kruyt and Dutilh, 1997, Otlogetswe 2007 and Atkins and Rundell 2008 for detailed discussions). A few excerpts serve to illustrate these lexicographic concerns.

Questions associated with 'representativeness' and 'balance' are complex and often intractable. (Kennedy 1998: 62.)

A general corpus is typically designed to be **balanced**, by containing texts from different genres and domains of use including spoken and written, private and public [...] For a corpus to be 'representative' there must be a clearly analysed and defined population to take the sample from. (Kennedy 1998: 20, 52.)

What we mean by *representative* is covering what we judge to be the typical and central aspects of the language, and providing enough occurrences of words and phrases for the lexicographers [...] to believe that they have sufficient evidence from the corpus to make accurate statements about lexical behaviour. (Summers 1993: 186, 190.)

[...] to be representative of general language. This is a bold ambition — some say one that is impossible to fulfil. (Summers *s.d.* [1996–1998]: 6.)

COBUILD have always insisted that it is impossible to create a corpus that is truly representative of the language, and have focused on size of corpus rather than balance. (Kilgarriff 1997: 150.)

Lexicographers traditionally aim at a 'representative' or 'balanced' corpus, that is, the corpus should be appropriate as the basis for generalizations concerning the language as a whole. (Kruyt and Dutilh 1997: 230.)

Scholars even differ in their interpretation of the terms. This debate, however, is beyond the scope of this article — the issue at stake here is simply whether a 1 000 million-word newspaper archive can be regarded as a suitable, main source for the compilation of major Afrikaans dictionaries.

The design of a pioneering corpus, such as the *Brown Corpus of Standard American English* and *Lancaster-Oslo/Bergen Corpus (LOB)*, was a carefully compiled selection of American English, totalling approximately a million words drawn from a wide variety of sources for which each contained 2 000 words. The corpus was sampled from 15 text categories given in Table 1.

PRESS: REPORTAGE (44 texts)	LEARNED (80 texts)
PRESS: EDITORIAL (27 texts)	FICTION: GENERAL (29 texts)
PRESS: REVIEWS (17 texts)	FICTION: MYSTERY (24 texts)
RELIGION (17 texts)	FICTION: SCIENCE (6 texts)
SKILLS AND HOBBIES (36 texts)	FICTION: ADVENTURE (29 texts)
POPULAR LORE (48 texts)	FICTION: ROMANCE (29 texts)
BELLES-LETTRES (75 texts)	HUMOR (9 texts)
MISCELLANEOUS: GOVERNMENT & HOUSE ORGANS (30 texts)	

Table 1: Design of the Brown and LOB corpora

It could be assumed that newspaper texts represent a specific, almost homogeneous subtype that can easily skew a balanced corpus if newspaper texts are added in large quantities (cf. MacLeod and Grisham (2000) for the case of adding a vast amount of newspaper data to the Brown Corpus). They indicate how an increase in the Brown Corpus of 1 329% (thus more than thirteen times) resulted in a skewed or inadequate corpus e.g. in the representation of business-related words, such as *sell*, *rise*, *buy*, *pay*, and *increase*. Newspaper texts also contain words belonging to a slightly higher register; cf. *arts* (instead of *dokter/geneesheer*) 'doctor', *baar* (*kraam/geboorte gee*) 'give birth'. On the other hand, these texts also contain words belonging to an informal register; cf. for example *herrie* (*oproer/rusie/ontevredenheid*) 'uproar/quarrel/dissatisfaction', *grondgryp* (*grondonteiening*) 'land seizure', and *blaser* (*skeidsregter*) 'referee'. Both these types are uncommon to everyday written and oral communication. The use of such words on newspaper banners or in headlines contribute to attracting attention and are usually shorter than their equivalents, fitting into limited space. They are apparently much less frequently used in non-newspaper corpora. Preliminary tests indicate that *herrie* is used ten times more in Media24 than in a 4 million-token test corpus consisting of Afrikaans literary works. Likewise no occurrence of *blaser* referring to a referee could be found in the test corpus. The real potential corpus-skewing factor of such words should however be determined by more detailed studies.

It could also be argued that a growing newspaper corpus, such as Media24, partially qualifies for what Atkins calls an *organic corpus*, at least as far as the 'growing part' is concerned.

A corpus builder should first attempt to create a representative corpus. [...] the corpus is enhanced by the addition or deletion of material [...] This is the way to approach a balanced corpus. One should not try to make a comprehensive and watertight listing [...] rather, a corpus may be thought of as organic, and must be

allowed to grow and live if it is to reflect a growing living language. (Atkins 1997, personal communication at Salex'97 (Atkins et al. 1997.))

Building neatly designed corpora, such as the Brown corpus, was also envisaged for African languages and Afrikaans when corpus creation for these languages commenced in 1990. For the African languages, it was not possible, because many of the categories, such as the three press sections, simply do not exist as most of the languages do not even have a single newspaper and some in fact have very limited printed matter. For these languages, a more organic approach (cf. Atkins et al. 1997) was followed. For Afrikaans, the situation was more conducive, but no attempt was ever made to build a large corpus, for example, along the lines of the Brown/LOB design. An organic corpus of 10 000 000 tokens was compiled at the University of Pretoria but this corpus is dwarfed by the Media24 newspaper archive estimated at more than 1 000 million tokens.

The question remains, however, as to what extent growing in size also means growing in representativeness or, in what Leech terms its *diversity*.

The value of a corpus as a research tool cannot be measured in terms of brute size. The **diversity** of the corpus, in terms of the variety of registers or text types it represents, can be an equally important (or even more important) criterion. (Emphasis in the original.) (Garside et al. 1997: 2.)

Regardless of the corpus size, a corpus that is systematically selected from a single register cannot be taken to represent the patterns of variation in an entire language; [...] corpora representing the full range of registers are required. [...] it is important to design corpora that are representative with respect to both size and diversity. However, given limited resources for a project, representation of diversity is more important for these purposes than representation of size. (Biber 1995: 131.)

What is important, therefore, is to estimate the value of the Media24 archive for Afrikaans lexicography. Is its 'brute size' also representative of the varieties of registers or non-newspaper categories in, for example, the design of the Brown Corpus?

The Media24 archive

The Media24 archive is a searchable database of Afrikaans media reports available at <http://152.111.1.251/cgi-bin/s.cgi>. Media24 contains among others the newspapers *Rapport*, *Beeld*, *Volksblad* and *Die Burger*, available in electronic format for the past two to three decades. A range of search functions such as basic words, fixed and semi-fixed phrases as well as the use of certain Boolean operators are allowed. Hits are presented as full reports as they were published in the newspapers, up to a maximum of 50 at a time. It also means that a report

can contain more than one occurrence of a word. As for the size of Media24, no authoritative figure is available. Evaluation of Media24 frequencies is therefore difficult if the size of the corpus is unknown. An attempt was made to calculate its approximate size in a simplistic way before comparisons with HAT and WAT were made.

A random selection of 18 words was chosen for the calculation of the size of the Media24 archive given in Table 2. Statistics used for this calculation were

- counts in a 750 million-word subsection of the corpus (exact size: 749 553 152 tokens); Column 2,
- number of newspaper reports containing each of these 18 words in the 750m subcorpus; Column 3, and
- number of newspaper reports in the entire archive containing each of these 18 words; Column 5.

First, the relation between the number of media reports in which a specific word occurs and the total number of occurrences of the word in all reports in the 750m subcorpus was calculated; Column 4. In the case of *die*, for example, the number of reports is only 6% of the total counts for *die*, i.e. *die* occurs very frequently in each report (more than 50 million times in less than 3 million reports).

This relation was then used to calculate the total count of each word in the entire Media24 archive based on the number of reports in the entire archive; Column 6.

A basic correlation value between counts in the 750m subcorpus and the total size 750m was then calculated for each word in the 750m subcorpus by dividing 750m with the total counts for each word; Column 7.

This correlation value was finally used to calculate the size of the Media24 archive by multiplying it with the calculated total counts in Column 6.

Thus for all of the 18 keywords, a corpus size slightly exceeding 1 000 000 000 (one thousand million tokens) was independently postulated; Column 8.

1	2	3	4	5	6	7	8
Word	750m Subcorpus counts	750m Subcorpus reports	750m Reports as % of counts	Media24 Reports	Media24 Calculated total counts	750m Correlation value	Media24 Calculated size
boek 'book'	100 787	55 096	55	85 881	157 102	7 437.00	1 168 367 472
die 'the'	51 184 148	2 889 785	6	4 079 349	72 253 819	14.64	1 058 102 558
drink	25 450	24 524	96	44 692	46 380	29 451.99	1 365 969 233
eet 'eat'	41 741	35 329	85	58 660	69 306	17 957.24	1 244 552 291
en 'and'	16 741 151	2 606 098	16	3 718 038	23 884 073	44.77	1 069 363 893
het 'has'	16 870 704	2 443 463	14	3 449 872	23 819 378	44.43	1 058 277 711
hond 'dog'	30 571	19 020	62	30 500	49 023	24 518.44	1 201 964 834
huis 'house'	360 993	226 869	63	345 317	549 467	2 076.36	1 140 893 845
kat 'cat'	17 614	13 384	76	20 797	27 370	42 554.40	1 164 708 376
loop 'walk'	132 839	132 328	100	200 486	201 260	5 642.57	1 135 624 458
mens 'human'	413 683	280 379	68	432 056	637 474	1 811.90	1 155 039 916

Table 2: Calculation of the size of the Media24 archive

HAT and WAT

HAT is the 5th edition of the *Verklarende Handwoordeboek van die Afrikaanse Taal* containing more than 50 000 lemmas. WAT, *Woordeboek van die Afrikaanse Taal*, is a multi-volume explanatory dictionary currently published up to the letter R (13 volumes).

Comparison of types in the Media24 archive to the lemmalists of HAT and WAT

For the first test, a random sub-stretch of the arbitrarily selected alphabetical stretch 'I' was selected i.e. *ideaal* to *idiot*. There are 153 lemmas strictly alphabetical¹ in this stretch in WAT and HAT taken together. WAT has 147, HAT 48 and they have 42 in common.

Lemmas given in both WAT and HAT with overall counts in Media24:

ideaal (31 557), idealis (298), idealiseer (163), idealisme (1 390), idealisties (673), idealiteit (2), idee (58 214), ideëassosiasie (0), ideëel (8), idée-fixe (2), ideëleer (0), ideëryk (7), ideëverering (0), ideëwêreld (13), idem (167), identiek (7), identies (2 345), identifikasie (3 185), identifiseer (25 680), identiteit (25 053), identiteitsbedrog (54), identiteitsbewys (45), identiteitskaart (420), identiteitsplaat(jie) (13), ideo- (142), ideofoon (0), ideografie (2), ideogram (7), ideolatrie (2), ideologie (5 898), ideoloog (231), ideomotories (0), idille (203), idillies (312), idio- (13), idiolatrie (0), idiomatiek (37), idiomaties (302), idiomorf (0), idiomorfie (0), idioom (2 838), idioot (923)

Lemmas given in WAT but not in HAT, with overall counts in Media24:

ideageen (0), ideasie (0), ideatief (0), **ideëdrama (7)**, ideëfonds (0), **idee-force (128)**, ideëgeskiedenis (0), ideëkuns (0), ideëliriek (0), ideëpoësie (0), ideerigting (0), ideëskat (0), ideëskrif (0), **ideëspel (1)**, ideëtragiek (0), ideëvlug (0), ideëwaarde (0), idemfaktor (0), idempotent (0), **identifieer (2)**, identifisering (0), identifiseringsparade (0), identifikasiebaken (0), **identifikasiebewys (4)**, identifikasiekaart(jie) (0), identifikasieletter (0), identifikasielig (0), **identifikasieparade (2)**, identifikasieplaat(jie) (0), **identifikasiesein (2)**, **identifikasieteken (1)**, **identifisering (3 534)**, identifiseringsparade (0), identiteitsafstand (0), **identiteitsbeginsel (2)**, identiteitsbrief (0), identiteitsselement (0), identiteitsfilosofie (0), identiteitshipotese (0), identiteitsstelsel (0), identiteitskenmerk (0), identiteitskyf (0), identiteitsmatriks (0), **identiteitsmerk (2)**, identiteitsoordeel (0), **identiteitsparade (16)**, identiteitsprinsipe (0), identiteitsreaksie (0), identiteitsteken (0), **identiteitstelsel (18)**, **identiteitsteorie (1)**, identiteitswet (0), ideofreen (0), ideofrenie (0), ideogeen (0), ideogenese (0), **ideograaf (2)**, ideografies (0), ideokineties (0), ideologieëleer (0), **ideologies (2 065)**, ideometabolies (0), ideometabolisme (0), ideomosie (0), ideomotoriek (0), ideomuskulêr (0), ideoplasie (0), ideoplastie (0), ideoplastiek (0), ideoplasties (0), ideoreleksie (0), ideosekretories (0), ideosensories (0), ideovaskulêr (0), ideovisueel (0), Idiacanthidae (0), Idiacanthus (0), idioadaptasie (0), idiobiologie (0), idioblas(t) (0),

idioblasties (0), idiochromaties (0), idiochromatine (0), idiochromidie (0), idiochromosoom (0), idiofonie (0), idiofreen (0), idiogaam (0), idiogamie (0), idiogeen (0), idiogenese (0), idioglossie (0), idioglotties (0), idiograaf (0), idiografies (0), idiogram (0), idiohipnose (0), idioïmbesiel (0), idiokinese (0), idiokineties (0), idiokrasie (0), idiolalie (0), idioma-tologie (0), idiomorfies (0), idiomuskulêr (0)

Lemmas given in HAT but not in WAT with overall counts in Media24:

ideëberaad (3), identikit (1 391), identiteitsdokument (3 124), identiteitskrisis (831),
idioëlektries (0), **idiolek (56)**

Media24 reflects counts for 55 of these 153 lemmas. In comparison to WAT, Media24 shows counts for 50 lemmas, i.e. roughly 30% and in comparison to HAT Media24 reflects 39 lemmas, i.e. 80%. Thus the value of Media24 for the compilation of a lemmalist for a multi-volume dictionary of the magnitude of WAT is substantially lower than for a single-volume major dictionary. The lemmas presented in WAT cover 165 993 tokens in Media24 and HAT covers 165 611. This is quite significant, i.e. that HAT, although having only one third of the lemmas compared to WAT, covers the same number of tokens in Media24. HAT fared well in comparison to WAT for lemmatising *identiteitsdokument* (3 124) 'identity document', *identikit* (1 391), *identiteitskrisis* (831) 'identity crisis' and *idiolek* (56) 'idiolect', which reflect high counts in the corpus. WAT on the other hand did well in comparison to HAT for lemmatising *identifisering* (3 534) 'identification', *ideologies* (2 065) 'ideological', *idee-force* (128) 'active idea', *identiteitstelsel* (18) 'identity system' and *identiteitsparade* (16) 'identity parade' which show high counts in the corpus.

From these comparisons, it is clear that the Media24 archive not only covers all frequently used lemmas in the dictionaries but also a significant number of low frequency lemmas. Some lemmas in the dictionary with zero occurrences in the archive could therefore be considered for omission in a forthcoming revision of the dictionary. Likewise, certain words in the archive could be considered for inclusion in the dictionary (given general considerations for lemma inclusion such as the self-explanatory nature of some morphologically complex words), such as *identiteitloos* (67) 'without identity', *identiteitloosheid* (21) 'state of being without identity', *identiteitlose* (62) 'being without identity', *identiteitsboek* (164) 'identity book', *identiteitsboeke* (125) 'identity books', *identiteitsboekie* (409) 'small identity book', *identiteitsboekies* (323) 'small identity books', *identiteitsfoto* (20) 'identity photo', *identiteitsfoto's* (48) 'identity photos', *identiteitsnommer* (690) 'identity number', *identiteitsnommers* (255) 'identity numbers'. These words could be lemmatised as *identiteitloos*, *identiteitsboek*, *identiteitsfoto* and *identiteitsnommer*. In order to gain an impression of their frequency trajectories over two decades, the total counts of these words are expressed per 50 million tokens for five-year periods ending in 1989, 1994, 1999 and 2003² respectively in Table 3 and graphically illustrated in Figure 1.

	89/50M	94/50M	99/50M	2003/50M
identiteitloos (lemma)	3	8	10	13
identiteitsboek (lemma)	29	85	84	51
identiteitsfoto (lemma)	3	4	5	4
identiteitsnommer (lemma)	52	47	73	67

Table 3: Total counts (lemmas and derivations) expressed per 50 million tokens in Media24

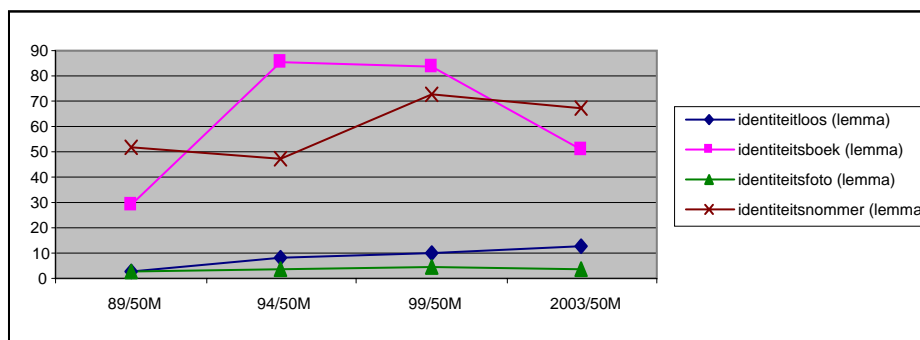


Figure 1: Trajectories of the total counts (lemmas and derivations) expressed per 50 million tokens in Media24

The frequency counts in Table 3 and the trajectories in Figure 1 suggest inclusion of these lemmas in major Afrikaans dictionaries.

The repetition factor in Media24

Frequency counts of words in a media corpus can be questioned on the basis of potential repetition of the same phrases in, for example, regional releases of the same reports or stereotypical repetitions of a word/phrase. In order to determine the extent and nature of repetition in Media24, concordance lines were generated for randomly selected words given in Table 5. For example, up to 53 repetitions of the line "... Die woorde hieronder kom voor in die **blok** met letters ..." '... The words below occur in the block with letters ...' occur as indicated in Table 4. The total number of concordance lines generated for each word from reports in Media24 were grouped and summed to determine the number of duplications. Consider the following instances of repetition of fixed phrases containing the word *blok* 'block' in Table 4.

Concordance lines for <i>blok</i>	Number of repetitions
Die woorde hieronder kom voor in die blok met letters. Hulle verskyn horisont	53
Nasionale Pers. Tenderaars moet die blok , die getal en die prys verstrek van	32

oop, of aandui dat die tender uit enige blok toegeken kan word. 'n Minimum va	23
estasies buite skoolgebied Verwys na Blok A: Teen die einde van die jaar nee	20
p die foto begrawe lê. Dui u die regte blok aan en u inskrywingsvorm is een v	12
inasies het om te wen. Indien jy 'n vol blok het (met ander woorde al die blokk	11
word. Sodra u al die nommers in die blok omkring het, kan u een van die pry	9

Table 4: Most frequent repeated concordance lines for *blok* in Media24

Concordance lines for six more randomly selected words were generated, grouped, summed and studied (cf. Table 5). The final column of Table 5 indicates that the percentage of duplication for these words range from 5%–14%.

Word	Total lines M24	Different lines	%	% duplication
identifisering 'identification'	2 333	1 997	86	14
ideologies 'ideological'	1 381	1 155	84	16
identiteitsnommer 'identity number'	691	620	90	10
identiteitsboekie 'identity book'	410	365	89	11
blok 'block'	8 905	7 427	83	17
borduur 'embroider'	562	521	93	7
steek 'stitch'	16 001	15 200	95	5
Average				11

Table 5: Duplication factor of words in Media24

Reports are repeated in sister newspapers, regional issues of the same newspaper or sequentially over a period of time. So, for example, concordance lines generated for *ideologies* 'ideological' rendered a number of identical lines for *ideologies* in the context *Die "ou manne" wat op alle samelewingsvlakke fanaties, krampagtig en ideologies apartheid bedink en bevorder het, moet gekonfronteer word* (The 'old men' who on all levels of coexistence, fanatically, desperately and ideologically conceptualised and promoted apartheid should be confronted) (*Beeld* 29 December 2000, *Die Burger* 30 December 2000, 2 October 2000 and *Volksblad* 29 September 2000). From a strict statistical point of view, it could be argued that these are repetitions skewing frequency counts. However, from a lexicographic perspective, a case could be made for *bona fide* use in multiple sources over large and different geographic areas, e.g. *Die Burger*, mostly southern regions of South Africa, *Beeld* mostly northern regions, etc., i.e. not true/basic repetition.

Evaluation of Media24 for categories *gardening*, *quilting* and *embroidery*

For the second test, the categories *tuinmaak* 'gardening', *laslappie* 'quilting' and *borduur* 'embroidery' were randomly selected as representatives of the catego-

ries Skills and Hobbies (cf. Table 1 above). These categories contain precise subject-specific terms and therefore pose a challenge in terms of coverage by a general newspaper corpus. Firstly, dedicated corpora were compiled for each of these categories from randomly selected sections of gardening (49 146 tokens); *Die Suid-Afrikaanse tuin* (Gilbert 1985); embroidery (64 968 tokens); *Borduursteke vir Suid-Afrika* (Eaton 1989) and quilting (28 131 tokens); *Die Suid-Afrikaanse boek van laslappie en appliekwerk* (Turpin-Delpont 1988). The Gardening Corpus and a combination of the Quilting and Embroidery Corpus were then compared to a 5.8-million token general Afrikaans corpus using the Keyness function of WordSmith Tools. The aim was to detect the so-called *positive keys*, i.e. words used in the Gardening, Quilting and Embroidery corpora that occur more frequently than expected in comparison to a general corpus.

N	WORD	KEYNESS	N	WORD	KEYNESS
1	NAT ('wet')	1 947.5	11	DEKLAAG ('upper layer')	727.8
2	PLANT	1 464.6	12	SAAI ('sow')	694.2
3	SNOEI ('prune')	1 367.9	13	EENJARIGES ('annuals')	681.0
4	PLANTE ('plants')	1 327.5	14	DAE ('days')	676.5
5	MAAL ('grind')	1 138.4	15	MAAND ('month')	657.9
6	KOMPOS ('compost')	1 085.9	16	BLARE ('leaves')	594.3
7	BESPUIT ('spray')	940.0	17	RADYSE ('radishes')	551.4
8	GROND ('soil')	831.1	18	MAAK ('make, prepare')	530.4
9	KYK ('see') ³	767.6	19	KAN ('can, may')	527.2
10	BEMES ('fertilise')	737.3	20	SAAD ('seed')	510.2

Table 6: The top 20 keys for gardening

All words in Table 6 typically used in the subject field of gardening occur frequently in Media24, i.e. *kompos* (2 601), *bemes* (612), *plant* (30 777), *snoei* (1987), *deklaag* (602), *saai* (13 762), *eenjariges* (155), *blare* (10 918), *saad* (8 708) and *bespuit* (1 845). Gardening Keys occurring more than once in the Gardening corpus but not occurring in Media24 were studied in more detail. The 225 words were lemmatised rendering 216 lemmas. Of these, 23 lemmas were found in WAT. For words occurring three times or more in the Gardening Corpus, WAT has the lemmas *poeiermeeldou*, *Heuchera*, *druifhiasint*, *Helianthemum*, *naelkruid*, *pondolandklimop*, *Cryptostegia*, *Holmskioldia*, *Habranthus*, *drakebloedboom*, *bloulangkop* and *klokblom*. For words occurring two or more times, HAT has the lemmas *wolfsboontjie*, *suurdoring*, *sonrosie*, *skeefblom*, *reseda*, *klokblom*, *kardinaalsmus*, *eschscholtzia* and *bordeauxmengsel*. Thus, for the treatment of these lower ranking Gardening Keys lemmatised in WAT and HAT, the lexicographer cannot count on Media24 for support.

N	WORD	KEYNESS	N	WORD	KEYNESS
1	GEWERK ('worked')	10 769.2	11	VERTIKALE ('vertical')	2 463.9
2	STEEK ('stitch')	10 543.0	12	HORISONTALE ('horizontal')	2 405.7

3	STEKE ('stitches')	7 140.9	13	VORM ('form')	2 292.4
4	WORD ('be (done)') ⁴	4 934.5	14	KRUISSTEEK ('cross stitch')	2 223.5
5	RYE ('rows')	3 464.5	15	GAAS ('canvas')	2 142.9
6	GARING ('cotton')	3 299.4	16	KETTINGSTEEK ('chain stitch')	2 125.8
7	EWEDRAADSTOF ('even weave')	3 030.0	17	P ('p.') ⁵	2 075.1
8	CM	2 805.5	18	VULLING ('filling')	1 866.2
9	EFFEBINDING ('simple (plain) weave')	2 761.2	19	GEBRUIK ('use')	1 856.7
10	STOF ('material')	2 647.7	20	BL. ('p.') ⁶	1 781.0

Table 7: The top 20 keys for *quilting* and *embroidery*

With the exception of *effebinding* (0), the other Quilting and Embroidery Keys were found in Media24, i.e. *steke* (3 459), *ewedraadstof* (2), *garing* (319), *kruissteek* (51), *kettingsteek* (11), *gaas* (191) and *vulling* (170). Quilting and Embroidery Keys occurring more than once in the Quilting and Embroidery Corpus but not occurring in Media24 were studied in more detail. The 473 words were lemmatised rendering 397 lemmas. Of these, 66 lemmas were found in WAT. For words occurring 10 times or more in the Quilting and Embroidery Corpus, WAT has the lemmas *effebinding*, *lynsteek*, *knoopsteek*, *komerssteke*, *graatsteek*, *glansgaring*, *legdrade*, *lussteek*, *applikee*, *kabelsteek*, *koraalsteek*, *buitelyinsteek*, *legwerk*, *krielseek*, *gekriel*, *kussingsteek*, *komerssteekvulling* and *diamantsteek*. For words occurring three times or more, HAT has the lemmas *visgraatsteek*, *veersteek*, *sluitsteek*, *siersteek*, *randsteek*, *oormekaarslaan*, *lussteek*, *knoopsgatsteek*, *kabelsteek* and *hegsteek*. As in the case of lower ranking Gardening Keys, these Quilting and Embroidery Keys lemmatised in WAT and HAT are not supported by Media24 texts.

Evaluation of the media corpus on microstructural level

The third test aims to determine whether concordance lines culled from a major media corpus could provide sufficient aid to the compiler of a major dictionary on a microstructural level. The focus is on the contribution towards sense distinction, authentic examples and collocations — all typically regarded as areas where the corpus gives valuable support to the lexicographer in compilation of the article (cf. De Schryver and Prinsloo 2000). The polysemous words *borduur*, *steek*, *patroon*, *knop* and *blok* were selected from the Quilting and Embroidery list and their treatment in HAT and WAT as well as their use in context in Media24 was studied.

HAT

borduur ww. (geborduur) [...] 1 Met naaldwerk versier: *Blomme op 'n kussing borduur.*
2 (fig.) Op oordrewe wyse opsier: *'n Verhaal borduur met romantiese verdigsels.*

borduur: ~draad, ~gaas, ~garing, ~kant, ~naald, ~patroon, ~raam, ~ster, ~werk, ~wol.

WAT

borduur

I s.nw. Borduurkatoen of -wol.

II ww. **1.** Bestaande stowwe met naaldwerk versier i/d vorm van rande, figure, festoene, ens. — in teenst. met weef, tapytwerk, ens., waarby dieselfde stof se drade gebruik word: 'n *Geborduurde kleedjie*. **2.** (fig.) Opsmuk op 'n oordrewe, verdigtende manier; uitbrei: *Die skrywer het verder daarop voortgeborduur*.

In the case of *borduur*, both HAT and WAT distinguish two basic senses, i.e. *met naaldwerk versier* 'decorate with needlework' and the figurative meaning *op oordrewe wyse opsier* 'elaborate in an exaggerated way'. The word *borduur* occurs 562 times in the Media24 archive. A random pick of 10 occurrences, i.e. every tenth line (keyword-in-context (KWIC) lines 10, 20, 30, 40, ... 100) are given in Table 8.

10	En op hierdie fototoonstelling borduur sy voort op dié onderwerp, met 'n
20	57) 352-9211 gerig word. Wedstryde borduur dwelms ...
30	051) 522-2130. Foto: Michelle Cahill borduur handwerk ...
40	het aan BBC Sport gesê: "Robinson borduur voort op Engeland se reputasie
50	am net in Engels agter op die trui te borduur . Die sterk Afrikaanse ondersteun
60	Los Angeles. Na die Grammys toe, borduur ek. "What are you going to do ther
70	borduur vir baba. Gesinsafdeling: borduur vir baba Arina du Plessis Maak ?
80	rksessies in Australiese kruissteek, borduur op wol, kralewerk op klere en eksp
90	wat jou hart begeer," sê hy. En hy borduur voort op wat hy haar sal gee en
100	gesorg, maar DEON GELDENHUYS borduur by. WAT het die gepeupel, geeste

Table 8: Concordance lines for *borduur*

The literal sense of *embroidering* is depicted by lines 30, 50, 70 and 80 and the figurative sense of *elaboration* in lines 40, 60, 90 and 100. Clear examples of usage for possible inclusion into the article of *borduur* in a new dictionary or a revised edition are available in abundance in the concordance lines, for example *borduur 'n naam/prentjie op die trui* 'embroider a name/picture on the jersey', *borduur voort op Engeland se reputasie* 'elaborate further on England's reputation'. Typical collocations such as *borduur voort/verder* 'elaborate further', *borduur en stik/brei* 'embroider and stitch/knit' can also easily be detected.

In the case of the lemma *blok* (noun), HAT distinguishes 9, and WAT 26 senses.

HAT

blok [...] s.nw. (-ke) **1** (Groot) stuk hout, metaal, ens.: 'n *Blok beton, marmer. Die seuntjie sit heerlik met sy blokkies en speel, speelgoed wat 'n reëlmatige vorm het.* **2** (fig.) Swaargeboude persoon: 'n *Blok van 'n vent.* **3** (hist.) Strafwerktuig om die bene of nek van 'n gevangene vas te sluit: *Iemand in die blok sit.* **4** Groep, afdeling: 'n *Hele blok huise. Blokke en blokke woonstelle langs mekaar in Hillbrow. 'n Politieke blok vorm.*

5 Stuk hout aan die poot van diere; ook fig., bv. *dis 'n blok aan my been*, 'n belemmering. **6** Vierkantige of reghoekige figuur of voorwerp: *Die blokke van 'n skaakbord, van rokmateriaal. 'n Skryfblok.* **7** Ruimte in 'n stad, dorp tussen strate: *Hy woon twee blokke verder.* **8** (*filat.*) Groep (van vier) seëls, gewoonlik in die boonste regterkant van 'n vel; kontroleblok. **9** (Son)blokker. UITDR.: *'n Blok aan die been hê*, iets wat jou belemmer, in jou vooruitgang strem. *Jou kop op 'n blok sit*, iets met klem bevestig.

WAT

blok. I s. 1. Swaar stuk hout, steen of metaal, min of meer reëlmatig van vorm, gew. met een of meer vlak sye: *'n Blok om vleis op te kap. 'n Blok marmar.* **2.** (*fig.*) Lywige persoon: *'n Blok van 'n kêrel. 'n Blok van 'n kind.* **3.** Swaar houtstrafwerktuig waarin eertyds die bene en soms ook die arms en nek van 'n gevangene gesluit is: *Iem. i/d blok sit (sluit).* **4.** 'n Stuk hout wat aan 'n perd of bees se voorpote gebind word om te belet dat die dier oor slote spring, ens. **5.** 'n Gerwehoop, netjies op die oesland gepak sodat dit bv. teen reën beveilig is. **6.** (*schoenm.*) Houtlees om aan 'n skoen sy vorm te gee, soms gebruik i.p.v. die ysterlees; ook genoem *leesblok.* **7.** Klein, kubusvormige stukkie hout, ens.: *Blokke (blokkies) uit 'n kind se speelgoeddoos. 'n Blok(kie) sjokolade.* **8.** Kompleks geboue; ook, groot, samegestelde gebou: *'n Blok woonstelsel, huise.* **9.** Ruimte, gew. min of meer reghoekig, in 'n stad of dorp, deur strate omgrens: *Ons woon drie blokke van die markplein af.* **10.** Vierkantige of langwerpige blaadjies papier wat a/d een kant op mekaar vasgeheg is vir skryf- of ander doeleindes: *'n Skryfblok.* **11.** Skrop: *Wiel- en skepblokke by dammakery.* **12.** Hout- of metaalliggaam waarin een of meer skywe om een as draai soos by 'n katrol: *Blok en loper vorm saam 'n takel.* **13.** (*spoorw.*) Een v/d aansluitende baangedeeltes waarin die baanvakke verdeel word i/d sisteem van beveiliging v/d treinloop, en waarin daar op enige tydstip nie meer as een trein hom op dieselfde spoor mag bevind nie. **14.** (*wam.*) **a.** 'n Stukkie yster met twee gate daardeur waar die twee punte van 'n klou deurgaan en a/d ander kant moere kry. **b.** Sien BRIEKBLOK. **15.** (*delw.*) Een v/d klein, min of meer kubusvormige, swart klippies in diamanthoudende gruis — gew. as vkw.: *Blokkies is 'n goeie teken van diamante.* Ook genoem *blokkiesbantom.* **16.** (*filat.*) 'n Aantal posseëls wat aan mekaar vas is i/d vorm van 'n reghoek en wat uit meer as een ry bestaan: *'n Blok van vier. 'n Blok van twaalf, in drie rye van vier elk. Twee posseëls kan nie 'n blok vorm nie.* **17.** Min of meer reghoekige stuk grond: *Ons het twee blokke geploeg met 'n wenakker aan elke ent.* Vgl. GEWEN. **18.** Plek waar die persoon wat "aan" is, sy oë toehou by aspaai en wegkruipertjie; ook genoem *bof.* **19.** (*tolspel*) Mislukte gooihou, d.w.s. sonder dat die tol draai: *'n Blok gooi.* **20.** Gevoellose persoon: *Die vrou is 'n koue blok.* **21.** Aaneensluiting van partye of groepe: *'n Politiek-ekonomiese blok. 'n Militêre blok.* **22.** Onderdeel van 'n skaaf waarin die beitel bevestig is. **23.** Vierkantige of reghoekige figuur of vak: *Die blokke in 'n blokkiesraaisel. Die blokke van 'n dambord. Blokke op 'n stuk geweefde goed, op plakpapier, ens.* **24.** (*landbou*) Werktuig bestaande uit drie houtballe i/d vorm van 'n gelykbenige driehoek wat gebruik word om omgeploegde grond gelyk te sleep. **25.** Sien DRUKBLOK. **26.** Groep plase wat gelyktydig uitgemeet en uitgegee is: *Die Hertzoglob in S.W.A. is met behulp van fondse van regeringsweë tot beskikking v/d Angola-Boere gestel.*

Sense	HAT	WAT	Media24 KWIC	HAT Sense no.	WAT Sense no.
Heavy object of wood/metal/brick	✓	✓	✓	1	1
Heavy person	✓	✓	✓	2	2
Heavy object of punishment (stocks)	✓	✓		3	3
Restraint for horses	✓	✓		5	4
Bulk of grain		✓			5

Last (shoemaking)		✓			6
Toy	✓	✓	✓	1	7
Group of apartments	✓	✓	✓	4	8
Street sections	✓	✓	✓	7	9
Stack of writing paper		✓			10
Dam scraper		✓			11
Block (and tackle)		✓			12
Railway section		✓			13
Object used in wagon building		✓			14
Black cubical stones (mining/digging)		✓			15
Group of stamps	✓	✓	✓	8	16
Square sections of ground	✓	✓		7	17
In hide-and-see game		✓			18
In top (toy) game		✓			19
Apathetic person		✓			20
Parties/groups	✓	✓	✓	4	21
Stock of a plane (tool)		✓			22
Square/rectangular figure	✓	✓	✓	6	23
Object used to flatten soil (agriculture)		✓			24
Pressure block		✓			25
Group of farms		✓			26
Prevention object	✓			9	
Rubik's cube			✓		

Table 9: Senses of *blok* in WAT and HAT and occurrences in Media24

Media24 occurrences were found in support of six of the nine senses given in HAT and eight of the 26 senses given in WAT. Once again the evidence suggests that the Media24 archive could be a sufficient tool for the compilation of a major dictionary but insufficient as sole corpus for the compilation of a dictionary of the magnitude of WAT.

Conclusion

It can be concluded that in the current situation where no large designed corpus for Afrikaans exists, the Media24 archive is an excellent substitute. In fact, its value goes far beyond a limited component of a corpus design pattern, i.e. 'press'. The Media24 archive is so vast and versatile that it can be regarded as a world of information in its own right and its success in terms of broad coverage can probably be attributed to the fact that virtually all aspects of modern life in South Africa are covered in the daily reporting of these newspapers.

The question could even be asked whether the stage has not been reached in corpus-based lexicography where media coverage is so comprehensive in reporting on all spheres of everyday life that mega newspaper corpora have indeed become a world in one medium, i.e. a corpus sufficient, or at least going a long way as a basis for the compilation of general dictionaries.

Endnotes

1. Related lemmas such as *idioties* ('idiotic'), *idiotisme* ('idiotism') falling alphabetically outside the stretch *ideaal–idiot* were not considered.
2. Strategy utilised by Prinsloo and Gouws (2006) to express the increasing number of tokens per 5-year period in the Media24 archive as equal comparable sections, i.e. frequency counts of words per 50 million tokens.
3. Key status as a result of its frequent use as a reference marker in the source text.
4. Key status as a result of (over)use of the passive in describing each activity.
5. Key status as a result of its frequent use as a reference marker in the source text.
6. Key status as a result of its frequent use as a reference marker in the source text.

References

Dictionaries and corpora

- BROWN = *Brown Corpus of Standard American English* http://www.essex.ac.uk/linguistics/clmt/w3c/corpus_ling/content/corpora/list/private/brown/brown.html.
- HAT = Odendal, F.F. and R.H. Gouws. 2005. *HAT. Verklarende Handwoordeboek van die Afrikaanse Taal*. Fifth Edition. Cape Town: Pearson.
- LOB = *Lancaster-Oslo/Bergen Corpus* http://www.essex.ac.uk/linguistics/clmt/w3c/corpus_ling/content/corpora/list/private/LOB/lob.html.
- WAT = Schoonees, P.C. (Ed.-in-chief). 1972. *Woordeboek van die Afrikaanse Taal. Volume 4 H-I*. Pretoria: Government Printer.

Other literature

- Atkins, B.T. Sue and Michael Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.
- Atkins, B.T. Sue, Michael Rundell and Edmund Weiner. 1997. *Salex'97. A Training Course in the Compilation of Monolingual Dictionaries*. Unpublished course material of a tutorial held at the Dictionary Unit for South African English, Rhodes University, Grahamstown, 15–26 September 1997.
- Biber, D. 1993. Using Register-diversified Corpora for General Language Studies. *Computational Linguistics* 19: 219–241.
- Biber, D. 1995. *Dimensions of Register Variation: A Cross-Linguistic Comparison*. Cambridge: Cambridge University Press.
- De Schryver, G.-M. and D.J. Prinsloo. 2000. Electronic Corpora as a Basis for the Compilation of African-language Dictionaries, Part 2: The Microstructure. *South African Journal of African Languages* 20(4): 310–330.
- Eaton, Jan. 1989. *Borduursteke vir Suid-Afrika. 'n Volledige gids*. Cape Town: Delos.
- Garside, Roger, Geoffrey Leech and Tony McEnery (Eds.). 1997. *Corpus Annotation: Linguistic Information from Computer Text Corpora*. London/New York: Longman.

- Gilbert, Zoë.** 1985. *Die Suid-Afrikaanse tuin. Maand vir Maand*. Second Edition. Johannesburg: Central News Agency.
- Kennedy, Graeme.** 1998. *An Introduction to Corpus Linguistics*. London/New York: Longman.
- Kilgarriff, Adam.** 1997. Putting Frequencies in the Dictionary. *International Journal of Lexicography* 10(2): 135-155.
- Kruyt, J.G. and M.W.F. Dutilh.** 1997. A 38 Million Words Dutch Text Corpus and its Users. *Lexikos* 7: 229-244.
- MacLeod, Catherine and Ralph Grishman.** 2000. The Influence of Corpora on Lexicons: Corpora Use in the Creation of COMLEX Syntax and NOMLEX. Heid, Ulrich et al. (Eds.). 2000. *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000, Stuttgart, Germany, August 8th–12th, 2000*: 141-148. Stuttgart: Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.
- Media24: <http://www.media24.co.za>.
- Prinsloo, D.J. and R.H. Gouws.** 2006. Fashion Words in Afrikaans Dictionaries: A Long Walk to Lexicographic Freedom or Just a Lexical Fly-by-Night? Corino, E., C. Marelllo and C. Onesti (Eds.). 2006. *Proceedings XII EURALEX International Congress. Turin, Italy, September 6th–9th, 2006*: 301-312. Alessandria: Edizioni dell'Orso.
- Otlogetswe, T.J.** 2007. *Corpus Design for Setswana Lexicography*. Unpublished Ph.D. Thesis. Pretoria: University of Pretoria.
- Summers, Della.** 1993. Longman/Lancaster English Language Corpus — Criteria and Design, *International Journal of Lexicography* 6(3): 181-208.
- Summers, Della.** s.d. [1996–1998]. Corpus Lexicography — The Importance of Representativeness in Relation to Frequency. *Longman Language Review* 3: 6-9.
- Turpin-Delpont, Lesley.** 1988. *Die Suid-Afrikaanse boek van laslappie en appliekwerk*. Cape Town: C. Struik.

Function-related Secondary User Needs and Secondary Data? A Critical Examination of Some Central Concepts in the Modern Theory of Lexicographical Functions

Marcin Overgaard Ptaszynski, *Centre for Lexicography, Aarhus School of Business, University of Aarhus, Aarhus, Denmark (maop@asb.dk)*

Abstract: The aim of this article is to contribute to the development of the modern theory of lexicographical functions by offering a critical examination of the following concepts associated with it: *primary needs, primary data, secondary needs, secondary data, function-related needs, and function-related data*. By way of introduction, a presentation of the basic tenets of the theory is offered, followed by a description of the general characteristics of secondary needs and data. Next, on the basis of both a critical analysis of the theory and an examination of selected data types in existing lexicographical products, it is argued that all user needs and all data that satisfy those needs are function-related. The distinction between the concepts *function-related* and *usage-related* is thus rejected. Since this has serious implications for the relation between secondary needs and data on the one hand and primary needs and data on the other, this relation is subsequently reconsidered. This leads to a redefinition of all the concepts examined. It is also explained why an ideal state of lexicography, where secondary needs and data do not exist, cannot be achieved in the real world.

Keywords: DICTIONARY, LEXICOGRAPHY, LEXICOGRAPHICAL FUNCTION, PRIMARY NEED, PRIMARY DATA, SECONDARY NEED, SECONDARY DATA, FUNCTION-RELATED NEED, FUNCTION-RELATED DATA, EXTRA-LEXICOGRAPHICAL SITUATION, INTRA-LEXICOGRAPHICAL SITUATION, COMMUNICATIVE FUNCTION, COGNITIVE FUNCTION, OPERATIVE FUNCTION, TEXT RECEPTION, TEXT PRODUCTION, TRANSLATION

Opsomming: **Funksieverwante sekondêre gebruikersbehoefte en sekondêre data? 'n Kritiese ondersoek na 'n aantal kernbegrippe in die moderne teorie van leksikografiese funksies.** Die doel van hierdie artikel is om by te dra tot die ontwikkeling van die moderne teorie van leksikografiese funksies deur 'n kritiese ondersoek aan te bied van die volgende begrippe wat daarmee verbind word: *primêre behoeftes, primêre data, sekondêre behoeftes, sekondêre data, funksieverwante behoeftes en funksieverwante data*. Ter inleiding word 'n aanbieding van die basiese beginsels van die teorie gegee, gevolg deur 'n beskrywing van die algemene eienskappe van sekondêre behoeftes en data. Vervolgens, op grond van sowel 'n kritiese ontleding van die teorie as 'n ondersoek van uitgesoekte datatipes in bestaande leksikografiese produkte, word daar geredeneer dat alle gebruikersbehoefte en alle data wat daardie behoeftes bevredig, funksieverwant is. Die onderskeid tussen die konsepte *funksieverwant* en *gebruikersver-*

want word sodoende verwerp. Omdat dit ernstige implikasies vir die verhouding tussen sekondêre behoeftes en data aan die een en primêre behoeftes en data aan die ander kant het, word hierdie verhouding gevolglik heroorweeg. Dit lei tot 'n herdefiniëring van al die ondersoekte begrippe. Daar word ook verduidelik waarom 'n ideale toestand van die leksikografie waar sekondêre behoeftes en data nie bestaan nie, nie in die werklike wêreld bereik kan word nie.

Sleutelwoorde: WOORDEBOEK, LEKSIKOGRAFIE, LEKSIKOGRAFIESE FUNKSIE, PRIMÊRE BEHOEFTE, PRIMÊRE DATA, SEKONDÊRE BEHOEFTE, SEKONDÊRE DATA, FUNKSIEVERWANTE BEHOEFTE, FUNKSIEVERWANTE DATA, EKSTRALEKSIKOGRAFIESE SITUASIE, INTRALEKSIKOGRAFIESE SITUASIE, KOMMUNIKATIEWE FUNKSIE, KOGNITIEWE FUNKSIE, OPERATIEWE FUNKSIE, TEKSONTVANGS, TEKSPRODUKSIE, VERTALING

1. Introduction

In his seminal work on the modern theory of lexicographical functions (hereafter referred to as the MTLF), Tarp defines the lexicographical function as "the satisfaction of the specific type of lexicographically relevant need that may arise in a specific type of potential user in a specific type of extra-lexicographical situation" (Tarp 2008a: 81). A lexicographically relevant need arising in an extra-lexicographical situation is invariably a need for information that can be satisfied by a lexicographical product. This would be, for example, the need to know what the Afrikaans word *woordeboek* means, which can be satisfied by looking up the word in, e.g., an Afrikaans–English dictionary.¹ A type of potential user is identifiable on the basis of carefully selected criteria, such as the user's mother tongue, their proficiency in a specific foreign language, and their knowledge of a specific subject or science (Tarp 2008a: 54 et seq.).

The current version of the MTLF distinguishes between four types of extra-lexicographical situations: communicative, cognitive, operative and interpretative. The communicative type, embracing situations in which a potential dictionary user is in need of linguistic assistance in order to solve a problem arising in an act of communication, is subdivided into the following subtypes: text reception, text production, text editing, and text revision (in all cases with regard to either the user's mother tongue or a particular foreign language), as well as translation (to and from the user's mother tongue or between two foreign languages). Situations of the cognitive type are those in which a potential user of a lexicographical product needs help with acquisition of new knowledge about a given subject (or with revision of what they already know). The communicative and cognitive types of situation are well-established concepts in the MTLF, and are described in a variety of publications, such as Tarp (2006, 2008a) or Bergenholtz and Tarp (2003). Recent developments of the theory have offered descriptions of two other situation types. The operative situation type is defined by Tarp (2008b: 126 et seq.) as a type of situation in which a potential user of a lexicographical product seeks "directions and instructions on how to proceed in specific situations".² The interpretative situation type has not been

given a definitive description yet, but is tentatively defined as a type of situation in which a potential user of a lexicographical product needs help with the interpretation of non-linguistic signs (Tarp 2009).

On the basis of the definition of a lexicographical function quoted and explicated above it is possible to construct an example of a functional relationship, i.e. a relationship between a type of lexicographically relevant need, a particular type of potential user and a specific type of extra-lexicographical situation. Thus, a Dane who is highly proficient in their mother tongue but less so in English may experience a problem with text reception in connection with reading a contemporary, non-specialised English text. The problem may, for example, consist in the fact that the Dane does not understand the English word *ambled* in the sentence *I just kept talking to them as they ambled by ignoring me.*³ The lexicographically relevant need that will arise in this situation is the need for information about the meaning of the word *ambled*. In order to satisfy this need, the user can consult a lexicographical product with an appropriate function, in this case an English–Danish dictionary or a monolingual English learner's dictionary. In either case it should be a general-purpose dictionary designed to help a Danish user with text reception in English. Assuming that the user knows that *ambled* is a verb and is the past tense form of *amble*, they will look up the infinitive form in the dictionary and find the Danish equivalent or an English definition, both of which will satisfy their need for information.

Certainly, the above scenario represents an idealised version of reality. It excludes the possibility of the user being unable to find the data they are looking for, e.g. if they do not know how to spell the word they are trying to look up. Likewise, it does not allow for the possibility that the user may be unable to interpret the data they have found, e.g. if they do not understand the meaning of some terms included in the definition of the word they have looked up. In both cases the user would develop another need for additional data that would help them find or interpret the data they were looking for before they consulted a particular lexicographical product. To account for such problems, the MTLF distinguishes between the so-called primary and secondary user needs. According to Tarp (2008a: 41 et seq.), primary needs always arise in an extra-lexicographical situation, i.e. before the consultation of a dictionary or another lexicographical product. They are to be distinguished from secondary user needs, which arise in an intra-lexicographical situation, i.e. during the consultation of a particular lexicographical product. Primary needs are satisfied by primary lexicographical data, while secondary needs are, correspondingly, satisfied by secondary lexicographical data. Since primary needs underlie the function of a lexicographical product, they are also known as function-related needs, whereas secondary needs are also referred to as usage-related needs (Tarp 2008a: 41). The binary opposition between *function-related* and *usage-related* applies not only to needs but also to data which are meant to satisfy these needs (Tarp 1998: 128 et seq.).

This article offers a critical examination of the distinction between the

notions of *function-related* and *usage-related*, which for a variety of reasons proves to be problematic. To support the view that all needs and data are related to the function of a lexicographical product, a critical analysis of the MTLF is undertaken with regard to the available works on the theory itself, as well as to selected data types. Consequently, the concepts of primary and secondary needs and data are re-examined and redefined with regard to the notion of the lexicographical function.

2. Secondary needs and data: general characteristics and examples

An analysis of the distinction between *function-related* and *usage-related* requires a thorough understanding of the concepts to which these terms apply. As mentioned above, in the MTLF the predicate *function-related* is used with regard to primary needs and primary data. The rationale behind this is clear: primary needs underlie the function of a given lexicographical product, and the function, in turn, decides the presence, form and content of primary data. Whatever determines or is determined by the function must be function-related. Therefore, the concepts of primary needs and data need no further elaboration in this context. The term *function-related* is not applied to the concepts of secondary needs and secondary data. In order to establish whether this is a felicitous approach, it is necessary to examine how these concepts are defined and used in the MTLF.

Tarp (2008a: 42) states that "*secondary user needs* (also known as usage-related needs) [...] do not arise until an actual lexicographical situation occurs, when the potential user turns into an *actual dictionary user* and may need guidance in the form of *secondary lexicographical data* to find and interpret the *primary lexicographical data* needed to satisfy their primary user needs". Secondary needs are "*both needs for information and needs for instruction and education*" (Tarp 2008a: 57 et seq.), the MTLF distinguishing three types of such needs:

- (1) "Secondary needs related to specific dictionary consultation", which are needs "for information that can help users to find and confirm the [primary] data required".
- (2) "Secondary needs related to the user's linguistic and subject-specific qualifications", which are needs for information that can help users to extract information out of primary data.
- (3) "Secondary needs related to general or specific dictionary usage", which include the needs for "general education in lexicography, general instruction in dictionary usage, information about the specific dictionary, and instruction in the use of specific dictionary".

Each type of secondary need can be satisfied by the corresponding types of data. Dictionaries thus contain "data that helps users to find the specific data

that is relevant in the situation in question, and data that may help users to interpret it — followed by data that users can use to extract general information about the dictionaries in question" (Tarp 2008a: 58), i.e. data types that satisfy the above-mentioned needs of type 1, 2, and 3 respectively. Other types of secondary data also exist, e.g. data concerning the content, quality and reliability of the dictionary, which the users may need to access before buying and using the dictionary (Tarp 2008a: 58).

The relation between types of secondary needs and lexicographical solutions satisfying these needs is illustrated by Tarp (2008a: 57 et seq.) with several examples. For instance, a secondary need related to specific dictionary consultation can be represented by a situation in which "a Danish user wishing to understand the Afrikaans word *skepe* needs to know that *skepe* is the plural form of *skip* and that additional information can be found under the lemma *skip*". As regards secondary needs related to the user's linguistic and subject-specific qualifications, "users who have very little knowledge about a specific discipline need simple lexicographical data about this discipline" in order to be able to "extract information out of complex data".

Secondary needs are also specified in Tarp's description of user needs in particular extra-lexicographical situations. For instance, in connection with the reception of non-specialised texts in their mother tongue, a Danish user may develop secondary needs with regard to the following information categories: lemmata, orthography, orthographical variants (and mistakes) as lemmata, irregular inflection forms as lemmata, word combinations as lemmata, word class, and gender (Tarp 2008a: 70 et seq.).

Examples of secondary data types that may be included in a dictionary are offered in an earlier work on the MTLF and include the preface, the user's guide and the consumer's label (Tarp 1998: 129).

Finally, two other important characteristics of secondary needs need to be mentioned. First, secondary needs can be satisfied "both within and outside the framework of specific dictionaries" (Tarp 2008a: 58). Second, it is possible to combine classes of information data that satisfy primary needs with those that are related to secondary needs, provided that the classes in question "are not in direct contradiction to each other" (Tarp 2008a: 76).

3. Critical examination of selected concepts in the MTLF

3.1 Function-related vs. usage-related needs and data

There appear to be good reasons to question the validity of this distinction, as it is possible to argue that not only primary but also secondary user needs and data are function-related. Several arguments in favour of such a view can actually be found in the existing works on the MTLF. Others can be drawn on the basis of an analysis of selected data types and the types of need they are meant to satisfy.

3.1.1 Evidence from the MTLF

Some incentive to reconsider the distinction between function-related and usage-related needs and data is provided by Tarp's example of a secondary need quoted above. On the one hand, it is certainly correct to maintain that the secondary data which inform the Danish user about the Afrikaans word *skepe* being the plural form of *skip* and redirect them to the lemma *skip* are usage-related, insofar as the user intends to look up the form *skepe* in the dictionary. On the other hand, it would be risky to claim that these data are not function-related. The situation described in the example is one of a user who wants to solve a problem connected with text reception and is seeking assistance in a dictionary having the corresponding function. It is solely in connection with the user's particular type of primary need, reflected in the function of the dictionary, that the secondary need appears and is satisfied by the presence of specific secondary data. If the function of the dictionary was other than that of helping a (specific type of) Danish user with text reception in Afrikaans, the user's secondary need to find out that *skepe* is the plural of *skip* would be neither predicted by the lexicographer nor satisfied by that dictionary. The presence of the secondary data in a dictionary and, in the first place, the lexicographer's prediction about the user's need that requires it, thus appear to be inextricably bound with the function of that dictionary.

This view is reinforced by the fact that secondary needs are specified by Tarp (2008a) in the description of user needs in particular extra-lexicographical situations. The secondary needs experienced by a user solving a problem concerning text reception in their mother tongue are, for example, different from the needs arising in connection with text production in the same user's mother tongue. Only the latter include, among others, the needs for information about the meaning of lemmata, the meaning of idioms and the meaning of proverbs (Tarp 2008a: 72). It follows thus that secondary needs are not universal but dependent on the extra-lexicographical situation. If this is the case, they are necessarily related to the function of a dictionary. Certainly, the above-mentioned relations between particular secondary needs on the one hand and information data categories on the other are only true if the analysis of user needs is restricted, in accordance with Tarp (2008a), to Danish, Danish texts and the use of non-specialist language. The needs for specific types of secondary data will vary across languages, and consequently, across dictionaries. For example, a secondary need for information about gender makes little sense with regard to dictionaries of Finnish, which has no grammatical gender, but is definitely conceivable with respect to dictionaries of Polish or German, both of which distinguish between three genders in nouns.

Further encouragement to consider secondary user needs to be function-related can be found in Tarp's comment on combining information classes related to primary and secondary needs. As mentioned earlier, the classes can be combined only if they are not in direct contradiction to each other. To

explain how this principle works, the following example concerning all communicative situations restricted to the user's mother tongue is offered: "either [...] the same demands must be made with regard to completeness in terms of the meaning of lemmata; or [...] it is at least possible to find a common denominator" (Tarp 2008a: 76). This implies that if the same demands *cannot* be made, or if a common denominator *cannot* be found, then information classes related to particular primary and secondary needs *are* in direct contradiction to each other and *cannot* be combined. In other words, in order to be included in a dictionary, information classes related to secondary needs must be compatible with those related to primary needs. And as primary needs determine the function of the dictionary and are function-related, it is thus hard to argue that the same does not apply to secondary needs. This observation is strengthened by the following statement from another publication on the principles of the MTLF: "no data whatsoever should be included in a dictionary if it cannot be argued on the basis of its respective functions" (Bergenholtz and Tarp 2003: 177).

Moreover, it seems that the above implication reveals a weakness of the MTLF. By determining the function of a dictionary and the scope of all classes of information to be included in it, primary needs necessarily determine the extent to which the dictionary can satisfy secondary needs. This is hardly reconcilable with the claim that the satisfaction of secondary needs is vital for the satisfaction of primary needs. If both statements are correct, then primary needs seem to determine the extent to which they themselves can be satisfied by a dictionary. This is clearly undesirable, since it allows for a situation where a primary need calls only for partial satisfaction.

3.1.2 Analysis of selected data types

An examination of data types in existing lexicographical products also provides a reason to question the division of needs and data into those which are function-related and those which are not. Since primary data are by definition function-related, the scope of the analysis will be restricted to data types that in accordance with the MTLF would be considered as secondary data. However, before any analysis is undertaken, it bears mention that it will exhibit an approach directly incompatible with the basic tenets of the MTLF. The point of departure adopted here is the specific data (types) that may or may not be present in particular dictionaries, rather than the specific needs that these data are supposed to address. This is what Tarp (2008a: 10) calls a *contemplative* approach, represented by various lexicographical theories (e.g. Hayashi (1978), Wiegand (e.g. 1984), and to a significant degree Svensén (2004) and Yong and Peng (2007)). By contrast, the MTLF is *transformative*, i.e. oriented towards generating rules that will govern the form and content of future dictionaries and other lexicographical products. Nevertheless, the contemplative approach will not be entirely dismissed here, as it does shed light on the problem of the dis-

tion between function-related and usage-related needs and data. Selected data types examined below are divided into two broad categories: data found in the outside matter of a dictionary and those included in the microstructure of a dictionary.

3.1.2.1 Data included in the outside matter

Two examples of data types often found in the outside matter will be considered here: the user's guide and the table of contents. Even in the absence of any lexicographical theory, both the lexicographer and the dictionary user would intuitively feel that these data types are somehow different from such data as, e.g. the pronunciations of headwords or the examples illustrating their use. It is difficult to imagine a situation in which a (specific type of) potential user would consult a dictionary to satisfy their primary need with the data offered in the user's guide or the table of contents.⁴

The user's guide is classified by Tarp (1998: 129) as secondary and thereby not a function-related data type. However, this claim is refuted by the content of existing dictionaries, some of which were actually compiled in accordance with the principles based on the MTLF. Examples in point include the Danish–English and English–Danish dictionaries of accounting (Nielsen et al. 2004 and 2007, respectively). In each case the user's guide is structured around the functions of the dictionary. The data it contains are grouped under the headings corresponding to the user's primary needs and, by extension, the dictionary functions, e.g. "Hjælp til at oversætte engelske tekster til dansk" ('Assistance in translating English texts into Danish') (Nielsen et al. 2007: 12). The function-oriented approach to the composition of the user's guide is also advocated in metalexicographical literature, e.g. in Nielsen (2006) and Ptaszynski (2008). Thus, in spite of being a secondary data type, the user's guide appears to be function-related.

A similar claim could be made with regard to the table of contents (if it is to be regarded as lexicographical data in the first place). While definitely usage-related, it presents the user with an overview of the entire content of the dictionary, most (if not all) of which is determined by the function of the dictionary. A dictionary illustrating it particularly well is Nielsen et al. (2007). The table of contents guides the reader, e.g. to the relevant, function-dependent section in the user's guide as well as to additional material in the back matter, such as examples of formats of financial statements in English. The primary function of the dictionary is to provide assistance in translating English accounting texts into Danish. If the function were different, the content of both the user's guide and the back matter, and, in consequence, the table of contents, would be different too.

At this point one might argue that the function of a dictionary clearly determines the form and contents but *not* the presence of the user's guide or the table of contents in it. Any dictionary, regardless of its function, normally con-

tains such data types. However, dictionaries are not the only kinds of lexicographical products. Tarp (2008b: 126 et seq.) incorporates handbooks, manuals, and how-tos into the domain of products which can include lexicographically structured data. Such products are meant to fulfil the operative function. This sheds more light on the status of such data types as the user's guide. They do not seem to appear in certain lexicographical products whose primary function is operative, e.g. manuals. By contrast, they are usually included in dictionaries, which are designed to assist the user in communicative and/or cognitive situations. This means that the user's guide is indeed function-related, as not only its form and content but also its presence is determined by the function of a particular lexicographical product. The case of the table of contents is more difficult. Although its form and content depend on the function of the specific lexicographical product, its presence is probably determined by other factors, such as the medium. For instance, printed reference works often contain tables of contents, while internet-based products rarely do (unless one admits the sitemap into this category).

3.1.2.2 Data included in the microstructure

A thorough examination of all possible data types included in the microstructures of a wide range of dictionaries would be clearly beyond the scope of the present discussion. It would also be unnecessary, as a single entry is sufficient to provide arguments in favour of the view that there exists a relation between secondary data and the function of the dictionary containing these data. All that is required is that the analyzed entry includes any data supposed to satisfy a secondary need, and that it is possible to determine the function of the dictionary containing that entry. Both conditions are met with regard to the entry *baffle* in *The Pocket Oxford Spanish Dictionary* reproduced below:

baffle /'bæfəl/ *vt* **(a)** (perplex) desconcertar* **(b)** (frustrate) <efforts> frustrar.

The primary function of the dictionary is to aid English-speaking learners of Spanish with translation (nothing is said about the direction of the translation, but since the dictionary is bidirectional, it possibly aims to assist the user in translation both to and from Spanish). Thus the relevant primary data in the entry *baffle* will be the Spanish equivalents of the headword, i.e. *desconcertar* and *frustrar*. By contrast, "**(a)**", "**(b)**", "(perplex)", "(frustrate)", and "<efforts>" are clear examples of secondary data. This can be established in two ways. First, it is inconceivable that any primary user need related to translation would be satisfied by these data. Second, the sole purpose of all these data appears to be to guide the user to the choice of the correct translation equivalent. The user's guide confirms that this assumption is correct with regard to the data type exemplified by "<efforts>". This data type is described as "words often used with the headword, shown to help select the correct translation for each context".

The reasons for including data types exemplified by "(a)" and "frustrate", which are termed "sense divisions" and "sense indicators" respectively are not specified (Carvajal and Horwood 2000: ix et seq.). However, they seem to be the same as those for the inclusion of the data type exemplified by "<efforts>". If this is the case, the data can only serve to help the user find the required primary data, and must, accordingly, be classified as secondary data.

There should be no doubt as to whether the Spanish equivalents of *baffle* given in the entry are function-related or not. All primary data are by definition related to the function of the dictionary. The same question is more difficult to address with respect to the secondary data mentioned above. A satisfactory answer must take into account the fact that the data types represented by "(a)", "frustrate" and "<efforts>" do not form a homogeneous category. The last of these data types appears to be more readily classifiable as function-related than the other two. As stated above, the user's guide reveals that the purpose of such data is to help the user select the correct translation. Since the function of the dictionary is that of helping the user with translation, these data must be function-related. This assumption is proved right if a corresponding entry in a dictionary with a different function is considered. For example, in a dictionary for Spanish-speaking users seeking assistance in text reception in English, such data would probably not be included, as the user does not need to know the collocational preferences of the headword (as exemplified by the relation of *efforts* to *baffle*) in order to grasp its meaning (cf. Tarp 2008a: 149).

The case of "(a)" and "frustrate" is slightly more complicated. Sense divisions and sense indicators also commonly appear in dictionaries with other functions than just that of helping the user with translation. However, it is still possible to establish the relation between these data types and the function of the lexicographical product that contains them. To begin by stating what is apparently obvious: sense divisions and sense indicators make sense only in lexicographical products addressing such user needs that arise either in connection with acts of verbal communication or during the acquisition of knowledge about the vocabulary of a particular language. In other words, their use is mostly confined to dictionaries. Such types of data would be absent from lexicographical products that aim to satisfy operative user needs, e.g. lexicographically structured manuals and handbooks.

Even if the scope of the discussion is restricted to dictionaries, the use of sense divisions and sense indicators still seems to depend on the function of a particular dictionary. One might argue that the presence of such data types depends in the first place on certain linguistic properties of the headword. After all, it is the fact that some words, like *book* or *cry*, have more than one established sense,⁵ but other words, like *stopcock*, *pneumonia* or *lemonade* do not, that makes it possible to use any kind of sense discriminators in some entries but not in others. However, even in the case of polysemous words, the decision to represent a particular number of senses in the specific entry lies ultimately with the lexicographer. The outcome of this decision depends exclusively on

the function of the dictionary. For instance, a dictionary for adult native speakers of English, designed primarily for the acquisition of knowledge about the English vocabulary, e.g. OED, may include four senses of *taxi* (n.), while a learner's dictionary intended for text reception and text production in English, e.g. LDOCE, will include only one. In the latter case, *ipso facto* no sense divisions are used. The use of sense indicators is subject to similar restrictions. They are often used in dictionaries for learners, such as LDOCE, but not, for instance, in dictionaries aiming to help users with text reception in their mother tongue, e.g. *Merriam-Webster Online Dictionary*. An adult native-speaker would normally be able to retrieve the relevant information about the particular sense of the headword from the definition or an illustrative example (Atkins and Rundell 2008: 214).

3.2 Secondary vs. primary needs and data

If, as argued above, secondary data seem to be function-related, are they definable as a category? In other words, what distinguishes them from primary data? An obvious answer, based on the definition of secondary needs presented earlier, is that the need to access secondary data does "not arise until an actual lexicographical situation occurs" (Tarp 2008a: 41). However, this statement appears to create more problems than it aims to solve. First, the typology of secondary needs includes "secondary needs related to general or specific dictionary usage", such as the needs for general education in lexicography and general instruction in dictionary usage. This type of secondary need could surely occur in an extra-lexicographical situation, independently of any primary need that would trigger an actual lexicographical situation with regard to a specific dictionary. The need for a course in dictionary use could be a part of someone's general education. It does not have to be a step on the way to satisfy their need to solve a particular problem by means of a specific dictionary. A related problem is that of a prospective user accessing data about a specific dictionary before actually buying and using it. Conscious buyers would carefully select a dictionary so that it would suit their needs, but that does not necessarily mean that they would do it only after they have experienced a specific problem that can be solved by a particular dictionary.

Second, as was observed earlier, secondary needs can be satisfied within the framework of specific dictionaries as well as outside this framework. If the latter is the case, assistance can be provided by general courses in dictionary use as well as by workbooks with exercises in the use of a specific dictionary (Tarp 1998: 128 et seq.).⁶ These are, however, not the only possibilities. For instance, even though an ideal dictionary should be thought to contain all the data necessary for the interpretation of the primary data included in it, the user may still decide to satisfy their secondary need by means of a different dictionary. And whenever the user turns to another lexicographical product — be it a dictionary, an encyclopaedia, or a lexicographically structured workbook or

manual — what was originally the secondary need becomes a new primary need, as the situation changes from intra-lexicographical to extra-lexicographical.

Third, the above definition of secondary needs does not take into account that these needs can often be satisfied by means of the primary data included in a given dictionary. For example, an advanced learner of English using an English monolingual dictionary in connection with text reception may experience difficulty with understanding the definition of the word they are looking up. If the problem is caused by the fact that the definition contains a word unfamiliar to the user,⁷ it can often be solved by looking up this word in the same dictionary. The definition provided in the relevant entry may be (and usually is) what the lexicographer intended to be a piece of primary data, included in the dictionary in order to satisfy needs arising in an extra-lexicographical situation. It is, of course, possible to object to such reasoning by arguing that it entails a mixture of two different perspectives. The same data are being viewed here in terms of their relation to the user on the one hand and to the lexicographer on the other. According to the adopted perspective, they are classified as data intended to satisfy secondary or primary needs. And if a general theory of lexicography, like the MTLF, is supposed to account for the needs-adapted access to data, the user's standpoint is certainly more important than the lexicographer's. Considered on the surface, such an argument seems valid. However, it fails to take into account that the user cannot access any data that would satisfy their need, until the lexicographer has physically placed and organized these data in a lexicographical product. A conscientious lexicographer certainly begins working on the dictionary project by conducting an examination of both primary and secondary needs the potential user of that dictionary might have. On this basis, all decisions regarding the inclusion and presentation of all primary and secondary data are made. Still, regardless of how careful the analysis of user needs may have been, it is impossible for the lexicographer to make absolutely correct predictions about what would be primary and secondary data for a specific user in all possible acts of dictionary use. To do so, one would virtually have to know all specific lexicographically relevant needs of that user and all facts relevant to establishing the user profile (and assume that all these variables are constant). Consequently, one would need to produce lexicographical products on an individual basis. Since this is not and never will be possible in practice, to produce even the best lexicographical product it is necessary to assume that the lexicographer's perspective matches that of the user's, despite the fact that these two standpoints only approximate each other. This assumption, in turn, allows for situations where a secondary need is satisfied by means of primary data.

From the above it follows that the difference between primary and secondary data is difficult, if not downright impossible to determine on the basis of the nature of the situation (extra-lexicographical vs. intra-lexicographical) that triggers the need for either kind of data. It seems more appropriate to speak

about the use of specific data as primary or secondary. But even this would only be true provided that (a) a given secondary need arises in connection with a specific primary need and (b) the user seeks the satisfaction of both needs in the same dictionary (or other kind of lexicographical product). It is, on the other hand, reasonable to say that some data types, e.g. the user's guide, the table of contents, sense discriminators, or labels indicating the part of speech, are more likely to satisfy secondary rather than primary needs. They may also be used to satisfy secondary needs more often than it would be the case with other data types, such as the explications of the meaning of the headword.

Some data types, e.g. the user's guide, are used to satisfy the user's secondary needs regardless of the function of the dictionary containing them (although, as was demonstrated earlier, the function still determines the contents, form and presence of the data). The use of other data types to satisfy the user's secondary needs will be determined by the function of the dictionary. This is e.g. the case with data concerning gender in dictionaries of Danish for native speakers. According to Tarp (2008a: 71 et seq.), such data are used to satisfy the user's primary needs in dictionaries intended for text production, and to satisfy the user's secondary needs in dictionaries for text reception. Lastly, data types like definitions of headwords may be used to satisfy both primary and secondary needs (though perhaps not with the same frequency) in the same dictionary.

4. Secondary needs as a lexicographical necessity

If secondary needs seem to be a volatile category and secondary data cannot be distinguished clearly from primary data even with respect to the same dictionary, why not reject both concepts? After all, this is what the MTLF would see as an ideal state of lexicography. In an ideal lexicographical situation, there are no secondary needs and, by extension, no secondary data. This conclusion can be drawn from the description of the ideal dictionary presented by Bergenholtz (2008). The dictionary in question is like the *Junior Woodchucks Guidebook*, used by Donald Duck's nephews Huey, Dewey and Louie. "The answer [to the user's problem] is always correct and always understandable." This means that secondary needs for interpreting the primary data never occur. Neither do users ever experience a secondary need in connection with accessing the primary data. "They find the answer without writing any question at all. The 'computer' is able to read their thoughts and tell them exactly what they need to know."⁸

The entire theory thus seems to rest on the axiom that for every need for information there are specific data which can satisfy this need exhaustively. This may seem obvious. In fact, however, the acceptance of the axiom as the supreme principle of the theory and practice of lexicography has two serious implications. First, it obliges lexicographers to strive towards the reduction of secondary needs to the absolute minimum. These needs appear only if primary needs cannot be satisfied directly, and in the ideal world this is never the case.

Second, whenever the interpretation of primary data in a particular dictionary poses a problem to the user, and the resulting secondary need cannot be solved by means of the same dictionary, the blame can conveniently be laid on the lexicographer (for not having implemented the axiomatic principle correctly) or on the user (for choosing an inappropriate tool to satisfy their need). Whichever is the case, the principle itself is by no means rendered invalid.

Nevertheless, as the absolute minimum of secondary needs cannot be defined universally, the above principle is not possible to observe in the real world. Two problems will help to illustrate the point. First, even in a lexicographical product perfectly geared to a specific user profile, it would still be impossible for the lexicographer to predict and absolutely avoid all situations in which the user finds the specific primary data difficult to understand. To do so, one would have to assume that the user's command of a particular language, their knowledge of a particular subject matter, or their ability to perform a particular operation never deteriorates. Unfortunately, this is not how the human mind works. The same primary data, e.g. the definition of a headword, can be fully understandable for the user on one occasion but not on another, if they happen to have forgotten the meaning of as much as a single word contained in it. In the latter case, a secondary need would appear, even though it may not have been predicted by the lexicographer or may not have been relevant for the user at the moment of purchasing the dictionary. Thus, to avoid the occurrence of such secondary needs, the user would need as many dictionaries as there are extra-lexicographical situations that would trigger their use. Such a scenario is obviously unrealistic.

Next, even if the user's cognition is assumed to be perfect, in many cases there would still be a need for data whose aim is to guide the user to the desired primary data. An example in point is the need to represent such data as, e.g. verb forms in Spanish. According to the adopted axiomatic principle, it should be possible to construct a lexicographical product that would help, for instance, an English user with little or no command of Spanish to understand the simple sentence *Hablamos español*. ('We speak Spanish.') The form *hablamos* is normally not lemmatized in dictionaries for text reception in Spanish. If it is, the users are, at best, redirected to the entry *hablar* to learn that the headword means 'speak', and are subsequently referred to the tables of verb conjugation in the grammar section appended in the back matter, where they learn that *hablamos* is the first person plural present of *hablar*. This is, however, not what the user was hoping for. Their primary need was to understand the meaning of *hablamos*, not of *hablar*, and certainly not to be directed to an exposition of Spanish grammar. In an ideal dictionary, all forms of all inflectable lexical items are lemmatized and supplied with all the relevant data the user requires to satisfy their primary need. This is certainly not possible in real dictionaries. It would overburden the macrostructure of every dictionary, even an internet-based one, where space considerations are of minor importance compared to printed dictionaries. Therefore, the number of forms lemmatized and given a

full entry must be reduced, and if such forms as *hablamos* are to be lemmatized at all, the entire entry will only guide the user to other data.

Moreover, the idea of minimizing the amount of secondary needs and data reveals another fact about the axiom. The principle based on it reflects an idealized world of essentially linguistic, communicative primary user needs. Secondary needs that appear in connection with these needs are relatively uncomplicated, because they can often be satisfied by simple data, e.g. a sense discriminator, a cross-reference to an entry for the canonical form of a verb, or definitions of the defining vocabulary in a monolingual dictionary for text reception. The same is not the case with respect to lexicographical tools designed to help the user in cognitive situations, especially those of the systematic type (cf. Tarp 2008b: 123 et seq.), where the user seeks assistance in connection with a regular study of a specific subject field. For example, a dilettante in music who wishes to understand the notion of the *musical scale* must become familiar with such concepts as *octave*, *interval* or *pitch class*. The lexicographer can, of course, provide cross-references to the corresponding entries (as is done, e.g. in the article *musical scale* in *Wikipedia*). This, however, would probably not satisfy the user's primary need. Having consulted the entries for the above-mentioned concepts, they might understand the definition of *musical scale*, but their knowledge of what a musical scale actually is would still be very far from exhaustive if they do not know the concepts of *whole tone*, *semitone*, *major*, *minor*, and *chromatic*. This looks like an endless chain of interrelated secondary needs, all of which should ideally be predicted by the lexicographer. The data necessary to satisfy them are complex (they comprise entire entries) and form an open set (it is not possible to predict how many entries the user will have to consult until they consider their knowledge of the subject matter studied to be satisfactory). These needs thus cannot be reduced to any minimum.

One could, of course, argue that such a view of systematic cognitive needs is inadequate. The primary need itself arises in a situation that presupposes a systematic use of a lexicographical product. Having to consult more than one entry should then perhaps be considered as part of the realization of the primary need rather than as a series of secondary needs following the first look-up. On the other hand, in terms of the very act of dictionary use, this scenario bears a resemblance to another one, in which the user needs to look up two (or perhaps several) entries in the dictionary in order to be able to arrive at the correct interpretation of the primary data satisfying the primary need for information in connection with, e.g. text reception in a foreign language. Why then should the MTLF be inclined to identify all the data searches following the first look-up as (aspects of) a primary need in the former case, but as secondary needs in the latter? The question seems even more appropriate to ask once one realizes that in both cases the data searched after the first look-up would otherwise be present in the lexicographical product because they satisfy the primary needs the product is intended to address. The only answer that appears reasonable is that what is called primary and secondary — needs and data alike

— is in fact closely related to the function of the specific lexicographical product. Moreover, if, in a systematic cognitive situation, the necessity to consult more than one entry is to be seen as part of the specific primary need, then it must be assumed that primary needs arise in an extra-lexicographical situation and then somehow continue to appear in actual lexicographical situations, until they are satisfied. This is in fundamental disagreement with the very core of the MTLF, according to which the difference between primary and secondary needs is based on their occurrence in, respectively, extra-lexicographical and lexicographical situations. Therefore, the concepts of primary and secondary needs and data need to be redefined, and their status in relation to the concept of the lexicographical function needs to be reconsidered. This is the topic of the next section.

5. Redefining the basic concepts

The above discussion of the concepts of secondary needs and data has revealed that the following distinctions postulated by Tarp are either questionable or inadequate if they are to be understood as choices between static, binary opposite categories:

- function-related vs. usage-related,
- primary need vs. secondary need, and
- primary data vs. secondary data.

The critical analysis of the MTLF, as well as of selected data types has yielded the following observations, all of which are in sharp contradiction with Tarp's definitions of primary and secondary needs and data:

- (1) Secondary needs are function-related. It follows from the fact that they vary according to the primary need, which itself is always function-related.
- (2) The satisfaction of primary needs depends on the satisfaction of particular secondary needs. However, at the same time primary needs determine the extent to which a particular lexicographical product can satisfy secondary needs.
- (3) The presence, form and content of secondary data types commonly found in the outside matter of a lexicographical product, e.g. the user's guide (and to a significant extent the table of contents too), depend on the function of this product. These data types are therefore function-related.
- (4) The presence, form and content of secondary data types commonly found in the microstructure of a lexicographical product, e.g. sense indicators or collocational specifications, depend on the function of this product. These data types are therefore function-related.

- (5) Some secondary needs, e.g. the need for a general course in dictionary use, can arise in extra-lexicographical situations, independently of any specific primary needs.
- (6) Secondary needs can become primary needs. This happens if the user attempts to satisfy their secondary needs by means of a different lexicographical product than the one chosen to satisfy the original primary need.
- (7) Secondary needs can be satisfied by primary data, e.g. the explications of meaning of the headword.
- (8) As the user's memory is naturally imperfect, it is beyond the lexicographer's ability to predict some secondary needs and guarantee the inclusion of secondary data that should satisfy them. Therefore, every lexicographical product will at some point of time prove to be imperfect.
- (9) Whether similar needs are classified as primary in one case and secondary in another depends on the kind of extra-lexicographical situation that triggers the first consultation of a particular lexicographical product.
- (10) Some primary needs, e.g. those related to systematic cognitive situations, can arise, at least partially, in intra-lexicographical situations.
- (11) The general, axiomatic principle that for every need for information there are specific data which can satisfy this need exhaustively is not observable in the real world, as the number of secondary needs cannot be reduced to any definable minimum.
- (12) The principle that for every need for information there are specific data which can satisfy this need exhaustively is applicable only to some lexicographically relevant needs.

In the light of these results, the notions primary and secondary user needs and data need to be reconsidered. Accordingly, a number of changes to the MTLF are proposed.

First, all lexicographically relevant needs are function-related. Likewise, all data that satisfy these needs are function-related. This is a corollary of observations 1–4 above. Accordingly, the distinction between *function-related* and *usage-related* is rejected.

Second, the distinction between primary and secondary needs is retained but modified. On the basis of observations 5, 6, 9 and 10 above, user needs are characterized as follows:

A *primary need* is a lexicographically relevant need for information. It typically appears in an extra-lexicographical situation. However, in user situations that by definition require multiple consultations of a lexicographical product a primary need can appear in an extra-lexicographical situation

and continue to reappear in an actual lexicographical situation until it is satisfied.

A *secondary need* is a lexicographically relevant need for information or instruction and education. It typically appears in an actual lexicographical situation as a result of difficulties with access to or interpretation of specific data meant to satisfy the user's primary need. However, a secondary need can also occur in an extra-lexicographical situation, independently of any specific primary need. The satisfaction of such a need will ideally facilitate the satisfaction of future primary needs.

The categories of primary and secondary needs are dynamic, because a secondary need can become a primary need if the user wishes to satisfy it by means of a different lexicographical product than the one chosen to satisfy their first primary need. Secondary needs can usually be related to specific primary needs although they do not necessarily have to result from them or be satisfied at the same time.

Third, the distinction between primary and secondary data as categories of information data that satisfy, respectively, primary and secondary needs is rejected. This follows from observations 7 and 8 above. Instead, a concept of the use of particular data (types) to satisfy the user's primary or secondary needs is proposed. Accordingly, rather than falling into a set of rigid categories with clear-cut boundaries, various data (types) yield permutations that can satisfy particular primary and secondary needs. Some, but not all, of these permutations can be predicted by the lexicographer on the basis of an analysis of the profile of a potential user and their lexicographically relevant needs. This does not contradict the view that some data types may reveal a propensity to be used to satisfy primary or secondary needs (or both). However, categorizing data types as such makes little sense without referring to a particular function requiring their presence in a lexicographical product.

Finally, two additional conclusions can be drawn from the analysis of secondary needs. First, in line with observations 8, 9 and 11, it must be admitted that the ideal state of lexicography in which the user experiences no secondary needs cannot be achieved. This is due to such factors as technical constraints on the macrostructure of a lexicographical product and the natural disposition of the human mind to forget knowledge and abilities. Although the impossibility to avoid secondary needs stems from these purely practical obstacles, it is only strengthened by the fact that on the theoretical level similar needs can be classified differently as primary or secondary, according to the extra-lexicographical situation triggering them. Second, in the light of observation 12, the MTLF is in need of further development. If a general principle for the theory and practice of lexicography is to be postulated at all, it has to apply to all lexicographically relevant needs of the user.

Notes

1. The choice of such a lexicographical product would obviously be preferred provided that the potential dictionary user is a native speaker of English or masters English better than Afrikaans.
2. This definition has recently been replaced by the following one: "An operative situation is defined as one in which a potential user needs help (instructions, directions, etc.) in connection with the completion of a physical or mental operation" (Tarp 2009).
3. *Protection when camping*, page 2. *City-Data Forum, Wyoming*. <http://www.city-data.com/forum/wyoming/605955-protection-when-camping-2.html> [accessed 15 April 2009].
4. Obviously, this assumption excludes the situation in which a metalexigrapher examines the user's guide in one or more dictionaries and subjects the data to a scholarly analysis. However, dictionaries, like all other utility tools, are made to be used rather than to be analyzed. Therefore, it is legitimate here to ignore the specific needs arising in such situations.
5. The dispute of whether a word can have more than one established sense at all, and if so, how to determine what a sense is (cf., e.g. Cruse 2004) will be left to semanticists. Likewise, an idealized state is assumed here, in which specific senses of a given word are universally recognized (although not necessarily represented in the same way in particular dictionaries).
6. It is taken for granted that such a workbook would necessarily have to reflect the function of the dictionary it is associated with. If the workbook is to be treated as a collection of secondary data (as the need it aims to satisfy is secondary), this provides further support for the view that such data are function-related.
7. It is assumed here that the dictionary is suited to the user profile in question and is indeed intended to help the user with text reception. However, even with the best intentions, the lexicographer is certainly not able to predict every word that would be unfamiliar to every user of the dictionary, and consequently, to avoid using these words in the definitions.
8. It should be noted here that not even the *Junior Woodchucks Guidebook* can satisfy these criteria. From the relevant entry in *Wikipedia*, one learns that access to data in the *Guidebook* is through "an extensive index". This is also confirmed in an earlier work by Bergenholtz (1999: 93). Thus, the *Guidebook* allows for the possibility of secondary needs arising in connection with accessing primary data.

References

Dictionaries

- Carvajal, C.S. and J. Horwood (Eds.). 2000. *The Pocket Oxford Spanish Dictionary*. Second Edition. Oxford: Oxford University Press.
- LDOCE = Summers, D. et al. (Eds.). 2005. *Longman Dictionary of Contemporary English*. Fourth Edition with Writing Assistant. Harlow: Pearson Education.
- Nielsen, S., L. Mourier and H. Bergenholtz. 2004. *Regnskabsordbogen dansk-engelsk. (Danish-English Accounting Dictionary)*. Copenhagen: Forlaget Thomson.
- Nielsen, S., L. Mourier and H. Bergenholtz. 2007. *Regnskabsordbogen engelsk-dansk. (English-Danish Accounting Dictionary)*. Copenhagen: Forlaget Thomson.
- OED = *Oxford English Dictionary Online*. <http://dictionary.oed.com> [accessed 13 March 2009]. Oxford University Press, 2009.

- Merriam-Webster Online Dictionary*. <http://www.merriam-webster.com> [accessed 13 March 2009]. Merriam-Webster, 2009.
- Wikipedia. The Free Encyclopedia*. <http://en.wikipedia.org> [accessed 13 March 2009].

Other literature

- Atkins, B.T.S. and M. Rundell.** 2008. *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Bergenholtz, H.** 1999. Das schlaue Buch. Vermittlung von Informationen für textbezogene und textunabhängige Fragestellungen. Zettersten, A., J.E. Mogensen and V.H. Pedersen (Eds.). *Proceedings of the Eighth International Symposium on Lexicography at the University of Copenhagen*: 93-110. Lexicographica. Series Maior 90. Tübingen: Max Niemeyer.
- Bergenholtz, H.** 2008. Concluding Remarks: Lexicographical Tools of Tomorrow. Abstract of a paper delivered at the conference *Lexicography at a Crossroads*, Aarhus, 19–21 May 2008.
- Bergenholtz, H. and S. Tarp.** 2003. Two Opposing Theories: On H.E. Wiegand's Recent Discovery of Lexicographic Functions. *Hermes* 31: 171-196.
- Cruse, A.** 2004. *Meaning in Language. An Introduction to Semantics and Pragmatics*. Second Edition. Oxford: Oxford University Press.
- Hayashi, T.** 1978. *The Theory of English Lexicography 1530–1791*. Amsterdam Studies in the Theory and History of Linguistic Science, Series III, 18. Amsterdam: John Benjamins.
- Nielsen, S.** 2006. A Functional Approach to User Guides. *Dictionaries. Journal of the Dictionary Society of North America* 27: 1-20.
- Ptaszynski, M.O.** 2008. Anmeldelse af Okoniewski, G. 1999. *Dansk-polsk juridisk-merkantil ordbog*. ('Review of Okoniewski, G. 1999. *Danish-Polish Dictionary of Law and Commerce*.) *LexicoNordica* 15: 301-316.
- Svensén, B.** 2004. *Handbok i leksikografi. Ordböcker och ordboksarbete i teori och praktik. (Handbook of Lexicography. Dictionaries and Dictionary Making in Theory and Practice.)* Second Edition. Stockholm: Norstedts Akademiska Förlag.
- Tarp, S.** 1998. Leksikografien på egne ben. (Lexicography on its Own Feet.) *Hermes* 21: 121-137.
- Tarp, S.** 2006. *Leksikografien i grænselandet mellem viden og ikke-viden. Generel leksikografisk teori med særlig henblik på lærerleksikografi. (Lexicography in the Borderland between Knowledge and Non-knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography.)* Doctoral Thesis. Aarhus: Aarhus School of Business.
- 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* 40: 117-131.
- Tarp, S.** 2009. Operative og interpretative funktioner. (Operative and Interpretative Functions.) PowerPoint presentation from the workshop held at the Centre for Lexicography, Aarhus, 3 March 2009.
- Yong, H. and J. Peng.** 2007. *Bilingual Lexicography from a Communicative Perspective*. Terminology and Lexicography Research and Practice 9. Amsterdam/Philadelphia: John Benjamins.
- Wiegand, H.E.** 1984. On the Structure and Contents of a General Theory of Lexicography. Hartmann, R.R.K. (Ed.). *LEXeter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9–12 September 1983*: 13-30. Lexicographica. Series Maior 1. Tübingen: Max Niemeyer.

How Effective is the Electronic Dictionary in Sense Discrimination?*

Kim Hua Tan, *Pusat Pengajian Bahasa dan Linguistik, Universiti
Kebangsaan Malaysia, Bangi, Selangor, Malaysia (kimmy@ukm.my)*

Abstract: This article compares the efficacy of the electronic dictionary with that of the print dictionary in helping learners differentiate senses of polysemous words in dictionaries. An adaptation of the mixed methodology proposed by Johnson and Christensen (2004), the research design in this article encompasses a qualitative phase and a quantitative phase in the overall research study along the dimensions of time order and paradigm emphasis. The element of 'comparison' is included resulting in a design of four paired comparison groups: (1) Group^{e-pre} and Group^{p-pre}, (2) Group^e and Group^p, (3) Group^{e-without} and Group^{e-with}, and (4) Group^{e-withoutLowMed} and Group^{e-withLowMed}. Findings show that the electronic dictionary is effective in helping Low to Medium Proficient students (Group^{e-LowMed}) in the electronic group after deliberate dictionary training in navigation and windows switching. This is indicated by improved scores regarding time taken (efficacy rate) and a significant correlation between actual efficacy and self-perceived efficacy. The results imply that dictionary users need to be given dictionary training based on specific problems they face. As a whole, however, the print dictionary group has higher efficacy than the electronic group but there was no discernible trend in the relationship between its actual efficacy and the self-perceived efficacy for both groups. This suggests that subjects' perceived efficacy beliefs are not good predictors of their performances.

Keywords: EFFICACY, ACTUAL EFFICACY, SELF-PERCEIVED EFFICACY, ELECTRONIC DICTIONARIES, PRINTED DICTIONARIES, POLYSEMOUS WORDS

Opsomming: Hoe doeltreffend is die elektroniese woordeboek by betekenisonderskeiding? Hierdie artikel vergelyk die doeltreffendheid van die elektroniese woordeboek met dié van die gedrukte woordeboek om aanleerders te help om die betekenis van polisemiese woorde in woordeboeke te onderskei. Die navorsingsontwerp in hierdie artikel wat 'n aanpassing van die gemengde metodologie voorgestel deur Johnson en Christensen (2004) is, omvat 'n kwalitatiewe fase en 'n kwantitatiewe fase in die totale navorsingstudie langs die dimensies van tydsvolgorde en paradigmatoneering. Die element van "vergelyking" wat ingesluit

* This article is based on research done for an unpublished Ph.D. Dissertation *Sense Differentiation in Electronic and Print Dictionaries: A Comparative Study of Dictionary Efficacy*, accepted by the Multimedia University, Cyberjaya, Selangor, Malaysia in 2007. It was presented as a paper at the Thirteenth International Conference of the African Association for Lexicography, organized by the Bureau of the *Woordeboek van die Afrikaanse Taal*, Stellenbosch, Republic of South Africa, 30 June–3 July 2008.

word, het 'n ontwerp van vier in pare gerangskikte vergelykingsgroepe tot gevolg: (1) Groep^{e-pre} en Groep^{P-pre}, (2) Groep^e en Groep^P, (3) Groep^{e-sonder} en Groep^{e-met}, en (4) Groep^{e-sonderLaeMed} en Groep^{e-metLaeMed}. Bevindings toon dat die elektroniese woordeboek doeltreffend is deur Lae tot Medium Bekwame studente (Groep^{e-LaeMed}) in die elektroniese groep te help na doelbewuste woordeboekopleiding in navigasie en vensterorskakeling. Dit word getoon deur die verbeterde puntestande ten opsigte van tyd geneem (doeltreffendheidsmaatstaf) en die belangrike verband tussen werklike doeltreffendheid en selfgeskatte doeltreffendheid. Die resultate toon dat woordeboekgebruikers woordeboekopleiding gegee moet word wat berus op die spesifieke probleme waarvoor hulle te staan kom. As 'n geheel egter het die gedruktewoordeboekgroep 'n hoër doeltreffendheid as die elektroniese groep, maar daar was geen waarneembare tendens in die verhouding tussen die werklike doeltreffendheid en die selfgeskatte doeltreffendheid van altwee groepe nie. Dit dui daarop dat proefpersone se geskatte doeltreffendheidsbeskouings nie goeie voorspellers is van hul prestasies nie.

Sleutelwoorde: DOELTREFFENDHEID, WERKLIKE DOELTREFFENDHEID, SELFGESKATTE DOELTREFFENDHEID, ELEKTRONIESE WOORDEBOEKE, GEDRUKTE WOORDEBOEKE, POLISEMIESE WOORDE

Introduction

This article looks at the potential of the e-dictionary in comparison to the print dictionary in helping English as a Second Language (ESL) learners in deciphering meanings of words. By and large, in the area of English language studies, particularly in an ESL learning context, one pertinent concern would be whether dictionaries, electronic or otherwise, can help learners in mastering the second language and, in particular, whether dictionaries can help in directing learners to the right sense of the word in the dictionary. If the words concerned are polysemous, they are known to be difficult to ESL learners. Two general questions guided the determination of dictionary efficacy in this research:

- (1) Do e-dictionaries have a higher efficacy in comparison to print dictionaries in providing easier access to sense differentiation?
- (2) Is there a relationship between dictionary users' actual efficacy and their self-perceived efficacy?

Past Research on Dictionary Comparison Studies

User studies with respect to electronic dictionaries are frequently linked to second language comprehension. Nesi (2000) attempted to define and categorize the electronic dictionary, and consider the skills associated with its use for second language comprehension. She also reviewed previous research and put forward new research approaches that involved electronic dictionaries.

In their research, Laufer (2000), Tono (2000) and Weschler and Pitts (2000) utilized controlled studies, comparing dictionary use under different conditions. Laufer (2000) investigated incidental vocabulary acquisition in two reading conditions: when unknown words are encountered in a paper text and glossed in the margin, and when they are read on computer screen and explained in an electronic dictionary. She collected data immediately after a reading task and again two weeks later. Results indicate that the electronic group performed significantly better than the paper text group. The superiority of the electronic dictionary over print was also pointed out by Leffa (1992a). The experiment, designed to test the performance of the electronic glossary as compared to traditional bilingual dictionaries, shows that the electronic glossary is superior, both in the number of idea units it allowed the subjects to retrieve from the text, and the time it demanded for the comprehension. Using the electronic glossary, the subjects read the passages faster and obtained more meaning from them. The difference was greatest with subjects whose language proficiency was lowest.

Speed is often perceived as an advantage of e-dictionaries but it has also been pointed out by Dillon (1992) that on-screen reading is 20%–30% slower than reading on paper in an experimental setting. It has been suggested that rapid reading must therefore be regarded as a deliberate activity. In a comparison of reading and skimming from 'books' and screens, Muter and Maurutto (1991) found that comprehension was higher when reading was done from screen. This could be accounted for by a speed-accuracy trade-off, as the skimming speed from paper was significantly faster than from screen. Another point to make is that different types of dictionaries are useful for different purposes. The electronic dictionary is conceivably useful for examinations, where speed is important, though the user might risk miscomprehension owing to the lack of sufficient explanation. On the other hand, the print dictionary is good for homework, where the aim is to learn, not to finish a task quickly.

Current efforts were also undertaken by Thinsan (2003) to compare the effects of the e-dictionary and paper-based dictionary in terms of the students' vocabulary learning retention and reading comprehension. Sharpe (1995: 50), who claims that 'one fear ... expressed by teachers ... is that the speed with which these electronic dictionaries retrieve information may not necessarily aid the memory's retention of the information for language learning purposes'. The reason for this, according to Sharpe, probably lies in the fact that while looking up a word in the print dictionary, the reader has to think about the word for a longer time since it takes longer to find the word than when working with the e-dictionary, but this point needs further research. In addition, the print dictionary enables the user to see the word family, thereby exposing him/her to more varieties and uses of the same word. The very fact that the user sees this variety can contribute to his/her retention of the meaning. It then appeared that print dictionaries can help to learn words, whereas most of the e-dictionaries help to find words. The retention of words is likely to be better with a print dictionary than with an e-dictionary.

Tono (2000) reported the effects of three different electronic interfaces on EFL learners' look-up behaviour. Subjects performed language tasks under three conditions: with a paper dictionary, a traditional electronic dictionary, and two non-traditional electronic dictionaries. He investigated ESL learners' dictionary use and three factors that might influence it: the interfaces of electronic dictionaries, language translation tasks, and repeated exposure to the interface. The three interfaces of electronic dictionaries he examined were the traditional interface, the parallel interface and the layered interface. In the traditional interface, the information provided is similar to that of a print dictionary and it is easy to import the dictionary file into an electronic format. Information in the parallel interface is provided in a parallel bilingual translation format. All the words and phrases are stored separately in individual entries with their translation equivalents in a parallel format. In the layered interface, information is organized by a menu using tabs. Microstructurally, the organization is such that each tab offers different information, ranging from basic to complex. Results showed that, at the macrostructural level at least, electronic dictionaries provide quicker access to the target entry than the paper medium. The parallel interface allows faster search than the other two interfaces, even in the case of the complex search of idioms and compounds. Weschler and Pitts (2000) examined look-up speed, investigating whether electronic dictionaries are faster to use than print dictionaries. This study only looked at the efficiency of the dictionaries without considering accuracy. The three studies, nevertheless, highlighted the potential of the electronic dictionary in helping the comprehension of English as a second language.

Other e-dictionary related areas looked into by researchers are the e-dictionary as a teaching and learning tool (Hartmann and James 2001), dictionary skills (Tickoo 1989; Nesi 1999), the e-dictionary and its users (Atkins 1998; Cowie 1999; Dolezal and McCreary 1999; Hartmann 1999; Rundell 1999; Scholfield 1999), dictionary training (Berwick and Horsfall 1996; Nesi 1999; Wright 1998) and the structure of e-dictionaries (Burke 1998; Dodd 1989).

The earlier dictionary comparison studies tended to centre on the use of bilingual dictionaries. This is because ESL learners use them more for decoding than encoding. Leffa (1992b) compared the efficacy of electronic dictionary glossaries with the traditional paper dictionary on text comprehension in a translation task among 20 university students enrolled for a course in beginner English for academic purposes. The hypotheses were that the electronic dictionary users would spend less time with translation and have a better comprehension. The results indicated that the electronic dictionary users understood more of the passage (86%) than the traditional dictionary users (62%) and needed 50% less time to translate the passage. There was little attempt to investigate the students' actual look-up behaviour and its effects on their performance. Aust, Kelley and Roby (1993), for example, also compared the use of a hyper-reference source (electronic dictionary) and a conventional print dictionary. Measures included consultation frequency, study time, efficiency and

comprehension. The efficacy of bilingual and monolingual dictionaries was also compared. But these studies again looked into the general behaviour patterns of learners in relation to reading skills with the help of glosses.

Clearly, dictionary comparison studies are not new, but very few of these deal with specific linguistic problems encountered by second language learners of English, such as the sense discrimination of polysemous words. There is an apparent need to add to this literature.

Method

The research design was a mixed method design comprising a paired experimental comparison group. This design utilized two control groups, the electronic group (Group^e) and the print group (Group^p). Subjects completed a task under two different conditions. Group^e used an electronic dictionary to complete a sense differentiation task and Group^p used a print dictionary to complete the same task. The research design was adapted from Johnson and Christensen's (2004) 2 x 2 matrix of qualitative and quantitative permutation but with changes so as to accommodate the element of 'comparison'. This is illustrated in Fig. 1 below. There are altogether four subdesigns of comparison groups that make up the whole research design for the study:

- (1) Comparison Design of Group^{e-pre} and Group^{p-pre} (Group^{e-pre}: Electronic Group pre-study, Group^{p-pre}: Print Group pre-study);
- (2) Group^e and Group^p (Group^e: Electronic Group, Group^p: Print Group);
- (3) Group^{e-without} and Group^{e-with} (Group^{e-without}: Electronic Group without dictionary training, Group^{e-with}: Electronic Group with dictionary training);
- (4) Group^{e-withoutLowMed} and Group^{e-withLowMed} (Group^{e-withoutLowMed}: Electronic Low Medium Proficiency Group without dictionary training, Group^{e-withLowMed}: Electronic Low Medium Proficiency Group with dictionary training).

In each subdesign, collection of data for each paired group is weighted either towards the qualitative or quantitative aspects (paradigm emphasis) or according to sequential or concurrent methods (time emphasis).

There are altogether three stages in the study, a pre-study, a main study and an extended study, iterative steps occurring at the main study and the extended study. This is shown in Fig. 2 below, which is a flow chart of the overall comparison design according to the stages of study. At the main study and extended study stages, both data collection and data analysis, and subsequent iterations, proceeded in tandem, repeatedly referring back to each other: 'It involves a weaving back and forth between data and theory' (Bryman 2001: 10).

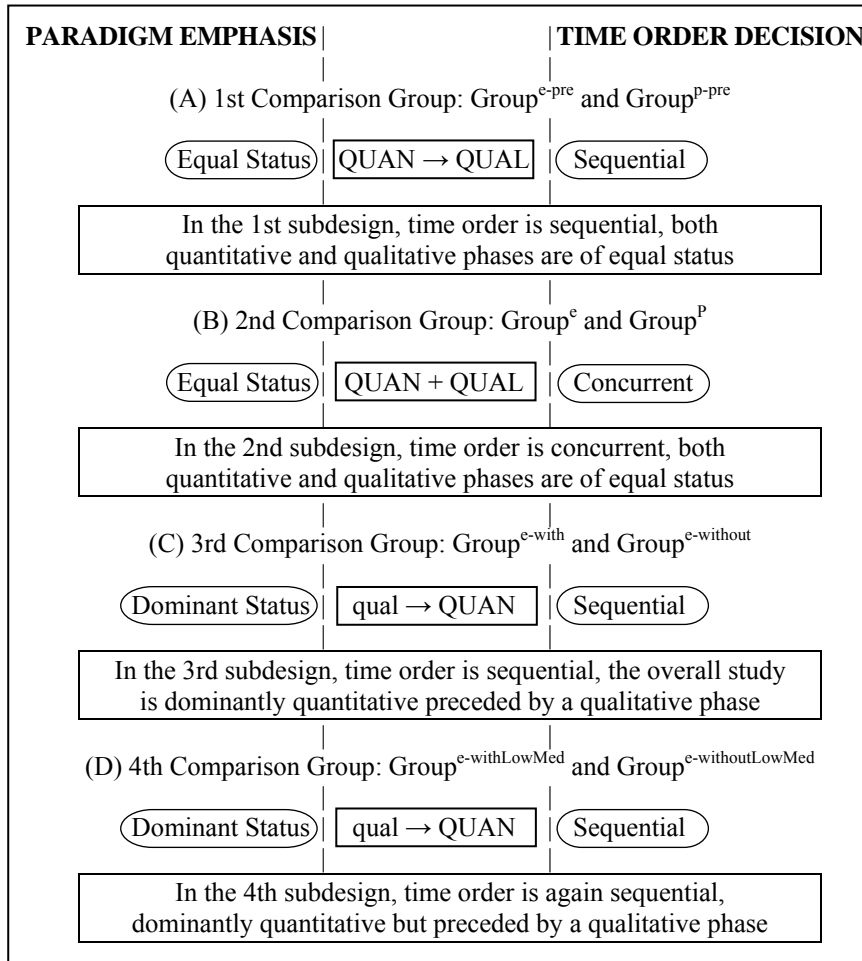


Fig. 1: Mixed-Method Comparison Design for Dictionary Efficacy Study (Tan 2007).

Key Notation: QUAL and qual both stand for qualitative research; QUAN and quan both stand for quantitative research; capital letters denote priority or increased weight; lower-case letters denote lower priority or weight; a plus sign (+) indicates the concurrent collection of data; an arrow (→) represents a sequential collection of data, for example qual → QUAN indicates dominant status, sequential design where the overall study is primarily quantitative but is preceded by a qualitative phase.

(Johnson and Christensen 2004)

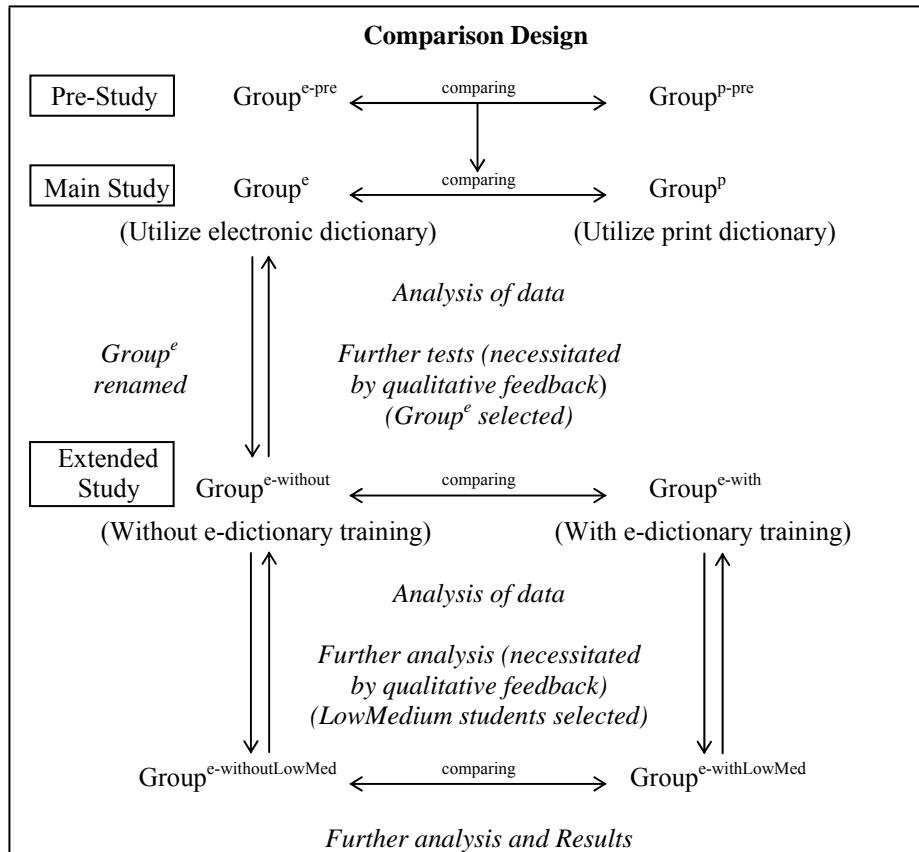


Fig 2: Comparison Design of Group^{e-pre} and Group^{p-pre}; Group^e and Group^p, Group^{e-without} and Group^{e-with}; Group^{e-withoutLowMed} and Group^{e-withLowMed} (Tan 2007)

The iterative nature of this study indicated two further significant steps in the procedure. Firstly, data analysis in the main study pointed towards and necessitated a pre- and a post-test for Group^e with dictionary training as a treatment procedure. This was to determine whether dictionary training was another possible factor in determining the subjects' efficacy. Secondly, data also indicated that further analysis on a selected group of subjects from Group^e, Low to Medium Proficient students (Group^{e-LowMed}) was pertinent and necessary.

Efficacy is the ability to complete a task accurately, independently and quickly. In this study, efficacy had two variants, namely actual efficacy and self-perceived efficacy. Actual efficacy is indicated by scores obtained over time. (See research instrument WoSIT below.) It is quantitatively measured. Self-perceived efficacy is the subjects' belief in their capabilities to produce effects, to effectively control specific events in their lives (Bandura 1997). In this

study, self-perceived efficacy refers to how users perceived the use of the dictionaries indicated by the perceptual ratings of the ease of use, satisfaction and usability of the dictionaries. (See research instrument Retro below.) The argument is that if subjects find the dictionaries easy, useful and satisfactory, by implication they also perceive their capability to be high, hence high self-perceived efficacy.

Subjects

The participants were university undergraduates majoring in English Language Studies at the Universiti Kebangsaan Malaysia, a public university in Malaysia. An electronic dictionary group, Group^e, consisting of 154 participants and a print dictionary group, Group^p, comprising 150 participants were formed from the total of 304 undergraduates selected for this study. (For Participants' Profile, see Table 1.) The proficiency level of the students was determined by their Malaysian University Entrance Tests (MUET) bandscores.

Table 1: Student Profile according to MUET bandscores

Electronic Dictionary n = 154 participants		Print Dictionary n = 150 participants	
MUET Band	% of students	MUET Band	% of students
6 High High	–	6 High High	–
5 Low High	22.08	5 Low High	20.7
4 High Medium	36.4	4 High Medium	66.7
3 Low Medium	35	3 Low Medium	12.56
2 High Low	6.52	2 High Low	0.04
1 Low Low	–	1 Low Low	–

Research Instruments

Three main research instruments were used in this study.

- (1) *WoSIT (Word Sense Identification Test)*. WoSIT consisted of ten sense identification items designed to test learners' ability to select the right sense or meaning of the underlined word in the test item. A slot for recording the total time taken to complete the whole task was also included at the end of the page. Students were required to identify and circle the correct sense of a particular word in a sentence from the given range of senses of that word. This task was comparable to the written component (Editing section) of the UETESOL (University Entrance Test in English for Speakers of Other Languages) previously known as NEAB (Northern Examinations and Assessment Board) that required candidates to match a word in context with one of several meanings in a given

dictionary entry. UETESOL is an established English proficiency entrance requirement equivalent or similar to IELTS or TOEFL for students enrolling for courses in British universities.

- (2) *Retro*. The Retro Sheets are self-completed questionnaires comprising two parts. Part A is a semantic differential scale. The items in the scale were designed to gauge the learners' perceptions and attitudes toward the use of the two types of dictionaries, particularly their perceived ease, satisfaction and usability by looking at their perceptual ratings. In Part B, the short response questions aimed to elicit information on the difficulties the participants faced in using the electronic dictionary for deciphering the senses of the words.
- (3) *COBUILD3*. Both the print and electronic versions of the third edition of *Collins COBUILD English Dictionary for Advanced Learners* (2001), designated COBUILD3, were used as references during the completion of the task in WoSIT. The electronic version of COBUILD3, called *Collins COBUILD on CD-ROM*, was based on the printed reference work. They are similar microstructurally (in terms of contents and sequencing of entries) but differ macrostructurally (in terms of means of access and navigation), because both print and electronic formats dictate different routes.

Findings

Generally, results of the study indicated that Group^P (print dictionary group) performed better in the task than Group^E (electronic dictionary group), the difference in efficacy being significant. This implied that the print dictionary was more effective in assisting subjects in sense differentiation than the electronic dictionary. Three reasons were put forward as possible factors affecting their performance. These were (1) familiarity with the print format for Group^P, (2) the nature of the test materials and (3) a lack of exposure to the electronic format for Group^E.

The efficacy of Group^P could be attributed to the fact that subjects were more familiar with the print format or paper-based dictionary and therefore used less time to complete the task. Actual efficacy was measured through performance indicated by accuracy rate (scores obtained over time). As time taken was of the essence, it is conceivable that Group^P had the upper hand in the task owing to familiarity with the dictionary format. Most Malaysian students have earlier experiences with print dictionaries in their schooldays in comparison to electronic dictionaries which only appeared locally in educational institutions less than a decade ago. It could be the case where there is a speed-accuracy trade-off. This refers to participants' trading of increases in speed for decreases in accuracy (and vice versa) over a substantial range (Lohman 1989). Speed is an advantage for Group^E because of the electronic medium while familiarity with the print format helped in accelerating the look-up proc-

ess for Group^P but if subjects were to prioritize accuracy to the exclusion of speed, then these are not advantages any more.

The nature of the materials used in the experimental task in this study could also have affected the relative performance of Group^e and Group^P. The fact that both the experimental test and the reference dictionary used by Group^P were in print mode could have contributed to Group^P's better performance. On the other hand, Group^e used the e-dictionary which was in a different mode from the printed experimental test. Though some of the subjects had indicated that they had no problems reading from a screen, it is worthwhile reiterating here that research had shown that reading from a screen in an experimental setting is 20% to 30% slower than reading from paper (Dillon 1992) which could have accounted for Group^e's lower efficacy. Although using an electronic dictionary in the completion of print work was also very much a real-life task, the differing modes or formats could have affected their performances. Other possible factors were familiarity with the computer keyboard, or even typing speed, though these factors were deemed insignificant because the words the subjects were required to type were on average only six letters in length.

Data had also suggested that the poor performance of Group^e could have been attributed to lack of exposure and training in using the electronic dictionary. Qualitative input indicated that more preparation was needed for Group^e before they attempted the task. Of Group^e, 25% claimed unfamiliarity with the format of the electronic dictionary, particularly navigational features. Although there was a short orienting phase before the collection of data, it was clearly not enough. Further investigation suggested that the orienting exercise consisting of a 10-minute exposure for Group^e was inadequate. Apparently, the novelty effect was far more overwhelming for most of the students in Group^e than Group^P. Only 0.9% of the subjects in Group^P indicated that they had difficulties using the print dictionary as a reference tool to complete the task. In comparison to Group^e which had 25%, 0.9% for Group^P was considered minimal.

Deliberate instruction or training, however, did not appear to improve the efficacy of Group^e. Although mean efficacy was higher after training, the results of the paired t-test showed that there was no significant difference in dictionary efficacy between Group^{e-without} and Group^{e-with} with the introduction of training.

However, data from the mean rating of the parameters of 'ease', 'satisfaction' and 'usefulness' indicated that 25% of the subjects, particularly those from the Low Medium Proficient group, Group^{e-LowMed}, showed encouraging responses and better perceptual ratings after more exposure to dictionary training. There was also improvement in their scores and efficacy. Training seemed to have positive effects on Group^{e-withLowMed}.

Subsequent analysis confirmed that deliberate instruction had in fact made a difference to this group of subjects and the e-dictionary was instrumental in helping them with their task. The pre-post paired t-test carried out

on $\text{Group}^{\text{e-LowMed}}$ is a commonly used inferential statistic, mainly intended to determine whether scores before training (pre-test) are significantly different from scores after training (post-test). Results from the analysis of the paired t-test on this particular paired group indicated a significant difference in the means of the efficacy between both groups, with the means of $\text{Group}^{\text{e-withLowMed}}$ higher than $\text{Group}^{\text{e-withoutLowMed}}$. It appeared that although training did not make a significant difference to the whole electronic group, Group^{e} , it did improve the efficacy of a subgroup, $\text{Group}^{\text{e-withLowMed}}$. This suggested that less proficient students may need more deliberate dictionary instruction, particularly when it comes to new modalities. Highly proficient students are generally better and quicker at grasping new ideas and skills regardless. Data pointed towards the fact that unnecessary training may have a less favourable effect on them. Training evidently should be tailored to the particular group of students lest it should be insufficient for the less proficient and needless for the more proficient.

Conclusion

In summary, the results of the study indicated that Group^{P} performed better in the task than Group^{e} and the difference in efficacy was significant. The three reasons being familiarity with the print format for Group^{P} , lack of exposure to the electronic format for Group^{e} and the nature of the test materials were proposed as possible factors affecting their performance. Even after training, results indicated that there was no significant difference in dictionary efficacy between the electronic group for both $\text{Group}^{\text{e-without}}$ and $\text{Group}^{\text{e-with}}$. As the qualitative data indicated, it appeared that training was unnecessary for the whole Group^{e} . It benefited only a small subgroup of Group^{e} , $\text{Group}^{\text{e-LowMed}}$. Subsequently, there was a significant difference in efficacy for $\text{Group}^{\text{e-withoutLowMed}}$ and $\text{Group}^{\text{e-withLowMed}}$ after training. It appeared that deliberate instruction was one of the factors affecting their performance.

The results indicated that except for $\text{Group}^{\text{e-withLowMed}}$, there was a weak correlation between actual efficacy and perceived efficacy within groups: Group^{e} , Group^{P} , $\text{Group}^{\text{e-without}}$, $\text{Group}^{\text{e-with}}$ and $\text{Group}^{\text{e-withoutLowMed}}$. It appeared that all the groups had higher confidence in their performance evidenced by the above average self-rating of their own performance. Generally, the relationship was weak to poor between their actual performance and their perception of their performance. $\text{Group}^{\text{e-withLowMed}}$ comparatively had a stronger relationship between its actual efficacy and perceived efficacy. There was, however, a discernible positive trend in the relationship between the actual efficacy and perceived efficacy for all the groups.

In conclusion, it is perhaps pertinent to say that expectations of the utility and functionality of electronic dictionaries among learners, particularly those of second language learners of English, have risen over the years. What this study has demonstrated is that both electronic dictionaries and print diction-

aries have merits. The two different modes are suitable for different purposes. It is perhaps too early to say that the electronic dictionary has the capability to replace the print dictionary as Sharpe (1995: 49) puts it: 'It is my view that the advantages of the electronic dictionary and the familiarity of today's young people with electronic devices will eventually relegate the printed notion of "dictionary" to a secondary sense.'

References

- Atkins, B.T.S. (Ed.).** 1998. *Using Dictionaries. Studies of Dictionary Use by Language Learners and Translators.* Lexicographica. Series Maior 88. Tübingen: Max Niemeyer.
- Aust, R., M.J. Kelley and W.B. Roby.** 1993. The Use of Hyper-reference and Conventional Dictionaries. *Educational Technology Research and Development* 41(4): 63-73.
- Bandura, A.** 1997. *Self-efficacy. The Exercise of Control.* New York: W.H. Freeman.
- Berwick, G. and P. Horsfall.** 1996. *Making Effective Use of the Dictionary.* London: Centre for Information on Language Teaching and Research.
- Bryman, A.** 2001. *Social Research Methods.* Oxford: Oxford University Press.
- Burke, S.M.** 1998. *The Design of Online Lexicons.* Unpublished Master's Thesis. Evanston, IL., Northwestern University. <http://www.netadventure.net/~sburke/bounce.cgi/ma/>.
- Cowie, A.P.** 1999. *English Dictionaries for Foreign Learners: A History.* Oxford: Oxford University Press/Clarendon Press.
- Dillon, A.** 1992. Reading from Paper versus Screens: A Critical Review of the Empirical Literature. *Ergonomics* 35(10): 1297-1326.
- Dodd, S.** 1989. Lexicomputing and the Dictionary of the Future. James, G. (Ed.). *Lexicographers and their Works:* 83-93. Exeter: University of Exeter Press.
- Dolezal, F.T. and D.R. McCreary.** 1999. *Pedagogical Lexicography Today: A Critical Bibliography on Learners' Dictionaries with Special Emphasis on Language Learners and Dictionary Users.* Lexicographica. Series Maior 96. Tübingen: Max Niemeyer.
- Hartmann, R.R.K. (Ed.).** 1999. *Dictionaries in Language Learning. Recommendations, National Reports and Thematic Reports from the Thematic Network Project in the Area of Languages, Sub-project 9: Dictionaries.* Berlin: Freie Universität. <http://www.fu-berlin.de/elc/TNPproducts/SP9dosier.doc>.
- Hartmann, R.R.K. and G. James.** 2001. *Dictionary of Lexicography.* London/New York: Routledge.
- Johnson, B., and L.B. Christensen.** 2004. *Educational Research: Quantitative, Qualitative, and Mixed Approaches.* Boston: Allyn and Bacon.
- Laufer, B.** 2000. Electronic Dictionaries and Incidental Vocabulary Acquisition: Does Technology Make a Difference? 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:* 849-854. Stuttgart: Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.
- Leffa, V.J.** 1992a. Reading with an Electronic Glossary. *Computers and Education* 19(3): 285-290.
- Leffa, V.J.** 1992b. Making Foreign Language Texts Comprehensible for Beginners: An Experiment with an Electronic Glossary. *System* 20: 63-73.

- Lohman, D.F.** 1989. Individual Differences in Errors and Latencies on Cognitive Tasks. *Learning and Individual Differences* 1: 179-202.
- Muter, P. and P. Maurutto.** 1991. Reading and Skimming from Computer Screens and Books: The Paperless Office Revisited? *Behaviour and Information Technology* 10(4): 257-266.
- Nesi, H.** 1999. The Specification of Dictionary Reference Skills in Higher Education. Hartmann, R.R.K. (Ed.). 1999. *Dictionaries in Language Learning. Recommendations, National Reports and Thematic Reports from the Thematic Network Project in the Area of Languages, Sub-project 9: Dictionaries*: 53-67. Berlin: Freie Universität. <http://www.fu-berlin.de/elc/TNPproducts/SP9dossier.doc>.
- Nesi, H.** 2000. *The Use and Abuse of EFL Dictionaries. How Learners of English as a Foreign Language Read and Interpret Dictionary Entries*. Lexicographica. Series Maior 98. Tübingen: Max Niemeyer.
- Rundell, M.** 1999. Dictionary Use in Production. *International Journal of Lexicography* 12(1): 35-53.
- Scholfield, P.** 1982. Using the English Dictionary for Comprehension. *TESOL Quarterly* 16(2): 185-194.
- Scholfield, P.** 1999. Dictionary Use in Reception. *International Journal of Lexicography* 12 (1): 13-34.
- Sharpe, P.** 1995. Electronic Dictionaries with Particular Reference to the Design of an Electronic Bilingual Dictionary for English-speaking Learners of Japanese. *International Journal of Lexicography* 8(1): 39-54.
- Tan, K.H.** 2007. *Sense Differentiation in Electronic and Print Dictionaries: A Comparative Study of Dictionary Efficacy*. Unpublished Ph.D. Dissertation. Multimedia University, Cyberjaya, Selangor.
- Thinsan, S.** 2003. *Effects of Electronic Dictionary on Vocabulary Retention and Reading Comprehension of EFL Undergraduate Students in a Thai University*. Unpublished Master's Thesis. Language Education, Indiana University.
- Tickoo, M.L. (Ed.).** 1989. *Learners' Dictionaries: State of the Art*. Singapore: SEAMEO Regional Language Centre.
- Tono, Y.** 2000. On the Effects of Different Types of Electronic Dictionary Interfaces on L2 Learners' Reference Behaviour in Productive/Receptive tasks. Heid, U., S. Evert, E. Lehmann and C. Rohrer (Eds.). *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000, Stuttgart, Germany, August 8th-12th, 2000*: 855-862. Stuttgart: Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.
- Weschler, R. and C. Pitts.** 2000. An Experiment Using Electronic Dictionaries with EFL Students. *The Internet TESL Journal* 6(8). <http://www.aitech.ac.jp/~iteslj/Articles/Weschler-ElectroDict.html>.
- Wright, A.** 1998. *Dictionaries. Resource Books for Teachers*. Oxford: Oxford University Press.

Reflections on Lexicographical User Research*

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

Abstract: The last decades have seen a growing interest in theoretical and practical problems related to lexicographical user research. Starting with a discussion of the concept of lexicographically relevant user needs, this contribution analyses, utilising the lexicographical function theory, various types of needs to be taken into account when doing this kind of research. It then discusses the positive and negative aspects of the various methods applied, i.e. questionnaires, interviews, observation, protocols, experiments, tests, and log files. With reference to both lexicographical and sociological literature, it raises a number of problems common to most of the lexicographical user research conducted until now and recommends the application of scientific methods in future research. Finally, it proposes a number of alternative methods in order to obtain more knowledge about the real user needs.

Keywords: LEXICOGRAPHY, FUNCTION THEORY, DICTIONARY USAGE, USER NEEDS, USER TYPOLOGY, USER SITUATIONS, USER RESEARCH, RESEARCH METHODS, QUESTIONNAIRES, INTERVIEWS, OBSERVATION, EXPERIMENTS, TESTS, LOG FILES

Opsomming: Gedagtes oor leksikografiese gebruikersnavorsing. Die laaste dekades het 'n groeiende belangstelling in teoretiese en praktiese probleme betreffende leksikografiese gebruikersnavorsing gesien. Deur met 'n bespreking van die konsep van leksikografiese tersaaklike gebruikersbehoefte te begin, ontleed hierdie bydrae, met gebruikmaking van die leksikografiese funksieteorie, verskillende soorte behoeftes wat in aanmerking geneem moet word wanneer hierdie soort navorsing gedoen word. Dit bespreek dan die positiewe en negatiewe aspekte van die verskillende metodes wat benut word, d.w.s. vraelyste, onderhoude, waarneming, protokolle, eksperimente, toetse, en loglêers. Met verwysing na sowel leksikografiese as sosiologiese literatuur, bring dit 'n aantal probleme ter sprake tipies van die meeste leksikografiese gebruikersnavorsing wat tot nou toe gedoen is en beveel dit die toepassing van wetenskaplike metodes in toekomstige navorsing aan. Ten slotte stel dit 'n aantal alternatiewe metodes voor om meer kennis te verkry oor die werklike gebruikersbehoefte.

Sleutelwoorde: LEKSIKOGRAFIE, FUNKSIETEOORIE, WOORDEBOEKGEBRUIK, GEBRUIKERSBEHOEFTE, GEBRUIKERSTIPOLOGIE, GEBRUIKERSITUASIES, GEBRUIKERSNAVORSING, NAVORSINGSMETODES, VRAELYSTE, ONDERHOUDE, WAARNEMING, EKSPERIMENTE, TOETSE, LOGLÊERS

* This is a modified and improved version of an article "Kan brugerundersøgelser overhovedet afdække brugernes leksikografiske behov?" published in *LexicoNordica* 15: 5-32, 2008.

1. Introduction

It may be no exaggeration to say that the greater number of surveys conducted today are a waste of time and money. Many are simply bad surveys. Samples are biased; questions are poorly phrased; interviewers are not properly instructed and supervised; and results are misinterpreted. Such surveys are worse than none at all because the sponsor may be misled into a costly area. (Sheatsley 1974)

The last decades have seen a growing theoretical interest in lexicography, and, parallel to this development, a similar interest has emerged in acquiring more knowledge of the usage of the practical products of lexicography, first of all dictionaries. This interest was especially aroused in the wake of a conference on lexicography convened in the U.S.A. in 1960, where one of the conclusions was that "dictionaries should be designed with a special set of users in mind and for their specific needs" (Householder 1967: 279). With this starting point it is understandable that a growing number of lexicographers focused on the users and their needs, and started organising research projects within this field.

Notwithstanding, the take-off was very slow. In a chronology of 220 published user researches, Welker (2006) only lists three contributions before 1980. Hence, it was not without reason that Wiegand (1977) called the user the *well-known unknown*. However, although the user was still treated as an *unknown creature* by Neubauer (1987), from that moment the situation began to change. The 1980s and 1990s were characterised by the publication of a growing number of articles on user research, mainly as individual contributions in books and journals. Starting in the late 1990s, a number of monographs, books and thematic issues exclusively dedicated to the topic were added to the list, for instance Atkins (1998) which contains a selection of contributions from various researchers, and Nesi (2000), Tono (2001), Wingate (2002), Thumb (2004), Lew (2004) and Dziemianko (2006), which are all monographs.

Apart from these generally available published contributions, various publishing houses also carried out their own user research which, however, was treated as business secrets to protect their competitive power. This practice is a violation of the ethic principle formulated by the sociologist Merton (1968), according to which scientific discoveries are always in one way or another the result of a collective effort and should therefore be considered public property to be freely used by anybody interested. Obviously the undisclosed results of these research projects cannot be subjected to a critical analysis contributing to the development of lexicographical theory, although some authors related to publishing houses, for instance Anthony P. Cowie, occasionally refer to them in their articles.

Meanwhile, the number of publications on user research has gradually reached proportions that make it increasingly difficult to keep up with them. The corresponding need to acquire an overview of these projects and easy access to their results was met by a synoptic work (Welker 2006) which con-

tains a short summary and index of 220 research projects published between 1962 and 2006.

Parallel to the publication of the various research projects, a corresponding theoretical literature emerged. Tono (1986), for example, calls for a more scientific approach to this type of research. Ripfel and Wiegand (1988) provide critical comments on the value of the research projects carried out until then. Hartmann (2001), Tono (2001), Thumb (2004), and other monographs contain valuable theoretical reflections on the planning and organisation of user research. The same holds true for Wiegand (1987, 1998) who also provides a systematic introduction to the action-theoretical basis of dictionary usage introducing a whole number of new categories and terms, of which many are very useful and ought to be assimilated by more scholars engaged in dictionary research, although others are perhaps too much integrated into his specific lexicographical theory and therefore less useful outside the framework of this theory (cf. Bergenholtz and Tarp 2002, 2003, 2004).

The following contribution refers to the ideas presented by various scholars and endeavours to develop them even further, utilising the lexicographical function theory. It will primarily discuss which types of needs users have, the advantages, disadvantages and limitations characterising the known types of research methods, and the value that can be assigned to their results. Finally, it will propose some alternative methods to supplement the ones already applied in lexicographical user research.

2. Dictionary usage

By way of introduction, it is necessary to make a few comments on the concept of *dictionary usage*, a concept which is, in fact, inappropriate in terms of lexicographical user research because it also incorporates other types of usage not relevant to lexicography. In this respect, Wiegand (1987: 197) distinguishes between five types of dictionary usage or usage-actions:

- the normal usage of dictionaries as reference works,
- the normal usage of dictionaries as reading books about language,
- the anomalous usage of dictionaries in order to learn something about dictionary usage,
- the anomalous usage of dictionaries when they are not used as dictionaries, and
- the usage of dictionaries in order to learn normal usage.

The above classification seems to be somewhat problematic. For instance, why is the wish to acquire knowledge about dictionary usage listed as anomalous when reading dictionaries is considered normal usage? An alternative classification based on the following three criteria could be suggested:

- Is the type of usage in question lexicographically relevant or not?

- Is the dictionary in question used as a reference work or in another relevant way?
- Is the dictionary consultation in question function-specific or not?

Of the five types of dictionary usage listed by Wiegand, it is only the fourth that is lexicographically irrelevant (and unpredictable), whereas the other four are relevant in one way or another. For instance, dictionary usage with a view to acquiring skills in terms of dictionary usage is very much relevant to lexicography. Within the framework of lexicographically relevant usage, it is first of all important to distinguish between *consultation of dictionaries as reference works* — precisely the type of genuine usage separating dictionaries from other kinds of texts where users also look for information — and *other types of usage* (items 2, 3, and 5 in Wiegand's classification). Finally, it is also necessary to distinguish between a *non-function-specific consultation*, i.e. a dictionary consultation in general, and a *function-specific consultation* which takes place when users look for assistance in a lexicographical work designed to meet exactly the type of needs which may occur for users of a type similar to themselves and in the same type of extra-lexicographical situation in which they find themselves. Of course, the precondition for a function-specific consultation is that the lexicographers have analysed and decided on the functions to be displayed by the dictionary and passed this information to the interested users.

Research into lexicographically irrelevant dictionary usage is, by its very nature, irrelevant. As a rule, the same holds true when dictionaries are not used as reference works. Both function-specific and non-function-specific dictionary consultations are obvious topics for lexicographical user research as it also appears from the previously published research projects. But it would be a problem if no clear distinction were made between research into the two types of consultation because this may lead to deceptive and contradictory results. Dictionary consultation takes place when users with a specific type of need occurring in a specific type of extra-lexicographical situation think that this type of need can be satisfied by consulting a dictionary and therefore take action in this direction. If this type of users consult dictionaries specifically designed to provide assistance in the respective situations, then they are more likely to have their needs met than if they used dictionaries not designed to provide this specific assistance.

Here it is also necessary to distinguish between two entirely different situations relevant to lexicographical user research, i.e. the *user situation* which is an extra- or pre-lexicographical situation where the need to consult a dictionary occurs for a *potential user*, and the *usage situation* where the user, now turned into an *actual user*, takes action to satisfy his/her need by consulting a dictionary or another lexicographical tool. Without a clear distinction between these two completely different types of situations, there is a considerable risk of obtaining deceptive and defective results. Research into the actual usage situation may, if carried out according to scientific standards, lead to reliable information about this type of situation, whereas — as will be argued — it can only

provide suppositions and vague ideas of the problems and needs occurring in the preceding extra-lexicographical user situation. In this respect, it is important to underline that for research into dictionary usage to be relevant, it should not only generate knowledge of *how* dictionaries are used, but also of *who* the users are, *where*, *when* and *why* they use dictionaries, and with *which* result. Hence, it is necessary to do research into:

- (a) the types of *user situations*,
- (b) the types of *users*,
- (c) the types of *user needs*,
- (d) the users' *usage* of a dictionary, and
- (e) the degree of *satisfaction* of the user needs.

These five categories are interrelated. Hence, without knowing a user's needs, it makes no sense to investigate the degree to which these needs have been satisfied. And the same applies to dictionary usage where it is necessary not only to know these needs but also the user's general experience in dictionary usage to draw relevant conclusions. Analogously, and as it will be discussed in the following sections, it makes no sense to speak about user needs if these needs are viewed in an abstract way without relating them to specific types of users and situations.

The types of relevant user situations are the *communicative* ones (production, reception, translation, text revision and marking) and the *cognitive* ones (systematic and sporadic), to which can also be added the *operative* ones covered by how-tos, handbooks and manuals (cf. Tarp 2007). However, it is far more complicated to establish a user typology because the criteria for such a typology until now discussed constitute an open list and vary from dictionary to dictionary (cf. Tarp 2008: 54-56). A typology of users depends both on the *user situation* — i.e. which types of users will find themselves in such a situation — and on their need for *differentiated lexicographical solutions*. For instance, if learners at a beginner's level need assistance to produce texts in a foreign language, they will, as a rule, need a bilingual solution; whereas advanced learners who think and express themselves directly in a foreign language will frequently manage with a monolingual solution in this language. This implies that it could be relevant to investigate or at least confirm the relevant criteria for establishing user typologies by means of user research.

3. User needs

The potential and actual dictionary users' lexicographical needs constitute a subset of the needs which human beings have in general. With a view to specifying them, it is therefore necessary to comment upon a number of problems related to human needs in general. To do so, the discussion in this section will take its starting point in the following oppositions:

- natural versus historical-cultural needs,
- recognised versus non-recognised needs,
- objective versus subjective needs, and
- genuine versus artificial needs.

Natural versus historical-cultural needs. All living creatures have a number of natural needs that have to be met, so that they can survive and propagate their species. Human beings are no exception. Their natural needs are related to a previous life in a state of nature and include, among others, air, water, minerals, proteins, fatty acids and other ingredients in their food. However, they are no longer living in a wild state, but in complex societies. Concurrent with this transition from natural state to social life, new types of needs of a historical-cultural character emerged and are continuously changing. In this process, the natural needs redress in historical-cultural forms. Human beings still need food, but the historical and cultural forms, which the food now adapts are radically different from previous times. While the natural needs are stable for long periods of time, the historical-cultural needs have changed dramatically during the last thousands of years revealing a clear tendency to accumulate from generation to generation in spite of the fact that human genes have almost not changed at all during the last 30 000 years (cf. Jensen 2007). This prodigious development cannot be explained genetically. Acquired characteristics are inherited from generation to generation, not genetically, but through social life, culture and communication. In this respect, Hobsbawn (2004) speaks of Lamarck's revenge on Darwin:

The changes in human life, collective and individual, in the course of the past 10,000 years, let alone in the past 10 generations, are too great to be explained by a wholly Darwinian mechanism of evolution via genes. They amount to the accelerating inheritance of acquired characteristics by cultural and not genetic mechanisms. I suppose it is Lamarck's revenge on Darwin via human history. [...] Cultural and biological inheritance don't work the same way.

Lexicographical needs clearly belong to the historical-cultural needs. In a historical perspective, according to Al-Kasimi (1977: 1), dictionaries have emerged as "practical tools" designed to satisfy specific needs observed within different cultures:

The major motives behind the rise of lexicography differ from one culture to another. Each culture fosters the development of dictionaries appropriate to its characteristic demands.

Like other similar needs, lexicographical needs will change in the course of time as a function of more profound changes in social life and culture. However, it is worth noting that their satisfaction may simultaneously influence and have a positive effect on the cultural and social development owing to the fact that dictionaries and other lexicographical tools are artefacts contributing to transfer acquired characteristics not only from one generation to another, but

also among the members of one and the same generation. Consequently, research into lexicographical needs must take into account that these needs are not static but continuously changing and developing. For instance, the needs of "the Assyrians who came to Babylonia about three thousand years ago" (Al-Kasimi 1977: 1) are surely not the same as the needs of the people living in modern South Africa. In a similar way, while it makes no sense investigating the usage of electronic dictionaries only 30 years ago, this situation has changed completely today.

Recognised versus non-recognised needs. The above example concerning food indicates that it is necessary to distinguish between recognised and non-recognised needs. Few people ignore their need for food, but this does not mean that they also recognise that the need for food includes a need for nutrients like minerals, vitamins and proteins. Such recognition is primarily a product of history and the development of science. What holds true for natural needs is to a great extent also valid for historical-cultural needs. In this regard, a person may, in an extra-lexicographical situation, have a lexicographically relevant need which he/she does not recognise and therefore does not try to solve consulting a dictionary, although dictionaries designed for this specific type of need may already exist. There may also be other lexicographically relevant needs which lexicographical theory has still not recognised although they could easily be met by the conception and production of the right dictionaries or other lexicographical tools. It could therefore be relevant to lexicographical user research to try to discover and describe both the recognised and the non-recognised needs. Analogously, there could be cases where a potential user actually recognises a need, but does not take lexicographical action because he/she does not think that it can be satisfied by means of dictionary consultation. In this respect, the recognised needs express themselves as needs that both trigger and do not trigger a lexicographical consultation, whereas the non-recognised needs never lead to any such consultation. As mentioned above, this indicates that it is not enough to do research on the actual usage situation when the purpose is to reveal the lexicographically relevant user needs.

Objective versus subjective needs. As a continuation of the previous opposition, it is also necessary to distinguish between the actual needs recognised by the potential users and the way these needs are reflected in their brains. Here it makes sense to consider an opposition or dialectical relation between the objective needs, i.e. the actual, genuine needs, and the subjective needs representing the reflection of these needs in the consciousness. This opposition or dialectical relation between objective and subjective needs is the result of a complex cognitive process including a number of pitfalls. Sometimes, the subjective and the objective needs may correspond, and in these cases the potential users are fully conscious of their genuine needs. But the users may frequently only have a vague or approximate idea of the objective needs; they may know that they need something, but not exactly what it is. In such cases, it makes no

sense to ask them about their real needs because the answers will be inaccurate and unreliable. And as the researchers will never know when the subjectively recognised and the objective needs correspond completely, it can be argued that user research exclusively built on the informants' own answers or their usage of dictionaries will have only little scientific value if the purpose is to discover the genuine or objective needs occurring before the consultation process, i.e. extra-lexicographically.

Genuine versus artificial needs. Publicity as a scientific discipline was invented in the U.S.A. about a hundred years ago when the American industry needed a bigger home market as the basis for its international expansion. In the beginning, the purpose was to defeat the resistance among the many Christian sects who defended a modest and Spartan life and considered any kind of extravagance and wastefulness to be the work of the Devil. This was the starting point for an industry trying not only to promote genuine, but unsatisfied needs, but also to create new and sometimes completely artificial needs. Although a complex dialectical relation exists between genuine needs and artificial, publicity-created needs, because the latter may sometimes, in the course of time, be transformed into genuine historical-cultural needs, artificial needs should nevertheless be considered as subjective needs for they do not, by definition, correspond to objective needs. The commercial houses publishing the big lexicographical works constitute no exception from the rule when it comes to the use of publicity and the priority of profits over the satisfaction of genuine human needs. In this respect, Hausmann (1989) speaks of a gulf between market and science. This opposition has big consequences for the planning of lexicographical user research. If such research projects have commercial purposes, they will typically try to reveal the subjective needs which the potential users (and buyers) themselves think they have, because this information may be helpful to expand the market capacity of future dictionaries. On the other hand, if the purpose of the research projects is to generate new scientific knowledge, then they ought to focus more on the discovery of genuine, objective needs. It is important to have a clear vision of the differences between the two types of research projects, both when planning a new research project and when comparing the results of a project already completed.

4. Lexicographical needs

The discussion in the previous section showed that lexicography as a scientific discipline first of all regards lexicographical needs as objective, historical-cultural needs, no matter whether these needs are recognised by the potential dictionary users, whereas commercial lexicography mainly focuses on the subjective needs that express themselves in the market. Of course, this does not mean that science-based lexicography is not interested in the subjective needs in as much as the distance separating the objective needs from the ways these are

expressed subjectively in the potential users' brain frames, the area where lexicographical education and training of the potential and actual users are needed.

However, lexicographical needs are much more. The most important point is that they are not abstract needs in general, but concrete, specific types of needs related to a specific type of user being in a specific type of extra-lexicographical situation as described above. It is evident that many different needs may occur in such a situation. For instance, the potential user may need new glasses when reading or a computer when writing. What separates lexicographical needs from these types of needs is that they can be met by consulting a dictionary or another lexicographical tool, no matter if this already exists or still has to be conceived. It is a matter of course that the various types of users may have very different lexicographical needs in the various types of situations, but common to all of them are that they are needs for information (cf. Tarp 2008: 56-58). However, dictionaries do not contain information, but only lexicographical data, which have been made accessible through structures in printed dictionaries and by means of links and search machines in electronic ones, and from which users with specific characteristics, through a complex mental process, may retrieve exactly the information required to meet their lexicographical needs (cf. Wiegand 1998, 2000, 2002).

In this regard, it is also relevant to distinguish between two basically different types of lexicographical needs, i.e. on the one hand, the so-called primary, *function-related needs* which are the objective needs occurring in an extra-lexicographical situation, and, on the other hand, the secondary, *usage-related needs* which only occur during the consultation process when the users need help to find and interpret the relevant lexicographical data. Obviously these two completely different types of user needs should not be confused when preparing and carrying out a user research project.

5. Research methods

Within lexicography, various methods have been used to conduct empirical research into users' needs and dictionary usage. The following methods are treated in lexicographical literature:

- questionnaires,
- interviews,
- observation,
- protocols,
- experiments,
- tests, and
- log files.

Of all these methods, questionnaires are the most commonly used, whereas log files constitute a relatively new method made possible by the introduction of

computers and electronic dictionaries. With the possible exception of log files, all these methods are well-known within social science, and there is a comprehensive body of literature discussing and evaluating them from a more theoretical angle. These methods and their biggest advantages and disadvantages will be briefly described and discussed below with references to both lexicographical literature and social science.

Questionnaires. As already mentioned, surveys by means of questionnaires are the most commonly used method to investigate the usage of dictionaries. There are two main types: the one with open and the other with closed questions. In the first type, the respondents or informants have to give the answers within certain predetermined categories, whereas in the second type, the respondents or informants may add categories of answers other than the predetermined ones. The advantage of questionnaires is mainly that it is possible to ask a large number of respondents and that it is relatively easy to analyse the answers, especially in the case of closed questions where no subsequent coding is necessary. But there are also serious disadvantages. This is illustrated by Welker (2006: 23) who classifies the questions typical in questionnaires into three main types:

- (a) about facts that can be easily remembered by the respondents (for instance: How many dictionaries do they have? When did they buy them? Why did they buy them?),
- (b) about the usage, and
- (c) about the user's opinion (for instance: Are you satisfied with the dictionaries? What type do you prefer? What improvements would you like?).

Welker (2006: 23) himself believes that the answers to questions of type (a) are reasonably reliable, whereas the answers to questions of type (c) are in all cases subjective "although they may guide lexicographers and publishers to a certain extent". This confidence in retrospective questions dealing with the past and requiring a good memory is not shared by sociologists such as Hansen and Andersen (2000: 146) who comment (my translation — S.T.):

Retrospective questions demanding a lot from the respondents' memory may [...] cause reliability problems. [...] However, this does not necessarily mean that such questions should be avoided when this type of reliability problems can be anticipated, but they should at least be taken into consideration when the results of the measuring are to be interpreted.

With this reservation in terms of retrospective questions, it is quite possible to agree with Welker (2006: 23), especially when he provides the following examples:

- What are you looking for in the dictionary?
- Under which conditions do you consult it most frequently?

- In which percentage of the look-ups does the dictionary provide assistance?
- Which part of the information is most useful?

Welker rightly comments that these questions, when put after the usage of the dictionary, only reveal the users' perception of their consultation, not the real usage. This scepticism corresponds to the classic criticism of questionnaires by Hatherall (1984: 184):

Are subjects saying here what they do, or what they think they do, or what they think they ought to do, or indeed a mixture of all three? Do they all define the categories in the same way — and in the same way as the researcher? When all is said and done, do we not, on this basis, arrive at a consensus on how subjects are likely to behave when faced with a particular questionnaire, rather than authentic data on what they use the dictionary for? [...] I conclude that, whatever the difficulties, the only reliable method of collecting data on dictionary user behaviour is by direct observation.

Although Hatherall's critical comments have often been quoted in lexicographical literature, many lexicographers still carry out user research by means of questionnaires, arriving at conclusions which even a modest sociological knowledge would show to have no scientific warranty.

Interviews. Within lexicography, it is normal to distinguish between questionnaires and interviews as two different methods to collect data. In contrast, within sociology interviews are frequently regarded as a special form of questionnaire (cf. Hansen and Andersen 2000: 98). However, it seems reasonable to maintain the distinction between questionnaires and interviews because lexicography does not only use the latter to conduct quantitative research where the questions are the same as in questionnaires, but also to carry out qualitative research where the respondents may speak their mind and where the interviewers do not base themselves on predetermined questions but, for instance, on an action pattern observed among the respondents. In this respect, Welker (2006: 26) speaks of "open interviews" in opposition to interviews undertaken by means of questionnaires. Zikmund (1997: 122) calls it an "in-depth interview" which he characterises as a "relatively unstructured, extensive interview".

Interviews have a number of clear advantages compared to questionnaires distributed to respondents to be answered at home, at work, or in a special room. For instance, the respondents cannot cheat the interviewer by consulting dictionaries in the course of the interview. Likewise, the interviewer may explain the meaning of questions not understood by the respondents, e.g. containing linguistic or lexicographical terms. On the other hand, interviews suffer from the disadvantage that they do not solve the dilemma mentioned by Hatherall (1984), i.e. whether respondents say what they do, or what they think they do, or what they think they ought to do, or a mixture of all three. Further-

more, the very presence of the interviewer or his/her way of asking may influence the informants and the answers given. Finally, interviews are time-consuming and therefore frequently expensive to undertake. This makes it more difficult to reach the number of informants necessary to be representative of the given population, i.e. the group of potential and actual users, and to provide statistically valid answers. Of course, this is especially a problem related to quantitative research, whereas other criteria are valid for in-depth interviews with regard to qualitative research where they are combined with other methods such as observation, protocols and log files.

Observation. Wiegand (1998) distinguishes between several types of observation with regard to dictionary usage. Hence, the observation may be open or hidden depending on whether the informant can see the observer. Likewise, the latter may participate actively with advice and instructions, or remain passive. The observation may be carried out directly in relation to the usage of the dictionary, or afterwards when the consultation process is videotaped. Furthermore, the observation may take place in a research laboratory or at the informant's normal work site. And finally, the results of the observation may be written down in a structured manner, or freely and unstructured (cf. Wiegand 1998: 570-583).

As seen above, Hatherall recommended observation as the most reliable method of collecting data on dictionary usage. There is no doubt that a well-trained and well-prepared observer may collect useful data through this method, and that the informant's chances to cheat are strongly reduced. But like interviews, observation is very time-consuming and expensive because it requires the presence of an observer who may also disturb the informant if it is a case of open observation as Wiegand calls it. Hatherall (1984: 184) is very much aware of this limitation:

Ideally, [...] the researcher would actually watch the users in action. But this, too, causes problems. Under such conditions it would probably be difficult for the subjects to behave normally as users. Also, it is unlikely that all the information the researcher needs would be retrievable via the visual medium. And finally, such an exercise is so time-consuming that the sample is likely to remain unrepresentatively small.

But apart from this legitimate criticism, it is also a big problem that observation only serves to study "the external aspects of the usage action", i.e. what happens, but not why it happens and with what result (cf. Wiegand 1998: 974). On the other hand, observation as method displays clear advantages compared to the other types of surveys discussed above. In this respect, Zikmund (1997: 265) writes:

The major advantage of observation studies over surveys, which obtain self-reported data from respondents, is that the data do not have distortions, inaccuracies, or other response biases due to memory error, social desirability, and so on. The data are recorded when the actual behaviour takes place.

Protocols. Wiegand (1998: 974) considers dictionary protocols to be a genuine metalexicographical method which embraces not only the external aspect of dictionary usage, but also the internal aspects as well as the "preceding and subsequent context". In this regard, he classifies protocols into two main types: written protocols and oral protocols. Written protocols are produced by the informants themselves, either during or after the dictionary consultation. These protocols may be verified or non-verified according to the presence of somebody to verify that the protocols actually reflect the consultation process. In addition to this, Wiegand distinguishes between structured, non-structured and semi-structured protocols depending on whether the informants have to embody their data in formulas with prepared fields, completely freely, or as a mixture of these two options.

Oral protocols are produced by means of the "think-aloud" method. The informants are invited to freely express which reflections and problems they have during the consultation process. These "thoughts" are tape-recorded and subsequently transcribed and written down in protocol form. Although one cannot "think aloud" as Wiegand (1998) rightly comments, the "think-aloud" method is nevertheless an excellent and highly appreciated method to go beyond the external aspects of the dictionary consultation process. As such, it gives the researcher an idea of the users' way of working as well as what is happening during the process, what users are looking for, what they think they are looking for, and which problems they face when trying to find and interpret the relevant data. A number of research projects performed with this method have provided valuable results, among others Wingate (2002) who did research into the usefulness of various types of definitions in learners' dictionaries, and Thumb (2004) who focused on the users' different look-up strategies and the problems they faced during the process.

A disadvantage of the protocol method — and especially the oral protocols that require further processing — is that it is very time-consuming and therefore frequently only involves a reduced number of informants. Moreover, it does not reveal the users' objective needs, but only the subjective needs they themselves believe to have.

Experiments. The purpose of an experiment is to see how the introduction of a certain factor influences the result. This factor may be teaching in dictionary usage, the introduction of a new type of definition, specific grammatical data or a group of such data, a different structuring of the dictionary article or of the dictionary as a whole, the introduction of new access routes in electronic dictionaries, etc. Only very few experiments of this kind have been conducted with regard to lexicography. Tono is probably the researcher who has worked most exhaustively with lexicographical experiments. He (Tono 2001: 70-72) classifies experiments in three different types:

- (a) a pre-experimental design: the one group pretest–posttest,

- (b) a quasi-experimental design: the non-equivalent control group design, and
- (c) a true experimental design: the pretest–posttest group design.

In the pre-experiment, a single group is tested in order to measure one variable. Then a new factor is introduced (e.g. teaching in dictionary usage), and a new test carried out to see how this factor influences the result. The problem related to this type of pre-experimental investigation is that it is not possible to control the external factors which may invalidate the result.

In the quasi-experiment, two different groups are tested as a start. Then a new test is performed where a new factor is introduced to one of the groups, the test group, whereas this factor is not introduced into the other group. When the two results are compared, it is easier to see the effect caused by the introduction of the new factor.

The true experiment is in many ways similar to the quasi-experiment but with the important difference that the two groups of test persons are selected at random. Tono (2001: 71) argues as follows:

In theory [...] selection controls all possible independent variables. In practice, of course, it is only when enough subjects are included in the experiment that the principle of randomisation has a chance to operate as a powerful control.

According to Zöfgen (1994: 50), most psychologists consider experiments to be the "royal road" of empirical research. The advantage of experiments is that it is possible to measure how the introduction of certain types of data or groups of such data in dictionaries and other lexicographical tools influence the result of the consultation by various types of users. A disadvantage is that they are very time-consuming and require a large number of informants to be representative, and that it is necessary to have knowledge of statistic methods or collaborate with statisticians with a view to determining whether the research results are statistically significant.

Tests. A test in the lexicographical sense of the word is a method to evaluate to what degree the consultation of a dictionary or other lexicographical tool can help users to satisfy their needs. As mentioned in the previous section, tests are an integrated part of experiments whose performance does not make sense if the results are not measured. Analogously, tests may contribute to avoid or mitigate some of the shortcomings mentioned with regard to other methods. As Nesi (2000: 32) rightly argues, this requires that they are properly performed:

The validity of test findings very much depends on correct test administration, and appropriate test design.

If this advice is followed, the advantage of tests is that they may provide reliable information about the results of dictionary consultation. In addition to

this, they are relatively easy to undertake, especially if multiple choice is used. However, if this method is not used, the assessment of the test results may be subjective if only one person is involved, for which reason more than one person is needed to evaluate them, an aspect that constitutes a clear disadvantage. In contrast, the application of some types of multiple-choice tests may "affect the validity of results by providing a further context for each word, and therefore facilitating contextual guessing as an alternative to dictionary use" (Nesi (2000: 32).

Log files. The research into log files is, as mentioned above, a relatively new method made possible by the introduction of electronic dictionaries. There are basically two different types of log files. The first type is the registration of all movements on the user's computer, i.e. activation of the keyboard and use of the mouse. The second type is the registration of all transactions between the user's computer and the database where the dictionary is located. In contrast to the previous type, the registration is made in the database facilitating the collection of data. From a lexicographical perspective, only research of the second type is known, but this does in no way imply that the first method is irrelevant to lexicography.

Although much research has been undertaken into the usage of electronic dictionaries, most of the research projects have been carried out by means of "traditional" methods. Only a few have been based on log files, among them De Schryver and Joffe (2004), De Schryver et al. (2006), Bergenholtz and Johnsen (2005, 2007) and Almind (2008). User research by means of log files is a new method and it is therefore reasonable to believe that all its potentialities have not by far been explored, also because electronic dictionaries themselves are rapidly developing and changing. However, based on the state of the art it is nevertheless possible to evaluate this method and its results, advantages and disadvantages so far.

One advantage of using log files is that they constitute a type of observation which does not influence the user and interfere in the lexicographical consultation process. It also provides easy access to a large amount of data representing the whole population of actual users, from which reliable information about consultation can be retrieved using the proper methods. In addition to this, the processing of the collected data is relatively easy in terms of quantitative research. The collected data may also be used for qualitative purposes, for instance to reveal the individual user's look-up pattern, although this type of research is very time-consuming (cf. Almind 2008). Finally, log files may also be used to obtain information on aspects related to the user typology and user situation, especially when the users are invited to define themselves and their extra-lexicographical situation before accessing the dictionary as such (cf. Bergenholtz and Johnsen 2007).

However, the use of log files as a lexicographical research method also suffers from a number of limitations and disadvantages. Like other forms of

observation, they only furnish data related to the external aspects of the consultation process, but no data capable of explaining why users do what they do. And neither do they provide data informing about the users' objective and subjective needs, or the results of the consultation. Hence, in order to acquire more qualified knowledge that may lead to improved lexicographical products, it is necessary to combine the analysis of the log files with one or more of the already discussed methods, for instance interviews that can follow up on the information obtained from the log files.

6. General problems

There are a number of general problems with regard to the majority of the lexicographical research projects published until now. This is especially the case with the quantitative user research which is generally characterised by the "left-hand work of the research institutions" criticised by the two sociologists Hansen and Andersen (2000: 23). As a rule, the informants have not been selected at random which is an indispensable requirement in this type of investigation. In his synoptic review of 220 lexicographical research projects, Welker (2006: 9) writes (my translation — S.T.):

The majority of the informants in the research projects were university students, among whom many studied foreign languages at the bachelor or licentiate level.

In this way, they have broken the golden rule of sociology that informants should never select themselves or be selected by the researchers, but that the selection should always be made at random. Of course, there may be cases where the whole population of dictionary users are university students and where the random sample to be investigated is necessarily made up by students. But this is the exception to the rule in the 220 research projects, of which the majority has also broken another golden rule, i.e. that the number of informants should be large enough to provide statistically significant (and relevant) results. This lenient way of taking samples implies that they are not representative of the total population, and that the results cannot be generalised. It is nevertheless a fact that the researchers behind many of these projects do not hesitate to add percentages and decimals to everything that their miniature world of informants have done, looked for, wanted, etc.

Another serious problem is the formulation of the questions used in various types of questionnaires and interviews. Also in this case, the lexicographical researchers have not followed the corresponding instructions put forward by sociology (cf. Hansen and Andersen 2000: 97-150). This is not only true of questions which are ambiguous, which the informants do not understand, or which require a good memory. It is also a matter of formulating the right questions. Even if the researchers have been meticulous with regard to data collection, analysis and validation — which is regrettably not always the case — they still run the risk of receiving not only wrong and dubious answers to the right

questions, but also "the right answer to the wrong question" (cf. Zikmund 1997: 96). These weaknesses can most probably be explained by poor preliminary work in the phase where the researchers are supposed to formulate the problem and reach clarity on which type of knowledge should be pursued and how it should be produced, as discussed by Hansen and Andersen (2000: 31) (my translation — S.T.):

The problem formulation refers to the process where the researcher, fixed upon the interest aroused by or the curiosity towards a phenomenon, chooses his research field, defines *what* he will investigate, and *which* aspects or problems within the field he will focus on. These considerations conclude with the formulation of the questions to be answered by the investigation. These questions should be formulated as precisely as possible so that they can form the basis of the evaluation of *which form of knowledge* the investigation should produce, and thus constitute the starting point for the decision on *how* this knowledge should be generated.

This preliminary work has to be done before deciding on the methods to be used in the research project and, if suitable, the formulation of a questionnaire. The aspects which previous lexicographical user research has tried to clarify can be divided into two main types. It seems that the best and most convincing results have been achieved when the research projects have focused on the usage situation and the different look-up strategies and search routes of the users (cf. Thumb 2004, and Bergenholtz and Gouws 2007). In addition to this, many projects have focused on where the users look for information and what information they look for. On the other hand, it is surprising that very few research projects endeavour to produce more knowledge on the users' needs. There is no known investigation that provides real information on the users' objective needs. And it is even more surprising that there are very few projects telling anything real and factual about the users' subjective needs, although they sometimes try to infer these needs from the generated data. However, this inference is most often built on a reconstruction of the users' needs according to the information they look for in dictionaries, i.e. an indirect determination of the user needs where all the lexicographically relevant needs not expressing themselves in any dictionary consultation are ignored. The dubious result of this circular arguing is that the users need exactly what has already been included in dictionaries. This is the same type of methodological problem mentioned by Mentrup (1984: 151) in his criticism of Wiegand. Humblé (2001: 53) has the following critical comment on this way of conducting research:

Investigating the link between needs and habits is useful only to find out how learners at present tackle the problems which current dictionaries are expected to solve. They reveal what the needs are to the extent that they are soluble by already existing means.

On the whole, it is worth considering the relevance of quantitative research to lexicography. For instance, how useful exactly is the information that such and

such a percentage of all dictionary users in such and such a percentage of their look-ups are searching for? What relevance has the information that 80 percent of all look-ups have to do with orthography and semantics? Of course, commercial lexicography would be delighted and would hasten to produce dictionaries and publicity material focusing on these types of data. On the other hand, scientific lexicography would above all be interested in knowing in which situations — e.g. reception and production — these needs may occur. Then it would set itself the task of uncovering the needs users have in the last 20 percent of the look-ups, i.e. in one out of five consultations. And it would not stop here, but would try to go even deeper into the problem in order to discover the needs that only show up in one out of a hundred or one out of a thousand consultations, or, even more rarely, in order to conceive dictionaries capable of meeting all the users' needs in specific types of situations. Where to else, if not to dictionaries, should users direct themselves when they look for assistance to satisfy lexicographically relevant needs?

Although the majority of the previous quantitative user research projects have not fulfilled their objectives, this does not mean that the generated results are useless if they are critically assimilated. For instance, Tarp (2008: 154-157) has studied the results of four of these research projects — Tomaszczyk (1989), Mackintosh (1998), Varantola (1998) and Nord (2002) — with a view to confirming one of the theses of the function theory, i.e. that translators frequently need to go directly to a dictionary in the target language instead of consulting a bilingual dictionary translating from the source language into the target language. Although the projects mentioned are all based on a very small number of informants — two of them with less than ten informants — they nevertheless provide an excess of percentages and decimals showing how often the informants are using one dictionary or another. It is evident that the small sample of informants is in no way representative of the total population, and that the results have no validity outside the narrow sphere of the projects. However, by means of a qualitative analysis of the generated data, it is nevertheless possible to conclude that the informants, in all research projects, sometimes pass directly to a dictionary in the target language. This proves that such a need — i.e. the consultation of a target language dictionary — exists with regard to translation, although it does not say anything about how frequently it may occur and if the need — as must be expected — is more pronounced with well-trained and experienced translators. However, it is worth noting that this conclusion was not the direct purpose of the research, but a deduced result that does not require a lot of percentages and decimals.

7. Alternative methods

It is evident from the above discussion that no known user research has produced real information on the objective user needs, i.e. the needs that may occur in the extra-lexicographical situation preceding the dictionary consulta-

tion. If this has to be accomplished, then it is necessary to move the focus from the dictionary usage situation to the extra-lexicographical user situation. In this respect, there are various methods, mostly with a view to obtain qualitative results. The only one applied so far is the deductive method used in the function theory and based on a complex of premises.

However, if more empirical data is needed to substantiate the function theory, other methods can also be used. Tests and interviews could be applied to investigate how much of a given text readers have understood and which reception problems they have experienced during the reading. Text revision and marking may be used to uncover non-recognised needs occurring during text production and translation. Likewise, the activation of Microsoft Word's spelling and grammar function may also reveal a number of non-recognised needs. Another option is the use of log files, eye-tracking and "think-aloud" protocols with regard to text production and translation, i.e. independent of any dictionary consultation. And finally, in-depth interviews may also be useful to clarify certain problems related to the other methods.

Except for the deductive method applied by the function theory, many of the other methods, which probably have to be combined to generate maximum results, are time-consuming and expensive to use. It may be difficult to obtain the necessary financing, and it is therefore a real question whether this is the way to be followed by lexicography with a view to projecting itself into the future.

8. Conclusion

At the beginning of this contribution, the survey expert Sheatsley (1974) was quoted for writing that "it may be no exaggeration to say that the greater number of surveys conducted today are a waste of time and money". It is difficult to judge how the general development on this front has been during the last 35 years, and which improvements and progress have been achieved. But if one focuses exclusively on lexicography, it seems that almost no qualitative progress has been made. There are of course positive aspects, but it is not difficult to reach the conclusion that the majority of the previous user research is in fact "a waste of time and money". This holds especially true for the quantitative research projects.

Future user research should learn from this experience. It should formulate clear objectives, incorporate the scientific methods used by modern social science, and do a meticulous planning. In this respect, the advantage of combining various methods within the framework of one and the same research project should be considered, as well as the degree to which quantitative research methods are at all relevant to lexicography. Finally, the research should be based on an advanced theory of lexicography capable of establishing scientific categories with regard to user needs, user typology, user situations, usage situations, access routes, etc., to be taken into account when planning and per-

forming research projects. Atkins and Rundell (2008: 4) synthesise their view of the scientific and theoretical status of lexicography:

This is not a book about 'theoretical lexicography' — for the very good reason that we do not believe that such a thing exists. But that is not to say that we pay no attention to theoretical issues. Far from it. There is an enormous body of linguistic theory which has the potential to help lexicographers to do their jobs more effectively and with greater confidence.

The deficient state of the art of lexicographical user research could perhaps be explained by the fact that many researchers still share the anti-theoretical approach defended by Atkins and Rundell and reject any attempt at establishing lexicography as an independent scientific discipline with its own concepts, methods, theory, and interdisciplinary vocation.

Literature

- Al-Kasimi, Ali M.** 1977. *Linguistics and Bilingual Dictionaries*. Leyden: E.J. Brill.
- Almind, Richard.** 2008. Søgemønstre i logfiler. *LexicoNordica* 15: 33-55.
- Atkins, B.T. Sue (Ed.).** 1998. *Using Dictionaries: Studies of Dictionary Use by Language Learners and Translators*. Lexicographica. Series Maior 88. Tübingen: Max Niemeyer.
- Atkins, B.T. Sue and Michael Rundell.** 2008. *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Bergenholtz, Henning and Rufus H. Gouws.** 2007. The Access Process in Dictionaries for Fixed Expressions. *Lexicographica. International Annual for Lexicography* 23: 237-260.
- Bergenholtz, Henning and Mia Johnsen.** 2005. Log Files as a Tool for Improving Internet Dictionaries. *Hermes. Journal of Linguistics* 34: 117-141.
- Bergenholtz, Henning and Mia Johnsen.** 2007. Log Files Can and Should Be Prepared for a Functionalistic Approach. *Lexikos* 17: 1-20.
- Bergenholtz, Henning and Sven 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, Henning and Sven Tarp.** 2003. Two Opposing Theories: On H.E. Wiegand's Recent Discovery of Lexicographic Functions. *Hermes. Journal of Linguistics* 31: 171-196.
- Bergenholtz, Henning and Sven Tarp.** 2004. The Concept of "Dictionary Usage". Dollerup, Cai (Ed.). 2004. *Worlds of Words. A Tribute to Arne Zettersten*. *Nordic Journal of English Studies* 3(1): 23-36.
- De Schryver, Gilles-Maurice and David Joffe.** 2004. On How Electronic Dictionaries are Really Used. Williams, Geoffrey and Sandra Vessier (Eds.). 2004. *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004, Lorient, France, July 6–10, 2004*: 187-196. Lorient: Faculté des Lettres et des Sciences Humaines, Université de Bretagne Sud.
- De Schryver, Gilles-Maurice, David Joffe, Pitta Joffe and Sarah Hillewaert.** 2006. Do Dictionary Users Really Look Up Frequent Words? On the Overestimation of the Value of Corpus-based Lexicography. *Lexikos* 16: 67-83.
- Dziemianko, Anna.** 2006. *User-friendliness of Verb Syntax in Pedagogical Dictionaries of English*. Lexicographica. Series Maior 130. Tübingen: Max Niemeyer.

- Hansen, Erik Jørgen and Bjarne Hjort Andersen.** 2000. *Et sociologisk værktøj. Introduktion til den koantitative metode.* Copenhagen: Hans Reitzels Forlag.
- Hartmann, Reinhard Rudolf Karl.** 2001. *Teaching and Researching Lexicography.* Essex: Pearson Education.
- Hatherall, Glyn.** 1984. Studying Dictionary Use: Some Findings and Proposals. Hartmann, R.R.K. (Ed.). 1984. *LEXeter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9–12 September 1983:* 183-189. *Lexicographica. Series Maior 1.* Tübingen: Max Niemeyer.
- Hausmann, Franz Josef.** 1989. Wörterbuchtypologie. Hausmann, Franz Josef, Oskar Reichmann, Herbert Ernst Wiegand and Ladislav Zgusta (Eds.). 1989–1991. *Wörterbücher. Ein internationales Handbuch zur Lexikographie/Dictionaries. An International Encyclopedia of Lexicography/Dictionnaires. Encyclopédie internationale de lexicographie:* 968-981. Handbücher zur Sprach- und Kommunikationswissenschaft 5.1–5.3. Berlin/New York: Walter de Gruyter.
- Hobsbawn, Eric.** 2004. Asking the Big Why Questions. History: A New Age of Reason. *Le Monde Diplomatique. English Edition.* December 2004. <http://mondediplo.com/2004/12/02why>.
- Householder, Fred W.** 1967. Summary Report. Householder, Fred W. and Sol Saporta (Eds.). *Problems in Lexicography:* 279-282. Bloomington: Indiana University.
- Humblé, Philippe.** 2001. *Dictionaries and Language Learners.* Frankfurt am Main: Haag und Herchen.
- Jensen, Peter K.A.** 2007. *Mennesket. Den genetiske arv.* Aarhus: Univers.
- Lew, Robert.** 2004. *Which Dictionary for Whom? Receptive Use of Bilingual, Monolingual and Semi-bilingual Dictionaries by Polish Learners of English.* Poznań: Motivex.
- Mackintosh, Kristen.** 1998. An Empirical Study of Dictionary Use in L2–L1 Translation. Atkins, B.T. Sue (Ed.). 1998: 121-149.
- Mentrup, Wolfgang.** 1984. Wörterbuchbenutzungssituationen — Sprachbenutzungssituationen. Anmerkungen zur Verwendung einiger Termini bei H.E. Wiegand. Besch, Werner, Klaus Hufeland, Volker Schupp and Peter Wiehl (Eds.). 1984. *Festschrift für Siegfried Grosse zum 60. Geburtstag:* 143-173. Göppingen: Kümmerle Verlag.
- Merton, Robert K.** 1968. Science and Democratic Social Structure. Merton, Robert K. 1968. *Social Theory and Social Structure:* 604-615. New York: The Free Press.
- Neubauer, Fritz.** 1987. Auf der Spur des "unbekannten Wesens" der Daf-Wörterbuchbenutzer. Zöfgen, Ekkehard (Ed.). 1987. *Wörterbücher und ihre Didaktik:* 216-235. Keimer/Zürich: Hebsacker.
- Nesi, Hilary.** 2000. *The Use and Abuse of EFL Dictionaries.* *Lexicographica. Series Maior 98.* Tübingen: Max Niemeyer.
- Nord, Britta.** 2002. *Hilfsmittel beim Übersetzen. Eine empirische Studie zum Rechercheverhalten professioneller Übersetzer.* Frankfurt am Main: Peter Lang.
- Ripfel, Martha and Herbert Ernst Wiegand.** 1988. Wörterbuchbenutzungsforschung. Ein kritischer Bericht. Wiegand, Herbert Ernst (Ed.). 1988. *Studien zur neuhochdeutschen Lexikographie VI. 2. Teilband. Germanische Linguistik 87-90:* 491-520. Hildesheim/Zürich/New York: Georg Olms.
- Sheatsley, Paul Baker.** 1974. Survey Design. Ferber, Robert (Ed.). 1974. *Handbook of Marketing Research:* 2-66. New York: McGraw-Hill.
- Tarp, Sven.** 2007. Lexicography in the Information Age. *Lexikos 17:* 170-179.
- Tarp, Sven.** 2008. *Lexicography in the Borderland between Knowledge and Non-Knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography.* *Lexicographica. Series Maior 134.* Tübingen: Max Niemeyer.

- Thumb, Jenny.** 2004. *Dictionary Look-up Strategies and the Bilingualised Learner's Dictionary*. Lexicographica. Series Maior 117. Tübingen: Max Niemeyer.
- Tomaszczyk, Jerzy.** 1989. L1-L2 Technical Translation and Dictionaries. Pöhl, E. and M. Snell-Hornby (Eds.). 1989. *Translation and Lexicography. Papers Read at the Euralex Colloquium Held at Innsbruck, 2-5 July 1987*: 177-186. Amsterdam: John Benjamins.
- Tono, Yukio.** 1986. A Scientific Approach toward Lexicography. *LEO. Journal of the Linguistic, Literary and Educational Organization* 15: 37-53.
- Tono, Yukio.** 2001. *Research on Dictionary Use in the Context of Foreign Language Learning. Focus on Reading Comprehension*. Lexicographica. Series Maior 106. Tübingen: Max Niemeyer.
- Varantola, Krista.** 1998. Translators and their Use of Dictionaries. Atkins, B.T. Sue (Ed.). 1998: 179-192.
- Welker, Herbert Andreas.** 2006. *O uso de dicionários: Panorama geral das pesquisas empíricas*. Brasília: Thesaurus.
- Wiegand, Herbert Ernst.** 1977. Nachdenken über Wörterbücher. Aktuelle Probleme. Drosdowski, Günther, Helmut Henne and Herbert E. Wiegand. 1977. *Nachdenken über Wörterbücher*: 51-102. Mannheim/Vienna/Zürich: Bibliographisches Institut.
- Wiegand, Herbert Ernst.** 1987. Zur handlungstheoretischen Grundlegung der Wörterbenutzungs-forschung. *Lexicographica. International Annual for Lexicography* 3: 178-227.
- Wiegand, Herbert Ernst.** 1998. *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie. 1. Teilband*. Berlin/New York: Walter de Gruyter.
- Wiegand, Herbert Ernst.** 2000. Wissen, Wissenrepräsentationen und Printwörterbücher. Heid, Ulrich, Stefan Evert, Egbert Lehmann and Christian Rohrer (Eds.). 2000. *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000. Stuttgart, Germany, August 8-12, 2000*: 15-38. Stuttgart: Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.
- Wiegand, Herbert Ernst.** 2002. Wissen in der Sprachlexikografie. Ein Plädoyer für einige immer noch notwendige Differenzierungen. Ezawa, Kennosuke, Wilfried Kürschner, Karl H. Rensch and Manfred Ringmacher (Eds.). 2002. *Linguistik jenseits des Strukturalismus. Akten des II. Ost-West-Kolloquiums Berlin 1998*: 265-281. Tübingen: Gunter Narr.
- Wingate, Ursula.** 2002. *The Effectiveness of Different Learners Dictionaries: An Investigation into the Use of Dictionaries for Reading Comprehension by Intermediate Learners of German*. Lexicographica. Series Maior 112. Tübingen: Max Niemeyer.
- Zikmund, William G.** 1997. *Business Research Methods*. Fort Worth: The Dryden Press/London: Harcourt Brace College Publishers.
- Zöfgen, Ekkehard.** 1994. *Lernerwörterbücher in Theorie und Praxis. Ein Beitrag zur Metalexikographie mit besonderer Berücksichtigung des Französischen*. Lexicographica. Series Maior 59. Tübingen: Max Niemeyer.

Die opvoedkundige waarde van woordeboeke: Voorstelle vir woordeboekonderrig in Suid-Afrika*

Michele F. van der Merwe, *Fakulteit van Onderwys en Sosiale Wetenskappe, Kaapse Skiereiland Universiteit van Tegnologie, Wellington-kampus, Wellington, Republiek van Suid-Afrika (vandermerwem@cput.ac.za)*

Opsomming: Die idee van woordeboekonderrig in die laerskool het nog nie baie aandag in die opvoedkunde en leksikografie in Suid-Afrika getrek nie. Tans word uitkomsgebaseerde onderrig as model in Suid-Afrikaanse skole gebruik en woordeboekonderrig kan baie goed binne dié opset geakkommodeer word. In die artikel word probeer om woordeboekonderrig te definieer en aangetoon watter uitkomst in verband met woordeboekgebruik deur leerders in die intermediêre fase bereik behoort te word. 'n Model om die prosesse van woordeboekonderrig te illustreer, word aangebied. Praktiese voorbeelde van toepaslike woordeboeke vir gebruik in die klaskamer word bespreek. Die verwagting word gestel dat sowel opvoeders as ouers bewus moet word van die voordele van woordeboekgebruik en leerders moet aanmoedig om woordeboeke te gebruik.

Sleutelwoorde: WOORDEBOEKONDERRIG, OPVOEDKUNDE, LEKSIKOGRAFIE, UITKOMSGEBASEERDE ONDERRIG, KURRIKULUMVERKLARING, TAALHANDBOEK, WOORDEBOEKGEBRUIK, TAALVAARDIGHEID, TAALONDERRIG, HULPMIDDEL, WOORDEBOEKVAARDIGHEDE, WOORDEBOEKKULTUUR, NASLAANVAARDIGHEDE, PRIMÊRE NASLAANVAARDIGHEDE, SEKONDÊRE NASLAANVAARDIGHEDE, WOORDEBOEFUNKSIES, WOORDEBOEKINLIGTING, BUITETEKSE, WOORDEBOEKUITKOMSTE, OPVOEDKUNDIGE DOEL, PEDAGOGIESE WOORDEBOEK, SKOOLWOORDEBOEK, AANLEERDERWOORDEBOEK, INTERNETWOORDEBOEK, SELFOONWOORDEBOEK, TEKSRESEPSIE, TEKSPRODUKSIE, KOMMUNIKATIEWE FUNKSIE, KOGNITIEWE FUNKSIE.

Abstract: The Pedagogical Value of Dictionaries: Suggestions for Dictionary Education in South Africa. The idea of dictionary education in primary school has not attracted much attention in education and lexicography in South Africa. At present outcomes-based education is used as model in South African schools and dictionary education can be accommodated well within this framework. In the article, it is attempted to define dictionary education and to indicate which outcomes regarding dictionary use should be achieved by learners in the intermediate phase. A model to illustrate the process of dictionary education is presented. Practical examples of the use of appropriate dictionaries in the classroom are discussed. The

* Hierdie artikel is 'n Afrikaanse bewerking van 'n lesing "The Pedagogical Purpose of Dictionaries in South Africa: A Perspective on Dictionary Education" wat aangebied is by die Dertiende Internasionale Konferensie van die African Association for Lexicography, georganiseer deur die Buro van die *Woordeboek van die Afrikaanse Taal*, Stellenbosch, Republiek van Suid-Afrika, 30 Junie–2 Julie 2008.

expectations are set that educators as well as parents should become aware of the advantages of dictionary use and motivate learners to use dictionaries.

Keywords: DICTIONARY EDUCATION, EDUCATION, LEXICOGRAPHY, OUTCOMES-BASED EDUCATION, CURRICULUM STATEMENT, LANGUAGE TEXT BOOK, DICTIONARY USE, LANGUAGE PROFICIENCY, LANGUAGE EDUCATION, TOOL, DICTIONARY SKILLS, DICTIONARY CULTURE, REFERENCE SKILLS, PRIMARY REFERENCE SKILLS, SECONDARY REFERENCE SKILLS, DICTIONARY FUNCTIONS, DICTIONARY INFORMATION, OUTER TEXT, DICTIONARY OUTCOMES, PEDAGOGICAL PURPOSE, PEDAGOGICAL DICTIONARY, SCHOOL DICTIONARY, LEARNER'S DICTIONARY, INTERNET DICTIONARY, CELL PHONE DICTIONARY, TEXT RECEPTION, TEXT PRODUCTION, COMMUNICATIVE FUNCTION, COGNITIVE FUNCTION.

Inleiding

Tydens 'n ouervergadering by 'n laerskool noem die graad 1-opvoeder dat dit tyd geword het vir leerders om hulle eie woordeboeke te begin gebruik. Dié woordeboek wat bestaan uit 'n lys alledaagse, algemene woorde waarmee leerders in die grondslagfase te doen kry, kan gebruik word om dié woorde se spelling na te slaan. Die leerders kan ook hulle eie woorde waarmee hulle spelprobleme ondervind, met die hulp van die opvoeder tot die woordeboek byvoeg. Al begin die leerders die woordeboek in graad 1 saamstel, kan hulle dit deur die hele grondslagfase gebruik en bywerk. Een van die ouers bevraagteken toe die gebruik van 'n woordeboek omdat dit leerders sou aanmoedig om nie die spelling van woorde te leer as hulle dit so maklik in die woordeboek kan naslaan nie! Die vraag kan gestel word of dié kommentaar tekenend is van baie ouers, en ook die Suid-Afrikaanse samelewing se houding teenoor woordeboeke weerspieël. Koen (2008: 28) beweer tereg dat leerders so vroeg moontlik, reeds tydens die grondslagfase, aan die gebruik van 'n woordeboek bekendgestel moet word. Die aanleer van woordeboekvaardighede word as 'n integrale deel van die ontwikkeling van leeselemente beskou. Koen (2008: 8) doen aan die hand dat 'n eenvoudige woordeboek deur opvoeders saamgestel word sodat leerders die alfabetiese stelsel kan aanleer. Sy definieer woordeboekvaardighede as die vermoë van leerders om letters en/of woorde te alfabetiseer, 'n woord in 'n woordeboek na te slaan en om gidswoorde te gebruik.

Die doel van hierdie artikel is om die opvoedkundige waarde van woordeboeke te ondersoek en om opvoeders, opvoedingsinstansies, asook woordeboekuitgewers daarvan bewus te maak. Die volgende aspekte word in hierdie artikel onder die loep geneem:

- Wat is woordeboekonderrig?
- Die status van woordeboekonderrig in Suid-Afrika
- Wat is die voordele van woordeboekonderrig?
- Die proses van woordeboekonderrig

- Praktiese voorbeelde van woordeboekonderrig
- Die rol van opvoeders in woordeboekonderrig

Die fokus van die artikel is op die onderrig van woordeboeknaslaanvaardighede in die laerskool. Voorvereistes hiervoor is kennis van die woordeboektipologie, die besef van die belangrikheid van naslaanvaardighede en die kennis van praktiese onderrigsituasies in die klaskamer. Maar het opvoeders die kennis, vaardighede en materiaal om woordeboekonderrig aan te bied? Taalhandboeke vir die laerskool voorsien nie inligting oor woordeboeknaslaanvaardighede nie en noem nie strategieë om woordeboekonderrig met taalkunde te integreer nie. Dié situasie kan die persepsie onder opvoeders laat ontstaan dat woordeboekonderrig nie so belangrik is soos taalverwerwing en die taalkunde-program nie. Voorstelle word gedoen vir praktiese, uitkomstgebaseerde maniere om woordeboekonderrig aan te bied, asook voorbeelde van woordeboeke om te gebruik.

Wat is woordeboekonderrig?

Tydens 'n taalonderrigprogram word 'n leerder se taalvaardigheid ten opsigte van praat, lees, skryf en luister verbeter. 'n Woordeboek is een van die hulpmiddels beskikbaar aan die taalopvoeder om leerders te help om hulleself te kan help (Hartmann 1987: 181). Béjoint (1989: 212) wys daarop dat leerders wat weet hoe om woordeboeke te gebruik, beter in taaltake presteer as dié wat nie weet hoe om woordeboeke te gebruik nie. Volgens Herbst en Stein (1987: 116) is woordeboeke naslaanwerke, en soos in die geval van enige naslaansisteem, moet gebruikers sekere naslaanvaardighede kan toepas. Naslaanvaardighede kan deur volgehoue oefening aangeleer word en op skool met direkte leer deur opvoeders ingeskerp word. Naslaanvaardighede ontwikkel dus nie natuurlik nie, maar moet aangeleer word.

Die onderrig van naslaanvaardighede ten opsigte van woordeboeke maak 'n integrale deel van woordeboekonderrig uit. Gouws (2006: 79) wys daarop dat daar veral ses strukture van kernbelang is vir die suksesvolle en optimale onttrekking van inligting uit 'n skoolwoordeboek. Hierdie strukture is naamlik die dataverspreidingstruktuur, die toegangstruktuur, die makrostruktuur, die mikrostruktuur, die adresseringstruktuur en die mediostruktuur. Indien die genoemde strukture in ag geneem word, kan twee vlakke van naslaanvaardighede onderskei word, naamlik primêre en sekondêre naslaanvaardighede. Dit behoort aan leerders tydens woordeboekonderrig geleer te word. Primêre naslaanvaardighede verwys na die algemene toegang tot 'n woordeboek, dit wil sê kennis en gebruik van die toegangstruktuur (alfabetiese struktuur in die meeste gedrukte skoolwoordeboeke), of die toepas van algemene soekprosedures in aanlynwoordeboeke. Vir jonger leerders in die intermediêre fase vir wie selfs die gebruik van 'n alfabetiese toegangstruktuur problematies kan wees, moet opvoeders kennis van die alfabetvolgorde doelgerig en planmatig oordra. Dit sluit kennis tot die derde alfabetlettervolgorde in.

Sekondêre naslaanvaardighede verwys na die verstaan van woordeboekspesifieke soekroetes om byvoorbeeld 'n toepaslike antwoord te vind op 'n vraag gevra deur 'n leerder of 'n probleem gestel deur 'n opvoeder. Die strukture verskil dus van woordeboek tot woordeboek. Dit sluit kennis van die dataverspreidingstruktuur, die makrostruktuur, die mikrostruktuur, die adreseringstruktuur en die mediostruktuur in. Voorbeelde van verskillende soekareas in 'n woordeboek is byvoorbeeld uitspraak-inligting, betekenisinligting, grammatikale inligting, gebruikseerinligting en etimologiese inligting. Tipes buitetekste in skoolwoordeboeke wat vir leerders van groot waarde kan wees, byvoorbeeld lyste name van die Suid-Afrikaanse provinsies, ampstale, en hoofstede, en die uiteensetting van mate en gewigte, is ook woordeboekspesifiek. Sekondêre naslaanvaardighede kan woordeboekspesifiek, asook taalspesifiek wees.

Baie leerders is onbewus van die verskillende beskikbare woordeboeksoorte en sal enige soort woordeboek gebruik om 'n vraag te probeer beantwoord. Hulle sal na enige woordeboeksoort verwys as "die woordeboek", heeltemal onbewus van die spesifieke eienskappe van so 'n woordeboek. Kipfer (1987: 46) het met behulp van 'n vraelys onder Amerikaanse senior sekondêre studente bevind dat min van hulle die verskil tussen 'n tweetalige en 'n eentalige aanleerderwoordeboek kon verduidelik, asook die verskillende groottes van 'n standaard eentalige woordeboek. Kennis van woordeboektipologie sal opvoeders en leerders help om 'n toepaslike woordeboek te kies om so doende hulle taalvrae op te los. Die aanleer van woordeboektipologie maak 'n belangrike deel van woordeboekonderrig uit.

Woordeboekonderrig kan dus beskryf word as die onderrig van primêre en sekondêre naslaanvaardighede spesifiek aan 'n woordeboeksoort, asook die onderrig van woordeboektipologie. Uitkomstebereik sluit die suksesvolle gebruik van woordeboeke deur leerders in, wat impliseer dat hulle bevredigende antwoorde op hulle vrae oor byvoorbeeld uitspraak, betekenis, grammatika, gebruikseer en etimologie kon verkry. Woordeboekonderrig behoort op 'n natuurlike wyse met betrekking tot taalprobleme en inligtingsprobleme onderrig te word, ondersteun deur die gebruik van woordeboektekste en woordeboekverbandhoudende tekste. Onderskeid kan tussen woordeboekonderrig en taalonderrig gemaak word — eersgenoemde verwys na die optimale gebruik van 'n woordeboek, terwyl laasgenoemde verwys na uitkomstebereik wat bereik moet word, byvoorbeeld in verband met korrekte taalgebruik, tydens die raadpleging van 'n woordeboek.

Die status van woordeboekonderrig in Suid-Afrika

Daar is tot dusver min navorsing oor woordeboekonderrig in Suid-Afrikaanse skole onderneem. Carstens (1995: 106) verwys na navorsing wat in Europa gedoen is en vergelyk dit met die toestand in Suid-Afrika. In 1995 was die situasie in Europese en Suid-Afrikaanse skole betreffende woordeboekonderrig min of meer dieselfde. Daar word in sillabusse gemeld dat leerders in staat

moet wees om algemene naslaanwerke te kan gebruik, maar daar word geen melding van woordeboekonderrig of die onderrig van naslaanvaardighede gemaak nie. Volgens onlangse mondelinge mededelings vind woordeboekonderrig nie formeel in die volgende Europese lande in skole plaas nie: Engeland (M. Rundell 2008), België (S. Verlinde 2009) en Denemarke (S. Tarp 2009).

Béjoint (1989: 208) wys daarop dat talle skrywers in verskillende lande en tale die behoefte aan woordeboekonderrig beklemtoon: Flaherty (1923), Mathews (1964), Beattie (1973), Croft (1973), Barone (1979), Tono (1984), Hausmann (1986), Hartmann (1988) en Kühn (1987).

Met die instel van die Nasionale Kurrikulumverklaring in 2002 (NKV 2002) het die situasie betreffende die inhoud van taalsillabusse sedert 1995 in Suid-Afrika verander. In die 1995-sillabus is daar slegs genoem dat leerders in staat moet wees om algemene naslaanwerke te gebruik. In die NKV word daar wel in die verklaring ten opsigte van tale in die intermediêre en senior fases melding gemaak van uitkomste van woordeboekgebruik, naamlik dat leerders byvoorbeeld in staat moet wees om 'n woordeboek te kan benut om spelling en betekenis te kontroleer. Die volgende assesseringstandaarde kom onder Taalstruktuur en -gebruik in die intermediêre fase voor:

Graad 4	Graad 5	Graad 6
Teken woorde in 'n persoonlike woordeboek op	Skep eie persoonlike spellys	Skep eie persoonlike spellys
Toon ontwikkelende kennis van die kompleksiteit van alfabetiese volgorde	Toon kennis van die kompleksiteit van alfabetiese volgorde waar die orde volgens die 1ste letter gegee word	Toon kennis van die kompleksiteit van alfabetiese volgorde waar die orde volgens die 2de letter gegee word
Gaan spelling en betekenis in 'n woordeboek na	Gaan spelling en betekenis in 'n woordeboek na	Gaan spelling en betekenis in 'n woordeboek of tesourus na

Woordeboeke en tesourusse word ook as aanbevole tekste in die kurrikulumverklaring vermeld. Dit is 'n verbetering op die vorige sillabus waar daar geen melding van woordeboekvaardighede gemaak is nie. Navorsing is egter nodig om vas te stel of die teorie in die praktyk toegepas word, hoe woordeboekgebruik in die klaskamer plaasvind, of woordeboekonderrig wel in die klaskamer voorkom en of taalhandboeke wat die NKV onderskryf, afdelings aan woordeboekonderrig afstaan.

Wat is die voordele van woordeboekonderrig?

Woordeboeke is waardevolle hulpmiddels in taalonderrig wat met vrag in die klaskamer aangewend kan word. Leerders kan hul taalvaardighede verbeter deur woordeboeke te ken en te gebruik. Opvoeders behoort 'n positiewe houding teenoor die gebruik van woordeboeke te kweek en leerders aan te moedig om woordeboeke aan te skaf en te raadpleeg. Verbeterde taalvaardigheid lei

noodwendig tot verbeterde kommunikasie. In 'n veeltalige land soos Suid-Afrika moet taal as inligtingsbron tussen gemeenskappe beklemtoon word en met die korrekte gebruik van veral vertalende woordeboeke kan hierdie uitkomst bereik word.

Woordeboeke kan ook as hulpmiddels vir leerders dien om hulle eie vordering en selfstandigheid te meet. Woordeboekgebruik bevorder onafhanklike denke en kweek selfvertroue gedurende die leerproses. Sodra 'n leerder 'n woordeboek onafhanklik kan gebruik en die korrekte inligting vind, weet beide leerder en opvoeder dat daar reeds 'n mate van vordering is. Indien 'n leerder byvoorbeeld in 'n redelike mate 'n tweede taal baasgeraak het, is dit vir hom/haar makliker om 'n woordeboek van die taal te raadpleeg as in die begin toe die taal aangeleer is.

Die doel van woordeboekonderrig is om leerders te leer om hulle woordeboeke op 'n gereelde grondslag te gebruik, asook vir die res van hulle lewens, om hulle sodoende vir lewenslange leer te bemagtig. Op so 'n wyse kan 'n woordeboekkultuur in Suid-Afrika gevestig word. Gouws en Prinsloo (2005: 42) wys daarop dat die meeste Suid-Afrikaners hulle in 'n prewoordeboekkultuur bevind. Volgens Klein (2007: 418) kan die ontwikkeling van 'n woordeboekkultuur ruweg in drie fases verdeel word:

- (1) lig mense in dat woordeboeke beskikbaar is;
- (2) onderrig mense om woordeboeke effektief te gebruik; en
- (3) aanvaar dat woordeboekgebruik iets normaal is, naamlik 'n alledaagse aktiwiteit wat leerders sowel by die skool as by die huis leer.

Die woordeboekonderrigscenario in Ghana is 'n goeie voorbeeld vir Afrika en Suid-Afrika om na te volg. In Ghana het elke leerder in primêre grade 4–6 (9–11 jaar) en elke student in senior sekondêre vlakke 1–3 (15–17 jaar) in Januarie 2008 'n woordeboek ontvang. Die Departement van Onderwys het 1,2 miljoen primêre en 600 000 senior sekondêre woordeboekuitgawes voorsien (J. Luther 2008, mondelinge kommunikasie). Wanneer leerders inligtingsbronne soos woordeboeke besit, kan hulle hulself beter bemagtig met kennis en die kans is beter dat hulle suksesvoller in hulle studie kan wees.

Die proses van woordeboekonderrig

Woordeboekonderrig kan volgens die skrywer as 'n proses beskou en soos volg geïllustreer word:

Identifiseer uitkomst van taalkurrikulum	Vind toepaslike woordeboek na 'n kritiese evaluering	Metode van onderrig Woordeboek- teks (insluitend tesourus) Taalteks	1. Woordeboeknaslaanvaardig- hede (primêr en sekondêr) 2. Woordeboektipologie 3. Woordeboekfunksies Uitkomst: Taalvaardighede
--	--	--	---

Die proses kan soos volg uiteengesit word: In die kurrikulumverklaring (NKV 2002) word sekere taaluitkomste genoem wat die opvoeder as uitgangspunt in die klaskamersituasie gebruik, aangesien dit die standaard is waarvolgens die leerder na onderrig geassesseer gaan word. Die opvoeder ondersoek verskillende woordeboeke op die mark en evalueer hulle krities om 'n toepaslike woordeboek te vind om in die klas situasie te gebruik. Die metode van onderrig in die klas sluit in die aanwending van woordeboektekste, taaltekste en ander tekste wat in die kurrikulum aanbeveel word (waar 'n woordeboek geraadpleeg kan word om inligting te verkry). Dit is belangrik om woordeboekonderrig op 'n natuurlike wyse te gee en om woordeboeke en taaltekste te gebruik sodat leerders taal in 'n werklike kommunikasiesituasie kan ervaar. Die proses van woordeboekonderrig kan vir die huistaal gevolg word, maar ook, met sekere aanpassings, vir die addisionele taal. Tydens woordeboekonderrig word primêre en sekondêre woordeboeknaslaanvaardighede, woordeboektipologie en woordeboekfunksies geleer. 'n Leerder wat hierdie kennis bemeester het, is bemagtig om aan die uitkomste van taalvaardighede te voldoen soos in die NKV gestel.

Behalwe die spesifieke "woordeboekuitkomste" wat in die kurrikulumverklaring genoem word, is daar ook verskeie taaluitkomste in die kurrikulum wat met behulp van 'n woordeboek bereik kan word. Opvoeders behoort daarvan bewus te wees en moet woordeboeke daarvoor in die klaskamer aanwend, asook leerders aanmoedig om tuis woordeboeke te gebruik om die uitkomste te bereik. Die volgende assesseringstandaarde kom onder Taalstruktuur en -gebruik in die intermediêre fase voor:

Graad 4	Graad 5	Graad 6
Spel dikwels gebruikte woorde korrek	Spel dikwels gebruikte woorde korrek en herken uitsonderings in spelpatrone	Spel dikwels gebruikte woorde korrek en herken uitsonderings in spelpatrone
Gebruik skryftekens doeltreffend in alledaagse woorde	Gebruik skryftekens doeltreffend in alledaagse woorde	Gebruik skryftekens doeltreffend in alledaagse woorde
Gebruik hoofletters korrek	Gebruik hoofletters korrek	Gebruik hoofletters korrek
Herken korrekte woordverdelings	Herken en gebruik korrekte woordverdelings	Herken en gebruik korrekte woordverdelings
Ontwikkel woordeskat deur woordfamilies en woorde uit dieselfde veld te herken	Ontwikkel woordeskat deur woordvorming, deur woordfamilies en woorde uit dieselfde veld te herken en deur betekenis uit die konteks af te lei	Ontwikkel woordeskat deur woordvorming, deur woordfamilies en woorde uit dieselfde veld te herken en deur betekenis uit die konteks af te lei
Ondersoek die herkoms van woorde op 'n algemene vlak (soos leenwoorde)	Begryp dat tale woorde by mekaar leen en gebruik nuutskoppings	Begryp dat tale woorde by mekaar leen en gebruik nuutskoppings doeltreffend

Gebruik alledaagse afkortings	Gebruik alledaagse afkortings en akronieme	Gebruik 'n verskeidenheid afkortings en akronieme
Verstaan en gebruik alledaagse adjektiewe: verbuiging, trappe van vergelyking, intensiewe vorme	Verstaan die funksie en gebruik van alledaagse adjektiewe: verbuiging, trappe van vergelyking, intensiewe vorme, letterlike en figuurlike betekenis	Verstaan die funksie en gebruik van alledaagse adjektiewe: verbuiging, trappe van vergelyking, intensiewe vorme, letterlike en figuurlike betekenis
Verstaan en gebruik alledaagse selfstandige naamwoorde, asook getal, geslag en verkleinwoorde	Verstaan en gebruik alledaagse selfstandige naamwoorde, asook getal, geslag en verkleinwoorde	Verstaan en gebruik alledaagse selfstandige naamwoorde, asook getal, geslag en verkleinwoorde
Gebruik alledaagse sinonieme, antonieme, homonieme en homofone	Verstaan en gebruik alledaagse sinonieme, antonieme, homonieme en homofone	Verstaan en gebruik alledaagse sinonieme, antonieme, homonieme en homofone
Gebruik die gepaste formaliteitsgraad (register)	Gebruik taal met die gepaste formaliteitsgraad (register) in minder bekende situasies en toon bewustheid van teikengroep	Gebruik taal met die gepaste register in minder bekende en meer uitdagende situasies
Gebruik eenvoudige idiomatiese uitdrukkings korrek	Toon kennis van die gebruik van bekende idiomatiese uitdrukkings	Toon kennis van die gebruik van 'n verskeidenheid idiomatiese uitdrukkings

Algemene taaluitkomst van woordeboekgebruik sluit dus in:

- Korrekte spelling
- Korrekte keuse van woorde
- Korrekte gebruik van idiome
- Korrekte uitspraak
- Korrekte gebruik van afkortings
- Uitbreiding van woordeskat
- Inligting oor die gebruik van 'n woord
- Semantiese inligting van 'n woord
- Etimologiese inligting van 'n woord
- Grammatikale inligting van 'n woord

Opvoeders, leerders en ouers moet bewus gemaak word van die groot hoeveelheid nuttige inligting in woordeboeke. Woordeboeke is houe van inligting, en kan dus effektief gebruik word om taal- en kommunikasieprobleme op

te los. Reeds in die primêre fase moet leerders bewus gemaak word van die probleemoplossende potensiaal van woordeboeke sodat woordeboeke lewenslank gebruik kan word. Sodoende kan 'n woordeboekkultuur asook die ideaal van lewenslange leer gevestig word.

Praktiese voorbeelde van woordeboekonderrig

Gouws (2006: 73) wys daarop dat, alhoewel verskillende woordeboeksoorte gerig is op behoeftes ter bereiking van 'n opvoedkundige doel, daar binne die heersende woordeboektipologie 'n kategorie bestaan wat pertinent op hierdie doel gerig is, te wete die kategorie pedagogiese woordeboeke. Hierdie woordeboekskategorie sluit twee subkategorieë in, naamlik skoolwoordeboeke en aanleerderwoordeboeke. Skoolwoordeboeke wat op skoolleerders gerig is, maak voorsiening vir sowel moedertaal- as niemoedertaalsprekers van die behandelde taal. Aanleerderwoordeboeke is primêr gerig op die behoeftes van gebruikers wat 'n nuwe taal aanleer.

Vervolgens word voorbeelde uit die *Nuwe Woordeboek Sonder Grense* en die *Oxford Afrikaans-Engels Skoolwoordeboek* bespreek. Al twee woordeboeke beskik oor werkboeke waarin leerders woordeboeknaslaanvaardighede kan aanleer en wat deur opvoeders in die klaskamersituasie gebruik kan word. (Daar is nie tans 'n toepaslike Afrikaanse skoolwoordeboek vir huistaalleerders op laerskoolvlak op die mark wat bespreek kan word nie. Die *Kernwoordeboek* word as woordeboek vir leerders bemark, maar dit is nie geskik vir laerskoolleerders nie. Vergelyk byvoorbeeld die ingewikkelde definiëring van woorde soos *rond* en *romp*, asook die opname van talle onverklaarde lemmas soos *ronddobber* en *rondswerf*.) Die *Oxford Afrikaans-Engels Skoolwoordeboek* se teikenbruikersgroep word nie duidelik in die woordeboek gespesifiseer nie, maar Gouws en Tarp (2008: 67) maak die afleiding dat die woordeboek deur skoolleerders gebruik kan word wat Engels of Afrikaans as moedertaal bemeester het of wat Engels of Afrikaans as addisionele taal aanleer. Die *Nuwe Woordeboek Sonder Grense* is 'n verklarende woordeboek ontwerp vir taalaanleerders van Afrikaans as addisionele taal. Dit ondersteun 'n handboekreeks vir graad 3 tot 12. Een van die hierbo genoemde taalkundige inligtingstipes gaan bespreek word. Grammatikale inligting wat op 'n konsekwente wyse deur die woordeboek verskaf word, korreleer met die spesifieke woordeboekverwante taaluitkomst wat in die NKV uiteengesit word, naamlik verbuiging, trappe van vergelyking by adjektiewe, asook getal, geslag en verkleinwoorde by selfstandige naamwoorde. In die *Nuwe Woordeboek Sonder Grense* word die terminologie gebruik soos deur die Departement van Onderwys voorgeskryf, naamlik *adjektief* en *selfstandige naamwoord*.

Wanneer die lemmas *moedeloos*, *moeg*, *moelik* en *mooi* in die onderstaande uittreksel uit die *Nuwe Woordeboek Sonder Grense* beskou word, kan gesien word dat die woorde as adjektiewe aangedui en die verbuiging en trappe van vergelyking telkemale verskaf word. (As 'n punt van kritiek kan genoem word dat

Mm

moed

132

moenie

moed

naamwoord ■

(-, -)

As 'n mens moed het vir iets, sien jy kans daarvoor en is jy nie bang nie.

⇒ *Die span het nie moed [courage/heart] verloor nadat hulle die wedstryd verloor het nie.* ⇒ *Die onderwyser het baie moed [hope] vir die klas, want hulle werk baie hard.*

moedeloos

adjektief ■

(moedelose; moedeloser, die moedeloosste)

Iemand wat nie kans sien of krag het om iets te doen nie, is moedeloos. ⇒ *Hy is baie moedeloos [dejected/despondent] omdat hy nie die somme regkry nie.* ⇒ *Die vrou se kinders is baie stout; daarom is sy gereeld moedeloos [dejected/discouraged].*

moeder

naamwoord ■

(moeders, moedertjie)

Moeder is die formele woord vir ma. ⇒ *Die seun se moeder [mother] het hom altyd bygestaan totdat sy oorlede is.*

kyk **ma** en **mamma**teenoor **vader****moeë**

adjektief ■

kyk **moeg****moeg**

adjektief ■

(moeë; moeër, die moegste)

① As 'n mens moeg is, het jy nie meer krag om aan te gaan nie. ⇒ *Ek is baie moeg [tired/exhausted] na die harde sokkeroefening.*

⇒ *Die seuntjie was so moeg [tired] dat hy onmiddellik aan die slaap geraak het.*

② As 'n mens nie meer lus het vir iets nie, is jy moeg daarvoor. ⇒ *Ek is moeg [tired] vir die ou rok.* ⇒ *Die onderwyseres is moeg [tired] daarvoor om goed te herhaal.*

moeilik

adjektief ■

(moeilike; moeiliker, die moeilikste)

① As iets moeilik is, sukkel jy om dit te verstaan of te doen. ⇒ *Wiskunde was nog altyd een van my moeilikste [most difficult] vakke.* ⇒ *Dit is 'n baie moeilike [difficult] onderwerp waarop ons in groepe moet gesels.*

teenoor **maklik**

② As iemand moeilik is, sukkel 'n mens om met hom/haar oor die weg te kom.

⇒ *Die skoolhoof sien gereeld moeilike [difficult] ouers en moeilike [difficult] kinders in sy kantoor.* ⇒ *My ma was baie moeilik [cross/upset] toe laat by die huis kom.*

moeilikhed

naamwoord ■

(moeilikhede, -)

① As 'n mens in die moeilikheid is, gaan jy probleme kry/hê. ⇒ *Rodney is al weer in die moeilikheid [trouble] omdat hy vrugte gesteel het.* ⇒ *Daar was groot moeilikheid [trouble] by die skool nadat daar 'n bakleiery was.*

② As iets vir 'n mens moeilikheid gee, gee dit jou baie probleme. ⇒ *My pa se ou bakkie wil omtrent nie meer ry nie en gee gereeld moeilikheid [trouble].* ⇒ *Goedkoop masjiene gee baie gou moeilikheid [trouble].*

③ As 'n mens moeilikheid het met iemand anders, kom julle nie oor die weg nie. ⇒ *Daar is altyd moeilikheid [trouble] as die twee broers bymekaar is.*

moeite

naamwoord ■

(-, -)

As 'n mens moeite doen, dan sit jy baie tyd en werk in iets. ⇒ *Die onderwyseres doen baie moeite [goes to a lot of trouble] met haar voorbereiding vir ons klasse.* ⇒ *Moenie so baie moeite [go out of your way] doen nie.* Dis net 'n informele kuier.

moenie

naamwoord ■

Moenie beteken jy mag/moet nie iets doen/sê nie. Dit is die verkorte vorm van moet nie. ⇒ *Moenie [don't] so haastig wees nie.* Ons het baie tyd. ⇒ *Moenie [don't] so raas nie.* Ek probeer slaap.

kyk **moet** ③

As jy **moenie** gebruik, moet daar nog 'n **nie** aan die einde van die sin kom. ⇒ *Moenie [don't] mors nie!* ⇒ *Moenie [don't] altyd stry as ek met jou raas nie!*

Mm

na 133 mooi

moes werkwoord ■
(verlede tyd)
kyk **moet**
⇒ *Jy moes [had to] gister al jou biblioteekboeke teruggeneem het.* ⇒ *Ek moes [should have] na my ma se raad geluister het.*

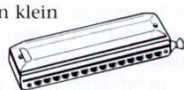
moet werkwoord ■
(moes)
① As 'n mens iets moet doen, dan het jy nie 'n keuse nie en is jy verplig om dit te doen. ⇒ *Jy moet [must] jou hande was voor jy kom eet.* ⇒ *Ek moes [had to] gister my biblioteekboeke terugneem.*
② As iets moet gebeur, is dit baie belangrik dat dit gebeur. ⇒ *Dit moet [must] nou vinnig reën, want al die damme is leeg.* ⇒ *Die vrou moes [had to] vinnig rem trap om die hond te mis.*
③ Moet word saam met nie gebruik om te wys dat iets nie mag gebeur nie. Moenie is die verkorte vorm. ⇒ *Moet [do not/don't] asseblief nie so raas nie. Die baba slaap.* ⇒ *Jy moet [must not] asseblief nie jou skooltas skip nie.*



Pasop: As jy moet asseblief nie gebruik, moet daar nog 'n nie aan die einde van die sin kom. ⇒ *Jy moet [must not/shouldn't] asseblief nie so laat gaan slaap nie.*

mond naamwoord ■
(monde, mondjie)
① 'n Mens se mond is die opening in jou gesig tussen jou lippe. ⇒ *Die vrou sit 'n happie kos op die vurk en dan in haar mond [mouth].* ⇒ *My boetie se mond [mouth] is altyd vuil nadat hy geëet het.*
② Die mond van 'n rivier is die plek waar die rivier in die see in loop. ⇒ *Ons swem altyd by die mond [estuary] van die rivier.*

mondfluitjie naamwoord ■
(mondfluitjies, -)
'n Mondfluitjie is 'n klein musiekinstrument wat 'n klank maak as jy met jou mond lug daarin blaas. ⇒ *My oupa het vroeër altyd op sy mondfluitjie [mouth-organ/harmonica] gespeel.* ⇒ 'n Mondfluitjie [mouth-organ/harmonica] en 'n kitaar word soms gebruik om lekker jazz te speel.



monster naamwoord ■
(monsters, monstertjie)
① 'n Monster is 'n karakter in sprokies of verhale wat gewoonlik lelik is en vir wie mens bang is. ⇒ *Daar woon 'n monster [monster] onder die koning se paleis.* ⇒ *Die monster [monster] in die sprokie eet klein kindertjies.*
② Iemand wat lelik lyk of optree, is 'n monster. ⇒ *Die arme kind se ma is 'n monster [monster].* ⇒ *Ons het hierdie jaar 'n monster [monster] vir 'n onderwyser.*

mooi adjektief ■
(mooi; mooier, die mooiste)
① As iemand/iets mooi is, hou 'n mens daarvan om na die persoon/ding te kyk. ⇒ *Sjoe, maar dis 'n mooi [pretty] meisie!* ⇒ *My ma het 'n baie mooi [beautiful] kers gekoop.*
② As 'n mens tevrede is met iets wat iemand gedoen het, gebruik jy mooi. ⇒ *Mooi [well done], Zolile. Jy het die oplossing gekry.* ⇒ *Die onderwyseres is trots op haar span, want hulle het mooi [well] gespeel.*
③ As dit 'n mooi dag is, is die weer lekker. ⇒ *Ons het nou drie mooi [fine/beautiful] dae na mekaar gehad.* ⇒ *Dit is so 'n mooi [fine/beautiful] dag. Kom ons gaan see toe.*

○ sinoniem **pragtig**
● teenoor **lelik**

daar in hierdie gevalle nie voorsiening gemaak is vir intensiewe vorme nie sodat woorde soos *doodmoeg* en *popmooi* nie opgeneem is nie.) Wanneer die lemmas *moed*, *moeder*, *moeilikheid*, *moeite*, *mond*, *mondfluitjie* en *monster* in die uittreksel uit die *Nuwe Woordeboek Sonder Grense* nagegaan word, blyk dit dat die woorde as selfstandige naamwoorde aangedui en die korrekte meervoudsvorm (of afwesigheid van 'n meervoudsvorm) en verkleiningsvorm telkens vermeld word. Dit is die soort inligting wat baie nuttig in 'n klaskamersituasie gebruik kan word.

Gegewe die vlak van die woordeboekkultuur in Suid-Afrika (Gouws 2006: 79) asook die beperkte toegang wat talle leerders tot woordeboeke het, gaan die leksikograwe in die beplanning van die meeste Suid-Afrikaanse skoolwoordeboeke van die standpunt uit dat die beplande woordeboek waarskynlik die enigste leksikografiese bron sal wees waarmee die tipiese lid van die teikengebruikersgroep in sy/haar skoolsituasie te doen sal kry. Gevolglik word skoolwoordeboeke nie slegs as monofunksionele bronne saamgestel nie. Die skoolwoordeboek wat op die behoeftes van die gemiddelde leerderteikengebruikersgroep gerig is, behoort volgens Gouws (2006: 79) sowel 'n kognitiewe as 'n kommunikatiewe funksie te hê. In die nakoming van die kommunikatiewe funksie is die fokus op beide 'n teksproduksie- en 'n teksresepsiefunksie.

Tans word uitkomsgebaseerde taalhandboeke in Suid-Afrikaanse skole gebruik. 'n Nuwe generasie skoolhandboeke is saamgestel wat spesifiek gerig is op die bereiking van uitkomst wat met uitkomsgebaseerde onderwys geassosieer word. Wanneer die nuwe generasie graad 4-taalhandboeke (Sadie e.a. 2004, Van Oort en Mentz 2004 en Bloemhof e.a. 2005) met die ouer generasie standaard 2-taalhandboeke (De Klerk en Sadie 1971, Knoetze en Els 1981 en Van Niekerk e.a. 1983) vergelyk word, is dit duidelik dat die meerderheid nuwe generasie taalhandboeke nie meer naslaanbronne vir taalkundige inligting is nie. Die ouer generasie taalhandboeke het wel oor voldoende naslaanmateriaal beskik om aan leerders goeie taalkundige leiding te verskaf. In die nuwer generasie taalhandboeke word slegs 'n paar voorbeelde van verbuiging, trappe van vergelyking, intensiewe vorme, meervoudsvorme en verkleiningsvorme gelys. Woordeboeke as houters van taalkundige inligting oortref dus die huidige generasie taalhandboeke as naslaanbronne. Leerders en hulle ouers en opvoeders behoort bewus te wees van die probleemoplossende potensiaal van woordeboeke.

Die Suid-Afrikaanse kurrikulum is baie gefokus op kommunikasievaardighede sodat leerders geleer word om goeie verbale kommunikeerders te wees, maar skryfvaardighede, veral spelvaardighede word afgeskeep. Daar word daaglik berig oor die lae geletterdheidsvlakke onder Suid-Afrikaanse leerders (Koen 2008). Koen (2008: 28) wys op die swak prestasie van Suid-Afrikaanse leerders ten opsigte van basiese leesvaardighede, soos uiteengesit in 2007 in die publikasie van die navorsingsresultate van die Universiteit van Pretoria. 'n Onlangse verslag deur die TIMSS and PIRLS International Study Center dui aan dat 80% van Suid-Afrikaanse kinders wat by die Sentrum vir Evaluasie en Assessering (UP) getoets is, nie oor basiese leesvaardighede

beskik nie (sien Koen 2008: 28). Suid-Afrika se graad 4- en graad 5-leerders het die swakste gevaar uit die 40 lande wat deelgeneem het.

Aangesien baie leerders spelprobleme ondervind, kan woordeboekgebruik, saam met ander opvoedkundige strategieë, 'n rol speel as hulpmiddel om sulke probleme op te los. Die *Oxford Afrikaans-Engels Skoolwoordeboek* bevat ook 'n naslaanafdeling met onder andere spel- en puntuasiereëls in Afrikaans. Dié naslaanafdeling wat naas die sentrale teks verskaf word, vorm deel van die middelteks van die woordeboek. Gesien die teikengroepgebruikers van die woordeboek wat jong leerders is wat leiding nodig het ten opsigte van spelreëls en spelbeginsels, kan dit van groot waarde vir hulle wees.

Woordeboeke is draers van inligting — hulle dien nie slegs as taalkundige handleidings nie, maar ook as buitetalige naslaanbronne. Gouws (2009) stel vakwoordeboeke vir laerskoolleerders voor waar vakwoorde 'n primêre kognitiewe funksie het. Die *Oxford Afrikaans-Engels Skoolwoordeboek* maak voorsiening vir vakterminologie wat in die laerskool gebruik word. In dié geval word dus sowel vakkundige as taalkundige inligting oorgedra. Dit het byvoorbeeld woorde soos *begroting*, *begroot*, *behoefte* en *belasting* opgeneem. Dit is terme wat in die Ekonomiese en Bestuurswetenskappekurrikulum vir graad 4 voorkom. Wanneer leerders met 'n term in die Ekonomiese en Bestuurswetenskappeklas gekonfronteer word wat hulle nie verstaan nie, kan die betekenis maklik in 'n woordeboek opgesoek word. Voorbeeldsinne soos die volgende in die *Oxford Afrikaans-Engels Skoolwoordeboek* tree verhelderend op:

Die skool het **begroot** om 'n swembad te bou.

Daar is nie geld in my **begroting** vir 'n DVD-speler nie.

Die skool het 'n **behoefte** aan 'n groter biblioteek.

'n Mens moet **belasting** betaal op jou salaris.

Internetwoordeboeke kan 'n groot rol speel om 'n woordeboekkultuur onder laerskoolleerders te vestig. Die jonger generasie hou van rekenaars, is daarmee vertrouwd en beskou dit as deel van hulle alledaagse lewe. Wanneer leerders internetwoordeboeke gebruik, het hulle ander soort naslaanvaardighede (soekvaardighede) nodig as wat die geval is by gedrukte woordeboeke. Die eindresultaat van die soektog is belangrik en die veronderstelling is dat die leerders die inligting sal kry waarvoor hulle soek. Talle uitgewers het internetwoordeboeke beskikbaar, wat in sekere gevalle selfs gratis geraadpleeg kan word (vergelyk byvoorbeeld Merriam-Webster in hierdie verband). Daar is nog nie internetwoordeboeke vir Afrikaanse laerskoolleerders beskikbaar nie. Taal oefeninge in die vorm van taalspeletjies dien as aantrekkingskrag vir jong gebruikers om hulle aan te moedig om woordeboeke te gebruik (vergelyk byvoorbeeld die webwerf van Merriam-Webster).

Die uitgewer Pharos het 'n nuwe dimensie aan leksikografie in Suid-Afrika verleen met 'n woordeboek wat op 'n selfoon gelaai kan word. Suid-Afrika se eerste tweetalige woordeboek vir selfone *Klein Woordeboek/Little Dictionary* is

in Julie 2008 bekendgestel. So 'n woordeboek kan maklik geraadpleeg word, aangesien dit oral saamgedra kan word. In veral plattelandse gebiede waar inligtingsbronne skaars is, kan 'n selfoonwoordeboek 'n nuttige inligtingsbron wees. Aangesien 90% van Suid-Afrika se bevolking toegang tot selfone het, kan dit 'n groot rol speel in die skep van 'n woordeboekkultuur (W. Smith-Muller 2008, mondelinge mededeling).

Die rol van opvoeders in woordeboekonderrig

Combrink e.a. (1997: 13) beskou visuele materiaal en naslaanbronne as basiese onderrig- en leermateriaal wat vir interaktiewe onderrig- en leerprogramme in elke klas beskikbaar behoort te wees. Onder naslaanbronne (Combrink e.a. 1997: 13) word die volgende genoem: die *Afrikaanse Woordelys en Spelreëls*, verklarende en vertalende woordeboeke, 'n uitspraakwoordeboek, 'n etimologiese woordeboek, 'n idioomwoordeboek en 'n voorsetselwoordeboek. Woordeboeke behoort in elke klaskamer teenwoordig te wees ten einde goeie taalonderrig te verseker. Die vraag is hoeveel klaskamers in Suid-Afrika oor selfs 'n enkele woordeboek beskik.

Is opvoeders bewus van die belangrikheid van woordeboekgebruik en woordeboekonderrig, en die rol wat dit in taal-, veral in taalkundeonderrig speel? In die Suid-Afrikaanse konteks van veeltaligheid vervul taalkundeonderrig 'n belangrike rol om suksesvolle kommunikasie tussen mense te verseker. Indien 'n leerder goed in sy huistaal onderrig word, word 'n goeie grondslag vir woordeboekgebruik gelê. Indien woordeboekgebruik in die klaskamer aangeleer word, word 'n goeie basis vir die verwerwing van 'n nuwe taal gevorm.

Sekere opvoeders in Suid-Afrika is nie voldoende gekwalifiseer om taalkunde aan te bied nie en die meerderheid is nie opgelei in basiese woordeboektipologie en -gebruik nie. Dit sal dus heeltemal verstaanbaar wees as hulle nie die noodsaaklikheid van woordeboekonderrig as deel van taalonderrig insien nie. Leksikografie vorm nie normaalweg deel van die universiteitsopleiding vir opvoeders nie. Dit beteken dat opvoeders nie opleiding ontvang ten opsigte van woordeboektipologie, woordeboekfunksies, leksikale velde (sino-nieme, antonieme) en leksikale items (subleksikale items en multileksikale items) nie. Dit kan ook die gevolg hê dat opvoeders nie geleer word om 'n kritiese instelling teenoor woordeboeke te ontwikkel nie. Gevolglik beskik hulle nie altyd oor genoeg kennis om 'n ingeligte keuse oor 'n gepaste woordeboek vir die klassituasie te maak nie. Opvoeders moet ook woordeboeke wat leerders tuis kan gebruik, aan ouers kan aanbeveel. Wanneer woordeboeke gekies word, moet opvoeders seker maak van die funksie van die woordeboek wat leerders benodig. Gaan 'n woordeboek vir teksontvangs, teksproduksie of vir vertaling gebruik word? Gouws (2006: 75) wys daarop dat daar twee tipes funksies binne die teorie van leksikografiese funksies bestaan, naamlik kognitiewe en kommunikatiewe funksies. Die doel van kognitiewe funksies (vroeër

bekend as kennisgerigte funksies) is om die gebruiker te help met algemene en ensiklopediese data, data oor spesifieke vakgebiede en data oor taal. Kommunikatiewe funksies ondersteun gebruikers met die oplos van probleme betreffende teksproduksie in die moeder- en vreemde taal, asook die vertaling van tekste uit die moeder- in die vreemde taal en omgekeerd.

Alhoewel uitkomst van woordeboekgebruik in die NKV van die intermediêre en senior fases voorkom, hoef dit nie noodwendig tot woordeboekgebruik en woordeboekonderrig in laerskole te lei nie. Allerlei ander praktiese probleme kan in die weg staan van suksesvolle woordeboekgebruik en woordeboekonderrig, naamlik die gebrek aan 'n geskikte skoolwoordeboek om in die klaskamer te gebruik, of die beperkte toegang tot bronne omdat nie alle leerders oor hulle eie woordeboeke beskik nie en nie alle klaskamers van woordeboeke voorsien is nie. Die moontlikheid bestaan ook dat dit die assesser van woordeboekuitkomst kan bemoeilik. Indien daar nie genoeg woordeboeke vir elke leerder in die klaskamer beskikbaar is nie, is assessering prakties byna onmoontlik.

Dit word sterk aanbeveel dat die Departement van Nasionale Opvoeding 'n opleidings- en bewusmakingsveldtog ten opsigte van woordeboekonderrig loods. Opvoeders, en hulle sluit volgens Klein (2007: 420) taalopvoeders en vakopvoeders in, behoort 'n deeglike kennis van woordeboeke te hê en behoort woordeboeke met selfvertroue te kan gebruik. Woordeboekonderrig behoort as deel van die opleiding van opvoeders op universiteitsvlak ingesluit te word en opvoeders in diens behoort ook opleiding te ontvang. Klein (2007: 420) is van mening dat inligting oor (a) die voordele van woordeboeke vir die individuele spreker, (b) inligting oor die verskillende woordeboektipies, (c) die inhoud van 'n woordeboek, (d) die prosedures om inligting uit 'n woordeboek te onttrek, en (e) die metodes om die inligting toe te pas deel van so 'n kursus behoort uit te maak. Die volgende kan daarby gevoeg word: inligting oor woordeboekfunksies en naslaanvaardighede, asook die uitkomst van woordeboekonderrig en die onderrigmetodologie van woordeboekgebruik.

Opvoeders kan 'n belangrike rol speel om 'n woordeboekkultuur in Suid-Afrika te vestig. Omdat hulle 'n groot invloed op leerders het, kan hulle leerders sodoende bewus maak van die waarde van woordeboeke. Woordeboekonderrig is belangrik in die laerskool waar die jong geslag leerders woordeboekvaardighede bygebring kan word sodat hulle die waarde van woordeboeke kan besef. Leerders behoort in die gewoonte te kom om gereeld woordeboeke te raadpleeg tydens teksresepsie (bv. lees) en teksproduksie (bv. praat en skryf, en woordspeletjies en blokkiesraaisels) — die begin van 'n lewenslange verhouding met woordeboeke.

Gevolgtrekking

Woordeboekonderrig vorm 'n belangrike aspek van taalonderrig wat aandag in die klaskamer moet kry, maar ook die aandag van opvoedkundiges, die De-

partement van Nasionale Opvoeding, uitgewers en leksikograwe verdien. Soos in die artikel aangetoon, vorm woordeboekonderrig reeds deel van die NKV. Nie net is woordeboekonderrig gedefinieer nie, maar daar is ook aangetoon watter uitkomst in verband met woordeboekonderrig in die laerskool bereik behoort te word. 'n Model ter illustrasie van die prosesse van woordeboekonderrig is voorgestel.

Woordeboeke as naslaanbronne is uitstekende bronne van kennis vir laerskoolleerders en dit is sowel opvoeders as ouers se plig om leerders daarvan bewus te maak. Sodoende kan 'n woordeboekkultuur in Suid-Afrika tot stand kom.

Bibliografie

- Barone, R.** 1979. On the Use of the *Advanced Learner's Dictionary of Current English*: From Theory to Practice. *Rassegna Italiana di Linguistica Applicata* 11(1-2): 187-194.
- Beattie, N.** 1973. Teaching Dictionary Use. *Modern Languages* 54(4): 161-168.
- Béjoint, H.** 1989. The Teaching of Dictionary Use: Present and Future Tasks. Hausmann, F.J., O. Reichmann, H.E. Wiegand en L. Zgusta (Reds.). 1989-1991. *Wörterbücher. Ein internationales Handbuch zur Lexikographie/Dictionaries. An International Encyclopedia of Lexicography/Dictionnaires. Encyclopédie internationale de lexicographie*: 208-212. Berlyn/New York: Walter de Gruyter.
- Bloemhof, F., S. Erasmus, B. Rau, J. Reid en J. van Lill.** 2005. *Leerplatform Afrikaans Graad 4*. Kaapstad: Nasou via Afrika.
- Carstens, A.** 1995. Language Teaching and Dictionary Use: An Overview. *Lexikos* 5: 105-116.
- Combrink, L., N. Faasen, H. Geysers en A. Kloppers.** 1997. *Taalkonteks*. Kenwyn: Juta.
- Cowie, A.P. (Red.)**. 1987. *The Dictionary and the Language Learner*. Tübingen: Max Niemeyer.
- Croft, K.** 1973. Dictionary Use in ESL Courses. *TESL Reporter* 7(1): 1-9.
- De Klerk, P.F. en J.J. van L. Sadie.** 1980. *Afrikaans my taal Standaard 2*. Kaapstad: Maskew Miller.
- Flaherty, M.C.** 1923. *How to Use the Dictionary*. New York: Ronald Press.
- Gouws, R.H., M. Stark en L. Gouws.** 2004. *Nuwe Woordeboek Sonder Grense*. 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.** 2006. Leerdergedrewe leksikografie. *Tydskrif vir Taalonderrig* 40(1): 72-84.
- 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.
- Gouws, R.H.** 2009. Geïntegreerde woordeboekgebruik van vakwoordeboeke vir aanleerders. *Lexikos* 19: 72-93.
- Hartmann, R.R.K.** 1987. Dictionaries of English: The User's Perspective. Bailey, R.W. (Red.). 1987. *Dictionaries of English: Prospects for the Record of Our Language*: 121-135.
- Hartmann, R.R.K.** 1988. The Learner's Dictionary: Traum oder Wirklichkeit? Hyldgaard-Jensen, K. en A. Zettersten (Reds.). 1988. *Symposium on Lexicography III. Proceedings of the Third International Symposium on Lexicography, May 14-16, 1986 at the University of Copenhagen*: 215-235. Tübingen: Max Niemeyer.

- Hausmann, F.J.** 1986. The Training and Professional Development of Lexicographers in Germany. Ilson, R. (Red.). 1986. *Lexicography. An Emerging International Profession*: 101-110. Manchester: Manchester University Press.
- Herbst, T. en G. Stein.** 1987. Dictionary Using Skills: A Plea for a New Orientation in Language Teaching. Cowie, A.P. (Red.). 1987: 115-128.
- Kipfer, B.** 1987. Dictionaries and the Intermediate Student: Communicative Needs and the Development of User Reference Skills. Cowie, A.P. (Red.). 1987: 44-54.
- Klein, J.** 2007. Implementing a Dictionary Culture in South Africa: An Attempt at a General Framework for the African Languages. *Lexikos* 17: 416-423.
- Knoetze, F.L. en J.S.M. Els.** 1981. *Keurige Afrikaans St. 2*. Tweede uitgawe. Goodwood: Nasou.
- Koen, M.** 2008. Bemagtig opvoeders vir onmiddellike mobilisasie van grondslagfase-leerders se geletterdheid. *Litnet Akademies* 5(2): 27-45.
- Kühn, P.** 1987. *Mit dem Wörterbuch arbeiten. Eine Einführung in die Didaktik und Methodik der Wörterbuchbenutzung*. Bonn/Bad Godesberg: Dürr.
- Louw, P. (Red.)** 2007. *Oxford Afrikaans-Engels/English-Afrikaans Skoolwoordeboek/School Dictionary*. Kaapstad: Oxford University Press.
- Mathews, M.M.** 1964. The Freshman and his Dictionary. Allen, H.B. (Red.). 1964. *Readings in Applied English Linguistics*: 372-375. New York: Appleton Century Crofts.
- Merriam-Webster** [aanlyn]. <http://WordCentral.com>.
- Odendal, F.F.** 2000. *Kernwoordeboek*. Kaapstad: Perskor.
- Republiek van Suid-Afrika. Departement van Onderwys.** 2002. *Hersiene Nasionale Kurrikulumverklaring Graad R-9 (Skole). Beleid Tale. Afrikaans Huistaal*. Pretoria: Departement van Onderwys.
- Sadie, J.J. van L., W. Louw, A. Engelbrecht en M. Mihai.** 2004. *Afrikaans ons taal graad 4*. Kaapstad: Maskew Miller Longman.
- Tono, Y.** 1984. *On the Dictionary User's Reference Skills*. Ongepubliseerde B.Ed.-skripsie. Tokio: Gakugei Universiteit.
- Van Niekerk, A.P., J.J. Steyn, W.H. Herman en J. Vosloo.** 1983. *Afrikaans vir vandag Standaard 2*. Kaapstad: Juta.
- Van Oort, R. en W. Mentz.** 2004. *Afrikaans in aksie Graad 4*. Kaapstad: Kagiso Education.

Volkslinguïstiek en dialectlexicografie in de zuidelijke Nederlanden¹

Jacques Van Keymeulen, *Vakgroep Nederlandse Taalkunde, Universiteit
Gent, België (jacques.vankeymeulen@ugent.be)*

Samenvatting: De professionele dialectlexicografie voor het zuidelijke Nederlands heeft drie grote regionale dialectwoordenboeken voortgebracht: het *Woordenboek van de Brabantse Dialecten* (WBD), het *Woordenboek van de Limburgse Dialecten* (WLD) en het *Woordenboek van de Vlaamse Dialecten* (WVD). Die drie woordenboeken zijn echter veeleer als thematisch geordende materiaalbasisen voor taalgeografisch onderzoek te beschouwen, dan als echte woordenboeken die op een nauwgezette manier semantische informatie verschaffen over het dialectlexicon. Daarom is de 'dialectrenaissance' in de Nederlanden en Vlaanderen, en de daarmee gepaard gaande lexicografische activiteit van plaatselijke amateurs een verheugend verschijnsel. Zaak is om de tientallen plaatselijke dialectwoordenboeken die de laatste decennia het licht hebben gezien in een efficiënt doorzoekbare database onder te brengen, die een zeer belangrijke aanvulling kan zijn bij de resultaten van de professionele dialectlexicografie.

Sleutelwoorden: LEXICOGRAFIE, TAALVARIATIE, AMATEUR, AMATEURLEXICOGRAFIE, DIALECTLEXICOGRAFIE, WOORDENBOEK, WOORDBEELD, REGIONAAL WOORDENBOEK, LOKAAL WOORDENBOEK, DIALECT, TRADITIONEEL DIALECT, VLAAMS, NEDERLANDS, NEDERLAND, BELGIË, METHODOLOGIE, TREFWOORD, GEBRUIKERSGROEP, BRUIKBAARHEID, CORRESPONDENTIEREGELS

Abstract: Folk Linguistics and Dialect Lexicography in the Southern Netherlands. Professional lexicography for the southern Dutch dialects has produced three major regional dialect dictionaries: the *Dictionary of the Brabantic Dialects* (WBD), the *Dictionary of the Limburgian Dialects* (WLD) and the *Dictionary of the Flemish Dialects* (WVD). The three dictionaries are, however, to be considered rather as thematically arranged collections of dialect words aimed at geographical research than as real dictionaries meticulously providing semantic information on the dialect lexicon. Therefore, the 'dialect renaissance' in the Netherlands and Flanders and the lexicographical activity of local amateurs which accompanies it, is a gratifying phenomenon. It is necessary to incorporate the numerous local dialect dictionaries published during the last decades in an efficient searchable database, which could serve as a very important addition to the results of professional dialect lexicography.

Keywords: LEXICOGRAPHY, LANGUAGE VARIATION, AMATEUR, AMATEUR LEXICOGRAPHY, DIALECT LEXICOGRAPHY, DICTIONARY, WRITTEN WORD, REGIONAL DICTIONARY, LOCAL DICTIONARY, DIALECT, TRADITIONAL DIALECT, FLEMISH, DUTCH, THE NETHERLANDS, BELGIUM, METHODOLOGY, HEADWORD, USER GROUP, USEFULNESS, CORRESPONDENCY RULES

0. Inleiding

0.1. Volkslinguïstiek en dialectlexicografie

De woordgroepen 'volkse linguïstiek'/'populaire linguïstiek' en de woorden 'volkslinguïstiek' of 'volkslinguïstisch' leveren in Google voor het Nederlands (en Afrikaans) geen of in elk geval geen significante hits op. Tegenwoordig mag dat als bewijs gelden dat de notie in het Nederlandse taalgebied zo goed als onbekend is. 'Folk linguistics' is blijkbaar wel bekend in de Angelsaksische wereld, voor een groot deel dank zij het werk van Preston en Niedzielski.² Met de term wordt het geheel bedoeld van volkse opinies en attitudes tegenover taal; het is de 'populaire taaltheorie'. De metatalige ideeën en percepties die bij het gewone volk leven, worden doorgaans tegenover het wetenschappelijke weten geplaatst en door de wetenschap als 'verkeerd' of in elk geval als onbelangrijk afgedaan. Preston en Niedzielski trachten met hun werk op overigens overtuigende wijze aan te tonen dat onderzoek naar de 'folk theory' nodig is voor een compleet beeld van de 'ethnography of language' en om processen van taalverandering beter te begrijpen. Volkse attitudes tegenover taal dienen door de wetenschap ernstig genomen te worden — ze zijn een valabel onderzoeksobject. 'Volkslinguïstiek' kan m.i. voor de dialectlexicografie op twee manieren belangrijk zijn: allereerst is inzicht in de populaire taalattitudes nodig bij lexicografisch veldwerk; in de tweede plaats zijn een aantal concrete resultaten uit die volkse taalattitudes voortgevloeid, nl. (plaatselijke) woordenboeken.

Een lexicograaf die werkt met mondelinge of schriftelijke enquêtes, zoals voor de verzamelaar van louter oraal overgeleverde taalvariëteiten noodzakelijkerwijs het geval is, moet uiteraard de reacties van de respondenten op zijn vragen kunnen voorzien. Die reacties worden gestuurd door de impliciete taaltheorie en de taalattitudes van de respondenten. In een 'diaglossische' taalsituatie als bijv. die in het huidige Nederlandstalige België (Vlaanderen), waar elke taalvariëteit een plaats heeft op een continuüm tussen traditioneel dialect en Standaardnederlands, wordt een vraag als "Hoe noemt u x in uw dialect " vandaag de dag niet door iedereen op dezelfde manier opgevat. Wat immers is voor een bepaalde respondent de inhoud van de term 'dialect' in een taalsituatie waarbij de aloude fijnmazige dialectgeografische differentiatie wordt afgelost door taalvariëteiten die horen bij de nieuwe maatschappij die in de jaren '60 van de vorige eeuw is ontstaan en gekenmerkt wordt door een verhoogde sociale en geografische mobiliteit en een steeds hogere scholingsgraad. In het huidige Nederlandstalige België³ bestaan naast elkaar: traditioneel dialect, regiolect, Vlaamse omgangstaal, Belgisch-Nederlands ... een aantal variëteiten zijn daarbij ontstaan onder druk van de nieuwe maatschappelijke omstandigheden en de introductie van de Nederlandse standaardtaal in brede lagen van de bevolking. Is 'dialect' dan voor de respondent de variëteit die zijn grootouders nog konden spreken (maar hij helaas niet meer) of de variëteit die hijzelf praat

in ongedwongen situaties in de eigen huiskring en door een academicus veel-
eer als regiolect of Vlaamse omgangstaal opgevat zal worden ... Voelt de
respondent zich vervolgens al dan niet gewapend om een dialectvragenlijst in
te vullen — de ervaring bij het *Woordenboek van de Vlaamse Dialecten* leert dat de
inhoud van het begrip 'dialect' bij respondent en lexicograaf wel eens ver uit
elkaar kan liggen.

In mijn bijdrage zal ik het echter niet hebben over de relevantie van de
volkslinguïstiek voor dialectlexicografisch veldwerk, maar wel over de ama-
teurlexicografie, die beschouwd kan worden als het resultaat van populaire
noties i.v.m. de oudste taallaag. In wat volgt, zal ik me dus toespitsen op de
regionale en lokale dialectwoordenboeken die zijn gemaakt in het zuidelijk-
Nederlandse⁴ dialectgebied, d.i. Nederlandstalig België en de drie zuidelijke
provincies van Nederland: Zeeland, Noord-Brabant en Limburg. In dat gebied
worden er vier dialectgroepen gesproken: Zeeuws (in Zeeland en op de Zuid-
Hollandse eilanden Goeree en Overflakkee), Vlaams (in Frans-Vlaanderen,
West-Vlaanderen, Oost-Vlaanderen en Zeeuws-Vlaanderen), Brabants (in
Vlaams-Brabant, Antwerpen en Noord-Brabant) en Limburgs (in Nederlands
en Belgisch Limburg). Het is uitdrukkelijk niet de bedoeling een analyse te
maken van alle verkeerdheden die in de amateurwoordenboeken aangetroffen
kunnen worden — en dat zijn er heel wat — maar wel om hun belang voor de
wetenschap voor het voetlicht te brengen.

1.1 Amateurdialectlexicografie

Met amateurlexicografie bedoelen we de lexicografie die niet door professio-
nele lexicografen — al dan niet aan een instituut of bedrijf verbonden — wordt
beoefend. Men kan de amateurlexicografie voor de Nederlandse dialectologie
laten beginnen met het werk van Hoeufft (1836), die een woordenboek voor het
dialect van Breda in het licht gaf. De kwaliteit van de amateurwoordenboeken
loopt zeer sterk uiteen. De auteur van het monumentale *Woordenboek der
Zeeuwse Dialecten* (WZD), dr. Hendrika C. Ghijsen, was een amateurlexicograaf,
maar dan wel een met een doctorstitel, die na vele jaren veldwerk een dialect-
woordenboek publiceerde waar voor het eerst alle opgegeven woorden gelo-
kaliseerd werden.⁵ Het WZD wordt daarom beschouwd als het begin van de
wetenschappelijke dialectlexicografie in de Nederlanden.⁶ Ook andere acade-
mici, zoals Debrabandere (1999), hebben regionale of lokale woordenboeken
gepubliceerd, die uiteraard kwalitatief gunstig afsteken tegen de producten
van taalkundig ongeschoolden.

Hoewel er goede tot zeer goede dialectwoordenboeken door amateurs zijn
gemaakt, is er ook heel wat kaf onder het koren. Heel wat lokale woordenboe-
ken en woordenlijsten worden ontsierd door pijnlijk onjuiste klankinleidingen
of door fantastische etymologieën. De inleidingen op de amateurwoordenboe-
ken zijn wel een rijke bron voor een volkslinguïstisch onderzoek naar attitudes
tegenover taal en dialect: het is immers daar dat de auteur zijn motivatie aan-

geeft om met een woordenboek te beginnen en dus zijn taalattitudes blootgeeft. De motivatie is altijd 'idealistisch'; commerciële overwegingen zijn normaal afwezig — wat niet wegneemt dat de meeste lokale woordenboeken bij de lokale bevolking gretig aftrek hebben gevonden.

De handleidingen van Cajot (1995) en van Van Keymeulen (2003) zijn erop gericht om de amateurlexicografie in goede banen te leiden. Van Keymeulen (2003) probeert te dien einde de amateur ervan te doordringen zich te beperken tot datgene wat hij — met enige begeleiding en oefening — wél aankan: het registreren van de woorden van een lokale taalvariëteit en het weergeven van hun belangrijkste betekenissen. Onvolledigheid is daarbij een geringer kwaad dan onjuistheid.

1.2 Amateurlexicografie voor de zuidelijke dialecten in de 19de en de eerste helft van de 20ste eeuw

Het valt op dat in het zuiden van het Nederlandse taalgebied, meer bepaald in Nederlandstalig België, op het einde van de 19de eeuw/begin 20ste eeuw enkele zeer belangrijke dialectlexicografische werken ontstonden: het *Algemeen Vlaamsch Idioticon* van Schuermans (1865–1870) beet de spits af. Het *Westvlaams Idioticon* van De Bo (1873), het *Waasch Idioticon* van Joos (1900), het *Antwerps Idioticon* van Cornelissen en Vervliet (1899–1903), het *Hagelandsch Idioticon* van Tuerlinckx (1886), het *Zuid-Oost-Vlaandersch Idioticon* van Teirlinck (1908–1924) en nog andere volgden snel. De belangrijkste woordenboeken waren regionaal opgezet, en moesten dus in de trefwoorden abstractie maken van de dialectische fonologie: de trefwoorden zijn noodzakelijkerwijs vernederlandst en vatten dus heel wat fonologische variatie samen.

De grote 19de-eeuwse werken en hun opvolgers in de eerste helft van de 20ste eeuw zijn typische producten van hun tijd. Ze zijn ontstaan vanuit een romantisch gedachtengoed, met een klemtoon op de natuurlijke taal waarin de 'volksziel' verondersteld werd tot uiting te komen en een hang naar een onbezoedeld agrarisch verleden. Bovenal kunnen ze niet los gezien worden van de Vlaamse taalstrijd tegen de Franstalige dominantie — De Bo's woordenboek kadert daarbij in het West-Vlaamse taalparticularisme. De grote dialectwoordenboeken bewezen in elk geval dat de volkstalen een rijke woordenschat hadden; de auteurs ervan hoopten ook dat hun verzamelingen in het *Woordenboek der Nederlandsche Taal* (WNT), een woordenboek waaraan vanaf 1852 werd gewerkt, terecht zouden komen, een verlangen dat maar gedeeltelijk is bewaarheid. Voor een kort overzicht van deze glorierijke periode van de amateurlexicografie verwijzen we naar Goossens en Van Keymeulen (2006: 74-80).

1.3 De jaren '70 van de 20ste eeuw en daarna: de dialectrenaissance

Met 'dialectrenaissance' wordt het verschijnsel bedoeld dat het dialect als voertaal gebruikt wordt voor allerlei (populair-)culturele uitingen: muziek,

theater, stripverhalen ... Dialect mag weer. Er worden in Nederland zelfs delen van de Bijbel in het dialect vertaald — iets waar niet iedereen even gelukkig mee is. Predikant en burgemeester durven zich in het dialect tot een publiek te richten — ook voor ernstige aangelegenheden. Het verschijnsel dateert van de jaren '70 van de vorige eeuw en is het eerst in Nederland waargenomen. Het Nedersaksisch en het Limburgs zijn er zelfs in geslaagd zich resp. in 1995 en 1997 tot streektalen te laten verklaren met een beroep op (deel II van) het Europees Handvest voor Regionale Talen of Talen van Minderheden.⁷

Het ontstaan van de dialectrenaissance kan zowel in Nederland als in Vlaanderen verklaard worden door een aantal maatschappelijke tendenzen van subjectieve en objectieve aard — die in beide gebieden een ongelijk verloop hebben gekend. Allereerst kan gewezen worden op de stijgende democratisering van de samenleving sedert de jaren '60 van de vorige eeuw en daardoor de informalisering van de sociale contacten, waardoor de standaardtaal aan prestige heeft moeten inboeten. Een tweede oorzaak is ongetwijfeld het verdwijnen van de traditionele dialecten, vooral wat de woordenschat betreft. De motivatie voor de oude 19de-eeuwse woordverzamelingen was ook al dat de dialecten verdwenen en dus dringend 'gered' moesten worden; in de jaren '60 van de 20ste eeuw is er in heel Europa echter wel een werkelijk grootschaliger taalverandering op gang gekomen, met het uitwissen van grote delen van de dialectwoordenschat (bijv. nagenoeg de totaliteit van de eeuwenoude landbouwwoordenschat), dan de relatief langzame veranderingen voordien. Een derde oorzaak is ongetwijfeld de zoektocht naar de eigen identiteit en het terugplooiën op regionale en lokale eigenheid, die een tegenwicht moet vormen tegen de voortschrijdende Europeanisering en globalisering van de samenleving.

De dialectrenaissance zal de oude dialecten niet doen herleven; dat is ook de bedoeling niet — men zou kunnen volhouden dat de standaardtaal (eventueel met regionale accenten) in Nederland en Vlaanderen zodanig sterk staat dat niemand het gebruik van dialect bij bepaalde gelegenheden of voor bepaalde culturele doeleinden nog als een bedreiging voor de standaardtaal aanvoelt. Het verschijnsel lijkt in Vlaanderen van recentere datum te zijn dan in Nederland, hoewel er in Vlaanderen altijd al een hoogstaande liedcultuur in het dialect is geweest. De liederen die bijv. Wannes Vandeveldde vanaf de jaren '60 van de vorige eeuw ten gehore brengt, zijn wel in het Antwerpse dialect, maar zijn van een literair en muzikaal niveau dat de vergelijking met de productie in de Nederlandse cultuurtaal moeiteloos kan doorstaan. Het gebruik van het Antwerps was voor Wannes overigens geen gevolg van een bewuste keuze voor het dialect, maar gewoonweg een evidentie voor een kunstenaar die het Algemeen Nederlands niet goed machtig was.

De verschillen tussen Nederland en Vlaanderen op het gebied van de dialectrenaissance en de attitudes tegenover standaardtaal en dialect zouden het onderwerp kunnen zijn van een boeiende studie. Ik heb het vermoeden dat dan zou blijken dat er geen renaissance is van het dialect zelf, maar enkel van de belangstelling ervoor — een omstandigheid die eigenlijk net aantoonde dat het

oude dialect verdwijnt, wat als een cultureel verlies wordt aangevoeld. Het is echter niet het onderwerp van mijn artikel. In wat volgt, zal ik het hebben over een belangrijk effect van de dialectrenaissance: het maken van lokale woordenboeken.

2. Het belang van de amateurlexicografie

2.1 Aantal en verspreiding

Op bijgaand kaartje en in de tabel wordt de productie van de lokale dialectwoordenboeken in de zuidelijke Nederlanden gevisualiseerd. We hebben ons gebaseerd op de recente bibliografie van Van Keymeulen e.a. (2007), waarin naast woordenboeken waarschijnlijk ook een beperkt aantal andere lexicografische producten zijn opgenomen. Titels als *Het dialect van Beveren en zijn deelgemeenten* (Cools 2000) of *Waur is d'n tijd naurtoe* (Moens 1988) maken het echter niet mogelijk uit te maken of het werk een grammatica, een woordverzameling of nog iets anders bevat. Het zijn overigens beide vrij uitgebreide woordenboeken. Uit de titel van maar een 10-tal werken kon worden opgemaakt dat ze waarschijnlijk uitsluitend verzamelingen uitdrukkingen, spreekwoorden (+ eventueel liedjes, rijmpjes e.d.) bevatten — ze werden in de tabel verwerkt. We waren helaas niet in de gelegenheid om alle werken te bekijken; vele lokale publicaties zijn in eigen beheer uitgegeven voor een plaatselijk publiek en zijn maar moeilijk op te sporen of te verwerven; het gaat daarbij soms om uitgebreide en belangwekkende werken als bijv. het recente *Woordenboek van het Schellebels dialect* (Van der Eecken 2007), waaraan meer dan 30 jaar is gewerkt en dat meer dan 500 bladzijden dik is. De datum post quem is 1836 met het verschijnen van het werk van Hoeffft.

Bij de interpretatie van de kaart en de cijfers moet met een aantal elementen rekening gehouden worden. Allereerst worden woordenboeken van zeer ongelijke omvang en kwaliteit in de tabel samengenomen: die omvang varieert van een paar tientallen bladzijden met enkele honderden woorden als het *Overmeers Woordenboekje* (Maes 1994), tot levenswerken als het driedelige *Gents Woordenboek* van Lievevrouw-Coopman (1952). Voorts is niet steeds heel duidelijk hoe 'plaatselijk' een bepaald woordenboek wel is. De titel (of zelfs de inleiding) op een woordenboek geeft niet steeds voldoende uitsluitel over het werkelijke geografische bereik van de gepresenteerde woordenschat. Soms is de titel van een woordenboek op geografisch gebied waarschijnlijk té ambitieus, zoals dat van Depraetere e.a. (2007).

Ook de woordenboeken die uitdrukkelijk de bedoeling hebben een hele regio te bestrijken, hebben overigens toch een bepaalde geografische focus. Het *Westvlaamsch Idioticon* van De Bo (1873) bijv. bevat maar weinig kustwoorden, maar is vooral gebaseerd op de zuidoostelijke hoek van West-Vlaanderen — niet toevallig de geboortestreek van De Bo (zie Devos 1973). Als een dialect-

woord in het *Westvlaamsch Idioticon* staat, betekent dat dus volstrekt niet dat het algemeen West-Vlaams is. Een woordenboek als dat van Rutten (1890), *Bijdrage tot een Haspengouwsch Idioticon*, bewerkte voornamelijk de gemeente Borgloon. De regionale woordenboeken die in hun titel toch een bepaalde lokaliteit als focus vermelden, zoals Hoefufft (1836) voor Breda of Cornelissen en Vervliet (1899–1903) voor Antwerpen, hebben we in de tabel hieronder opgenomen.

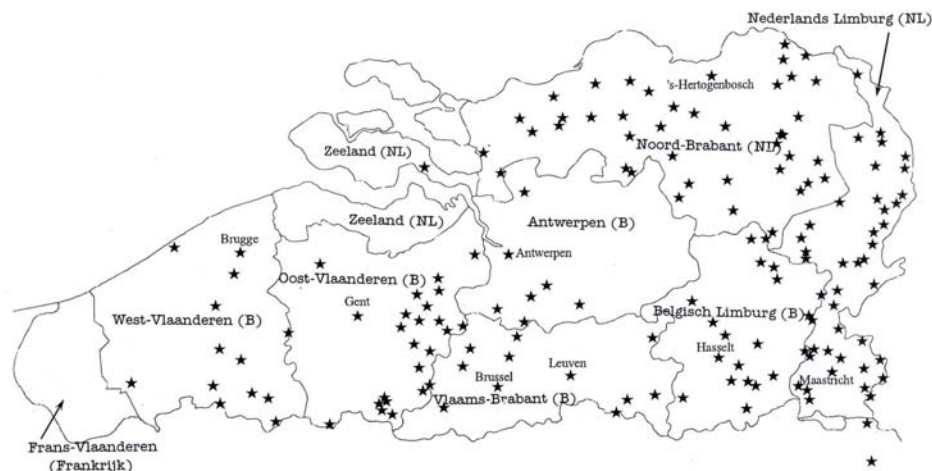
Men moet er ook rekening mee houden dat een plaatsnaam meestal als een administratieve aanduiding opgevat moet worden, die een aantal deelgemeenten met verschillende dialecten kan overkoepelen. Sedert de fusieoperatie voor de Belgische gemeenten in 1977 dekken de meeste gemeentenamen nu andere administratieve realiteiten dan voordien. In het woordenboek van Cools (2000), *Het dialect van Beveren en zijn deelgemeenten*, wordt dat netjes in de titel aangegeven; dat is echter lang niet altijd het geval.

	West-Vl.	Oost-Vl.	Antw.	Vl.-Br.	Nrd-Br.	Bel.-Lbg	Ned.-Lbg	
	B	B	B	B	NL	B	NL	
1830–1840					1			1
1880–1890							1	1
1890–1900			1					1
1900–1910								0
1910–1920							2	2
1920–1930							1	1
1930–1940				1	1			2
1940–1950								0
1950–1960		1			2		1	4
1960–1970							1	1
1970–1980	1	1		1	1	1	4	9
1980–1990	1	5	1	3	6	6	12	34
1990–2000	7	10	3	9	16	8	14	67
2000–2007	7	7	2	3	14	5	16	54
Totaal	16	24	7	17	41	20	52	177

Productie van lokale dialectwoordenboeken in de zuidelijke Nederlanden

Inleidingen op dialectwoordenboeken moeten dus steeds zeer goed nagelezen worden om het geografische bereik van het woordenboek te kennen, in het besef dat heel wat auteurs de neiging hebben in dat verband te overdrijven.

We hebben bij de telling de (soms sterk herwerkte en vermeerderde) herdrukken niet meegerekend, maar hebben het jaartal van het 'eerste initiatief' van de auteur genomen om het woordenboek bij een bepaald decennium in de tabel te plaatsen. Op die manier kwamen we in totaal aan niet minder dan 177



Verspreiding van de lokale woordenboeken in de zuidelijke Nederlanden⁸

'plaatselijke' lexicografische werken die niet aan instituten door professionele lexicografen zijn gemaakt, maar door plaatselijke dialectliefhebbers. De cijfers in de tabel staan voor aantallen woordenboeken, niet voor aantallen gemeenten met een woordenboek. Sommige gemeenten — meestal steden — hebben inderdaad meer dan één woordenboek, waarbij het ene soms op het andere voortbouwt, al dan niet door dezelfde auteur. Voor Maastricht, hoofdstad van Nederlands Limburg, tellen we niet minder dan 7 verschillende lexicografische werken, wat ongetwijfeld te maken heeft met het hoge Limburgse taalbewustzijn (zie ook verder); voor de metropool Brussel, hoofdstad van Vlaanderen, België en Europa, zijn er vier, waaronder het vertalende *Marollien-Français/Français-Marollien*⁹ van Starck en Claessens (1988, 1995²). Brussel is een geval apart: het aloude inheemse Nederlandse stadsdialect is sedert de 19de eeuw zeer sterk onder de druk van het Frans komen te staan.

Hoewel zowel de kaart als de tabel ongelijksoortig lexicografisch werk samenbrengt — er zijn zowel grote verschillen in kwantiteit (aantal bladzijden!) en kwaliteit als in macro- en microstructureel opzet — vallen toch zowel op chronologisch als op geografisch gebied een paar zaken onmiddellijk op. Op chronologisch vlak is het overduidelijk dat vanaf 1980 het aantal lexicografische initiatieven van plaatselijke amateurs plots toeneemt: bij een vergelijking van de periode 1970–1980 met 1980–1990 zien we een sprong van 9 naar 34 publicaties (in 1960–1970 was er maar één!). Het volgende decennium laat 67 woordenboeken zien. Die opbloei heeft alles te maken met de reeds genoemde dialectrenaissance, die haar wortels heeft in de maatschappelijke veranderingen van de jaren '60 van de vorige eeuw, en die in Nederland blijkbaar krachtiger is geweest dan in Vlaanderen.

We stellen ons voor dat het in de loop van de jaren '60 voor de bevolking bewust of onbewust duidelijk werd dat de traditionele dialecten onder grote druk kwamen te staan; en in de jaren '70 namen een aantal dialectliefhebbers, meestal leden van plaatselijke culturele verenigingen, dan het initiatief om alleen of samen met anderen dialectwoorden te beginnen te verzamelen, een activiteit die dan in de loop van de jaren '80 tot een publicatie leidde. We schatten dat het verzamelen en ordenen van dialectwoordschat en het publiceren ervan in een dialectwoordenboek van enige omvang als éénmansonderneming in het pre-computertijdperk gemakkelijk een 10-tal jaar in beslag kon nemen. De dialectrenaissance is blijkbaar nog niet ten einde: sedert 2000 tellen we al 54 woordenboeken.

Op geografisch vlak laat de kaart met de verspreiding van de plaatselijke initiatieven allereerst zien dat er voor Frans-Vlaanderen en Zeeland blijkbaar geen plaatselijke dialectlexicografie bestaat. Voor Frans-Vlaanderen heeft die afwezigheid uiteraard te maken met de specifieke taalsituatie van het Nederlandse dialect in Frankrijk; het oude Vlaams is al te ver verdwenen om nog plaatselijke initiatieven te kunnen inspireren en bovendien zijn de personen die zo'n initiatief zouden kunnen nemen in het Frans geschoold, en kunnen geen Nederlands lezen of schrijven. Gelukkig bestaat er sedert 2005 een regionaal *Woordenboek van het Frans-Vlaams*, door Moeyaert e.a. (2005), een woordenboek waarvoor het materiaal door Belgische Vlamingen is samengebracht. De bijna totale afwezigheid van plaatselijke woordenboeken in Zeeland heeft dan weer een andere oorzaak; de omvang, kwaliteit en bekendheid van het *Woordenboek der Zeeuwse Dialecten* (+ het Supplement erop) — een werk dat voor het Zeeuws als het ware symbool staat — zijn zodanig hoog dat geen enkel individu de noodzaak heeft gevoeld om daarnaast nog met een lokaal woordenboek te beginnen. Alle dialectgeografische activiteit van de talrijke Zeeuwse vrijwilligers werd decennialang naar hetzelfde product toegeleid doordat alle dialectliefhebbers van de provincie via de Zeeuwse Vereniging voor Dialectonderzoek werden ingeschakeld bij de materiaalverzameling. Wel zijn er op initiatief van mevr. E.J. Van de Broeke-de Man negen zgn. regioboeken verschenen, monografieën van de verschillende Zeeuwse eilanden en gebiedsdelen, waarin aanvullingen op het WZD in een thematische ordening werden gepresenteerd in de vorm van verhalen en gedichten over de geschiedenis van het betrokken gebied. Alle nieuwe woordschat uit die regioboeken kwam achteraf in het Supplement op het WZD terecht.

De concentratie van tekentjes op de kaart is hoger in de Nederlandse provincies (behalve dus Zeeland) dan in de Belgische. Nederlands-Limburg neemt het voortouw met 52 woordenboeken; Noord-Brabant volgt met 41 publicaties; dan komen de Belgische provincies met 20 voor Belgisch Limburg, 24 voor Oost-Vlaanderen, 17 voor Vlaams-Brabant, 16 voor West-Vlaanderen en 7 werken voor de provincie Antwerpen. Hoewel we in het licht van bovenvermelde beperkingen m.b.t. de tabel voorzichtig moeten zijn, komt er toch een patroon naar voren. Het lijkt duidelijk dat de amateurdialectlexicografie in Nederlands-

talig België later op gang komt en tot vandaag minder krachtig is dan in Nederland. Voor nagenoeg elk decennium scoren de twee Nederlandse provincies beter dan de vijf Belgische (één keer, in 1980–1990 scoren Noord-Brabant en Belgisch Limburg gelijk).¹⁰ Dat kan verklaard worden door het feit dat de traditionele dialecten in Vlaanderen langer bleven doorleven dan in Nederland, waardoor de amateurlexicografen later gemotiveerd werden om een dialectwoordenboek te schrijven. Ook is het zo dat bij de massale invoering in de jaren '60 van de Nederlandse standaardtaal in de Vlaamse scholen via de zgn. ABN-campagnes,¹¹ de strijd voor het Algemeen Nederlands meestal verward werd met een strijd tegen de dialecten. De dialecten werden gestigmatiseerd als minderwaardige taalvariëteiten, die geen ernstige aandacht waard waren. Tegenwoordig heeft men zowel in Vlaanderen als in Nederland tegenover de dialectische (en andere) taalvariëteiten een meer ontspannen en tolerantere houding.

De concentratie aan woordenboeken in de Limburgse provincie van beide landen is opvallend: samen 72 lexicografische producten. Die drukke bedrijvigheid is ongetwijfeld te verklaren door het regionale taalbewustzijn van de Limburgers — ook over de staatsgrens heen — dat zich ook vertaalt in een goed georganiseerd verenigingsleven op dialectgebied en de erkenning van het Limburgs als 'streektaal' (althans in Nederland, niet in België). In de andere provincies van Nederlandstalig België kunnen we geen klaar geografisch patroon zien. Er blijkt een concentratie te zijn in het oosten van Oost-Vlaanderen, die misschien te maken heeft met een relatief hoog dialectbewustzijn: het Vlaamse dialectgebied (Frans-, West- en westelijke 2/3 van Oost-Vlaanderen) en het Brabantse (Vlaams-Brabant, Antwerpen, Noord-Brabant) raken daar elkaar en vormen een isoglossenbundel die een aantal markante taalverschijnselen bevat. De relatief sterke dialectische differentiatie in dat gebied heeft misschien meegebracht dat de gemiddelde aandacht voor taal bij de bevolking hoger is dan elders. Op andere plaatsen is het voorkomen van een plaatselijk woordenboek waarschijnlijk gewoon te danken aan de toevallige motivatie van een dialectliefhebber of een culturele vereniging. Er zijn meer woordenboeken voor grote dan voor kleine gemeenten (al zijn er uitzonderingen), wat uiteraard te maken heeft met de grotere bevolking en met de aanwezigheid van (sterkere) culturele verenigingen die een lokaal lexicografisch initiatief kunnen nemen.

Het bestaan van een ouder regionaal dialectwoordenboek blijkt het plaatselijke enthousiasme niet te kunnen temperen. De grote 19de-eeuwse woordenboeken voor bijv. West-Vlaanderen (De Bo 1873) of het Waasland (= noord-oostelijk Oost-Vlaanderen) (Joos 1900) waren zeer uitgebreid en — de tijd waarin ze ontstaan zijn in acht genomen — van uitstekende kwaliteit. Ze werden overigens op het einde van de 20ste eeuw herdrukt (ook een teken van dialectrenaissance!) en waren dus opnieuw beschikbaar. Toch ontstonden net in het Waasland de uitvoerige en kwaliteitsvolle woordenboeken van Pieters (1995) voor de stad Lokeren en Cools (2000) voor Beveren. Beide auteurs zijn

overigens op de gelukkige gedachte gekomen om de hulp van de universiteit in te roepen bij hun onderneming. In hetzelfde dialectgebied kreeg ook het dorp Belsele (nu deelgemeente van de stad St.-Niklaas) zijn woordenboek met De Belie 2006). Het *Waausch Idioticon* van Joos heeft blijkbaar veeleer stimulerend dan remmend gewerkt op de plaatselijke amateur.

2.2 Macro- en microstructuur in de amateurwoordenboeken

2.2.1 Macrostructuur

De macrostructuur van een plaatselijk dialectwoordenboek wordt in eerste instantie bepaald door de motivatie en taalattitude van de amateurlexicograaf, meer dan door de gebruikersbehoeften. Die motivatie is — al dan niet uitdrukkelijk in de inleiding van het woordenboek meegedeeld — in alle gevallen het vastleggen voor het nageslacht van de met verdwijnen bedreigde dialectwoordenschat, die veelal als typisch voor de plaatselijke gemeenschap wordt beschouwd. Uit heel wat inleidingen blijkt dat de auteur de teloorgang van het dialect betreurt, die evolutie wel onvermijdelijk vindt, maar toch een poging wil ondernemen om tenminste een deel van de taalidentiteit van de lokale gemeenschap te redden. Alle lokale woordenboeken zijn dan ook in principe *contrastief* opgezet;¹² woorden die enkel fonologisch van het Algemeen Nederlands verschillen, krijgen normaal geen aandacht, tenzij ze dialectische betekenissen dragen. Veel amateurs zijn erg op zoek naar wat 'typisch' is voor het eigen dialect, en zullen dan ook curiosa of erg oude dialectwoorden koesteren.

Nagenoeg alle amateurwoordenboeken zijn alfabetisch geordend. Er zijn erg weinig uitzonderingen: maar een paar titels duiden op een andere aanpak, als het *Thematisch Woordenboek van het Tungelroys* (Kooijman 1985) of het woordenboek voor Reusel van Van Gompel (2002–2006), dat in verschillende delen is opgedeeld: De mens, de wereld, enz. ... Wel is het zo dat een aantal werken verhalen bevat waarbij de dialectwoordenschat omtrent een bepaald onderwerp wordt aangebracht door een volkscultureel onderwerp (bijv. streekgerechten) te beschrijven. Dat is bijv. de werkwijze van de Zeeuwse regioboeken. De alfabetische ordening wordt ongetwijfeld als vanzelfsprekend ervaren doordat de meeste auteurs enkel met dat type woordenboeken vertrouwd zijn, hoewel de doelstelling die uit heel wat inleidingen spreekt — een beeld schetsen van het leven van vroeger — een thematische ordening aannemelijker maakt. Ik denk dat een plaatselijk dialectwoordenboek, een werk dat stevast gretig aftrek vindt bij de lokale bevolking, slechts zelden als naslagwerk gebruikt wordt, maar veeleer als een boek waarin gegrasduind wordt op zoek naar mooie, oude woorden of pittige uitdrukkingen.

De macrostructuur van een amateurwoordenboek kan nog heel wat meer bevatten dan alleen maar woorden. Vele auteurs zijn bijv. tegelijk ook amateurhistorici of hebben ook andere culturele belangstellingen, die ze in hun

woordenboek op een of andere manier kwijt proberen te raken. Een aantal woordenboeken bevat bijv. familienamen, bijnamen, rijmpjes en kinderliedjes ... die eventueel in appendices worden ondergebracht, samen met lijsten van Franse leenwoorden, scheldwoorden en vloeken, spreekwoorden en zegswijzen. Sommige auteurs kunnen daarbij maar moeilijk maat houden.

Wat de macrostructuur van de amateurwoordenboeken van na de jaren '80 van de vorige eeuw betreft, kunnen we samenvattend stellen dat vanuit een conserverende motivatie en de vertrouwdheid met alfabetische standaardtaalwoordenboeken, het allergrootste deel van de amateurlexicografie de klemtoon legt op het contrastieve deel van de woordenschat van de oudste taallaag in het dialect, in een alfabetische volgorde.

2.2.2 De microstructuur

De microstructuur van de meer dan 177 woordenboeken in de bibliografie van Van Keymeulen e.a. (2007) loopt uiteraard zeer uiteen, naargelang van de opzet/omvang van het woordenboek en de (taalkundige) scholingsgraad van de auteur. Nagenoeg alle woordenboeken bevatten dialectische trefwoorden en betekenisomschrijvingen met het Standaardnederlands als metataal; ze kunnen dus beschouwd worden als vertaalwoordenboeken dialect → Algemeen Nederlands.¹³ De meeste plaatselijke woordenboeken proberen ook de dialectische uitspraak op een of andere manier in de microstructuur weer te geven.¹⁴ In wat volgt, beperk ik mij tot de vorming van de trefwoorden en de manier waarop de dialectische uitspraak wordt weergegeven.

Er zijn in principe twee manieren om dialectische trefwoorden te vormen in een alfabetisch geordend woordenboek: de eerste manier is een dialectspelling te ontwerpen en de trefwoorden in die spelling neer te schrijven; de tweede manier is de standaardtaal als referentie te gebruiken en de dialectwoorden te 'vernederlandsen'. Met 'vernederlandsen' wordt bedoeld dat de fonemen van het dialectwoord worden vervangen door hun AN-pendant en vervolgens volgens de Nederlandse spellingregels worden neergeschreven. De dialectuitspraak kan dan na het trefwoord toegevoegd worden, in IPA of op een andere manier. Een dergelijke werkwijze werkt uiteraard het best als het dialectwoord in het AN een lexicale pendant heeft (eventueel met een andere betekenis), zoals in bijv. dialect [o.ʔrdəχ], dat vernederlandst kan worden als <aardig>. In de meeste zuidelijke dialecten betekent het woord *aardig* 'eigenaardig'; in het AN echter 'lief'. Die werkwijze blijkt echter ook te werken voor het formeel-contrastieve deel van de dialectwoordenschat. We gaan er immers vanuit dat de dialectsprekers, die heden ten dage alle het AN tenminste passief beheersen, hun dialectwoorden intuïtief kunnen vernederlandsen. In elk geval blijken de respondenten tot op grote hoogte in staat woorden die wel in hun dialect, maar niet in het AN voorkomen, 'op z'n Nederlands' neer te schrijven. Met de vernederlandste dialectwoorden worden als het ware de intuïtieve 'woordbeelden' gevat die bij een dialectspreker aanwezig zijn. Die 'woordbeel-

rang zn. v.; mv. -en; dim. -sken: rij. (school) *In de rang staan in de rij staan. Een gieël rang schoolkadéjjen.* [rang].

rank zn. m.; mv. -en; dim. -sken: 1. hopperank. *Zoe lank a(l)s nen hoeprank.* (uitdr.) *Van de rank krâigen slaag krijgen.* (uitdr.) *Hè és de rank ni wêd veu(r) zèn âigen op t'hangen.* 2. (mv.) slimme streek, list. (uitdr.) *Ranken in zèn voor hémnen vol (heimelijke) streken zitten; (var.) ranken in zèn lâif hémnen; (var.) ranken inhémnen.*
z. rank.

ranseln regelm. ww.; ranseldn, geranseld: eventjes ter sprake brengen vooral om er meer over te weten; over iets in het voorbijgaan reppen. *Ranselt daa(r) neki naa(r).* 'k *Hém der neki van geranseld maa(r) hè zâ niks.* [ranselen; verw. met randen, ranten (praten)].

ransènjeern regelm. wederk. ww.; ransènjeedn, geransènjeed: inlichtingen inwinnen. 'k *Gon ma neki ransènjeern.* [renseigner (Fr.)].

rap (1) bn.: 1. vlug. (uitdr.) *Zoe rap a(l)s de blôrn dat wôn zo vlug als de blaren die waaien; (var.) zoe rap a(l)s tén!; (var.) zoe rap a(l)s iet.* 2. slim, schrander. *Ne rapen.* [rap].

rap (2) zn. v.; mv. -pen; dim. -peken: roof, korst. *Dê van As goenken naa(r) Blaazerus veu(r) teegen de rappen (KP).* [rap].

rapiet bn.: vlug. (zgw.) *Sâs rapiet* gezegd van eten waaraan bij gebrek aan tijd onvoldoende zorg was besteed. *Ne rapietn, een rapiete* een vluggerd. [rapide (Fr.)].

rappeloem zn. altijd mv.: (Mazenzele) volksnaam voor bosanemoon die de volkswijsheid kwalijke eigenschappen toeschreef; wie de anemoon aanraakte zou

rappen (korsten van wonden) op de ogen krijgen. [rap(pe)+bloem].

rappelen zn. o.: (gez.) *Op een rappeken* gezegd van snel cafébezoek. [rappeken].

rappelen regelm. ww.; rappeleedn, gerappeleed: 1. herinneren. *A(l)s ik het vergeet, goint ma rappeleern?* zul je er mij aan herinneren als ik het vergeet? 2. (wederk.) *Ik rappeleer ma daa(r) niks van!* ik herinner er mij niets van. [rappeler (Fr.)].

rapte zn. v.: gauwte. (samenspr.) A. *Gommen ieëne pakken?* B. *Allêi, in de rapte!* [rapte].

râr zn. m.; mv. -s: mannetjeskonijn. (samenspr.) A. 't *Ês verniet!* B. *Verniet?* (spr.) *Verniet gon de katn naa(r) de râr!* (uitdr.) *Ne goeje râr mag ni vait staan!* [rijder; FFP 25.2.2].

ras zn. o.: *een vies ras* gemeen volk. [ras].

râschittink zn. v.; mv. -gen: (boogschutter) rouwschieting ter ere van overleden lid. [FFP 22].

rastieël zn. o.: ruif voor paarden. [rasteel; Ofr. rastelier].
z. hoeë.

rat zn. v.; alleen als dim. -teken: 1. snel, bijdehands kind. 2. (kleine) bijverdienste. *Hè gaat hie(r) èn daa(r) nog een ratteken doên* gezegd van gepensioneerde die ergens nog wat gaat bijverdienen. 3. (dim.) koosnaam voor meisje. [ratteken].

ratatoeil zn. m.: 1. etensresten die bij mekaar geklutst zijn. 2. minderwaardig volk, plebs. [ratatouille (Fr.); FFP 8].

rateern regelm. w.; rateedn, gerateed: missen. 'k *Hém main tram gerateed!* *Wâln hémnen makannere gerateed!* *Dat gewêir*

ze de mèlk u:t euren ieëmer deur 'n mèlkzei in de mèlkkiet. - Synoniemen: **melkzeef** [mèlkzeef] en **melkzift** [mèlkzift].

Niet in KA, ME; in KI: [mèlkzè:].

Melsele [Meiëlsen]: Melsele, een der acht deelgemeenten van Beveren. *Ik zin i Meiëlse gebooren, èn ik voel mi noch àlt wa ne Meiëlseneiër, è pertàng ik wieën der àl guuël lâ:nk nemiëër... Moaër já:, 't terp woaërdache geboore zit, da trèkt ee... - Afleidingen: Melsenaar* [Meiëlseneiër]; **Melsenas** [Meiëlsenas]; **Melses** [Meiëlses].

In DO, KA: [Mèlsen]. - Afleidingen: in DO: [Mèlsenèr, Mèlsenas, Mèlses]; in KA: [Mèlseneiër, Mèlsenas, Mèlses]; in KI: [Mè:lsenjè:ër, Mè:lsenas, Mè:lses]; in VE: [Meiëlsenjeiër].

mem [mèm]: vrouwenborst. *Azoe ne vaveur, die ee nogàl mèmnen.* In DO ook: [mam].

memorie [meemoorie]: geheugen. *'k Weet nie wa ta ta is, moaër min meemoorie is àl neich vermindert, da's à:t woaëre zeeker?...* In KI: [memoore].

meneer [menieër]: mijnheer. - Uitdrukkingen: *ne groeëte menieër* = iemand die het ver geschopt heeft; *ne schoeëne menieër* = een mooie mijnheer (met pejoratieve bijklank). - Afleiding: **meneren** [menieëren]: iemand met "mijnheer" aanspreken. *Ge moet azoeë àlt nie menieëren, ge wit toch oe da'k ieët ee, è 'k zi kik oeëk moaër u:t 'n warkmansbroek geschut...*

menen [mieënen]: denken, voor waarheid aannemen, voor mening hebben. *Ik mieën dache da verkieërt veur èt.* - Uitdrukkingen: *da mieënde nie* = wat je daar vertelt is toch wel niet waar zeker?; *ik mieënet* = ik lach er niet mee, het is ernst; *Mieënen is 'n stat, èn die lich vaar van o gat, è ge kunter nie*

op zitten: repliek op een bewering die werd ingeleid met *ik mieën da...*, waarmee gezegd wil worden dat men die bewering niet gelooft of bijtreedt. - Afleiding: **menens** [mieënes]: gemeend. *Is da mieënes?...*

In HA ook: [maanen]. - Afleiding: in HA ook: [maanes].

mengen [mèngen, mingen]: 1. bemoeien, tussenkomen. - Uitdrukking: *Mingd o: doaër nie mee* = bemoei je daar niet mee. - 2. synoniem van **mengelen** [mèngelen, mingelen] (zie aldaar).

mengelen [mèngelen, mingelen]: mengen, onder elkaar doen. *Ache zwart è wit mèngelt dè krijchde grijs.* - Afleiding: **mengeling** [mèngeling, mingeling]. - Synoniem: **mengen** [mèngen, mingen].

mengelwerk [mingelwaarek]: vervolgverhaal in dag- of weekblad. In "De Wo:se Klok" *stienger in den tijt àltit e mingelwaarek, va Konsjâ:ns of azoe-iet, da lâ:zen de meinse geieën.* Niet in DO, KI, VE, VR.

mennegat [mènnegat]: 1. plaats via dewelke men (desgevallend over ingebuisde sloot) op en af de akker rijdt. *Wi zummen ieës wa stieëngru:s op da mènnegat moette kappen, want da's doaër ieëne slijkboel.* - 2. synoniem van de akker zelf bij een eventuele telling van het aantal akkers. *Witte gei oevvel mèn-nego:ten da tieëm boe:r wèl eet!* - Zie ook: **dam** [dam] en **dam(me)gat** [dam(me)gat].

In HA, ME, VE, VR: [minnegat]. - Betekenis 2: niet in DO, KI.

mennen [mènnen]: de oogst binnenhalen. *'t Is ie:r druk in de poolder no:, de boe:re zin on 't mennen.* - Zie ook: **oogsten** [oesten].

mens [meins]. - Uitdrukkingen: *àlle meinse groeët* = zeer groot; *ammoaër*

den' komen tot stand op basis van de synchrone correspondentieregels die tussen het dialectische fonologische systeem en dat van de Nederlandse standaardtaal bestaan. Het is hier echter niet de plaats om uitvoering uit te weiden over vernederlandsingen (en de problemen daarbij); ik verwijs hiervoor naar Van Keymeulen 1992 en 2004 en naar Rys en Van Keymeulen 2009. In wat volgt, wil ik vooral wijzen op een opmerkelijke, typisch 'volkslinguïstische' tendens, die zo sterk is dat het impliciete of expliciete doel van het lokale woordenboek zelfs voorbijgeschoten wordt.

Zeer veel amateurs stellen er namelijk prijs op dialectische trefwoorden te vormen in een zelfgemaakte dialectspelling. Het trefwoord ziet er dan erg 'authentiek' uit; de uitspraak van het woord wordt in de vorm van het trefwoord zelf aangegeven. Een spelling ontwerpen voor een dialect is echter niet makkelijk; het vergt inzicht in de fonologische structuur ervan — een inzicht dat bij veel amateurs door gebrek aan scholing afwezig is. Een en ander loopt dan ook soms verkeerd af. Voor Brabantse of Vlaamse dialecten kan een amateur zich nog min of meer uit de slag trekken; de Limburgse toondialecten bijv., met soms meer dan 40 verschillende klinkerfonemen, kunnen echter voor zeer veel problemen zorgen. Er zijn ten behoeve van dialectliefhebbers dan ook richtlijnen geschreven over hoe men een dialectspelling tot een goed einde kan brengen; de bekendste voor de zuidelijke dialecten is waarschijnlijk de *Referentiespelling* voor alle Brabantse dialecten (1999) (zie echter ook Cajot 1995 en Van Keymeulen 2003).

Hoewel de bestaande richtlijnen hopelijk een gunstig effect hebben gehad op de kwaliteit van de dialectspellingen, valt het op dat veel auteurs toch een of andere persoonlijk toets willen geven aan de spelling. Ze kennen de bestaande richtlijnen wel, maar wensen er toch aan te sleutelen. Die neiging vloeit m.i. vooral voort uit de behoefte zich het dialect a.h.w. 'toe te eigenen' via het ontwerpen van de spelling ervoor. Het schrijven van een woordenboek wordt blijkbaar ook voor een dialect aanvoeld als een codificerende activiteit — een aantal auteurs drukt overigens de wens uit dat de spelling ook zal kunnen dienen bij het schrijven van dialectliteratuur.

Opmerkelijk is dat heel wat amateurs er niet bij stilstaan dat een dialectisch trefwoord in een zelf ontworpen dialectspelling impliceert dat men het dialect goed moet kennen en kunnen uitspreken, voor men een woord kan opzoeken — aangenomen dat de gebruiker de moeite genomen heeft om de uitleg omtrent de soms complexe spelling in de inleiding op het woordenboek te lezen en te memoriseren. Veel auteurs schrijven dus een woordenboek voor degenen die er het minste behoefte aan hebben, nl. de goede dialectsprekers, die uit de aard der zaak zowel het woord als de uitspraak ervan nog beheersen. De dialectwoordenschat is dan wel 'voor het nageslacht' bewaard, maar is helaas niet voor dat nageslacht ontsloten. Daarnaast zijn er ook nog andere nadelen aan de werkwijze verbonden: niet alleen worden de gebruikers geconfronteerd met zeer ongewone schriftbeelden, zeker als er veel diakritische tekens gebruikt worden, maar ook is de alfabetische volgorde van de trefwoor-

den niet steeds evident. Een woord opzoeken is in sommige plaatselijke woordenboeken m.i. ook voor de dialectspreker zelf geen eenvoudig karwei.

Een minderheid van amateurwoordenboeken vernederlandst de dialectwoorden. Daardoor zijn de trefwoorden toegankelijk voor een grotere groep gebruikers (bijv. die van de veronderstelde dialectloze toekomst) doordat ze beter aansluiten bij de schriftbeelden die men gewend is uit de standaardtaal. Informatie over de uitspraak kan overigens gemakkelijk in de microstructuur toegevoegd worden, bijv. in IPA-schrift, al zullen amateurs ook dan geneigd zijn terug te grijpen naar zelfgemaakte spellingen. De procedure veronderstelt echter inzicht in de klankstructuur van het dialecten en inzicht in de etymologie van het betreffende dialectwoord, iets wat wel eens boven de kracht gaat van ongeschoolde lexicografen. De intuïtieve correspondentieregels die hierboven werden geponeerd, zorgen er echter wel voor dat de weinige plaatselijke woordenboeken die de dialectwoorden als vernederlandste trefwoorden presenteren het er toch vrij goed vanaf brengen. Een mooi voorbeeld is Cools (2000), die zich tot vernederlandsingen verplicht zag, aangezien zijn woordenboek een (klein) gebied bestrijkt, waar toch heel wat fonologische differentiatie voorkomt.

De meeste woordenboeken gebruiken dus dialectwoorden als trefwoorden, in al dan niet vernederlandste vorm. Het is markant dat veel auteurs niet inzien dat op die manier het dialect voor de toekomst niet ontsloten wordt. Men moet in beide gevallen het dialect immers kennen om een woord op te zoeken. De handleiding van Van Keymeulen (2003) probeert de aandacht op die tegenstrijdigheid te vestigen en roept op om het woordenboek via een AN → dialectregister om te draaien,¹⁵ zodat het woordenboek ook een antwoord geeft op de vraag 'Hoe zeg ik x in het dialect'. Ook in het geval van dialectwoordenboeken is een grondige reflectie over de noden van de gebruiker een noodzakelijke voorwaarde om tot een goed woordenboek te komen.

3. Het belang van de amateurlexicografie voor de wetenschap

De plaatselijke dialectwoordenboeken vormen een bont gezelschap. De overgrote meerderheid ervan is geschreven door taalkundig ongeschoolde lexicografen. Er zijn echter ook woordenboeken van hoge kwaliteit, die bewijzen dat de lexicografie binnen het bereik ligt van verstandige dialectliefhebbers, die zich willen laten voorlichten (o.a. door de richtlijnen van de *Referentiespelling* (1999), Cajot (1995) of Van Keymeulen (2003)), die de grenzen van hun kunnen inzien en die zich laten leiden door het voorbeeld van een bestaand goed woordenboek. Sommige woordenboeken zijn de vrucht van tientallen jaren werk en dwingen door hun omvang respect af voor de werkkraft van de lexicograaf. In wat volgt, probeer ik aan te geven wat het wetenschappelijke belang kan zijn van die 177 woordenboeken. Eerst heb ik het over de opzet (§ 3.1) en de beperkingen (§ 3.2) van de drie grote regionale dialectwoordenboeken van het zuidelijke Nederlands; daarna doe ik een voorstel om de gegevens in de amateurlexicografie te ontsluiten (§ 4).

3.1 De drie grote regionale woordenboeken (GRW) van het zuidelijke Nederlands

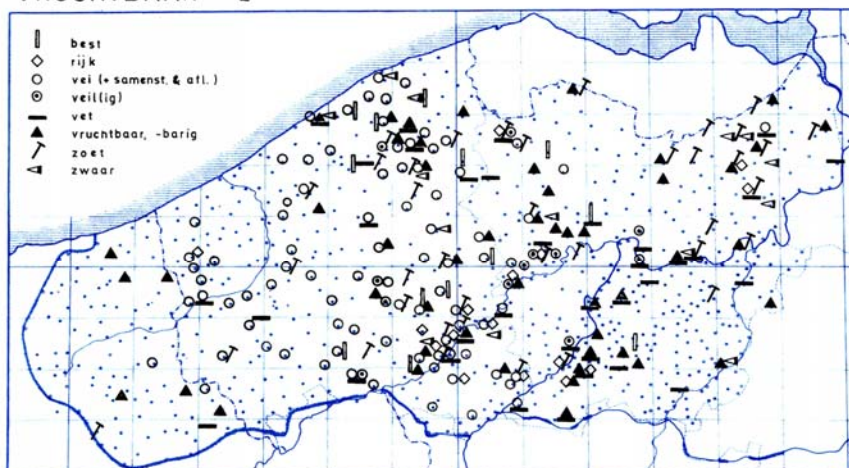
De opzet van de drie grote dialectwoordenboeken van het zuidelijke Nederlands was reeds het onderwerp van een artikel door Kruijssen en Van Keymeulen in *Lexikos* (1997). We vatten zeer kort samen. In 1960 nam prof. A. Weijnen aan de Nijmeegse universiteit het initiatief om de dialectwoordenschat te verzamelen van het Brabantse en Limburgse dialectgebied, d.i. resp. de dialectwoordenschat van de provincies Noord-Brabant (Nederland), Antwerpen en Vlaams-Brabant (allebei België), en de provincies Nederlands en Belgisch Limburg. Het *Woordenboek van de Brabantse Dialecten* (WBD) werd in 2005 afgerond; het *Woordenboek van de Limburgse Dialecten* (WLD) in 2008. In 1972 volgde prof. W. Pée aan de Universiteit Gent Weijnens voorbeeld met het *Woordenboek van de Vlaamse Dialecten* (WVD) voor Frans-, West-, Oost- en Zeeuws-Vlaanderen. De drie woordenboeken bestrijken drie geografisch complementaire gebieden en zijn alle drie volgens hetzelfde principe opgezet.

Weijnen koos voor een onomasiologische ordening.¹⁶ De 3 GRW zijn dus thematisch opgezet: binnen drie delen (I. Landbouwwoordenschat, II. Vakwoordenschat en III. Algemene Woordenschat) worden afleveringen gepubliceerd die telkens aan een bepaald onderwerp zijn gewijd: bijvoorbeeld 'Landrijen en Behuizing', 'Het paard', 'Flora' ... Elke aflevering bestaat uit een reeks woordenboekartikelen die alle dezelfde structuur hebben (zie de voorbeeldbladzijde hierna): na een titel in de standaardtaal volgt een omschrijving van het behandelde begrip (+ eventueel nog andere toelichtingen), de bronnenlijst en dan de reeks vernederlandste trefwoorden (met vanaf deel III eventueel nog lexicale varianten), gevolgd door aanduidingen omtrent frequentie en geografisch bereik van het woord. Het WVD heeft van bij het begin elke aflevering in twee parallelle publicaties uitgebracht: een woordenboektekst voor het 'grote publiek' en een parallel 'Wetenschappelijk Apparaat' (met dezelfde informatie, maar dan meer gedetailleerd en in code) voor de wetenschapper (voor meer details: zie Kruijssen en Van Keymeulen 1997). De drie woordenboeken zijn alle gebaseerd op (schriftelijk) materiaal dat door de redactie zelf verzameld is (het zgn. 'eigen materiaal') en op gegevens uit oudere vragenlijsten en gepubliceerd materiaal (het zgn. 'vreemde materiaal').

Zowel WBD/WLD als WVD putten ook gegevens uit de plaatselijke amateurwoordenboeken; het WVD beperkt zich daarbij van bij het begin tot vier regionale dialectwoordenboeken (nl. De Bo, Joos, Teirlinck en vanaf WVD I,4 ook het WZD) en twee grote stadswordenboeken (Desnerck voor Oostende en Lievevrouw-Coopman voor Gent). Het WBD en WLD maken meer gebruik van de amateurlexicografie; vanaf deel III hebben de redacties een lijst van zgn. 'canonieke' (d.i. 'goedgekeurde') lokale woordenboeken opgesteld, waaruit gegevens gehaald werden wanneer het andere materiaal wat zwak uitviel (zie de inleidingen op deel III). De informatie van de plaatselijke (of regionale) woordenboeken wordt door de 3 GRW zeer onvolledig verwerkt. Lang niet alle

— 1.1.0.1.0.

VRUCHTBAAR I



VRUCHTBAAR

[Fertile, productief — fertile, productive — fruchtbar, ertragreich.]
Goed geschikt om gewassen te telen, een goede opbrengst gevend.

WVD 1, 23; 4, 24-25-55.

I. VRUCHTBAAR (Kaart).

best: het type BEST(E) LAND, — GROND, — PARTIJ e.d. s p o r. WV oost en OV west.

dragend: s p o r. Meetjesl *dròòant*, ook Lede *gòet dròògant*.

eerst: Noordpene *eeestè lant*.

fel: s p o r. FV, ook Eeklo.

fertil: Ebblingem *firtiel*.

goed: het type GOED LAND, GOEDE GROND, GOED STUK e.d. alg. WV, OV; f r e k w. FV (zfd).

groeizaam: Destelbergen (zfd).

jeugdìg: z e l d z. WV oost, ook St.-Martens-Lierde.

kweekzaam: Gent (zfd).

mals: Mater.

open: St.-Pieters-op-den-Dijk.

rijk: f r e k w. omg. Kortrijk *riekè, rikè*; z e l d z. OV *raek, rek, reik*; ook Wulveringem, Vinken.

vei: alg. WV; z e l d z. FV *vei*, Meetjesl en omg. Nevele *vaej*.

Wdb : DB.

veiachtig: Ruiselede.

veilig: z e l d z. WV *veiech, veiech, veigè*, ook St.-Gillis-Waas *vaejgè gront*.

veil: s p o r. WV oost en OV west *veil, vaejl*.

Wdb : Loq : St.-Martens-Lerne.

veilig: Melle, Eine (zfd).

veischot: Zedelgem, Ieper *veischotè*.

veischotìg: Snaaskerke *veischootèch*.

veischottìg: z e l d z. WV *veischotèch, veischotiech*, ook Vinkt *vaejschotèch*.

veizaam: Oostkamp *veizòòm*.

Wdb : DB.

veizamìg: Oostkamp *veizòòmèch*.

vet: f r e k w. OV; s p o r. WV; ook Meteren.

vettìg: Melle, Lede.

vruchtbaar: f r e k w. OV en WV oost; s p o r. FV *vrugbòòr*.

vruchtbarìg: s p o r. FV *vrugbòòrèch, vrugbòòriech*; ook Sinaai, Meldert *vrugbòòrèch*.

weeldìg: Alveringem, Poperinge *wil-diech*.

zoet: s p o r. WV, OV; ook Ebblingem *zòet, zoet*.

17

Een voorbeeldbladzijde van de 'woordenboektekst' uit WVD I,1: 17¹⁷

woorden werden geëxcerpeerd en de redactie beperkte zich doorgaans in voorkomend geval tot de mededeling van de attestatie van de woordvorm (soms met een korte semantische toelichting).¹⁸

3.2 Beperkingen op lexicaal en semantisch gebied van de 3 GRW

De onomasiologische aanpak van Weijnen, waarbij hij uitging van 'begrippen' en daarbij dan per begrip de dialectwoorden vermeldde die bij het refereren aan dat begrip gebruikt konden worden, is erg vruchtbaar gebleken op geografisch gebied. Aangezien alle dialectwoorden voor één begrip netjes bij elkaar worden gepresenteerd, ligt het materiaal klaar om een taalkaart te tekenen. Op semantisch vlak echter heeft de onomasiologische ordening belangrijke beperkingen veroorzaakt. In de jaren '60 van de vorige eeuw hebben Weijnen en De Tollenaere, hoofdredacteur van het WNT, een boeiende wetenschappelijke discussie gevoerd omtrent de vraag of het mogelijk is 'betekenis' weer te geven in een onomasiologisch georganiseerd woordenboek (zie Weijnen 1961, 1963, 1967, De Tollenaere 1960, 1968). Volgens De Tollenaere is dat enkel mogelijk in een semasiologisch woordenboek, waarbij men uitgaat van het woord zelf, en daarvan in de microstructuur de betekenis(sen) beschrijft. In onomasiologisch ingerichte woordenboeken als WBD/WLD/WVD kunnen enkel gebruikswijzen van de woorden meegedeeld worden, geen betekenissen. De Tollenaere (1968: 199 vlg.) zegt hierover: "De ideologische lexikografie kan dus principieel geen bemoeienis hebben met wat men betekenis noemt. Haar beoefenaars zijn in dit opzicht te benijden".

De uitslag van de discussie was dat Weijnen zich bij de visie van De Tollenaere aansloot. In de 3 GRW kan noch de titel van een woordenboekartikel (in ons voorbeeld VRUCHTBAAR), noch de semantisch-encyclopedische omschrijving die daarop volgt (in ons voorbeeld 'Goed geschikt om gewassen te telen ...'), in principe als een betekenisomschrijving voor alle daarop volgende dialectwoorden beschouwd worden. Titel en omschrijving duiden enkel op een begrip waaraan door de daarna opgelijste woorden gerefereerd kan worden; die woorden kunnen echter een ruimer toepassingsbereik hebben (niet een enger!) dan dat dat door de titel of omschrijving wordt aangegeven. Dat is in ons voorbeeld voor zeer veel woorden het geval: het is duidelijk dat o.a. *best*, *goed*, *rijk* en *vet* een ruimer toepassingsbereik hebben dan enkel maar 'vruchtbaar'. Pas wanneer het woordenboek af is, en men via de alfabetische registers kan opzoeken in welke woordenboekartikelen een bepaald woord voorkomt, kan men een beter zicht krijgen op de betekenisinhoud van dat woord. Kortom, hoewel de 3 GRW het woord 'woordenboek' in hun titel dragen, zijn het vooral onomasiologisch geordende materiaalverzamelingen die als bouwstoffen voor een woordenboek kunnen dienen.

De 3 GRW zijn door de ordening van de macrostructuur dus in de eerste plaats woordatlassen; ze geven in principe geen betekenissen weer. Ze zijn op te vatten als thematisch geordende thesauri van gebruikswijzen van dialect-

woorden in de zuidelijke dialecten van het Nederlands. Die geografische focus, samen met de onomasiologische opvraagtechniek, heeft in de praktijk ook meegebracht dat de klemtoon ligt op de woorden voor concreta; woorden met abstracte betekenis zijn zowel moeilijk op te vragen als moeilijk in onomasiologische woordenboekartikelen in te delen.¹⁹ De 3 GRW zijn dan ook vooral rijk aan substantieven en werkwoorden, in mindere mate aan adjectieven. Woorden van de gesloten woordklassen ontbreken nagenoeg volledig.

3.3 De bijdrage van de amateurlexicografie

Uit het bovenstaande is ongetwijfeld duidelijk geworden dat de 3 GRW de dialectwoordenschat van de zuidelijk-Nederlandse dialecten niet 'in zijn volheid' inventariseren. Niet alleen is er een semasiologische benaderingswijze nodig om de betekenis van woorden weer te kunnen geven, maar ook is de microstructuur in de 3 GRW relatief arm aan informatie. Behalve de begripsomschrijving, krijgen we op werkelijk systematische wijze enkel trefwoorden en de lokalisatie daarvan. Het is ook de bedoeling (geweest) om van elk dialectwoord de uitspraak weer te geven. De behandeling van de fonetische component verschilt tussen WBD/WLD en WVD; WVD heeft overigens voor deel III zeer sterk op de fonetische component bespaard (voor details hierover: zie Kruijsen en Van Keymeulen 1997). De microstructuur bevat geen morfologische of syntactische informatie. De 3 GRW zijn ook zeer arm aan collocaties — er zijn bijv. geen voorbeeldzinnen; spreekwoorden of uitdrukkingen worden overigens niet systematisch opgevraagd.

Gelukkig zijn er dus de amateurwoordenboeken, die nagenoeg alle semasiologisch geordend zijn en waarvan er zeer veel een heel wat uitgebreidere microstructuur hebben dan de 3 GRW. Hoewel de kwaliteit van de amateurwoordenboeken heel sterk kan variëren, leveren ze toch een schat aan additionele gegevens bij de 3 GRW. Zelfs de kwalitatief mindere woordenboeken kunnen aan de wetenschap nuttige informatie leveren omtrent het voorkomen van een bepaald woord, met tenminste een aanduiding omtrent de betekenis ervan. Zaak is dat de gevestigde lexicografie de amateurs ondersteunt en begeleidt. Die begeleiding heeft twee aspecten: lexicografische basisnoties bij de amateurs aanbrengen en hen hun wetenschappelijke beperkingen (op het vlak van fonetica en historische taalkunde) leren aanvaarden.

4. Naar een Woordenboek van de Nederlandse Dialecten (WND)

Een digitaal doorzoekbare database die de gegevens van alle amateurwoordenboeken zou bevatten, zou dus een zeer welkome aanvulling zijn bij de gegevens van de 3 GRW. Voor het zuidelijke Nederlands is er nu een bibliografie voorhanden (Van Keymeulen e.a. 2007); ik betwijfel of er een bibliotheek is waar de hele verzameling bijeenstaat. Heel wat van de plaatselijke woordenboeken zijn immers in eigen beheer en maar in beperkte oplage uitgegeven —

ze zijn daardoor soms moeilijk op te sporen of te verwerven. Het inscannen en verder bewerken van de papieren teksten vormt technologisch geen onoverkomelijke problemen — eventueel zijn er wel auteursrechtelijke hindernissen te overwinnen.²⁰

Een database met daarin alle artikelen van de verschillende regionale en lokale dialectwoordenboeken als aparte bestanden zou wel in elk geval verrijkt moeten worden met vernederlandsste trefwoorden (het liefst met aanduiding van de morfologische geleiding en van de woordsoort) om een efficiënt gebruik mogelijk te maken — een werk dat noodzakelijkerwijs manueel zal moeten gebeuren. Ideaal zou zijn de bestanden ook via de Nederlandse standaardtaal toegankelijk te maken door het vernederlandsste dialectwoord ook naar het Algemeen Nederlands te vertalen en/of het te voorzien van een thematische markering — in beide gevallen uiteraard op grond van een goed overdacht en eenvormig systeem.²¹

Voorbeeld uit: Pieters 1995: 319 (*Woordenboek van het Lokers dialect*):

Oorspronkelijk woordenboekartikel:

slapersmisse: de zondagse elfuurmis, vroeger de laatste misviering op zondag.
De slapersmisse dè wār veur de dee die dèn nie vroeg uit (h)odder bēde kosten. Zie ook *rijkemēnsemisse*.

Manueel toe te voegen verrijkingen:

slaper-s-mis (= vernederlandsing, met morfologische indeling)

zelfstandig naamwoord (= woordsoort)

elfuurmis (= vertaling in het Algemeen Nederlands)

religie (= thematische aanduiding / semantisch veld)

Het spreekt vanzelf dat er nog andere verrijkingen aangebracht kunnen worden, en ook kan de database verder verfijnd worden, naargelang van de aard van de microstructuur van de onderscheiden woordenboeken. De meest essentiële verrijking is wel de vernederlandsing van de dialectische trefwoorden — indien dat niet gebeurt, is de database ontoegankelijk en dus zinloos.

5. Conclusie

De geschiedenis van de Nederlandse (en Afrikaanse!) woordenschat kan maar geschreven worden wanneer we de beschikking hebben over inventarissen van drie soorten woordenschat: de woordenschat van de hedendaagse standaardtaal, de historische woordenschat en de uitsluitend oraal overgeleverde woordenschat. De inventarissen van de twee eerste soorten woordenschat worden vooral (en voor de historische periode uitsluitend) op schriftelijke bronnen gebaseerd. De registratie van de louter oraal overgeleverde woordenschat, de

woordenschat dus van de traditionele dialecten, is een dringende taak, wegens het verdwijnen (sedert de jaren '60 van de vorige eeuw) ervan onder de druk van de veranderende maatschappij en door de introductie van het AN in brede lagen van de bevolking.

De optekening van het lexicon van de dialecten is een gigantische taak, niet alleen wegens de uitgebreidheid en geografische differentiatie ervan — zowel op formeel als semantisch gebied — maar ook wegens de moeizame verzamelmethode. Alle gegevens moeten immers bij de oudste generatie dialectsprekers geëliciteerd worden, meestal via mondelinge of schriftelijke enquête. Hoewel de 3 GRW in termen van menskracht en financiën zeer grote ondernemingen zijn, hebben ze toch zeer veel moeten laten liggen — mede doordat de opzet van de woordenboeken veeleer taalgeografisch dan semantisch gericht was.

Gelukkig is er nu de dialectrenaissance. Om de registratie van de verdwijnende/reeds verdwenen Nederlandse dialectwoordenschat tot een goed einde te brengen, is het nodig dat de professionele dialectlexicografie een verbond sluit met de amateurlexicografie. De begeleiding van amateurlexicografen is m.i. niet enkel een vrijblijvende dienstverlening, maar is ook een zaak van wetenschappelijk belang. Het zijn de amateurs die de tijd en de motivatie hebben om plaatselijk een diepgaandere inventarisatie te verrichten dan een professionele dialectlexicograaf, die een heel gebied moet overzien, ooit zal kunnen doen. Zaak is om alle gegevens die door de amateurlexicografie zijn verzameld, in een efficiënt doorzoekbare database, een *Woordenboek van de Nederlandse Dialecten (WND)*, onder te brengen.

Eindnoten

1. Het typoscript voor dit artikel werd afgesloten eind 2007.
2. Zie onder andere Niedzielski en Preston (2000, 2003) en recenter werk als Preston (2005) en Preston en Niedzielski (2008).
3. België heeft drie officiële talen: Nederlands (voor de Vlamingen), Frans (voor de Walen), Duits (in de zgn. Oostkantons). Brussel is officieel tweetalig; 85 à 90% van de bevolking bekent zich echter tot de Franstalige cultuur (we maken abstractie van de talrijke immigranten met Europese of niet-Europese thuistalen). De numerieke verhoudingen zijn (afgerond): Vlaanderen 6 000 000; Wallonie: 3 400 000 (met inclusief de 75 000 Duitstaligen), Brussel 1 000 000.
4. Met de term zuidelijk-Nederlands proberen we het woord Zuid-Nederlands te vermijden. Zuid-Nederland is immers een lastige geografische aanduiding. Men kan er Nederlandstalig België mee bedoelen, het zuiden van Nederland of die twee gebieden samen.
5. Het valt op dat er nagenoeg geen vrouwelijke amateurlexicografen zijn. Mevr. Ghijsen was de uitzondering die de regel bevestigde.
6. In 1994 werd het WZD gedigitaliseerd o.l.v. Rinus Willemsen; in 1997 werd begonnen met het inlezen van de voorbeeldzinnen door een honderdtal dialectsprekers. Het WZD is met de

publicatie van de CD-rom in 1998 waarschijnlijk het eerste 'sprekende' dialectwoordenboek ter wereld.

7. Het Fries heeft in Nederland een erkenning op een 'hoger' niveau, nl. op basis van deel I van het Handvest.
8. Het kaartje is genomen uit Van Keymeulen e.a. 2007 en is van de hand van Ronny Keulen (met dank aan Valerie Bouckaert voor het overtekenen en aanvullen ervan). De twee sterretjes in het zuidoosten ten zuiden van de taalgrens staan voor Gemmenich en Eupen, twee Duitstalige gemeenten in de Belgische provincie Luik. Het kaartje laat geen aantallen zien, maar geeft aan of er voor een bepaalde plaats tenminste één woordenboek is.
9. Het dialect van de nu grotendeels verdwenen volkswijk 'Marollen', *Marols* of *Marollien* genoemd, zou het 'échte' Brussels geweest zijn.
10. Strikt genomen zou men bij een vergelijking ook het inwonersaantal/aantal gemeenten van de provincies moeten verrekenen; maar de verhoudingen springen ook nu al in het oog.
11. ABN staat voor *Algemeen Beschaafd Nederlands*. Het is tekenend voor de verandering in de tijdgeest dat men het woord *Beschaafd* tegenwoordig weglaat; de standaardtaal heet nu *Algemeen Nederlands* (AN). Men vond achteraf namelijk — overigens terecht — dat men ook in het dialect 'beschaafd' kan zijn.
12. De technische term voor een contrastief dialectwoordenboek is *idioticon* (afgeleid van het Griekse *idios* 'eigen'). Het woord wordt tegenwoordig nauwelijks nog gebruikt wegens de associatie met het woord *idiot*. Een zeldzame uitzondering is het woordenboek voor Tienen en Hoegaarden van Kempeneers (2004).
13. Een zeldzame uitzondering hierop is het woordenboek van het Hasselts van Staelens (1982), waar het AN-woord voorop gaat en in het Hasseltse dialect vertaald wordt, gevolgd door een betekenisomschrijving in het AN.
14. In woordenboeken die een relatief groot geografisch gebied bewerken, is het moeilijk om recht te doen aan de fonogeografische differentiatie. De 19de-eeuwse regionale dialectwoordenboeken geven de uitspraakvariatie niet weer (eerste verdienstelijke poging is Tuerlinckx 1886).
15. We gaan hier niet in op de problemen die dat omdraaien kunnen meebrengen. Voor zeer veel dialectwoorden is er meer dan één AN-vertaling mogelijk, waartussen het lastig kiezen is.
16. Weijnen liet zich daarbij wellicht inspireren door de beschrijvingen van het volksleven in een aantal dorpen in de drooggelegde Zuidzeepolders, waarbij de opgetekende woordenschat in een systematische opzet werd opgenomen (zie Goossens en Van Keymeulen 2006: 78).
17. De vertalingen van de titel in Frans, Engels en Duits werden vanaf WVD I,2 weggelaten.
18. Ik ga hier niet in op de verschillen tussen het WBD/WLD en het WVD in de redactionele behandeling van de gegevens uit bestaande woordenboeken.
19. De afleveringen bij het WVD over de karaktertrekken en de gevoelens (WVD III, 4 en 5) leggen getuigenis af van de worsteling van de redacteurs met onomasiologische indelingen van abstract-emotionele begripvelden.
20. Enkele proeven met de 19de-eeuwse woordenboeken hebben wel laten zien dat scannen en automatisch in een tekstbestand omzetten geen eenvoudige klus zal worden wegens de slechte druk kwaliteit van de oudere werken.
21. Ik ga hier niet in op de problematiek m.b.t. het vertalen van de dialectwoorden in het AN of het onderbrengen van de betekenissen in thematische betekenisvelden. De concrete uitwerking van een en ander is overigens ongetwijfeld erg complex.

Referenties

- Cajot, J.** 1995. *Hoe maak ik een dialectwoordenboek. Een handleiding voor Limburgers en anderen die dialectwoorden willen spellen, verzamelen en beschrijven*. Mededelingen van de Vereniging voor Limburgse Dialect- en Naamkunde nr. 78/79. Hasselt: Vereniging voor Limburgse Dialect- en Naamkunde.
- Cools, H.** 2000. *Het Dialect van Beveren en zijn deelgemeenten. Met toelichtingen over afwijkingen in Doel, Haasdonk, Kallo, Kieldrecht, Melsele, Verrebroek en Vrasene door Abdon Van Bogaert*. Beveren: Gemeentebestuur.
- Cornelissen, P.J. en J.B. Vervliet.** 1899–1903. *Idioticon van het Antwerpsch dialect (stad Antwerpen en Antwerpsche Kempen)*. Gent: Siffer. *Aanhangsel* (1906). Gent: Siffer. *Bijvoegsel* (1936–1938). Turnhout: Koninklijke Vlaamse Academie voor Taal- en Letterkunde.
- De Belie, A.** 2006. *Woordenboek van het Belseels dialect*. Belsele: Culturele Kring Baudelo.
- De Bo, L.** 1873. *Westvlaamsch idioticon*. Brugge: Gailliard. Heruitgave door Joseph Samyn. 1890–1892. Gent: Siffer. Herdruk 1970, 1976, 1984. Handzame: Familia et Patria.
- Debrabandere, F.** 1999. *Kortrijks Woordenboek*. Kortrijk/Brugge: De Leiegouw/Van de Wiele.
- Depraetere, E. e.a.** 2007. *Aé's ge dat lui'è leest teus hoorde hoe da'me wilder klapp'n al de kant'n van Anzegem. Klanken, woorden en zegswijzen uit de streek van Anzegem, Kruishoutem, Oudenaarde, Waregem (Potegem), Wortegem-Petegem, Zingem, Zwevegem (Moen, Heestert, Otegem), en verder tot in ... Deerlijk, Zulte, Aarsele, Ruiselede ...* Anzegem: Davidsfonds.
- Desnerck, R.** 2006. *Oostends woordenboek*. De Haan: Roland Desnerck.
- De Tollenaere, F.** 1960. *Alfabetische of ideologische lexicografie? Bijdragen tot de Nederlandse Taal- en Letterkunde*. Uitgegeven vanwege de Maatschappij der Nederlandse Letterkunde I. Leiden: E.J. Brill.
- De Tollenaere, F.** 1968. Problemen van het dialectwoordenboek. Theorie en praktijk. *Tijdschrift voor Nederlandse Taal- en Letterkunde* 84: 197-212.
- Devos, M.** 1973. De Bo en het Noordwestvlaams. *Album Willem Pée*: 131-140. Tongeren: Michiels.
- Ghijzen, H.C.M.** 1964. *Woordenboek der Zeeuwse Dialecten*. Den Haag: Van Goor. [Fraanje, K. e.a. 2003. *Supplement Woordenboek der Zeeuwse Dialecten*. Krabbendijke: Van Velzen]
- Goossens, J. en J. Van Keymeulen.** 2006. De geschiedenis van de Nederlandse dialectstudie. *Handelingen van de Koninklijke Commissie voor Toponymie en Dialectologie* 78: 37-97.
- Hoefufft, J.H.** 1836. *Proeve van Bredaasch taal-eigen of Lijst van eenige in de stad en den lande van Breda gebruikelijke en in sommige oorden van ons vaderland min gewone woorden en spreekwijzen, verzameld en toegelicht*. Breda: Sterk.
- Joos, A.** 1900. *Waesch idioticon*. Gent/St.-Niklaas: Siffer/Strijbol. Herdruk 1979. St.-Niklaas: Danthe.
- Kempeneers, P.** 2004. *Tiens en Hoegaards idioticon*. Tienen: Drukkerij Peeters.
- Kooijman, G.** 1985. *Thematisch Woordenboek van het Tungelroys*. Amsterdam: Rodopi.
- Kruijzen, J. en J. Van Keymeulen.** 1997. The Southern Dutch Dialect Dictionaries. *Lexikos* 7: 207-228.
- Lievrouw-Coopman, L.** 1950–1952. *Gents woordenboek*. Koninklijke Vlaamse Academie voor Taal- en Letterkunde Reeks VI, nr. 68. Gent: Erasmus. Herdruk 1974. Gent: Broers.
- Maes, L.** 1994. *Overmeers Woordenboekje (herzien en uitgebreid)*. Overmere: Heemkundige Kring.
- Moens, B.** 1988. *Waur is d'n tijd naurtoe? Geïllustreerd Waasmunsters Woordenboek*. Sint-Niklaas: Ten Bos.

- Moeyaert, C. e.a.** 2005. *Woordenboek van het Frans–Vlaams. Dictionnaire du flamand de France*. Leuven: Davidsfonds.
- Niedzielski, N.A. en D.R. Preston.** 2000. *Folk Linguistics*. Berlijn: Mouton de Gruyter. [Een herziene paperbackversie verscheen in 2003.]
- Pletinckx, L.** 2003. *Woordenboek van het Asses. Bijdrage tot de studie van de West-Brabantse streektaal*. Asse: Koninklijke Heemkring Ascania.
- Pieters, M.** 1995. *Woordenboek van het Lokers Dialect*. Lokeren: Oelbrandt.
- Preston, D.R.** 2005. What is Folk Linguistics? Why Should You Care? *Lingua Posnaniensis. Review of General and Comparative Linguistics* 47: 143-162.
- Preston, D.R. en N.A. Niedzielski.** 2008. Trends in Folk Linguistics. Achard-Bayle, G. en M.A. Paveau. *Pratiques. La linguistique populaire ou la valeur des savoirs profanes*. (december 2008.) Referentiespelling (1999) = *Hoe schrijf ik mijn dialect. Referentiespelling voor alle Brabantse dialecten*. Uitgegeven door de Stichting Brabantse Dialecten (Vlaanderen) en het Noord-Brabants Genootschap (Nederland). Leuven: Acco.
- Rutten, A.** 1890. *Bijdrage tot een Haspengouwsch Idioticon*. Antwerpen: Zuidnederlandsche Maatschappij van Taalkunde, Boucherij. [Met een Bijvoegsel in 1904.]
- Rys, K. en J. van Keymeulen.** 2009. Intersystemic Correspondence Rules and Headwords in Dutch Dialect Lexicography. *International Journal of Lexicography* 22: 129-150.
- Schuermans, L.W.** 1865–1870. *Algemeen Vlaamsch Idioticon*. Leuven: Van Linthout. *Bijvoegsel aan het Algemeen Vlaams Idioticon*. 1883. Leuven: Fonteyn. Herdruk met inbegrip van het bijvoegsel. 1984. Torhout: Flandria Nostra.
- Starck, O. en L. Claessens.** 1988, 1995². *Dictionnaire Marollien–Français/Français–Marollien*. Brussel: Lojipe.
- Staelens, X.** 1982, 1982², 1989³. *Dieksjenèr van 't (H)essels*. Eerste en tweede, verbeterde druk: Hasselt: Heideland. Derde, verbeterde druk: Hasselt: De Langeman.
- Teirlinck, I.** 1908–1924. *Zuid-Oostlaandersch Idioticon*. Gent: Siffer. Herdruk 1986. Handzame: Familia et Patria.
- Tuerlinckx, J.F.** 1886. *Bijdrage tot een Hagelandsch Idioticon*. Gent: Zuidnederlandsche Maatschappij van Taalkunde.
- Van der Eecken, G.** 2007. *Woordenboek van het Schellebels Dialect. Met Heemkundige aantekeningen*. Schellebelle: Heemkring Schellebelle.
- Van Gompel, W.** 2002–2006. *Reusels Woordenboek. Deel I: De mens; Deel II: De wereld; Deel III: Beroepen en andere bezigheden; Deel IV: Inleiding, schets van het dialect, aanvullingen, registers*. Reusel: Heemkunde Werkgroep Reusel.
- Van Keymeulen, J.** 1992. *De Algemene Woordenschat in de grote dialectwoordenboeken (WBD, WLD, WVD). Een methodologische reflectie*. Ongepubliceerd proefschrift. Gent: Universiteit Gent.
- Van Keymeulen, J.** 2003. Dialectwoorden verzamelen. Een praktische handleiding. *Handelingen van de Koninklijke Commissie voor Toponymie en Dialectologie* 75: 383-506. [Ook als aparte overdruk beschikbaar. Tongeren: Michiels.]
- Van Keymeulen, J.** 2004. Trefwoorden en lexicale varianten in de grote regionale dialectwoordenboeken van het zuidelijke Nederlands (WBD, WLD, WVD). De Caluwe, J. e.a. (Reds.). 2004. *Taeldeman, man van de taal, schatbewaarder van de taal*: 897-908. Gent: Academia Press.
- Van Keymeulen, J., R. Keulen, J. Swanenberg en T. van de Wijngaard.** 2007. Dialectlexicografie in de zuidelijke Nederlanden: bibliografie van de regionale en plaatselijke woordenboeken. *Handelingen van de Koninklijke Commissie voor Toponymie en Dialectologie* 79: 399-428.

- WBD = **Weijnen, A. e.a.** 1967–2005. *Woordenboek van de Brabantse Dialecten*. Assen: Van Gorcum/Groningen: Gopher.
- Weijnen, A.** 1961. De semantische en syntactische problematiek van het dialectwoordenboek. *Tijdschrift voor Taal- en Letterkunde* 78: 81-95.
- Weijnen, A.** 1963. Het dialectwoordenboek. *Woordenboek en Dialect. Bijdragen en Mededelingen der Dialectencommissie van de Koninklijke Nederlandse Akademie van Wetenschappen te Amsterdam* 24: 34-52.
- Weijnen, A.** 1967. De waarde van een dialectwoordenboek. *Mededelingen van de Nijmeegse Centrale voor Dialect- en Naamkunde* 6: 5-11.
- WLD = **Weijnen, A., J. Goossens e.a.** 1983–. *Woordenboek van de Limburgse Dialecten*. Assen: Van Gorcum/Groningen: Gopher.
- WNT = *Woordenboek der Nederlandsche Taal*. 1864–1998. Digitale versie 1999. Rotterdam: AND Publishers b.v.
- WVD = **Devos, M. e.a.** 1979–. *Woordenboek van de Vlaamse Dialecten*. Gent/Tongeren: Michiels.
- WVD I, 1 = **Devos, M. en H. Ryckeboer.** 1979. *Woordenboek van de Vlaamse Dialecten. Deel I Landbouwoordenschat. Afl. 1 'Akkerland en Weiland'*. Gent/Tongeren: Michiels. [Met apart uitgegeven Wetenschappelijk Apparaat.]
- WVD III, 4 en 5 = **De Pauw, T. en M. Devos.** 2005 en 2007. *Woordenboek van de Vlaamse Dialecten. Algemene Woordenschat. Afl. 4 'Karakter' en Afl. 5 'Verstand en Gevoel'*. Gent/Tongeren: Michiels.
- WZD = zie Ghijssen, H.C.M. 1964.

A Corpus-based Survey of Four Electronic Swahili–English Bilingual Dictionaries

Guy De Pauw, *CLiPS — Language Technology Group, University of Antwerp, Antwerp, Belgium; and School of Computing and Informatics, University of Nairobi, Nairobi, Kenya (guy.depauw@ua.ac.be),*

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

Peter Waiganjo Wagacha, *School of Computing and Informatics, University of Nairobi, Nairobi, Kenya (waiganjo@uonbi.ac.ke)*

Abstract: In this article we survey four different electronic bilingual dictionaries for the language pair Swahili–English. Aided by a data-driven morphological analyzer and part-of-speech tagger, we quantify the coverage of the dictionaries on large monolingual corpora of Swahili. In a second series of experiments, we investigate how applicable the dictionaries are as a tool in the development of a machine translation system, by evaluating bilingual coverage on the parallel SAWA corpus. At the same time we attempt to consolidate the dictionaries into a unified lexicographic database and compare the coverage to that of its composite parts.

Keywords: LEXICOGRAPHY, EVALUATION, MORPHOLOGY, LEMMATIZATION, PARALLEL CORPORA, MACHINE LEARNING, MACHINE TRANSLATION, SWAHILI (KISWAHILI), ENGLISH

Samenvatting: Een corpusgebaseerde evaluatie van vier bilinguale elektronische woordenboeken Swahili–Engels. In dit artikel evalueren we vier verschillende elektronische woordenboeken voor het talenpaar Swahili–Engels. Met behulp van automatische morfosyntactische analyse, kwantificeren we de dekking van de woordenboeken op basis van grote monolinguale corpora voor het Swahili. In een tweede reeks experimenten onderzoeken we de toepasbaarheid van de woordenboeken als hulpmiddel bij de ontwikkeling van automatische vertaalsystemen, door hun bilinguale dekking te meten op basis van het parallelle SAWA corpus. Tegelijkertijd proberen we de woordenboeken te integreren in een overkoepelende lexicografische databank en vergelijken we de dekking ervan met die van de samenstellende delen.

Sleutelwoorden: LEXICOGRAFIE, EVALUATIE, MORFOLOGIE, LEMMATISERING, PARALLELE CORPORA, AUTOMATISCHE LEERTECHNIEKEN, AUTOMATISCH VERTAALLEN, SWAHILI (KISWAHILI), ENGELS

1. Introduction

Bilingual dictionaries are typically used as linguistic aids, providing support for translators, travellers, foreign language students and comparative linguists. In recent years, bilingual dictionaries have also become essential components in the field of machine translation, particularly for the data-driven approaches. Bilingual dictionaries help establish correct word alignment patterns between words in the source and target language, which enables the automated creation of machine translation systems on the basis of language-independent techniques. For this kind of language technological purpose, the bilingual dictionary not only needs to be available in electronic format, but more importantly, needs to have sufficient coverage of the language pair in question.

Unfortunately, not enough research efforts survey and quantitatively compare dictionaries for this kind of task. In the context of ongoing research on machine translation English ↔ Swahili (De Pauw et al. 2009), we set out to compare and evaluate electronically available bilingual dictionaries for this language pair. This will not only enable us to pick the best candidate(s) for the job at hand, but also to consolidate the information sources into a uniform lexicographic database that can serve as a machine translation aid.

We begin this article by reviewing a previous survey of Swahili dictionaries in Section 2. We then provide a brief description and assessment of the currently available electronic dictionaries Swahili–English in Section 3. The lemmatizer used to perform the lookup procedures in the corpus-based evaluation, is described in Section 4. A quantitative assessment of the coverage of the dictionaries is then given in Section 5, after which we conclude this article with a discussion of the results in Section 6.

2. Previous work

This article updates and complements a previous attempt at surveying Swahili dictionaries using a computational method. Hurskainen (2004) considers five different dictionaries, which are converted into *worsened* finite-state morphological analyzers. Their generative power is consequently evaluated on three different corpora and their coverage is compared to that of SWATWOL (Hurskainen 1992), a *comprehensive Swahili parser*.

The publication, however, does not make it clear how the dictionaries were obtained or converted into digital format, nor does it provide any insight into how the morphological information contained in them is translated into a morphological analyzer. That said, it seems counter-intuitive to evaluate a dictionary as a morphological description of a language, rather than as a lexical one.

Our survey employs an alternative computational and corpus-based evaluation technique, one which can easily be replicated and one which addresses some of the issues apparent in Hurskainen (2004). We focus on readily

available digital dictionaries, dictionaries that can easily be converted into a unified database format. As such, only one dictionary is covered by both Hurskainen (2004) and our survey. We add to our comparison three recently published dictionaries, including the expansive *Internet Living Swahili Dictionary*, which was strangely absent from Hurskainen (2004).

Our evaluation method uses a single, comprehensive Swahili lemmatizer, which is used to retrieve lemmas for word forms in a large Swahili corpus. The lemmatizer that was used in our experiments, allows us to simply evaluate the dictionaries in terms of how many lemmas in the corpus they cover, regardless of the morphological information they encode. Contrary to Hurskainen (2004) we also focus our evaluation of bilingual dictionaries on their potential as tools in machine translation, by comparing their coverage on a parallel corpus.

3. Digitally available bilingual dictionaries Swahili–English

A fair number of bilingual dictionaries Swahili–English have been published over the years. They range from early colonial attempts at Swahili lexicography, to simple tourist phrase books, to fully-fledged translation dictionaries. While the source files for most of these dictionaries are typically not digitally available, the current major dictionaries are electronically accessible and can therefore be included in our survey. In this section, we briefly describe the electronically available dictionaries in terms of development history and features, and provide a first qualitative assessment.

3.1 The *Internet Living Swahili Dictionary* [ILSD]

One of the most famous Swahili–English dictionaries is not only available online, but is also largely developed there: the *Internet Living Swahili Dictionary* at KamusiProject.org. Development on this dictionary started in the early 1990s. Apart from the inclusion of Rechenbach (1967), this dictionary is conceived as a community effort, allowing non-expert users to create and update dictionary entries, which are reviewed by an editorial team. The dictionary is not available in print format, but — like the Freedict dictionary (cf. Section 3.2) — the *Internet Living Swahili Dictionary* (henceforth ILSD), has an open development architecture and the data is readily available for download.

The formatting of the entries in the downloadable files is illustrated in Figure 1. A Swahili word is associated with an English translation equivalent, a part-of-speech tag, and possible inflections and derivations. Some entries also include terminology and taxonomy fields. Many dictionary entries also feature example sentences. Even though this format does not rule out cross-referencing as such, it is currently not an active feature in ILSD.

[Swahili Word] -anguka
[English Word] fall
[Part of Speech] verb

[Derived Word] angika V
 [Swahili Example] Theluji ikianguka, hatutakwenda baharini.
 [English Example] If snow falls, we won't go to the beach.

Figure 1: Dictionary entry for *anguka* in ILSD

Thanks to the community effort, which has Swahili speakers from around the globe contributing lexical entries, the ILSD is by far the largest Swahili–English dictionary available with more than 60 000 entries. These entries are not comparable to dictionary 'articles', however, as each sense of each lemma is given a separate 'entry' — many of which unfortunately overlap. Also, quite a few inconsistencies and untidy entries can be observed in the dictionary. These include some obvious trial entries still remaining in the database, a sloppy definition of the field "Derived Word" (pointing to derivations and inflections alike, while at other times referring to taxonomy or dialectal features), and an inaccurate attribution of dialectal features to words. Furthermore, translation equivalents are often paraphrased, potentially hampering the use of those entries as an aid for word alignment in the context of machine translation.

3.2 The *Freedict Swahili–English Dictionary* [Freedict]

The *Freedict Swahili–English Dictionary* is an attempt to unify and homogenize existing bilingual dictionaries (Bański and Wójtowicz 2009). It is based on a previously published electronic dictionary (Dict 2009) and also includes entries from a Freedict dictionary (Freedict 2009), and a Swahili–Esperanto–English dictionary (Ergane 2009). It uses the open-source Freedict architecture for development and dissemination and sources are therefore freely downloadable.

The latest version includes 2 600 entries, associated with an English translation equivalent and a part-of-speech tag. Figure 2 illustrates the typical layout of the entries. While the dictionary itself is very small and the information provided is scarce, the developers seem to have tried as much as possible to provide single-word translation equivalents that bode well in a machine translation environment.

anguka
 fall
 (v)

Figure 2: Dictionary entry for *anguka* in Freedict

3.3 The *TshwaneDJe Swahili–English Dictionary* [TeDJe-SED]

The *TshwaneDJe Swahili–English Dictionary* (Hillewaert, Joffe and De Schryver 2009), for short TeDJe-SED, is the most recently published work in our survey. It includes about 16 000 entries and features morphological decomposition, corpus-based example phrases, and an intricate system of cross-references. It

also includes a tool that provides translation equivalents in Microsoft Word, which indicates the developers had (human-aided) machine translation in mind during development.

The dictionary was created using TshwaneDJe's in-house lexicography tool *TshwaneLex* (Joffe et al. 2009). It can be accessed on-line through a web interface and is available as a stand-alone download as well. The actual source files can only be accessed through *TshwaneLex*. Even though it features a fairly limited number of lemmas, it provides by far the most detailed lexicographic information of all the dictionaries in this survey, as illustrated in Figure 3.

-anguka

-anguka *verb, + stative* Root -angua

1 fall (down), crash
mti ulianguka
 the tree fell down

2 come down, drop

3 fail (business), fail (exams)

Has References To: >>

-angua *verb*

1 drop, knock down, throw down

2 hatch

-angua mayai *verb*
 hatch eggs

>> Has References From:

anaanguka *inflected verb, cl. 1* Root -anguka

1 he/she falls down

2 he/she comes down

3 he/she drops

4 he/she fails

anguko *noun 5/6 (It/ya-)* Derived from -anguka
 fall

kuanguka *infinitive* Root -anguka

Figure 3: Dictionary entry for *anguka* in TeDJe-SED

3.4 The TUKI *Swahili-English Dictionary* [TUKI]

The *Kamusi ya Kiswahili-Kiingereza / Swahili-English Dictionary* (TUKI 2006²) was developed at the then Institute of Kiswahili Research of the University of Dar es Salaam, in Tanzania, over the course of three years and was first published in hard copy in 2001. It is currently the best-selling paper bilingual dictionary Swahili-English. It includes lemmas, part-of-speech tags, translation equivalents, sporadic example phrases, and derivations.

In 2003 a digital version of the dictionary became available on CD-ROM. It consists of formatted HTML files with the same content as the hard copy. In principle, the HTML formatting tags should allow us to extract the required information for the purpose of our experiments, as illustrated in Figure 4. Unfortunately, a large number of inconsistent formatting issues, not present in the hard copy, can be observed in the digital version.

anguk.a *kt* [*sie*] 1 come down, fall down, drop, crash. 2 lose in a business. 3 fail: *Ame~ mitihani yake* he failed his examinations. (*ide*) **angukia**, (*iden*) **angukiana**, (*idew*) **angukiwa**, (*tdk*) **angukika**. **anguko** *nm*.

```
<p align="JUSTIFY"><b>anguk.a</b> <i>kt</i>
[<i>sie</i>] 1 come down, fall down, drop, crash.
2 lose in a business. 3 fail: <i>Ame~ mitihani
yake</i> he failed his examinations. <i>(ide)</i>
<b>angukia</b>, <i>(iden)</i> <b>angukiana</b>,
<i>(idew)</i> <b>angukiwa</b>; <i>(tdk)</i>
<b>angukika</b>. <b>anguko</b> <i>nm.</i></p>
```

Figure 4: Dictionary entry for *anguka* in TUKI

[NB! Dictionary article: commas vs. semicolons to separate the run-ons as in original.]

Since no updated copies were available, we decided to semi-automatically clean up the HTML files. A range of scripts scanned for unusual patterns in the HTML formatting. These were automatically converted and consequently proof-read by a human annotator. This resulted in a cleaner and more consistently formatted electronic dictionary, which can be converted into the database format required for our experiments.

3.5 Consolidation of sources

To perform a quantitative survey, we had to convert the four different lexicographic sources to a uniform database format. For this we propose the format illustrated in Table 1. The main field is the lemma, from which we removed root indications (e.g. hyphens in ILSD and TeDJe-SED; dots in TUKI). We add part-of-speech tag information and noun class information where applicable. We include a field with different English translations for the lemma at hand. If inflections of a lemma are listed, they are included in the "Related words" field. Finally, each database record is associated with its source. Note that multi-word expressions are not included in the consolidated database.

Table 1: Consolidated lexical database for four dictionaries

Lemma	POS-tag	Class	English translation	Related words	Source
geuko	noun	ma-	change, transformation		TUKI
geuza	verb	–	change, modify, ...	aligeuza, anageuza, kugeuza, niligeuza, ...	TeDJe-SED
geuzo	noun	5/6	change	mageuzo	ILSD

Table 2 shows some quantitative information about the converted dictionaries.

Table 2: Quantitative information for four dictionaries and consolidated database (approximate numbers)

	ILSD	Freedict	TeDJe-SED	TUKI	ALL
Swahili entries	61 000	2 600	15 500	14 500	—
Number of lemmas	17 000	2 500	2 500	13 000	21 000
Unique lemmas	8 000	100	150	3 900	—

The first row displays the number of dictionary entries as advertised by the developers. The second row shows the number of orthographically distinct lemmas per dictionary (not taking into account homographs with different morphosyntactic or semantic features). The last row shows how many lemmas are exclusive (i.e. unique) to that dictionary. Both the larger dictionaries as well as TeDJe-SED and Freedict include a fair number of unique lemmas, so unifying the different sources can lead to a rich lexicographic database.

4. Swahili morphological analyzer

In Section 5 we will compute the coverage of the dictionaries on the basis of large corpora. Given the rich morphological features of Swahili, however, we first need to lemmatize the word forms in the corpus to be able to match them to the lemmas in the database (as illustrated in Table 1). In De Pauw and De Schryver (2008) we introduced the first data-driven morphological analyzer for a Bantu language. We described how the lemmatized Helsinki Corpus of Swahili (Hurskainen 2004a) can be used as an information source that powers a lemmatizer using the machine learning technique of memory-based learning. We proceeded to show in a number of different experiments how the analyzer can be observed to significantly outperform a meticulously designed rule-based approach.

Since then, we continued development of the system and made some significant changes, including the introduction of trigram-based classification, which has previously shown to be beneficial for morphological processing (Van den Bosch and Daelemans 2005). The consolidated database described in Section 3.5 also yielded extra "word form — lemma" pairs that further complemented the training data. Finally, we replaced the machine learning technique of memory-based learning with that of maximum entropy learning. All of the above tweaks serve to further improve the accuracy of the lemmatizer from 88% (De Pauw and De Schryver 2008: 312) to 91% for morphologically complex words.

5. Computing the coverage

In this section we attempt to quantify the coverage of the respective dictionaries on the basis of a large monolingual Swahili corpus (Section 5.1). We also

investigate the usability of the dictionaries as a tool in machine translation, by looking at their coverage on a parallel corpus English–Swahili (Section 5.2).

5.1 Monolingual corpus

We used several textual sources to compute the coverage of the dictionaries. These include:

- The Helsinki Corpus of Swahili, HCS (Hurskainen 2004a) consisting of more than 9 million words.
- The TshwaneDJe Kiswahili Internet Corpus, TeDJe-KIC (De Schryver and Joffe 2009) of more than 20 million words.
- The Swahili part of the parallel SAWA corpus (De Pauw et al. 2009), containing ± 0.5 million words.
- Wikipedia in Swahili: almost 12 000 Internet pages, good for more than 1 million words.

We pre-processed the texts by uniformly converting them into UTF-8, tokenizing the data and lemmatizing them using the automatic morphological analyzer described in Section 4. The data was also part-of-speech tagged using the method described in De Pauw et al. (2006). The Helsinki Corpus of Swahili already has lemmatization and part-of-speech tag information available. We nevertheless chose to process it again using our own techniques, for reasons of annotation accuracy and consistency across the data sets.

We then proceeded to compute coverage. We used a purely quantitative approach for this, which checks for each word in the corpus whether its lemma (for the given part-of-speech) can be retrieved in the dictionaries. Table 3 displays the scores for the different dictionaries and corpora.

Table 3: Coverage scores on monolingual corpora (in %)

	ILSD	Freedict	TeDJe-SED	TUKI	ALL
HCS	87.9	50.7	60.4	73.4	90.0
TeDJe-KIC	88.2	51.0	61.2	74.1	90.5
SAWA corpus	85.5	50.2	60.2	71.9	89.9
Wikipedia	83.4	48.9	59.4	69.8	85.2
ALL DATA	87.9	50.8	60.9	73.7	90.2

The ILSD has the highest coverage across the board, but loses a lot of coverage for the more recent texts in the recently developed and noisier Wikipedia pages. TUKI follows the same trend, while the TeDJe-SED dictionary hardly loses coverage. The latter's smaller set of lemmas consistently covers the most frequent words in the corpus and is therefore not as vulnerable to change of

register and publication date. The complete consolidated dictionary database performs quite well with an overall coverage of about 90.2%.

To study the effect of publication date, we calculated the coverage per year of the periodicals included in the Helsinki Corpus of Swahili (1990 → 2002, no data for 1995, 1996, 1997). The downward trend in coverage is visible for all four dictionaries; see Figure 5. The most frequent items not covered by the dictionaries are named entities, foreign words and IT terminology. The need to update dictionaries consistently is therefore high. The open architectures of the Freedict and ILSD projects are in this sense suitable solutions.

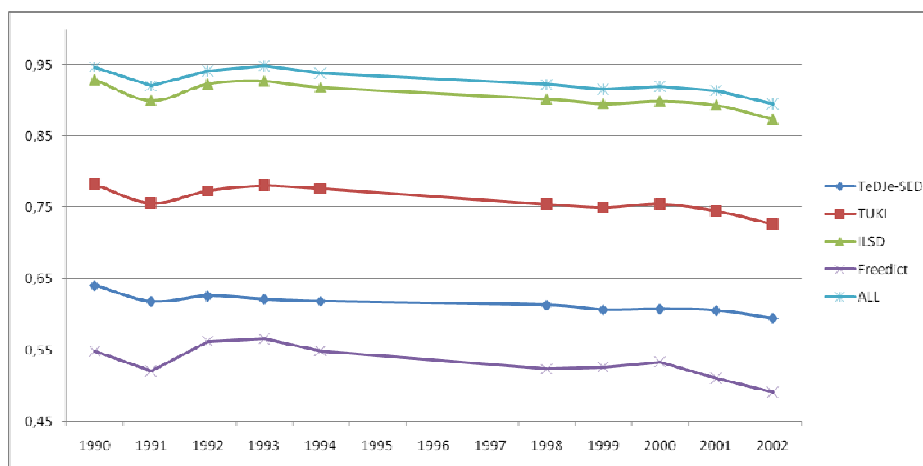


Figure 5: Coverage shift over time

Interestingly, the scores reported here differ significantly from those in Hurskainen (2004). Overall, coverage scores are lower than those reported in the previous survey, which may be due to differences in evaluation metrics. Stranger however is that Hurskainen (2004) observes higher recall scores for more recent documents, whereas Figure 5 shows a definite downward trend over time. These discrepancies warrant further investigation.

We also calculated the coverage of the dictionaries disregarding the frequency of the lemma. In this calculation, covering a highly frequent word like *lakini* 'but; however; nevertheless' scores the same as covering a hapax. Table 4 shows that in this experiment ILSD is trailing TUKI, indicating that even though ILSD contains many more entries, TUKI seems to cater for a wider range of words.

A final experiment counts for how many lemmas in the respective dictionaries evidence can be found in the corpus. The last line of Table 4 shows that TeDJe-SED has all lemmas covered by real-world data. About 30% of the entries in TUKI are not found in the data, while only two thirds of ILSD is covered by the corpus. Comparing the data in Tables 3 and 4 shows that while TUKI trails in comparison to ILSD in terms of raw coverage (Table 3), it does

seem to strike a better balance in terms of both lexical richness and empirical evidence (Table 4).

Table 4: Coverage scores for unique lemmas and reverse coverage (in %)

	ILSD	Freedict	TeDJe-SED	TUKI
ALL DATA	21.7	5.7	7.8	22.7
Reverse coverage	67.4	95.5	100	70.5

Owing to the massive amount of data, it is impossible to check whether the lemma retrieved in the dictionary is indeed the one intended in the text. It might indeed be the case that a particular lemma-tag combination retrieved in the dictionary, does not describe the correct meaning in its actual context. We manually checked a small section of the corpus ($\pm 2\ 000$ words) and found only two occasions of such an error. We are therefore confident that our scores are reliable in the context of the comparison between the dictionaries.

In De Pauw and De Schryver (2008) we presented our morphological analyzer as a way to unearth undiscovered lemmas in the corpus data. Our approach indeed has the distinct advantage that it is not dependent on a preset list of roots or lemmas, and is thus capable of lemmatizing word forms for previously unseen lemmas. The experiments outlined in this section have further underlined this property, as we now have at our disposal a list of word forms and associated candidate lemmas (roughly put the remaining 10% not covered by the consolidated dictionary) that need to be lexicographically described.

5.2 Parallel corpus

So far we have only evaluated the dictionaries in a monolingual context. We have calculated the raw coverage of the dictionaries, but this does not provide any insight into the suitability of the dictionary as a bilingual information source for (machine) translation purposes. To properly estimate this, we used the parallel SAWA corpus (De Pauw et al. 2009).

The SAWA corpus is a one million word bilingual corpus, consisting of political documents, religious texts, movie subtitles, investment reports and other documents in both English and Swahili. They were semi-automatically sentence aligned and a small portion was manually word aligned.

The third line in Table 3 gives us some insight into how many lemmas are covered in the Swahili portion of the corpus. We then counted how many times the *actual translation* provided by the dictionary can be found in the associated English part. This gives us some insight into how useful the bilingual dictionary is in the context of machine translation and how appropriate the translation equivalents are that are provided by the respective dictionaries from an empirical point of view.

The first row of Table 5 shows that the small TeDJe-SED dictionary out-

performs the other dictionaries in terms of accuracy. The ILSD has a surprisingly low score. While it covers more than 85% of the lemmas in the SAWA corpus (cf. Table 3), this is only useful 59% of the time in the context of machine translation. Even TUKI's translation equivalents seem to be better suited to the task. The smaller dictionaries have a higher score, since they tend to cover more frequent words, which they describe better.

In a further experiment we take the subset of lemmas that is shared by all dictionaries and investigate their bilingual coverage in the SAWA corpus. Hereby we level the playing field in terms of dictionary size and only compute the usability of the respective bilingual dictionaries as a machine translation tool. The results can be found on the last row in Table 5. While the differences are definitely more narrow, Freedict and TeDJe-SED surprisingly keep the edge over the larger dictionaries, indicating they truly provide more useful translation equivalents.

Table 5: Bilingual coverage scores

	ILSD	Freedict	TeDJe-SED	TUKI
SAWA corpus (all)	58.5%	67.6%	69.4%	60.5%
SAWA corpus (common subset)	64.5%	67.6%	69.4%	65.6%

This experiment shows that evaluating the coverage of bilingual dictionaries needs to be performed on different fronts. Raw coverage scoring on monolingual data does indeed give us some insight into the scope of the dictionary and in this sense both TUKI and ILSD score very well. In terms of how useful the dictionary is as a tool in machine translation, computing the coverage of bilingual documents provides an interesting, alternative insight into the matter, and here Freedict and especially TeDJe-SED score admirably.

6. Discussion

In this article we compared four different electronically available dictionaries for Swahili: one ported from a standard hard copy dictionary (TUKI), two developed and distributed electronically as a community effort (Freedict and ILSD), and a small but accurate electronic one developed by a lexicographic company (TeDJe-SED). From the various analyses presented, it is clear that none of the dictionaries by themselves offer a one-stop solution for machine translation work.

Of course, it is important to keep in mind that none of these four dictionaries was conceived with the aim to function as a component in an automated translation system. The four dictionaries in this study mostly have different aims and target user groups in mind, apart from being compiled in quite different ways. Therefore, comparing them from a very specific angle, an angle that was not intended by the compilers, does not really do justice to any of them. As such, the fact that the reverse coverage of TUKI is 'only' 70% for

example (cf. Table 4) may be seen as a positive aspect, as the compilers surely attempted to cover as wide a range of vocabulary as possible, and may even have included obsolete terms on purpose. Conversely, the 100% reverse coverage for TeDJe-SED is exactly a design feature, given the compilation of that dictionary is directly inspired by corpus facts (cf. De Schryver et al. 2006).

We nevertheless made several attempts at unifying the different dictionaries into a consolidated lexicographic database. This indeed improved monolingual coverage to more than 90%. To compute bilingual corpus coverage of the consolidated database, we needed to first resolve conflicts between database fields for similar lemmas. Our experiments showed that the order of preference TeDJe-SED → Freedict → TUKI → ILSD yielded the best results, with about 70% of word pairs retrieved.

We believe that the consolidated database will be of great value to our machine translation system, as it helps link the English words to the associated Swahili words in the translation pairs. Reversely, the parallel corpus also contributes to the discovery of new, previously unrecorded translation pairs. Future research will investigate how this iterative procedure can be maximally exploited in a lexicographic, as well as a language technological context.

The biggest challenge however remains the development of a large coverage and effective machine translation system for Swahili. Even the recently released Google Translation System for Swahili seems to suffer from some apparent gaps in the vocabulary. We are confident that a machine translation system built using the consolidated database described in this article can significantly alleviate this problem.

It is actually interesting to note that Swahili, the widest spoken Bantu language, still does not have a fully functional bilingual dictionary available that is applicable in the context of machine translation. The vast coverage of ILSD is somewhat hampered by the noise in the database fields and the often impractical translation equivalents. TeDJe-SED seems highly accurate and lexicographically sound, but is so far lacking in terms of raw coverage. Freedict is a simple and effective dictionary, but suffers from its limited scope. Finally, TUKI strikes a nice balance between size and lexicographic scope, but is seemingly in arrested development, which is unfortunate in light of the graphs displayed in Figure 5. We are however confident that the electronic availability of these dictionaries and our evaluation thereof might help lead the way for lexicographers to develop a new, accurate and large-coverage bilingual Swahili–English dictionary, one that does not only serve as a human translation aid, but distinctly moves forward towards machine translation as well.

Acknowledgements

Guy De Pauw is funded as a Postdoctoral Fellow of the *Research Foundation — Flanders* (FWO). Gilles-Maurice de Schryver would like to thank *Ghent University* for its continued support of his field trips in Africa.

References

- Bański, P. and B. Wójtowicz.** 2009. A Repository of Free Lexical Resources for African Languages: The Project and the Method. De Pauw, G. et al. (Eds.). 2009: 89-95.
- De Pauw, G. and G.-M. de Schryver.** 2008. Improving the Computational Morphological Analysis of a Swahili Corpus for Lexicographic Purposes. *Lexikos* 18: 303-318.
- De Pauw, G., G.-M. de Schryver and L. Levin (Eds.).** 2009. *Proceedings of the EACL 2009 Workshop on Language Technologies for African Languages*. Athens: Association for Computational Linguistics.
- De Pauw, G., G.-M. de Schryver and P.W. Wagacha.** 2006. Data-Driven Part-of-Speech Tagging of Kiswahili. *Lecture Notes in Artificial Intelligence* 4188: 197-204.
- De Pauw, G., P.W. Wagacha and G.-M. de Schryver.** 2009. The SAWA Corpus: A Parallel Corpus English–Swahili. De Pauw, G. et al. (Eds.). 2009: 9-16. [SAWA corpus]
- De Schryver, G.-M. and D. Joffe.** 2009. *TshwaneDJe Kiswahili Internet Corpus*. Pretoria: TshwaneDJe HLT. [TeDJe-KIC]
- De Schryver, G.-M., D. Joffe, P. Joffe and S. Hillewaert.** 2006. Do Dictionary Users Really Look Up Frequent Words? — On the Overestimation of the Value of Corpus-based Lexicography. *Lexikos* 16: 67-83.
- Dict.* 2009. The DICT Development Group [online]. <http://dict.org>
- Ergane.* 2009. A Multilingual Translation Dictionary for Windows [online]. <http://download.travlang.com/Ergane>
- Freedict.* 2009. Free Bilingual Dictionaries [online]. <http://freedict.org> [Freedict]
- Google Translate.* 2009. Google's free online language translation service instantly translates text and web pages [online]. <http://translate.google.com> [Google]
- Hillewaert, S., P. Joffe and G.-M. de Schryver.** 2009. *Kamusi ya Kiswahili–Kiingereza Katika Mta-ndao/Online Swahili–English Dictionary* [online]. <http://africanlanguages.com/swahili/> [TeDJe-SED]
- Hurskainen, A.** 1992. A Two-Level Computer Formalism for the Analysis of Bantu Morphology: An Application to Swahili. *Nordic Journal of African Studies* 1: 87-119.
- Hurskainen, A.** 2004. Computational Testing of Five Swahili Dictionaries. Karlsson, F. (Ed.). 2004. *Proceedings of the 20th Scandinavian Conference of Linguistics, Helsinki, January 7–9, 2004*. Department of General Linguistics Publications No. 36 [online]. Helsinki: University of Helsinki. <http://www.ling.helsinki.fi/kielitiede/20scl/proceedings.shtml>
- Hurskainen, A.** 2004a. HCS 2004 — Helsinki Corpus of Swahili. Compilers: Institute for Asian and African Studies (University of Helsinki) and CSC. [HCS]
- Joffe, D. et al.** 2009. *TshwaneLex Suite* [online]. <http://tshwanedje.com/tshwanelex/>
- Kamusi Project.* 2009. The Internet Living Swahili Dictionary [online]. <http://kamusiproject.org> [ILSD]
- Rechenbach, C.W.** 1967. *Swahili–English Dictionary*. Washington: Catholic University of America Press.
- TUKI.* 2006². *Kamusi ya Kiswahili–Kiingereza/Swahili–English Dictionary*. Dar es Salaam: Taasisi ya Uchunguzi wa Kiswahili, Chuo Kikuu cha Dar es Salaam. [TUKI]
- Van den Bosch, A. and W. Daelemans.** 2005. Improving Sequence Segmentation Learning by Predicting Trigrams. *Proceedings of the Ninth Conference on Natural Language Learning*: 80-87. Ann Arbor: Association for Computational Linguistics.
- Wikipedia.* 2009. Wikipedia in Swahili [online]. <http://sw.wikipedia.org> [Wikipedia]

Towards a Corpus of South African English: Corralling the Sub-varieties*

Leela Pienaar, *Extended Studies Unit (l.pienaar@ru.ac.za)*
and
Vivian de Klerk, *Dean of Students (v.deklerk@ru.ac.za),
Rhodes University, Grahamstown, Republic of South Africa*

Abstract: Within the last twenty years, the use of a corpus for language research has become the *sine qua non* in many areas of linguistic enquiry. This trend is particularly evident in lexicography, a discipline which has become increasingly and overtly 'corpus-driven'. This article draws on research from a Master's project which involved the collection of a small corpus of Indian South African English (ISAE), an acknowledged component or sub-variety of South African English (SAE). The discussion highlights the importance of aiming for a balanced representation of the known sub-varieties of a language when compiling corpora for lexicographic and linguistic investigation. Since ISAE is primarily an oral dialect, specific focus is given to the methodological challenges involved in compiling a spoken corpus. Methodological insights from local as well as international corpus research were used to guide and inform the process. These include the Xhosa English Corpus, the New Zealand Corpus of Spoken English and the Hong Kong Corpus of Conversational English. The various stages in the research process are described, together with explanations of how problems such as ways of corpus design, the selection of corpus contributors, the data-collection process and developing guidelines for consistency during the corpus compilation were addressed. The article provides a keyhole view of the main lexical and syntactic features of ISAE exemplified in the corpus and juxtaposes these against the backdrop of general SAE and trends in World English. The article concludes with a proposal for the collection of parallel corpora of other sub-varieties of SAE which will provide an objectively compiled repository of language in use to enable researchers to discern the linguistic features at the core and periphery of SAE. It is argued that the establishment of corpora of the various known sub-varieties of SAE could constitute an important step towards the creation of a truly representative large corpus of SAE and ultimately towards a better definition and understanding of SAE.

Keywords: CORPUS, SPOKEN CORPUS, DESIGN, SOUTH AFRICAN ENGLISH, SUB-VARIETIES, INDIAN SOUTH AFRICAN ENGLISH, LEXICOGRAPHY

**Opsomming Op weg na 'n korpus van Suid-Afrikaanse Engels: Die byme-
kaarbring van die subvariëteite.** Gedurende die laaste twintig jaar het die gebruik van 'n

* This article is a revised version of a paper presented at the Thirteenth International Conference of the African Association for Lexicography, organized by the Bureau of the *Woordeboek van die Afrikaanse Taal*, Stellenbosch, Republic of South Africa, 30 June–2 July 2008.

korpus vir taalnavorsing die *sine qua non* op baie gebiede van taalondersoek geword. Hierdie neiging is veral te sien in die leksikografie, 'n vakgebied wat toenemend en merkbaar "korpusgedrewe" geword het. Hierdie artikel maak veral gebruik van navorsing vir 'n Meestersprojek wat die versameling behels het van 'n klein korpus Indiese Suid-Afrikaanse Engels (ISAE), 'n erkende komponent of subvariëteit van Suid-Afrikaanse Engels (SAE). Die bespreking beklemtoon die belangrikheid om na 'n gebalanseerde weergawe van die bekende subvariëteite van 'n taal te streef wanneer korpusse vir leksikografiese en linguïstiese ondersoek saamgestel word. Omdat ISAE primêr 'n mondelinge dialek is, word spesifiek gefokus op die metodologiese uitdagings gepaardgaande met die samestelling van 'n gesproke korpus. Metodologiese insigte van sowel plaaslike as internasionale korpusnavorsing is gebruik om leiding en vorm aan die proses te gee. Dit sluit die Xhosa English Corpus, die New Zealand Corpus of Spoken English en die Hong Kong Corpus of Conventional English in. Die verskillende stadiums in die navorsingsproses word beskryf, saam met verduidelikings van hoe probleme soos maniere van korpusontwerp, die keuse van korpusbydraers, die dataversamelingsproses en die ontwikkeling van riglyne vir konsekwentheid gedurende die korpussamestelling gehanteer is. Die artikel verskaf 'n intieme blik op die belangrikste leksikale en sintaktiese eienskappe van ISAE soos beliggaam in die korpus en plaas dit teen die agtergrond van algemene SAE en neigings in Wêreldengels. Die artikel sluit af met 'n motivering vir die versameling van parallelle korpusse van ander subvariëteite van SAE wat 'n objektief saamgestelde bron van taal in gebruik sal verskaf om navorsers in staat te stel om taalkundige eienskappe in die kern en periferie van SAE te onderskei. Daar word geredeneer dat die totstandbrenging van korpusse van die verskillende bekende subvariëteite van SAE 'n belangrike trap kan vorm tot die skep van 'n werklik verteenwoordigende groot korpus van SAE en uiteindelik tot 'n beter omskrywing en begrip van SAE.

Slutelwoorde: KORPUS, GESPROKE KORPUS, ONTWERP, SUID-AFRIKAANSE ENGELS, SUBVARIËTEITE, INDIESE SUID-AFRIKAANSE ENGELS, LEKSIKOGRAFIE

South African English in the canon of World Englishes

Recent thinking on the global uses of English has acknowledged that there is not just one English language but rather a family of 'World Englishes'. The umbrella-term 'World Englishes' provides a conceptual framework to accommodate the different varieties of English which have evolved from linguistic cross-fertilization caused by colonization, migration and trade, which resulted in the transplantation of the original 'strain' or variety. Linguistic models such as Kachru's concentric circle model (1985) have done much to legitimize varieties of English around the world. Although the model has been lauded for acknowledging the pluralistic nature of English, it has also been criticized: firstly, for entrenching an essentially Anglocentric view of English by situating historically-native varieties of English (e.g. British and American) at the centre while relegating other varieties to the periphery, and secondly for failing to capture the dynamics of exchange that occur between varieties in the different circles.

In multilingual South Africa where English is one of 11 official languages, the model treats South African English as a unified whole and does not ade-

quately take account of the socio-linguistic complexities of the South African situation and of the different varieties of English that have evolved as a result. A closer look at English in South Africa (henceforth 'South African English' or 'SAE') reveals a complex situation where the English used here is anything but monolithic. Its speakers include those for whom it is a first language, those for whom it is an additional language and those for whom it is a replacement language. Census data which elicit information about home language do not tell the whole story as they fail to capture such complexities. For example, data derived from Census 2001 which merely indicate that 8.2% (roughly 3.7 million people) of the population of 44 million are English mother-tongue speakers, do not acknowledge that roughly 45% of the South African population (with different home languages) also use English to varying degrees in the domains of government, education, commerce, industry and in the media.

Linguistic research into SAE has identified various sub-varieties: Afrikaans English (Watermeyer 1996), Black South African English (Gough 1996), Cape Flats English (Malan 1996), Xhosa English (De Klerk 2002a, 2006) and Indian South African English (Mesthrie 1996), and has hinted at the existence of others. However to date there has been no real attempt to explain the dynamics either between the constituent sub-varieties themselves or between them and the variety of English spoken as a mother tongue in South Africa. There are ideological problems too with marking the sub-varieties of SAE with racial qualifiers (such as Black, Indian or Coloured) as it entrenches their status as ethnolects that are 'other' while leaving the 'colonial' variety of English unmarked. It has been argued that this practice actually affirms the position of a variety that is spoken by less than 10% of the population and sets it up as the standard against which all other English varieties in South Africa are measured (De Kadt 2001, cited in Coetzee-Van Rooy and Van Rooy 2005). In South Africa's rapidly-changing linguistic environment it is therefore useful to take a more realistic and inclusive view of SAE. Therefore any attempt to document and define SAE should strive to incorporate all known sub-varieties in a balanced way in order to provide a reliable representation of this Southern Hemisphere variety of English.

Corpora

One of the ways of creating a balanced and representative sampling of a language is through the establishment of a corpus. The term *corpus* is used here to mean an organized 'collection of pieces of language text in electronic form, selected according to external criteria to represent as far as possible a language variety as a source for linguistic research' (Sinclair 2005: 16). Thus defined, the use of a corpus for language research has become the *sine qua non* in many areas of linguistic enquiry (including lexicography) within the last twenty years and there are corpora for a host of major national as well as minority languages, ranging from Arabic to Walloon. In addition, there are corpora for different varieties of English: the British National Corpus (BNC) and the Bank of

English (BOE) in Britain; the American National Corpus; the Wellington Corpora of Spoken and Written New Zealand English; the Australian and the Macquarie Corpora; the Kolhapur Corpus of Written Indian English; and the enormous International Corpus of English (ICE) which, when completed, will comprise parallel corpora of regional varieties of World English (Meyer 2002). Yet, to date, there is no large corpus to represent South African English (SAE). A South African component of the ICE project (henceforth ICE-SA) has been in preparation since the early 1990s (Jeffery 2003) but the corpus has not yet been completed or released to the research community. Constructed according to ICE specifications, ICE-SA will be a relatively small contemporary corpus of 1 million words (500 texts of 2 000 words each) and is unlikely to be able to provide a balanced representation of the different varieties of English in South Africa. A corpus which is to serve as standard reference for a language would need to be much larger in order to reflect its sub-varieties and the domains in which these are used. In tacit acknowledgment that SAE is a constellation of various sub-varieties, De Klerk (2002a: 35) has argued for differentiated corpora of SAE on the grounds that 'linguists need ... a database which carefully distinguishes speakers of English on the basis of their background MT [mother tongue], ethnicity and geographical location'. To this end, she pioneered the development of a 500 000-word spoken corpus of Xhosa-English which was completed in 2005.

The idea of creating a corpus of Indian South African English (ISAE) was conceived in the wake of the development of the afore-mentioned corpus of Xhosa English. The corpus of ISAE, which involved the collection of conversational data from a narrow age band, is intended as the first building block towards a full corpus of ISAE that will ultimately be demographically and contextually balanced. When completed, the full corpus of ISAE could serve a dual purpose: firstly as a standard for referencing research into the sub-variety; and secondly as a component in a truly representative large corpus of SAE. The collection of spoken ISAE is envisaged as the foundational section towards the construction of a fuller corpus of ISAE, and despite its modest size, it could provide useful initial data for comparison with earlier studies of ISAE and Indian English worldwide. When complete, the full corpus of ISAE would constitute an important building block in a comprehensive corpus of SAE, with the sub-varieties represented in equitable ratios. A SAE mega-corpus thus constituted could represent a valuable standard reference for determining the salient features of this important variety of world English. Sub-corpora such as the ISAE data at the centre of this study would provide ready linguistic repositories for testing theories of language variety and for assessing the effect(s) of the country's official policy of desegregation since 1994 on ethnically-based taxonomies of SAE.

It is hoped that such a corpus, which encompasses all the known sub-varieties of SAE, would create a more nuanced understanding and definition of this important variety of World English. This enterprise will obviously depend on collaboration amongst South African researchers.

What is ISAE?

Indian South African English or South African Indian English (henceforth ISAE) refers to a variety of English spoken by the 1.1 million (or 2.5% of the population) South Africans of Indian descent (Census 2001). The majority (71.6%) of this self-classified group reside in KwaZulu-Natal, descendants mainly of indentured labourers, together with some traders and missionaries who arrived in the former province of Natal between 1860 and 1911. The original immigrants were a linguistically diverse group from different geographical regions in India, and very few had a command of English. However, just over 100 years after the arrival of the first Indian immigrants to South Africa, English has ousted the ancestral Indian languages and become a replacement first language or mother tongue for this community. Mesthrie describes ISAE as a complex example of a 'language-shift' variety of English where English replaced the Indian languages 'as the main (and often sole) language of daily interaction' (Mesthrie 1992b: 3). According to the 2001 census 95.8% of Indian South Africans listed English as their first language or home language. The shift to English by this community is linked to various factors such as the lack of a common Indian language, diminished contact opportunities with India during the apartheid regime, the desire for economic advancement through proficiency in English, and finally, an education system that did not support the vernacular needs of minority communities (Mesthrie 1992b: 32-33).

ISAE as a sub-variety of SAE is well-documented (Bughwan 1970, Crossley 1987 and Mesthrie 1992a, 1992b, 1996, 2002a, 2002b). It is an ethnolect of SAE spoken by South Africans of Indian extraction. It is largely distinct from 'Indian English' as spoken on the Asian sub-continent, the latter being broadly characterized by ornate lexis and stylistically formal constructions (Kachru 1994). ISAE has retained residues of lexis and syntax rooted in the ancestral Indian languages, but it shares several features with other sub-varieties of SAE. ISAE has absorbed lexical items such as *robot* (traffic light), *dagha* (mud), *baba-laas* (a hangover) and *tickey-line* (cheap or of poor quality) from general SAE; and in turn ISAE has enriched the lexis of general SAE with contributions such as *bunny-chow* (a hollowed out half-loaf of bread filled with curry), *char-ou* (Indian person), *larney/lahnee* (one's boss or a wealthy person), *ballie* (an old man or person) (Silva et al. 1996). ISAE also features additional senses for general English words (Mesthrie 1992a), as the following sentences illustrate:

My uncle's got *sugar* (= diabetes).
 I went to visit my *future* (= fiancé or fiancée).
 She's so *independent* (= haughty or aloof).

In the absence of sufficient contextual information to prime comprehension, an outsider to the ISAE speech community would be challenged clearly to discern the sense of the lexical examples *sugar*, *future* and *independent* quoted above.

For the most part, the lexis of general ISAE does not feature the numerous

Afrikaans-based items such as *handlanger* (an untrained assistant), *lappie* (a rag) and *skelm* (a rascal), which are common in several other sub-varieties of SAE, as its geographical base has been KwaZulu-Natal, where English rather than Afrikaans has dominated in official and public domains. The notable exception to this generalization is the slang register in ISAE where the shift away from community-based norms is discernible in the liberal use of Afrikaans-based lexis such as *ou* ('chap'), *graaf* ('work'), *lakker* (from *lekker*, 'nice'), and *vaai* (from *waai*, 'go'), alongside Zulu-based words such as *make* ('money' from *amatshe* meaning 'stones'), *chebe* ('a beard', from *intshebe*), *gane* ('a child', from *ingane*), *skatul* ('a shoe', from *isiscathulo*), and *pozi* ('a house', from English army slang 'pozzie', a shortening of 'position', which could conceivably have been a dug-out or shelter). There are also a few slang lexical items traceable to Indian languages such as *mota* ('rich', from Hindi *mota* meaning big or fat) and *ballie* ('old man', from Hindi *balig* meaning an adult). As with many other forms of slang (Burchfield 1985), ISAE slang is governed by gender and age boundaries with usage prolific in the speech of young males (under 25 years of age), but extending to older males from socio-economic groups and occupations which favour a very informal style of speech.

While the lexis of ISAE was influenced to some extent by contact with local languages such as English, to a lesser extent Afrikaans, and the pidginized Fanagalo used on the mines, it was preserved and fossilized by the social isolation caused by the South African government's apartheid policies enforced between 1948 and 1994.

Methodology used to create the corpus of ISAE

Structure and design

There is no 'one-size-fits-all' corpus design, as each corpus is determined by socio-linguistic factors relating to the population under consideration and by the purpose which the corpus will serve. In planning this module of the ISAE corpus, the design features of significant earlier corpora such as the BNC and ICE, as well as corpora of the spoken language such as the LLC, the New Zealand Spoken Component of ICE (ICE-NZ), the Hong Kong Corpus of Conversational English (HKCCE) and the Xhosa-English Corpus (XE Corpus) were explored to find a suitable framework.

Size and boundedness of the corpus

There is also no ideal corpus size, only an optimum corpus size determined by the research needs and pragmatic considerations such as the availability of resources. In terms of size, the corpus for this research did not strive to be in the same league as the mega-corpora of hundreds of millions of words such as the BNC (100 million words), the Oxford English Corpus (1 billion words) or

the continually growing BOE Corpus (450 million words). It is a small corpus of finite length, rather than a large unconstrained or continually-growing monitor corpus. As a sample corpus collected within a narrow age and education band, any inferences drawn from the results would need to be interpreted with those parameters in mind.

Since the research aim was to collect a corpus of conversations, the ICE specification of 2 000 words for private, direct conversations between two people was taken as a useful benchmark of size (Nelson 1996: 29). With one researcher carrying sole responsibility for all the conceptual and labour-intensive aspects of supervising corpus collection and transcription, a scaled-down version of the ICE conversational component of 180 000 words was realistic. Thus a feasible target for this module of the ISAE corpus was a third of that in size or 60 000 words. Apart from the time- and labour-intensive aspects of converting spoken data into machine-readable form, there are other arguments in favour of modest-sized, well-balanced corpora. De Klerk (2003: 467), quoting McCarthy (1998), argues for smaller, well-designed corpora 'of spoken material which contain authentic and reliable representative data, [that] can be analyzed exhaustively in a variety of ways'. It is tempting to accumulate vast quantities of data on the assumption that corpus analysis is largely computerized, but Kilgarriff et al. (2004: 106) caution against accumulation of a welter of data that makes even the analysis of simple features like word occurrence difficult and time-consuming: 'If there are five hundred [occurrences of a word], [analysis] is still a possibility but might well take longer than the editorial schedule permits. Where there are five thousand, it is no longer viable. Having more data is good — but the data then needs summarizing.' (It must nonetheless be acknowledged that software is available which sifts and organizes data semantically, thus simplifying the searching of huge quantities of data.)

The 60 000-word corpus of ISAE is made up of thirty 'texts' or speech samples of approximately 2 000 running words or 'tokens'.¹ Following the ICE model, texts of 2 000 running words constitute the building blocks of the ISAE corpus, as they provide reliable linguistic samples for analysis, while being manageable in size. In fact, Biber and Finegan (1991: 212-213) maintain that a component of even half that size (1 000 words) is adequate to deliver data that will reveal the main linguistic characteristics in a text. Each ISAE text segment is a self-contained unit of roughly 2 000 words extracted from one thirty-minute dialogue, rather than a composite constructed from several short verbal exchanges.

Type of corpus: spoken rather than written

This research selected spoken English as the starting point because ISAE is 'primarily [an] oral dialect' (Mesthrie 1992b: 35). Previous research into ISAE has also focused on the spoken variety of the language (Bughwan 1970, Crossley 1987, Mesthrie 1992, 1996). In the broader South African linguistic context, there has been a proposal to collect spoken corpora for nine of the official

languages of South Africa (Allwood and Hendrickse 2003) and in terms of SAE in particular, much research has already been done towards a corpus of spoken Xhosa-English (De Klerk 2002a, 2003, 2006) which, it is hoped, will ultimately form part of a larger corpus of Black South African Englishes. Viewed against these national linguistic research initiatives, this first building block towards a corpus of ISAE could facilitate comparative studies of different sub-varieties of spoken English in the South African context and Indian English worldwide.

Internationally there are more examples of written than spoken corpora and even in the BNC the bias towards written data as opposed to spoken data is in the ratio of 9:1. Leech et al. (2001: 1) explain that this imbalance was a result of practical considerations. They acknowledge that the spoken language is 'the primary channel of communication', and that on these grounds it should have been allocated a greater proportional share of the corpus. However, they explain that this was not done because 'it is a skilled and very time-consuming task to transcribe speech into the computer-readable orthographic text that can be processed to extract linguistic information'. Compiling a corpus of spoken language is comparatively more difficult, labour-intensive and expensive than compiling a similarly-sized corpus of written language. The reasons for this are located in the basic differences between speech and writing. Writing is already in a mode visible for study, but speech (an audio medium) has to be converted to writing (a visible medium) before it can be studied and analyzed. Casual spontaneous speech is also 'messy' and not well-behaved syntactically: incomplete sentences are the norm, as are false starts, latched (or simultaneous) utterances and hesitations.

For the corpus of ISAE, the transfer of the spoken data to the written mode involved listening to the recording several times and manual word-for-word transcription. Although speech-recognition technology has been developed to handle the automated transcription of formal, clearly articulated speech such as broadcast monologues and dialogues, as yet there is no reliable program to deal with the unpredictable nature of spontaneous speech. Transcribing the recordings therefore was the most arduous part of the research project with a 2 000-word text segment taking roughly fifteen hours to transcribe, mark up with simple annotations and proof-read. The ICE-USA team have reported similar experiences of 15 to 20 hours from transcription to proof-reading of a 2 000-word multi-party conversation (Meyer 2002: 71).

Classifying 'text types' or genres for spoken corpora

In addition to written corpora being more numerous, there are also established systems for classifying written data. Although there is no agreed taxonomy for categorizing genres of spoken language, two broad approaches to classifying spoken data for corpora exist: one demographically-motivated and the other context-governed or task-oriented. The BNC, for example, distinguishes between private conversation (40%) and public, task-oriented aspects of speech (60%), and classifies the latter in four domain-specific areas, designated educa-

tional and informative, public and institutional, business, and leisure. Within each domain, verbal interactions are identified as being either monologues (such as lectures, speeches, sermons) or multi-party activities (such as classroom interactions, meetings, chat shows) (Leech et al. 2001: 2-3). In the conversational component, the speech interactions are all spontaneous and informal with a demographically-motivated approach controlling variables such as social class, gender, age and geographical distribution across samples. From the BNC experience it would appear that both approaches are valuable for determining text types for spoken corpora. De Klerk (2002b: 27) recommends the context-governed approach for classifying spoken text types, on the grounds that it strives for a balance 'between speaker, environment, context and recurrent features' and because it facilitates subsequent analysis from 'different [speaker and contextual] perspectives'. This argument holds, provided that demographic considerations are also accommodated within the defined linguistic contexts.

Since such detailed sub-types would have been impractical in a small corpus, as they would not have generated sufficient data for generalizable linguistic patterns or the formulation of reliable conclusions, a decision was taken to confine data-collection to one demographic band, namely 'young adults'. The spoken data comprised only casual face-to-face conversations between two people. The research took casual conversation as a starting point because firstly, spontaneous informal dialogues in private settings exemplify the kind of naturally-occurring language that everyone engages in daily. Casual conversation has also been described as the quintessence of language, a kind of 'pre-genre' in the development of language since all other forms of language, whether spoken or written, trace their genesis to this genre (Swales 1990). Cheng and Warren (1999: 6-7), in their study of inter-cultural conversations of Hong Kong English, argue that 'conversations are a benchmark for other spoken discourses, and that by more fully describing conversational English ... we will better understand the ways in which other spoken discourses differ from it'. Secondly, in the case of ISAE, it is in informal, private settings, rather than in public speech situations that the features of ISAE are most observable. In this regard Mesthrie observes that 'in public it is the ISAE accent which is its clearest marker; but in private situations or informal situations involving ISAE speakers mainly, the lexical carry-over and use of basilectal syntax increases' (Mesthrie 1992: xviii).

The contributors to the corpus

There were 49 South African-born contributors to the corpus, all of whom were of Indian extraction. Contributors supplied biographical details and information about their linguistic background on the *Personal Details and Consent Form* (Appendix A). Although the questionnaire did not use the racial labels designed by the apartheid government (Black, White, Coloured and Indian), respondents indicated their alignment with the group 'Indian' by selecting the

substrate Indian language or cultural group with which they identified (Question 8 Appendix A). The corpus excluded anyone who had not been born or raised in South Africa, such as Indian nationals and members of the Indian diaspora in general. The corpus also excluded anyone who had spent more than 12 months outside South Africa within the last three years in order to eliminate linguistic features which might be the result of recent contact with other languages or other varieties of World English (Question 5 Appendix A).

The first small group of data collectors and contributors to the corpus comprised family members and friends of the researcher, all of whom were students at Rhodes University. Since ISAE is not generally used in public discourse and members of this speech community tend to adopt 'more careful and formal styles in public interactions' (Mesthrie 2002a: 341), the fieldworkers used their access to existing social networks to identify other contributors to the corpus. This measure was guided by significant earlier research (Gumperz 1970, Milroy 1987) who recommend that a researcher who might be perceived as an outsider should avoid interaction with the targeted social group, and who also found that using a member of the 'in-group' was effective in securing access to a range of vernacular and non-standard codes which are often eschewed in groups specifically constituted for research and observation. In other research, Schmied (1996: 186) refers to the 'famous sociolinguistic paradox' of the observer effect where the presence of a researcher who is perceived as an outsider to the 'in-group' causes the participants in an observed conversation to speak in ways that are not natural, in a bid for standard or prestigious forms. In a further attempt to secure naturalistic data, field-workers were instructed not to structure the social interactions as interviews as they could potentially control the discourse and determine the elicitation topics and techniques. Instead, field-workers maintained their status as members of the 'in-group' by functioning as active participants with equal speaker rights. It is hoped that all these precautions assisted in modifying the observer effect and contributed to the procurement of representative linguistic data.

Substrate language groups represented in the corpus

In order to yield a reliable sample, the corpus was structured to be proportionally representative of the five main Indian language groups found in South Africa. Although many terms are shared by all groups, other terms especially those from culinary, kinship, clothing and religious domains are specific to different linguistic or cultural groups. There are no recent national statistics for ancestral language affiliation to Indian language groups in South Africa but by using the 1960 census records as a reference point it was possible to establish very broad guidelines for symbolic attachment to the main language groups. The year 1960 appears to have been a linguistic watershed for the Indian community in South Africa as there was a marked decline in the use of Indian languages from that date (Mesthrie 2002b: 165). Thus the ancestral language dis-

tribution in the corpus was closely aligned to language data recorded for Indian South Africans in the 1960 census (Table 1).

Table 1: A comparison between the 1960 census and the substrate Indian language groups represented in the ISAE Corpus

Language	1960 census %	ISAE corpus %	ISAE corpus actual number
Hindi	32%	37%	18
Tamil	36%	27%	13
Gujarati	14%	14%	7
Telugu	9%	10%	5
Urdu	9%	4%	2
Other	0.5%	8%	4

Gender

In order to avoid a gender bias in the corpus and facilitate future comparisons of language use in equally-weighted gender group configurations, attention was given to achieving a 50:50 gender distribution. In addition, conversations between same-sex dyads and mixed-sex pairs were distributed as follows: ten conversations between women only, ten between men only and ten in mixed gender groups.

Age and geographical distribution

The contributors to the corpus were all young adults ranging in age from 18 to under 29, making it a highly-focused collection of ISAE speech with potential value for comparison with similarly-profiled corpora, such as the locally-collected Xhosa English Corpus (De Klerk 2002a, 2006) and the Corpus of London Teenagers (COLT) abroad (Stenström et al. 2002). All contributors to the corpus were born and educated in KwaZulu-Natal, the province with the largest concentration of South African Indians (Census 2002). Previous significant studies on the use of English by Indian South Africans have all used population samples from KwaZulu-Natal (Bughwan 1970, Crossley 1987 and Mesthrie 1992, 1996). This research has the potential, therefore, to provide useful data for comparison with these studies.

Time frame for data collection

With the research focus on *contemporary* ISAE, the time frame for data collection was limited to eighteen months (October 2004 to April 2006) in order to build a reliable synchronic corpus and thus exclude, or at best minimize, variables related to language change. This is even narrower than the time-frame of five to ten years for synchronic corpora recommended by Meyer (2002: 46).

Equipment and recordings

The conversations were recorded on small, battery-powered analogue tape recorders with built-in flat microphones which were unobtrusive, unthreatening and manageable in informal, private settings. The decision to make analogue recordings rather than digital ones was a pragmatic one, influenced by financial constraints and technological availability. If required, at a later stage the analogue recordings could be digitized for preservation using specially designed software, such as Syntrillium's 'Cool Edit' program.² Out of a total of 37 recorded conversations, only 30 were eligible for inclusion in the corpus and in retrospect, the audio quality would have been enhanced if participants had been wired up with small lapel microphones, as has been successfully implemented in other research, notably COLT (Stenström et al. 2002).

Transcription and storage

Contributors to the corpus were anonymous, so each speaker was assigned a core identity number from 1–49 (labelled \$01–\$49) and biographical details such as gender, age group and ancestral language group were encoded together with the speaker numbers. Thus \$10M1H would be interpreted as follows: \$10 = core speaker identity number; M = male; 1 = age group 16–19; H = Hindi. In line with the practice followed by the LLC (Svartvik 1990), pseudonyms of equivalent gender and number of syllables were substituted for personal names of third parties, addresses, telephone numbers and names of clubs or groups mentioned in the conversations. However, it was not deemed necessary to protect the identities of figures which exist in the public domain. The principle guiding these decisions was to disguise only details traceable to private individuals. The thirty recordings used in the corpus were given file numbers ranging from #01 to #37, and each file was prefaced by additional header information with encoded details about the material in the file:

FILENAME: (e.g. #21)
RECORDING: (e.g. 10B)
DATE RECORDED: (e.g. 25/09/2005)
DATE TRANSCRIBED: (the date when the transcription was completed
e.g. 7/10/2005)
NO. OF WORDS: (e.g. 2 014)

This header information is stored separately, as it is not part of the speech text itself, but it does provide a useful 'handle' for identification and retrieval of material.

In line with other corpora such as the XE Corpus, ICE-SA and the proposed spoken language corpora for the nine official African languages of South Africa, data was orthographically transcribed with no prosodic or phonetic mark-up. Transcription was done in lower case with no punctuation apart from

question marks, apostrophes for enclitic forms and possessives and hyphens for hyphenated terms. The initial transcription used word-processing software (MS Word) and converted the data to plain text to ensure electronic compatibility between different operating systems and different programs. There was minimal annotation to ensure that the corpus files are not 'bloated' and that the raw corpus is available in a simple form. The following notation conventions were used:

- = incomplete words e.g. *wed* for 'wedding'
- <,> short one-second pause
- <, , ,> pause over three seconds
- [sniffs] or [doorbell rings] non-verbal features essential to making sense of the conversation
- <??> unfamiliar words for which an approximate spelling was used
- [unclear] undecipherable utterances
- 21 numbers transcribed in full e.g. 'twenty one'

The conversation was laid out like the script of a play with each speaker-turn on a new line and changes in speaker-turns sequentially numbered in multiples of 5 to allow for corrections or insertions during the checking phase. Latched utterances were enclosed between brackets thus: <{> </}> and each speaker's words within the overlapping segment were enclosed by brackets <[></]>. This system of mark-up does not strive for iconicity in transcription but preserves each speaker-turn as a unit, while using the mark-up to indicate the position and extent of the overlap (Nelson 1996: 41).

The following excerpt is a typical extract of a transcription taken from the ISAE Corpus:

- <#03:\$07F2G:930> but like, I was gonna wear it today, but look at the weather it's like, <{><[>warm.</[></}>
- <#03:\$09F2T:935> <[>tomorrow's twenty one.</[></{>

Analysis and findings

Wordsmith Tools was used to produce alphabetical and frequency-based word lists of the corpus data. Despite the absence of lemmatization, it was possible to discern inflected forms of words from the alphabetically-arranged lists. The frequency-based lists yielded information on the scope and range of vocabulary within the ISAE corpus, while at the same time delivering data to facilitate objective comparisons with other spoken corpora. Mesthrie's extensive research into ISAE (1992a, 1992b, 2002 etc.) has established it as a distinct variety of SAE, so features selected for analysis were referenced firstly against prior research into the sub-variety (Mesthrie 1992a, 1992b, Crossley 1987 and Bughwan 1970) and then against the *DSAE Hist.* (Silva et al. 1996). Recurrent features in the corpus for which no explanation or discussion could be found in ISAE- or SAE-

related literature were investigated in terms of the context in which they manifested themselves and according to the type of spoken language they typified. These investigations involved research into slang, discourse markers, the language of adolescents and briefly, and admittedly, only superficially, diachronic linguistics.

Although small (approximately 60 000 words), and representing a narrow age band of young adults, the resulting corpus of spoken data confirmed the existence of robust features identified in prior research into the sub-variety. These features include the use of *y'all* as a second person plural pronoun, of *but* in a sentence-final position, and of *lakker* /'lAkə/ as a pronunciation variant of *lekker* (meaning 'good', 'nice' or 'great'). An examination of lexical frequency lists revealed examples of general South African English such as the colloquially pervasive *ja*, *bladdy* (for bloody) and *jol(ling)* (for partying or enjoying oneself) together with neologisms such as *eish*, the latter previously associated with speakers of Black South African English. An extraction of frequency lists facilitated cross-corpora comparisons with data from the BNC and the Corpus of London Teenage Language and similarities and differences were noted. The study also used discourse analysis frameworks to investigate the role of high-frequency lexical items such as *like* in the data. In recent times, *like* has emerged globally as a lexicalized discourse marker, and its appearance in the corpus of Indian South African English confirms this trend.

Future developments

In considering what the ideal speech corpus should be like, Williams (1996) makes several important recommendations which have relevance for the addition of 'building blocks' to extend this modest initiative towards a corpus of ISAE. In an ideal world, a fundamental consideration is that the speech corpus should include a range of speech forms (e.g. monologues as well as dialogues) contributed by a demographically representative range of speakers (taking account of age, gender, geographical location and occupation) across a range of styles and functions. In addition, it is vital that future spoken language corpora ought to be available in at least two forms: audio and written, to take account of the fact that the transcription of speech in orthographic form results in the loss of much essential information. The recordings should be publicly available, together with different versions of the transcript (all in electronic form): orthographic, grammatically tagged, as well as one with prosodic mark-up. In addition, the corpus should be phonetically segmented and labelled. Naturally, developing such a corpus is a formidable task and 'a major undertaking' (Williams 1996: 19), and for this reason, building a corpus is usually a team endeavour requiring the kind of investment in terms of human resources and capital outlay that is beyond the scope of single individuals, and even well-resourced university research departments. The scale of such an undertaking requires the support of large agencies with appropriate financial, human and technological

resources. In addition, since the development of a corpus is usually a lengthy process, it requires a long-term commitment to language research and development. Successful international precedents for this type of collaboration are the BNC, created by an academic/industrial consortium which included Lancaster University, Oxford University Press, Longmans, the British Library and the British Academy, and the BOE which is jointly owned by the University of Birmingham and the publishing house HarperCollins. In the South African context, such an initiative could involve government structures such as the Pan South African Language Board (PanSALB) in combination with university-based language research centres and business enterprises such as publishing houses.

In the case of SAE, what is needed is a system that will acknowledge the permeability of the boundaries between what has been defined as 'Inner Circle' and 'Outer Circle' varieties. Lee (2001) and Nelson (2006) have successfully demonstrated the value of conceptualizing parallel corpora of varieties of World English as a set of overlapping Venn diagrams in order to discern the essential items at 'the core' of World English and the items which radiate out towards 'the periphery'. This model has potential relevance for SAE, which is composed of several sub-varieties and where the notion of what constitutes the standard is constantly under review. It could also provide useful methodology for editors and lexicographers whose job it is to decide on usage norms and the degree of assimilation of various lexical and syntactic options. However, the key to harnessing that methodology is the establishment of parallel corpora of the existing and emerging sub-varieties of SAE. The idea of establishing exactly what constitutes 'the core' of SAE and noting the degree of closeness or distance of different lexical items and grammatical features from this core, seems to offer a really objective method of classifying constituent sub-varieties of the language.

Notes

1. A token is 'an individual occurrence of any word form' (Barnbrook 1996: 53).
2. Share-ware freely available from <http://www.syntrillium.com/cooledit/index.html>

References

- Allwood, J. and A.P. Hendrickse.** 2003. Spoken Language Corpora for the Nine Official African Languages of South Africa. *Southern African Linguistics and Applied Language Studies* 21(4): 189-201.
- Barnbrook, G.** 1996. *Language and Computers: A Practical Introduction to Computer Analysis of Language*. Edinburgh: Edinburgh University Press.
- Bughwan, D.** 1970. *An Investigation into the Use of English by the Indians in South Africa with Special Reference to Natal*. Unpublished Ph.D. Thesis. Pretoria: University of South Africa (UNISA).
- Burchfield, R.** 1985. *The English Language*. Oxford: Oxford University Press.

Census 2001: http://www.info.gov.za/otherdocs/2003/census01_key.pdf

Cheng W. and M. Warren. 1999. Facilitating a Description of Intercultural Conversations: The Hong Kong Corpus of Conversational English. *ICAME Journal* 23: 5-20.

Coetzee-Van Rooy, S. and A. van Rooy. 2005. South African English: Labels, Comprehensibility and Status. *World Englishes: Journal of English as an International and Intranational Language* 24 (1): 1-19.

Crossley, S. 1987. *The Syntactic Features of South African Indian English among Students in Natal, with regard to Use and Attitudes towards Usage*. Unpublished M.A. Thesis. Durban: University of Durban-Westville.

De Kadt, E. 2001. *What's in a Name? Labelling English in South Africa*. Paper presented at the 8th International Association for World Englishes Conference, Potchefstroom, Republic of South Africa, 1 December 2001.

De Klerk, V. 2002a. Starting with Xhosa English ... Towards a Spoken Corpus. *International Journal of Corpus Linguistics* 7(1): 21-42.

De Klerk, V. 2002b. Towards a Corpus of Black South African English. *Southern African Linguistics and Applied Language Studies* 20: 25-35.

De Klerk, V. 2003. Towards a Norm in South African Englishes: The Case for Xhosa English. *World Englishes. Journal of English as an International and Intranational Language* 22(4): 463-481.

De Klerk, V. 2006. *Corpus Linguistics and World Englishes: An Analysis of Xhosa English*. London/New York: Continuum.

De Klerk, V. (Ed.). 1996. *Focus on South Africa*. Varieties of English around the World 15. Amsterdam/Philadelphia: John Benjamins.

Gough, D. 1996. Black English in South Africa. De Klerk, V. (Ed.). 1996: 53-77.

Greenbaum, S. (Ed.). 1996. *Comparing English Worldwide: The International Corpus of English*. Oxford: Clarendon Press.

Gumperz, J.J. 1970. Sociolinguistics and Communication in Small Groups. Pride, J.B. and J. Holmes (Eds.). 1972. *Sociolinguistics: Selected Readings*: 203-224. Harmondsworth: Penguin.

Jeffery, C. 2003. On Compiling a Corpus of South African English. *Southern African Journal of Linguistics and Applied Language Studies* 21(4): 341-344.

Kachru, B. 1985. Standards, Codification and Sociolinguistic Realism: The English Language in the Outer Circle. Quirk, R. and H.G. Widdowson (Eds.). 1985. *English in the World: Teaching of Learning of Language and Literature*: 11-16. Cambridge: Cambridge University Press.

Kachru, B. 1994. English in South Asia. Bolton, K. and B. Kachru. 2006 *World Englishes: Critical Concepts in Linguistics. Volume 2*: 255-310. London: Routledge.

Kachru, B.J., Y. Kachru and C.L. Nelson (Eds.). 2006. *The Handbook of World Englishes*. Malden, MA: Blackwell.

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.

Lee, D.Y.W. 2001. Defining Core Vocabulary and Tracking its Distribution across Spoken and Written Genres. *Journal of English Linguistics* 29(3): 250-278.

Leech, G., P. Rayson and A. Wilson. 2001. *Word Frequencies in Written and Spoken English: Based on the British National Corpus*. Harlow/London: Pearson Education.

- Malan, K.** 1996. Cape Flats English. De Klerk, V. (Ed.). 1996: 125-148.
- McCarthy, M.** 1998. *Spoken Language and Applied Linguistics*. Cambridge: Cambridge University Press.
- Mesthrie, R.** 1992a. *A Lexicon of South African Indian English*. Leeds: Peepal Tree Press.
- Mesthrie, R.** 1992b. *English in Language Shift*. Johannesburg: Witwatersrand University Press.
- Mesthrie, R.** 1996. Language Contact, Transmission, Shift: South African Indian English. De Klerk, V. (Ed.). 1996: 79-98.
- Mesthrie, R.** 2002a. From Second Language to First Language: Indian South African English. Mesthrie, R. (Ed.). 2002: 339-355.
- Mesthrie, R.** 2002b. Language Change, Survival and Decline: Indian Languages in South Africa. Mesthrie, R. (Ed.). 2002: 161-176.
- Mesthrie, R. (Ed.)**. 2002. *Language in South Africa*. Cambridge: Cambridge University Press.
- Meyer, C.F.** 2002. *English Corpus Linguistics: An Introduction*. New York: Cambridge University Press.
- Milroy, L.** 1987. *Observing and Analyzing Natural Language: A Critical Account of Sociolinguistic Method*. Oxford: Blackwell.
- Nelson, G.** 1996. The Design of the Corpus. Greenbaum, S. (Ed.). 1996: 27-53.
- Nelson, G.** 2006. The Core and Periphery of World Englishes: A Corpus-based Exploration. *World Englishes. Journal of English as an International and Intranational Language* 25(1): 115-129.
- Schmied, J.** 1996. Second-Language Corpora. Greenbaum, S. (Ed.). 1996: 182-196.
- Silva, P., W. Dore, D. Mantzel, C. Muller and M. Wright (Eds.)**. 1996. *A Dictionary of South African English on Historical Principles*. Oxford: Oxford University Press.
- Sinclair, J.M.** 2005. Corpus and Text: Basic Principles. Wynne, M. (Ed.). *Developing Linguistic Corpora: A Guide to Good Practice*: 1-16. Oxford: Oxbow Books. <http://ahds.ac.uk/linguistic-corpora/> [Accessed 18 May 2008].
- Stenström, A., G. Andersen and I.K. Hasund.** 2002. *Trends in Teenage Talk*. Amsterdam/Philadelphia: John Benjamins.
- Svartvik, J. (Ed.)**. 1990. *The London-Lund Corpus of Spoken English: Description and Research*. Lund: Lund University Press.
- Swales, J.M.** 1990. *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.
- Watermeyer, S.** 1996. Afrikaans English. De Klerk, V. (Ed.). 1996: 99-124.
- Williams, B.** 1996. The Status of Corpora as Linguistic Data. Knowles, G., A. Wichmann and P. Alderson (Eds.) 1996. *Working with Speech: Perspectives on Research into the Lancaster/IBM Spoken English Corpus*: 3-19. London/New York: Longman.

Appendix A

Personal Details and Consent

1. Gender

Male Female

2. Age group:

16–19 20–24 25–29 30–34 over 35

3. Were you born in South Africa?

Yes No

Province:..... Town/City:.....

Where did you grow up?

4. Where did you go to school ?

Province:.....

Name of school(s):.....

Highest standard or grade passed:.....

5. Have you spent more than 12 months in total overseas in the last 3 years?

Yes No

If yes, please state which country

6. Which language does/did your mother use most often in your home?

Tamil Telugu Hindi Gujarati Urdu English

Afrikaans Other (please specify).....

7. Which language does/did your father use most often in your home?

- Tamil Telugu Hindi Gujarati Urdu English
Afrikaans Other (please specify).....

8. Which language/cultural group do you identify with?

- Tamil Telugu Hindi Gujarati Urdu
Other (please specify)

9. Which language(s) did you first speak at home? (You may tick more than one).

- Tamil Telugu Hindi Gujarati Urdu English
Afrikaans Other (please specify).....

10. Which language(s) do you still speak at home?

- Tamil Telugu Hindi Gujarati Urdu English
Afrikaans Other (please specify).....

11. Apart from English and Afrikaans which languages can you read and write?

- Tamil Telugu Hindi Gujarati Urdu
Other (please specify)

I give permission for the recording of my voice to be included in a corpus of South African English (which may be released on CD) to be used for linguistic research.

Signed: Date:.....

Lexikos at Eighteen: An Analysis

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

Abstract: At eighteen, *Lexikos* became a major player in the field of linguistics, by being awarded an Impact Factor. This article presents a double analysis of the foundation that led to this success. On the one hand a thorough statistical study is undertaken with regard to all contributors and their contributions to *Lexikos*. To this end a metadata database was designed, with the aim to answer the question: 'Who publishes what type of material from where and when?' On the other hand a content analysis is carried out which focuses on the actual topics (i.e. 'keywords') in *Lexikos*. To this end an all-inclusive text corpus containing all the *Lexikos* material was built, with the aim to answer the question: 'What are the major trends in *Lexikos*?'

Keywords: LEXIKOS, LEXICOGRAPHY, METALEXICOGRAPHY, DICTIONARIES, LEXICOGRAPHERS, METADATA DATABASE, TEXT CORPUS, CONTRIBUTIONS, CONTRIBUTORS, AFFILIATIONS, STATISTICS, TRENDS, ENGLISH, AFRIKAANS, BANTU

Samenvatting: *Lexikos* op achttienjarige leeftijd: Een analyse. Op achttienjarige leeftijd werd *Lexikos* een speler van wereldformaat binnen de linguïstiek, door het behalen van een Impact Factor. Dit artikel stelt een dubbele analyse voor van de grondslag die tot dit succes leidde. Aan de ene kant wordt een gedetailleerde statistische studie ondernomen met betrekking tot alle auteurs en hun bijdragen tot *Lexikos*. Daarvoor werd een metadata-databasis ontworpen, waarmee de volgende vraag wordt beantwoord: 'Wie publiceert welk soort materiaal, van waaruit en wanneer?' Aan de andere kant wordt een inhoudsanalyse uitgevoerd die zich toespitst op de feitelijke onderwerpen (m.n. 'sleutelwoorden') in *Lexikos*. Daarvoor werd een alles-inclusief tekstcorpus met al het *Lexikos* materiaal gebouwd, waarmee de volgende vraag wordt beantwoord: 'Wat zijn de voornaamste trends in *Lexikos*?'

Sleutelwoorden: LEXIKOS, LEXICOGRAFIE, METALEXICOGRAFIE, WOORDENBOEKEN, LEXICOGRAFEN, METADATA-DATABASIS, TEKSTCORPUS, BIJDRAGEN, AUTEURS, AFFILIATIES, STATISTIEKEN, TRENDS, ENGELS, AFRIKAANS, BANTOE

1. *Lexikos*: From newborn, to toddler, to adult

In the foreword to the very first issue of the present journal, the genesis of *Lexikos* is explained thus: "a formal discussion on lexicography between the editorial staff of the *WAT* and a group of prominent linguists got off the ground in Stellenbosch in 1989, [and] the staff is attempting to extend this discussion to include the rest of Africa" (Harteveld 1991: xii). From humble but

enthusiastic beginnings at the Bureau of the *Woordeboek van die Afrikaanse Taal* (WAT) two decades ago, *Lexikos* has been on a straight trajectory, indeed extending its discussion throughout, not only in geographical terms but also in terms of the ever-growing variety and quality of the lexicographic topics discussed. *Lexikos* has now been granted its place amongst the major journals in linguistics, through its inclusion in the Thomson Reuters Web of Science Citation Index since 2005, and the calculation of its impact, by means of an annual Impact Factor, as from last year's issue (De Schryver 2009). It is fitting, therefore, to look back at this point in time — '*Lexikos* at Eighteen' — to see how the toddler turned into a teenager, before it will now continue life as an adult.

Lexikos received some stimuli along the way; the most effective of these was the result of the establishment of the African Association for Lexicography (AFRILEX) in 1995, and the subsequent agreement whereby membership of AFRILEX includes a copy of the annual journal *Lexikos*. As such, *Lexikos* became the official mouthpiece of AFRILEX.

As a newborn, *Lexikos* had to find its way, and learn to walk. The first issue, virtually entirely in Afrikaans, was actually a cross between a journal and a Festschrift (cf. also Du Plessis 2008: xi), and was even reviewed (as if it were a book) for the *International Journal of Lexicography* by the father of modern lexicography: Ladislav Zgusta (1993). From a single file of articles in 1991, the journal morphed into one with multiple subsections, in an attempt to keep abreast of the ever-changing and expanding discipline of lexicography. From Volume 4 to Volume 14, for example, articles were mostly grouped into 'Research Articles' vs. 'Contemplative Articles', in addition to 'Review Articles' and 'Reviews' (of books). In the last volume, Volume 18, by contrast, there are no fewer than twelve categories, from 'Articles' to 'Lexicosoftware' to 'Projects' to 'Lexicosurvey', etc. Other changes include the introduction, in Volume 4, of a 'Review Editor', an 'Advisory Board', an 'Editorial Committee', and a statement of the 'Editorial Objectives' in three languages (Afrikaans, English, and German). Starting with Volume 6, each issue has included 'A Few Words from AFRILEX' (in both Afrikaans and English). 'Instructions to Authors' (again in three languages) first appeared in Volume 8. E-mail addresses for all authors were added for the first time in Volume 11. As of Volume 12, finally, both the 'Editorial Objectives' and the 'Instructions to Authors' have also been presented in a fourth language (French).

The overwhelming majority of the contributions to *Lexikos* are original studies without any prior traces of publication.¹ In order to study the first eighteen years of *Lexikos*, two databases were built. On the one hand a *Lexikos* Metadata Database was designed, in which all the main bibliometric details of each and every contribution were brought together. On the other hand a *Lexikos* Corpus was built, containing the full text of each and every contribution that appeared during the first eighteen years. In what follows, both databases are subjected to various queries, with the results of those queries summarized in tables and figures.

2. The *Lexikos* Metadata Database

2.1 Number of contributions and number of pages

There are a total of 6 587 pages in *Lexikos* 1 (1991) through *Lexikos* 18 (2008), on average 366 pages per year. The actual number of pages per volume may be seen from the line in Figure 1: from about 300 pages in the first few volumes, to about 500 pages in the last few volumes. The number of pages thus continuously rose over the years. In parallel, also the number of contributions rose, as may be seen from the bars in Figure 1: from under 20 contributions in the first two volumes, to about 35 in the last few volumes. There are exactly 543 contributions in all, on average 30 per volume. The Pearson correlation coefficient r between the number of pages and the number of contributions is 0.74, so the fit is not so good. This, on the whole, points to a widely varying average length of the contributions.

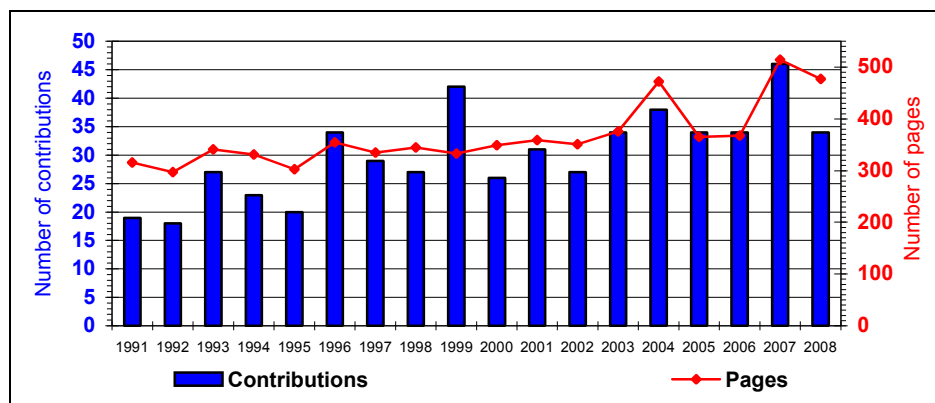


Figure 1: Number of contributions and number of pages in each *Lexikos* issue

Lexikos is a multilingual journal, and contributions in Afrikaans, English, Dutch, German, and French have all been accepted; distributed as seen in Figure 2. Two conclusions may be drawn from Figure 2: (a) Contributions in languages other than Afrikaans or English are by far in the minority; (b) English has rapidly taken over from Afrikaans as the main language of discourse. The latter may also be confirmed from Figure 3, where the number of Afrikaans and English contributions has been expressed as a percentage of the total per *Lexikos* issue. While 84% of the contributions were in Afrikaans in the first issue, this proportion decreases over the years to stabilize to around 20% (cf. the Afrikaans polynomial trend line). In parallel, while only 16% of the contributions were in English in the first volume, this proportion increases over the years to stabilize to about 70% (cf. the English polynomial trend line). (The remainder, $100\% - 20\% - 70\% = 10\%$, accounts for the 'other languages'.)

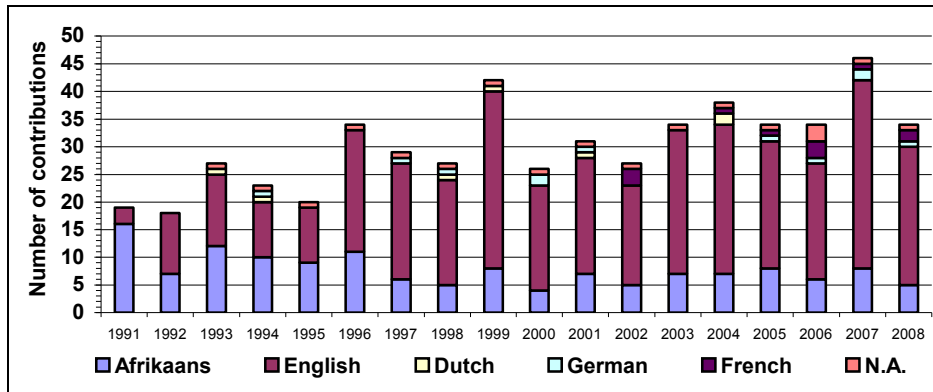


Figure 2: Distribution of the languages in each *Lexikos* issue

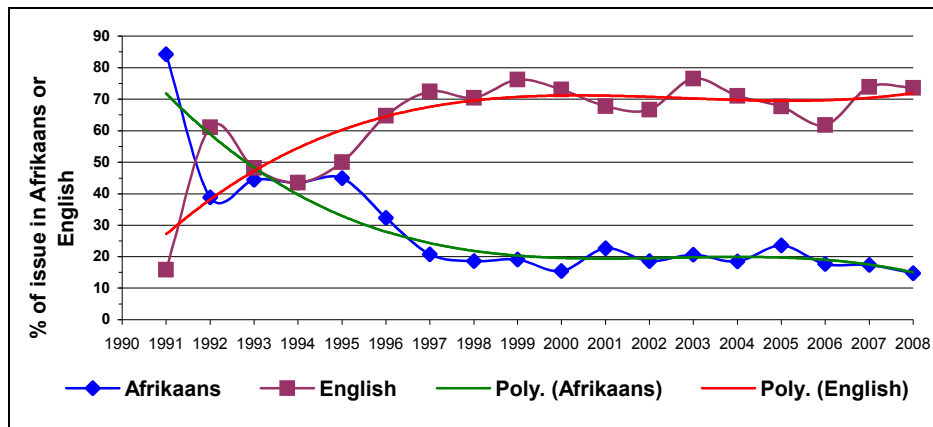


Figure 3: Afrikaans vs. English in each *Lexikos* issue (expressed as % of total)

A detailed view of the exact number of contributions and exact number of pages per *Lexikos* issue is shown in Table 1. From that data one sees that, overall, about 26% of the contributions in *Lexikos* are in Afrikaans, 65% in English, 1% in Dutch, 2% in German, and another 2% in French. A further 3% cannot be linked to particular languages, as they consist of lists of publications. Once English and Afrikaans had been 'established' as the discourse languages of *Lexikos*, the next (logical) language that was added was Dutch (from 1993 until 2004), followed by German (since 1994), and then French (since 2002).

Looking at the three main 'peaks' in Figures 1 and 2, finally, one can observe that all three are the results of one-time 'additions': in 2007 a special section was added, called 'Lexicofocus', in which the lexicographic activities at ALRI (in Harare, Zimbabwe) were surveyed; in 2004 a special section on 'Learners' Dictionaries' was added; while in 1999 the full proceedings of a meeting, 'Lexicography in a Multilingual South Africa', was added. Page-wise,

especially the 2007 and 2004 additions had a great effect (by contrast, the 1999 papers were all short); expressed in number of contributions, all three additions of course had a considerable effect.

Table 1: Detail of the number of contributions and number of pages in each *Lexikos* issue (with A = Afrikaans, E = English, NL = Dutch, DE = German, FR = French)

Year	A	%	E	%	NL	%	DE	%	FR	%	N.A.	%	Items	Roman pages	Arabic pages	Sum pages
1991	16	84.21	3	15.79	-	-	-	-	-	-	-	-	19	xix	297	316
1992	7	38.89	11	61.11	-	-	-	-	-	-	-	-	18	x	287	297
1993	12	44.44	13	48.15	1	3.70	-	-	-	-	1	3.70	27	x	331	341
1994	10	43.48	10	43.48	1	4.35	1	4.35	-	-	1	4.35	23	xi	320	331
1995	9	45.00	10	50.00	-	-	-	-	-	-	1	5.00	20	xiii	290	303
1996	11	32.35	22	64.71	-	-	-	-	-	-	1	2.94	34	xvi	339	355
1997	6	20.69	21	72.41	-	-	1	3.45	-	-	1	3.45	29	xiv	321	335
1998	5	18.52	19	70.37	1	3.70	1	3.70	-	-	1	3.70	27	xiv	331	345
1999	8	19.05	32	76.19	1	2.38	-	-	-	-	1	2.38	42	xv	318	333
2000	4	15.38	19	73.08	-	-	2	7.69	-	-	1	3.85	26	xiv	335	349
2001	7	22.58	21	67.74	1	3.23	1	3.23	-	-	1	3.23	31	xv	344	359
2002	5	18.52	18	66.67	-	-	-	-	3	11.11	1	3.70	27	xvi	335	351
2003	7	20.59	26	76.47	-	-	-	-	-	-	1	2.94	34	xv	361	376
2004	7	18.42	27	71.05	2	5.26	-	-	1	2.63	1	2.63	38	xv	457	472
2005	8	23.53	23	67.65	-	-	1	2.94	1	2.94	1	2.94	34	xv	350	365
2006	6	17.65	21	61.76	-	-	1	2.94	3	8.82	3	8.82	34	xiii	355	368
2007	8	17.39	34	73.91	-	-	2	4.35	1	2.17	1	2.17	46	xvi	498	514
2008	5	14.71	25	73.53	-	-	1	2.94	2	5.88	1	2.94	34	xiii	464	477
	141	25.97	355	65.38	7	1.29	11	2.03	11	2.03	18	3.31	543	ccliv	6 333	6 587

2.2 Types of contributions

The second series of queries one can submit to the *Lexikos* Metadata Database are those concerning the types of contributions. Using the *Lexikos*-internal categorization, the data shown in Table 2 is obtained.

Table 2: Types of contributions in *Lexikos* (according to *Lexikos*-internal categories)

Type	N	%	Type	N	%
Review	111	20.44	Meeting	14	2.58
Article	106	19.52	Lexicofocus	10	1.84
Contemplative Article	68	12.52	Lexicosoftware	6	1.10
Editorial Material	64	11.79	Lexiconews	2	0.37
Research Article	49	9.02	Lexicotribute	2	0.37
Lexiconote(s)	26	4.79	Report	2	0.37
Review Article	23	4.24	Lexicobibliography	1	0.18
Project(s)	20	3.68	Lexicohonour	1	0.18
Lexicovaria	19	3.50	Lexicosurvey	1	0.18
Publication Announcements	18	3.31		543	100.00

Obviously, some of the categories seen in Table 2 should be collapsed, as they are the result of diverging and converging attempts to categorize the various contributions over the years. In regrouping the types, we were inspired by the Web of Science categorization on the one hand, as well as the section on 'type and content of articles' as found in the *Lexikos* 'Instructions to Authors' (where the main dichotomy is one between articles subjected to strict anonymous refereeing vs. those that are not). Doing so, we obtained seven categories: ARTICLE (i.e. Article, Contemplative Article, Research Article, Lexicofocus, Lexicosoftware, and Lexicosurvey), REVIEW ARTICLE (idem), BOOK REVIEW (i.e. Review), PROCEEDINGS PAPER (i.e. Meeting), OTHER (i.e. Lexiconote(s), Project(s), Lexicovaria, Lexiconews, Lexicotribute, Report, Lexicobibliography, and Lexicohonour), EDITORIAL MATERIAL (idem), and PUBLICATION ANNOUNCEMENTS (idem). The ARTICLE and REVIEW ARTICLE sections are subjected to strict anonymous refereeing; the others must all meet the requirements of academic writing. The resulting distribution is tabulated in Table 3.

Table 3: Types of contributions in *Lexikos* (inspired by Web of Science categories)

Type	N	%
Article	240	44.20
Review Article	23	4.24
Book Review	111	20.44
Proceedings Paper	14	2.58
Other	73	13.44
Editorial Material	64	11.79
Publication Announcements	18	3.31
	543	100.00

As may be seen, as many as 44% of all contributions to *Lexikos* are true articles, a further 4% are review articles, and 20% are book reviews.²

Given that an analogous metadata study was performed for the *International Journal of Lexicography* (IJL) for the period January 1988 up to December 2008 (cf. De Schryver 2009a), a comparison of the values is instructive. In IJL, one finds 31% articles, 2% review articles, and 43% book reviews. While, in quantitative terms, IJL tends to focus on book reviews, the reverse is true for *Lexikos*, where most attention goes to articles and review articles.

2.3 Contributors and their contributions

A total of 555 names (i.e. name tokens) appear under the 543 *Lexikos* contributions. These two values are very close, and indeed, most contributions are authored by just one scholar, as may be seen from Table 4. As many as 83% of all contributions to *Lexikos* are written by a single author, a further 12% by two. There are very few with three, and just two with four — the maximum. By

contrast, in IJL one finds contributions with up to 12 authors. The single- and co-authored publication values in IJL are, however, very similar to those in *Lexikos*: 81% for sole authorship, 13% for co-authorship.

Table 4: Author position in the *Lexikos* Metadata Database

Author position	N	%
1	459	82.70
2	64	11.53
3	10	1.80
4	2	0.36
(no author)	20	3.60
	555	100.00

The 555 names that appear under the 543 contributions may be collapsed into 260 different (unique) names.³ In other words, a total of 260 different scholars have written material for *Lexikos* over the past eighteen years. The distribution of the number of contributions for each of those is as shown in Table 5.

Table 5: Number of contributions per author in the *Lexikos* Metadata Database

# contributions	N	%	# contributions	N	%
1	171	65.77	10	1	0.38
2	42	16.15	12	1	0.38
3	12	4.62	14	1	0.38
4	10	3.85	17	1	0.38
5	6	2.31	18	1	0.38
6	3	1.15	20	1	0.38
7	3	1.15	32	1	0.38
8	2	0.77	32	1	0.38
9	3	1.15		260	100.00

Two thirds (66%) of the contributors to *Lexikos* only contributed once, 16% twice, 5% thrice, etc. In contrast, in IJL as many as three quarters (74%) contributed only once, 11% twice, 6% thrice, etc. *Lexikos*, then, has far more repeat contributors than IJL. However, the maximum number of contributions for a single contributor stands at around thirty for both journals.

Studying who is behind the high number of contributions, one obtains the data shown in Tables 6 and 7. In Table 6 the focus is on the peer-reviewed contributions: articles and review articles (as regrouped above). The top five here is: Gouws, De Schryver, Prinsloo, Wiegand, and Louw. In Table 7 the focus is on the book reviews. The top five there is: Abecassis, Smit, Louw, Nkabinde, and Swanepoel. The overall top five brings together the *Lexikos* Editor and the four AFRILEX Presidents so far, rather surprisingly in their tenure order: Gouws (with 32 contributions), Du Plessis (32), Prinsloo (20), Alberts (18), and De Schryver (17). Of course, the high overall ranking for the first four is helped by the editorial material each wrote.⁴

Table 6: Top contributors to articles and review articles in *Lexikos* (with A = article and review article, B = book review, P = proceedings paper, O = other, E = editorial material; -, ✓ or ☑ = 0, 1 or more than 1 article or review article per year)

#	Author	Lexikos 1 (1991) through 18 (2008)	A	B	P	O	E	N
1	Gouws RH	✓-☑✓-☑✓✓✓✓✓✓✓✓✓✓✓✓✓✓	20	2	-	-	10	32
2	De Schryver G-M	-----✓✓☑✓✓☑☑☑☑	16	1	-	-	-	17
3	Prinsloo DJ	- ✓-----✓✓- ☑✓☑✓✓- ✓--	11	-	1	-	8	20
4	Wiegand HE	-----✓-----✓-✓✓✓✓✓✓✓✓	9	-	-	-	-	9
5	Louw PA	-----✓✓☑☑-----✓✓--	8	4	-	-	-	12
6	Otto AN	---✓✓-----✓✓- ✓- ✓- ✓- ✓-	7	1	-	-	-	8
7	Swanepoel PH	✓✓-- ✓-----✓-----✓--✓	6	3	-	-	-	9
8	Carstens A	-----☑☑- ✓- ✓-----	6	1	-	-	-	7
9	Alberts M	- ✓-----✓✓- ✓✓✓-----	6	-	-	4	8	18
10	Botha W	✓✓✓- ✓✓-----✓-----	6	-	-	1	-	8
11	Taljard E	-----✓✓- ☑-- ☑-	6	-	-	-	-	6
12	Mavoungou PA	-----✓✓✓- ✓✓-----	5	1	-	1	-	7
13	Chabata E	-----✓- ✓- ✓✓- ✓- ✓-	5	-	-	1	-	6
14	Tarp S	-----✓-----☑-- ✓✓	5	-	-	-	-	5
15	Smit M	-----✓- ✓✓✓-----	4	6	-	-	-	10
16	Beyer HL	-----✓- ✓-----✓✓--	4	1	-	-	-	5
	Feinauer I	-----✓-----☑✓	4	1	-	-	-	5
18	Lombard FJ	✓✓✓✓- ✓-----	4	-	-	-	-	4
	Mdee JS	--- ✓- ✓- ✓-----✓-----	4	-	-	-	-	4
20	Potgieter L	-----✓✓✓- ✓✓- ✓	3	1	-	-	-	4
21	Al-Ajmi H	-----✓-----✓- ✓	3	-	-	1	-	4
	Bergenholtz H	-----✓-----✓✓-	3	-	-	1	-	4
	Hadebe S	-----✓- ✓- ✓- ✓-	3	-	-	1	-	4
	Hartmann RRK	--- ✓ - ✓-----✓-----	3	-	-	1	-	4
25	Gauton R	-----✓-----✓✓	3	-	-	-	-	3
	Joffe D	-----☑✓-----	3	-	-	-	-	3
	Kotzé EF	- ✓-----✓- ✓-----	3	-	-	-	-	3

Table 7: Top contributors to book reviews in *Lexikos*

#	Author	Lexikos 1 (1991) through 18 (2008)	A	B	P	O	E	N
1	Abecassis M	-----✓☑☑- ✓	2	7	-	-	-	9
2	Smit M	-----✓✓- ✓☑✓	4	6	-	-	-	10
3	Louw PA	-----✓☑-----✓-----	8	4	-	-	-	12
4	Nkabinde AC	--- ☑-----	1	4	1	-	-	6
5	Swanepoel PH	-----✓-----☑-----	6	3	-	-	-	9
6	Heliel MH	- ✓-----✓- ✓-----	2	3	-	-	-	5
	Prinsloo AF	-----✓-----✓- ✓-	2	3	-	-	-	5
8	Du Plessis JCMD	-----✓- ✓- ✓- ✓-	1	3	2	2	24	32
9	Ridge E	--- ✓✓-----✓-----	-	3	-	-	-	3

Also shown in Tables 6 and 7 are the particular volumes of *Lexikos* in which the articles and review articles on the one hand, and the book reviews on the other, appeared. From these distributions across time, one may conclude that the

repeat contributors are also 'faithful' contributors to *Lexikos*, in the sense that once they start contributing, they also tend to continue to contribute over the years. In Table 6 there are only four uneven distributions (Carstens, Smit, Lombard, and Kotzé), and in Table 7 three (Nkabinde, Heliel, and Ridge). Note, further, that the two tables are complementary for three contributors (Smit, Louw, and Swanepoel), whereby a 'gap' in one category actually means that efforts have been put in another category (this is especially the case for Smit, who switched her attention from articles to book reviews over the years). Given that the ticks also tend to cluster near the right-hand side of the time scale, one may further conclude that *Lexikos* attracts ever more repeat contributors. This finding is clearly a very welcome one, as it bodes well for the future of *Lexikos*.

2.4 Continents, countries, cities and affiliations of contributors

A total of 649 affiliations (i.e. affiliation tokens) accompany the 543 *Lexikos* contributions. These will now be analysed, starting with a world view down, zooming in on the most productive region(s) at each step.

Although contributions are submitted from every corner of the world, there is a clear bias towards researchers who work from Africa, as seen in Table 8. As many as three quarters (74%) come from Africa, and 18% from Europe. Tiny numbers come from the Americas, the Middle-East, Asia and Australasia.

Table 8: Continents of contributors in *Lexikos*

Continent	N	%
Africa	478	73.65
Europe	116	17.87
Americas	12	1.85
Middle-East	10	1.54
Asia	7	1.08
Australasia	5	0.77
(no affiliation)	21	3.24
	649	100.00

Within Africa, over 86% of the contributions come from South Africa, 6% from Zimbabwe, 2% from Namibia, etc. See Table 9. The top three for Europe is: Belgium, Germany, and the UK, each with about one-fifth of the contributions from Europe. For the Americas, 10 come from the USA and 2 from Canada; for the Middle-East, 8 come from Kuwait and 2 from Egypt; for Asia, 4 come from Hong Kong and 3 from Japan; and for Australasia, 5 come from Australia.

With about 200 countries in the world, that means that about 15% is represented in *Lexikos*. With about 50 countries each in Africa and Europe, about 20% of the African as well as 20% of the European countries are represented. The other continents cannot really be said to be represented; rather, they are sampled.

Table 9: Countries of contributors in *Lexikos*

#	AFRICA	N	%	#	EUROPE	N	%	#	AMERICAS	N	%
1	South Africa	413	86.40	1	Belgium	25	21.55	1	USA	10	83.33
2	Zimbabwe	31	6.49	2	Germany	24	20.69	2	Canada	2	16.67
3	Namibia	11	2.30	3	UK	23	19.83			12	100.00
4	Tanzania	6	1.26	4	Denmark	16	13.79	# MIDDLE-EAST		N	%
5	Gabon	5	1.05		Netherlands	16	13.79	1	Kuwait	8	80.00
6	Botswana	3	0.63	6	Poland	4	3.45	2	Egypt	2	20.00
	Kenya	3	0.63	7	Sweden	3	2.59			10	100.00
8	DR Congo	2	0.42	8	Norway	2	1.72	# ASIA		N	%
	Ghana	2	0.42	9	Finland	1	0.86	1	Hong Kong	4	57.14
10	Nigeria	1	0.21		France	1	0.86	2	Japan	3	42.86
	Zambia	1	0.21		Switzerland	1	0.86			7	100.00
		478	100.00			116	100.00	# AUSTRALASIA		N	%
								1	Australia	5	100.00
										5	100.00

Zooming in on South Africa, where 64% of all *Lexikos* contributions come from, one obtains the breakdown in terms of cities shown in Table 10.

Table 10: South African cities of contributors in *Lexikos*

#	City	N	%	#	City	N	%
1	Stellenbosch	190	46.00	13	Potchefstroom	3	0.73
2	Pretoria	122	29.54	14	Alice	2	0.48
3	Johannesburg	18	4.36		Isipingo	2	0.48
4	Bellville	14	3.39		Mankweng	2	0.48
5	Polokwane	11	2.66	17	Jeffreys Bay	1	0.24
6	Bloemfontein	9	2.18		Letaba	1	0.24
7	Port Elizabeth	8	1.94		Linden	1	0.24
8	Grahamstown	6	1.45		Mafikeng	1	0.24
	KwaDlangezwa	6	1.45		Thornville	1	0.24
10	Cape Town	5	1.21		Upington	1	0.24
11	Durban	4	0.97		Wellington	1	0.24
	Soshanguve	4	0.97				
						413	100.00

Within South Africa, 46% emanate from Stellenbosch; the second spot is for Pretoria, with 30% of all contributions; the third for Johannesburg, with just 4%. Zooming in on Pretoria, as done in Table 11, one sees that several groups at the University of Pretoria as well as at the University of South Africa are very actively contributing to *Lexikos*, with one third (33%) from the Department of African Languages at the University of Pretoria alone.

Table 12, lastly, zooms in on Stellenbosch. The two biggest contributing centres here are the Department of Afrikaans and Dutch at the University of Stellenbosch with 39%, and of course, with 45%, the Bureau of the WAT itself — the birthplace of *Lexikos*, and also where the journal is still being edited and printed.

Table 11: *Lexikos* contributors from Pretoria

Affiliation	N	%
<i>University of Pretoria</i>	54	44.26
Department of African Languages	40	32.79
Department of Computer Science	2	1.64
Unit for Language Skills Development	2	1.64
Department of Afrikaans	1	0.82
Northern Sotho NLU – Pretoria Branch	1	0.82
School for Languages	1	0.82
"University of Pretoria"	7	5.74
<i>University of South Africa</i>	25	20.49
Department of Afrikaans (and Theory of Literature)	9	7.38
Department of African Languages	7	5.74
Department of Linguistics	3	2.46
Department of Art History, Visual Arts and Musicology	2	1.64
Department of Criminal and Procedural Law	1	0.82
Department of Criminology	1	0.82
Department of Public Administration	1	0.82
Department of Social Work	1	0.82
TshwaneDJe HLT	14	11.48
Pan South African Language Board (PanSALB)	11	9.02
Department of Arts and Culture	6	4.92
National Terminology Services, Department of National Education	5	4.10
(<i>independent scholar</i>)	5	4.10
Translate.org.za	1	0.82
Unit for African Languages, Human Sciences Research Council	1	0.82
	122	100.00

Table 12: *Lexikos* contributors from Stellenbosch

Affiliation	N	%
<i>University of Stellenbosch</i>	101	53.16
Department of Afrikaans and Dutch	75	39.47
Department of Music	11	5.79
SU Centre for Language and Speech Technology (SU-CLaST)	2	1.05
Department of African Languages	1	0.53
Department of Curriculum Studies	1	0.53
Department of Didactics	1	0.53
Department of English	1	0.53
Department of Logistics	1	0.53
Law Faculty	1	0.53
Medicine Faculty	1	0.53
Unit for Afrikaans, Language Centre	1	0.53
Unit for isiXhosa, Language Centre	1	0.53
"University of Stellenbosch"	4	2.11
Bureau of the WAT	86	45.26
(<i>independent scholar</i>)	3	1.58
	190	100.00

3. The *Lexikos* Corpus

In order to transcend the mere bibliometric details discussed in Section 2, a corpus was built consisting of every single contribution in *Lexikos* so far, except for the 18 lists of 'Publication Announcements'. It thus contains $543 - 18 = 525$ files, and after conversion to plain text the size of that corpus is 17MB. The *Lexikos* Corpus can be loaded into and queried with a corpus tool, just like any other corpus. We used WordSmith Tools (Scott 2009). The total size of the *Lexikos* Corpus is about 2.5 million running words, with about 100 000 distinct words (2 493 250 tokens; 103 069 types). A subsection of this full corpus was extracted, namely all English contributions, 355 files in all (as there are 170 non-English files, cf. Table 1), good for 1 744 536 tokens and 65 485 types. As a comparison, the English part of the IJL corpus (covering the period January 1988 up to December 2008) contains 592 files, good for 3 042 282 tokens and 111 903 types (cf. De Schryver 2009a).

Why undertake to build such a corpus? The more traditional approach to study journal contents is to assign keywords to each contribution, and then to draw up various tables that show where (i.e. in which issues, and thus which years) certain topics are being discussed. From such manually drawn-up tables one can then attempt to deduce certain trends. Such an approach has been followed for *Lexikos* before, in an M.A. dissertation (De Schryver 1999: 16-17). Although manual approaches are most definitely valuable, in today's day and age the use of automated approaches with which unlimited amounts of data may be manipulated, and with which that data may be processed and reprocessed ad infinitum, is fast becoming the norm. Querying the *Lexikos* Corpus thus enables one to 'look' at the full contents of *Lexikos* all at once.

The main 'tool' that will be used below to do so is WordSmith's KeyWords function, a tool that automatically extracts words that 'stand out' in a first corpus by means of a comparison with the relative frequencies of those same words in a much larger reference corpus. The reference corpus used is the 100-million-word BNC World Edition.⁵ Moreover, in order to uncover trends, the English section of the *Lexikos* Corpus was partitioned in three, as follows:

- Phase 1: 1991–1996 (six years, 69 files)
- Phase 2: 1997–2002 (six years, 130 files)
- Phase 3: 2003–2008 (six years, 156 files)

In Addenda 1 to 3, the top 300 keywords in each of these three *Lexikos* sub-corpora are listed. In analysing those, it is important to recall that the *full text* of each and every (English language) contribution is included: title, affiliation(s), abstract, keywords, entire article text, endnotes, references, and addenda — as applicable. Given that all these levels are included, one obtains a better picture of what is truly happening in *Lexikos*, especially compared to a study that would merely take the stated keywords into account, for example.

The Addenda are a real source of information, and only a few selected aspects can be touched upon within the space of a single journal article. Each of the lists of keywords contains at least seven types of terms:

- Academic speak (cf. next paragraph)
- Case study terms (cf. next paragraph)
- Non-English words (cf. next paragraph)
- Dictionaries (cf. Section 3.1)
- Proper names (cf. Section 3.2)
- Languages (cf. Section 3.3)
- True keywords (cf. Section 3.4)

'Academic speak' refers to the academic formulae, sentence and thought connectors, as well as abbreviations typically used in academic writing. These include: according, also, as such, based, cf., Ed., Eds., et al., etc., given, in, is, namely, non-*x*, of, or, presented, regard, should, specific, therefore, these, thus, various, and vs. 'Case study terms' are words that have been used in particular case studies and are by consequence very frequent in (some sections of) the corpus. Examples include: kaffir/kaffer, oneself, manifestable, considerate, thoughtful, etc. 'Non-English words' are often also the result of particular case studies or of the text in foreign examples used, such as: *bohlale, dipuku, ka, ke, le, morutiši, mosadi, mpša, na, sa, wa, ya, ...* in Northern Sotho, or *achtung, angst, freude, gefühl, überwinden, ...* in German. These are not particularly interesting for the purposes of the current study. The other categories are, and are looked into in the following sections.

3.1 Dictionaries in *Lexikos*

Extracting the dictionary names from the Addenda one obtains Table 13.

Table 13: Dictionaries in the three *Lexikos* sub-corpora

Phase 1 (1991–1996)		Phase 2 (1997–2002)		Phase 3 (2003–2008)	
Keyword	Keyness	Keyword	Keyness	Keyword	Keyness
WAT	3039.17	WAT	3318.27	WOORDEBOEK	2994.28
WOORDEBOEK	1384.41	WOORDEBOEK	2667.21	WAT	2895.51
AFRIKAANSE	1181.81	AFRIKAANSE	1515.44	ISN	1624.42
OXFORD	411.96	COBUILD	850.31	AFRIKAANSE	1530.81
WNT	360.16	DICTIONNAIRE	706.04	DURAMAZWI	1396.02
COBUILD	330.31	DURAMAZWI	575.86	DICTIONNAIRE	1164.80
ECD	300.46	DGC	538.35	OXFORD	1120.67
DICTIONNAIRE	270.04	RECHISHONA	515.24	COBUILD	1083.54
OED	267.87	MACQUARIE	492.82	RECHISHONA	866.49
		GROOT	458.70	LONGMAN	829.72
		ISN	444.37	ISICHAZAMAZWI	625.79
		VERKLARENDE	434.42		

The picture is rather stable, both in terms of number of keywords and keyness values. The WAT (*Woordeboek van die Afrikaanse Taal*) and COBUILD dictionaries appear throughout. During Phase 1, the WNT (*Woordenboek der Nederlandse Taal*), the ECD (*Explanatory Combinatorial Dictionary*), and the OED (*Oxford English Dictionary*) also receive considerable attention. Moving to Phase 2, Oxford dictionaries disappear out of the top keyword list, while three Zimbabwean dictionaries enter — *Duramazwi reChiShona*, *DGC (Duramazwi Guru reChiShona)*, and *ISN (Isichazamazwi SesiNdebele)* —, as well as the *Macquarie Dictionary* from Australia, the *Groot Noord-Sotho woordeboek*, the *Groot Woordeboek/Major Dictionary*, and the *HAT (Verklarende Handwoordeboek van die Afrikaanse Taal)*. This state is consolidated in Phase 3. Throughout, a variety of French dictionaries is also referred to.

3.2 Lexicographers in *Lexikos*

There are two types of names in the corpus: personal and place names. Extracting all personal names from the Addenda one obtains Table 14.

Table 14: Lexicographers in the three *Lexikos* sub-corpora

Phase 1 (1991–1996)		Phase 2 (1997–2002)		Phase 3 (2003–2008)	
Keyword	Keyness	Keyword	Keyness	Keyword	Keyness
HARTMANN	1301.68	WIEGAND	7095.03	GOUWS	3321.64
WIEGAND	984.95	GOUWS	2121.63	SCHRYVER	3163.63
HAUSMANN	766.13	PRINSLOO	1665.49	WIEGAND	3051.72
BLOUNT	686.81	HAUSMANN	1543.78	BERGENHOLTZ	2464.73
BRANFORD	559.64	WIEGAND'S	1353.79	TARP	2134.20
ZGUSTA	495.23	SCHRYVER	1261.29	PRINSLOO	2133.77
NADEN	478.79	ALBERTS	682.21	HARTMANN	1518.58
GOUWS	472.71	GILLES	653.54	HESELENG	1232.34
OSSELTON	270.12	BERGENHOLTZ	636.47	PHILIPPSON	1213.09
TONO	269.36	ZGUSTA	626.37	GILLES	1105.95
ILSON	258.87	KRIEL	599.05	HAUSMANN	1047.65
MEL'CUK	247.61	HARTMANN	592.92	CHIMHUNDU	991.65
		TARP	585.90	SCHOLTZ	983.35
		CLUVER	575.86	ZGUSTA	847.23
		LOUW	557.01	MAVOUNGOU	808.72
		LOUWRENS	545.55	HADEBE	711.44
		WYK	541.63	BOSMAN	683.80
		SMIT	525.56	SVEN	683.58
		SAGER	474.41	NIENABER	655.69
		KONERDING	444.52	HÜLLEN	654.68
				MAROLE	606.54
				TOIT	599.28
				RAPONDA	548.77
				MAURICE	547.90
				HENNING	545.84
				JOFFE	544.51

One firstly observes that the number of lexicographers competing for keyword status grows with each new phase (those scholars thus displace true lexicographic keywords, cf. Section 3.4). While the 'big shots' Hartmann, Wiegand, Hausmann and Zgusta are represented throughout, the overpowering presence of Wiegand in Phase 2 (note his keyness value there!) is astonishing. If anything, this points to *Lexikos's* openness towards metalexicography. With each volume of *Lexikos*, Gouws is gathering more and more attention; see how he rises through the phases. Important here is that his keyness values also increase at each step. The fact that Zgusta seems to go down from one phase to the next is thus misleading: his keyness values too, are increasing at each step; cf. the highlights in Table 14.

3.3 Languages in *Lexikos*

Extracting the languages from the Addenda one obtains Table 15.

Table 15: Languages in the three *Lexikos* sub-corpora

Phase 1 (1991–1996)		Phase 2 (1997–2002)		Phase 3 (2003–2008)	
Keyword	Keyness	Keyword	Keyness	Keyword	Keyness
AFRIKAANS	11942.42	AFRICAN	7485.60	ENGLISH	9489.79
ENGLISH	8755.20	ENGLISH	6987.90	SHONA	8319.34
AFRICAN	4861.51	AFRIKAANS	5313.14	NDEBELE	7895.39
DUTCH	3783.88	SOTHO	3946.56	AFRIKAANS	6101.93
SWAHILI	2202.69	HAUSA	3424.94	AFRICAN	5608.01
YIDDISH	1664.82	SEPEDI	3212.81	SOTHO	4607.16
ZULU	1365.02	SHONA	3058.44	SWAHILI	3051.10
JEWISH	1310.92	ZULU	1648.46	SESOTHO	3029.74
ARABIC	1151.52	NDEBELE	1485.13	LEBOA	2512.87
XHOSA	1144.14	XHOSA	1335.53	ZULU	2401.48
NETHERLAND	1057.53	SWAHILI	1289.27	SETSWANA	1778.25
SOTHO	945.81	SESOTHO	1122.08	FRANÇAIS	1771.51
BENG	826.54	FRANÇAIS	990.08	VENDA	1409.96
MALAYU	810.37	ARABIC	879.79	BANTU	1169.46
AMERICAN	692.96	BANTU	834.20	ISIXHOSA	982.02
KISWAHILI	480.59	CILUBÀ	788.02	SHANGANI	953.14
HOTTENTOT	478.79	ISIXHOSA	616.27	TSHIVENDA	943.51
MUSLIM	478.49	ISIZULU	606.17	SEPEDI	905.00
ISIXHOSA	472.71	DUTCH	519.55	YILUMBU	895.37
BANTU	441.47	SETSWANA	518.29	KHWE	779.84
ASHKENAZIC	384.80	NOORD	515.43	DEUTSCH	690.33
SHONA	358.27	NORTHERN	456.66	HAUSA	685.63
KISI	337.65	VENDA	454.28	KISWAHILI	669.92
NORTHEASTERN	331.40	SWEDISH	442.12	YIPUNU	616.17
MALAYO	292.63			NGUNI	605.85
AUSTRALIAN	261.75			DANISH	591.96
OTJIHERERO	247.61			ZEZURU	577.66
NGUNI	239.39				

Note that, with 'Languages in *Lexikos*', we are not referring to the languages in which the contributions are written (this aspect was dealt with above, cf. Figure 2), but to the languages that are the objects of the investigations. Those languages are overwhelmingly English and Afrikaans, and a good number of Bantu languages. Most key Bantu languages are the (official) South African ones, in addition to: Swahili (Kiswahili), Shona (with dialect Zezuru), Cilubà, Otjiherero, Shangani, Yilumbu, and Yipunu. African non-Bantu languages include: Beng, 'Hottentot', Kisi, Hausa, and Khwe. Also key are: Dutch, New Netherland Dutch, and Cape Muslim Afrikaans; (Northeastern) Yiddish, Jewish, and Arabic; American English, Ashkenazic English, and Australian English; French, Swedish, German, and Danish; as well as Malayu/o.

Although this is an interesting distribution, one cannot escape the fact that these language lists are biased towards the official South African languages, so this warrants a more in-depth study. To do so, eighteen different sub-corpora were built, one for each *Lexikos* volume, from which all non-English contributions were again subtracted. Comparing each of these eighteen sub-corpora with the BNC, keyness values can now be calculated 'per year', and set out graphically.⁶ In Figure 4 this is done for the keywords Afrikaans, English and Bantu.

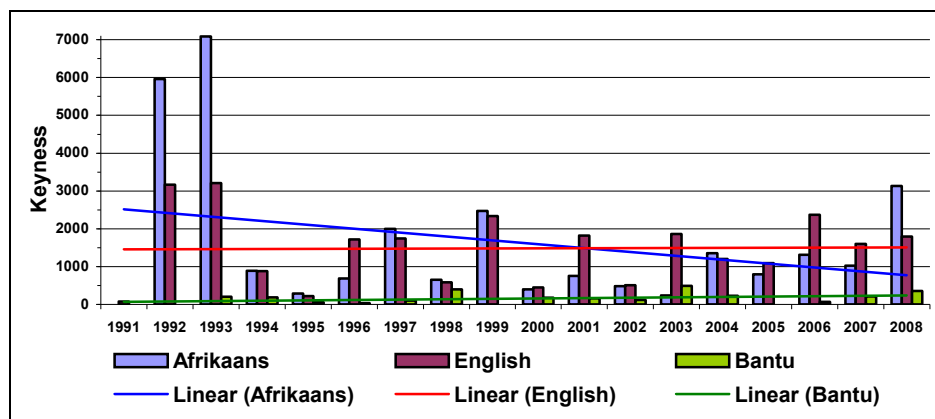


Figure 4: Keyness values for *Afrikaans*, *English* and *Bantu* in each *Lexikos* issue

On the whole, the keyness value of Afrikaans is decreasing over the years, that of English is mostly stable, while that of Bantu rises only slightly. The term 'Bantu' being stigmatized in South Africa, even with reference to the language family, it is not entirely surprising that the term is not really picking up speed.

In order to unravel the real trends, then, one needs to go down to the language level. In Figure 5 this is done for the so-called Sotho group of southern Bantu languages. Note, however, that 'Sotho' can refer to the group itself, but also to two of its members (Northern Sotho and Southern Sotho), while 'Sesotho' can refer to Southern Sotho or even Northern Sotho when that language is referred to in Northern Sotho itself, viz. as 'Sesotho sa Leboa'. Sepedi, although

actually but one of the dialects of Northern Sotho, normally unambiguously refers to Northern Sotho. The third member of the Sotho family is unproblematic: it is either referred to as Tswana (in English) or Setswana (in Tswana). Regardless of these ambiguities, it is clear from Figure 5 that the Sotho languages started to be increasingly discussed from 1997 onwards, and peaked for about seven years, from 2001 until 2007. The main language within the family being discussed is Northern Sotho (with often, for *Lexikos* 12 for example, all research articles devoted to it).

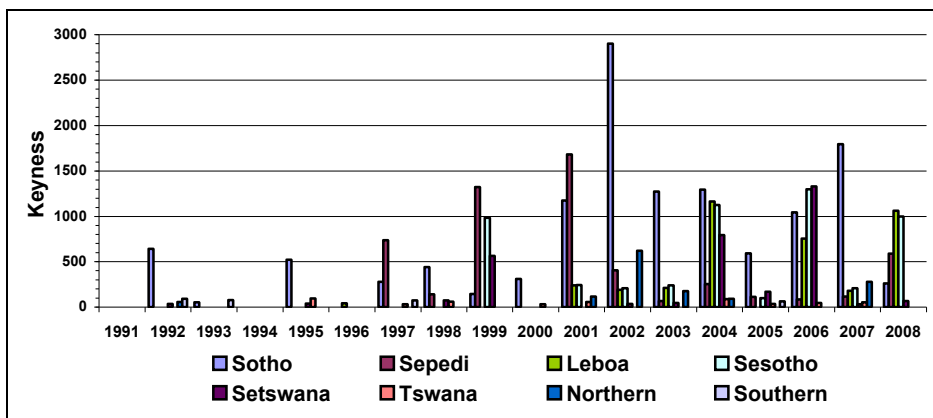


Figure 5: Keyness values for 'Sotho languages' in each *Lexikos* issue

Moving to the Nguni family of southern Bantu languages, one notes that one member is ambiguous: 'Ndebele', as it can refer to the Ndebele spoken in South Africa (isiNdebele in the language) or Zimbabwean Ndebele. The extreme values for Ndebele in Figure 6 are actually mostly for Zimbabwean Ndebele.

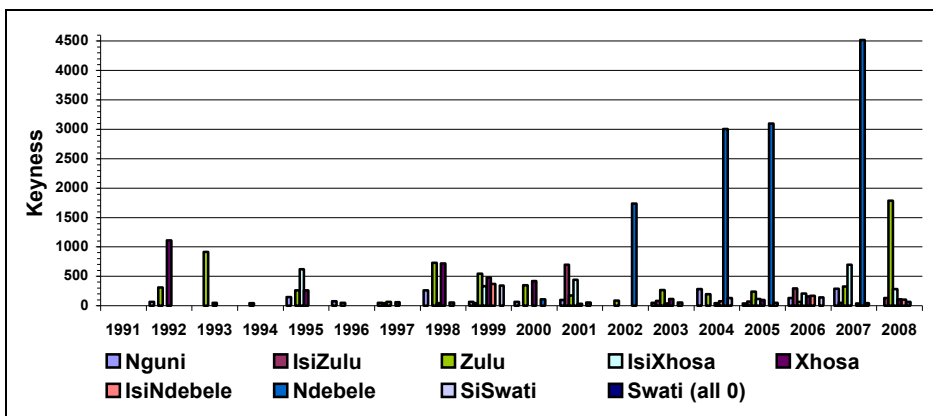


Figure 6: Keyness values for 'Nguni languages' in each *Lexikos* issue

Compared to the Sotho family, then, the Nguni family of languages is not frequently discussed in *Lexikos*, except for the period 1998–2001, and starting with 2008 for Zulu. 1992 was also a good year for Xhosa.

To complete the picture for the official South African languages, Figure 7 displays the keyness values for Venda (Tshivenda) and Tsonga (Xitsonga). Both these languages are not often discussed, with their keyness values on the whole (much) lower than for the other official South African languages.

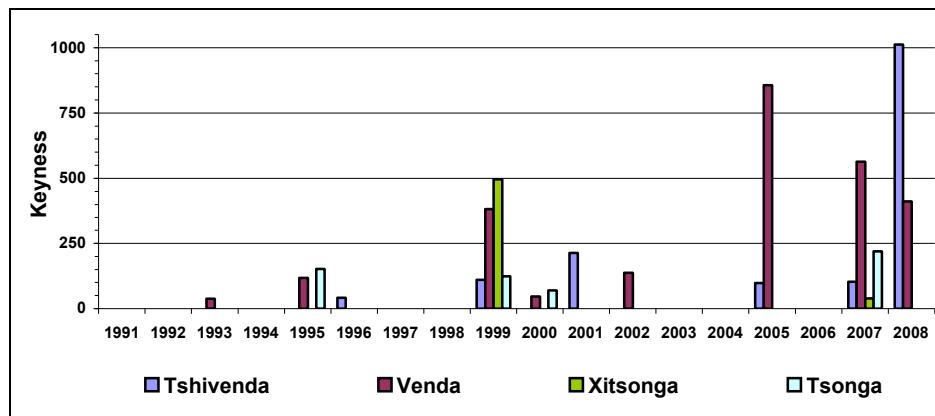


Figure 7: Keyness values for *Venda* and *Tsonga* in each *Lexikos* issue

3.4 Key topics in *Lexikos*

Most of the remaining keywords in Addenda 1 to 3 are true keywords. Analysing those brings out the real features of a journal like *Lexikos*. A few groups of related keywords and their patterns will be looked into below and, where relevant, comparisons will be made with the corresponding patterns found in IJL (cf. De Schryver 2009a).

3.4.1 Parts of Speech (POSS)

In Table 16 the parts of speech (POSS) as well as a few related terms relevant for the study of Bantu languages — all excerpted from the addenda — have been tabulated. The differences between the three phases are striking. In Phase 1, only two POSS stand out: *verbs* and *nouns* (together with the related *plural*), and for the Bantu languages also *stem*, *reflexive*, *ideophone*, and *prefix*. Going to Phase 2, *copulative* (and also *singular*) is added to the basic core; and going to Phase 3, *adverb*, *demonstrative*, *adjective*, and *pronoun* are further added, as well as *suffix* and *cl.* (class) for the Bantu languages, and even *POS* (part of speech) itself. The list of word classes studied from a lexicographic point of view most definitely grows with the years in *Lexikos*, and in that it is the opposite from the

IJL pattern. Actually, following the IJL study we 'predicted' that such lists of POSs would only grow whenever 'new' languages would become the object of lexicographic investigation. This is the case in *Lexikos*, where many of the languages are being described for the very first time.

Table 16: Parts of speech (POSs) in the three *Lexikos* sub-corpora

Phase 1 (1991–1996)		Phase 2 (1997–2002)		Phase 3 (2003–2008)	
Keyword	Keyness	Keyword	Keyness	Keyword	Keyness
VERB	2440.16	VERB	2335.04	VERB	1791.75
VERBS	1782.64	NOUN	2218.57	NOUN	1481.39
NOUN	1305.44	NOUNS	1327.65	ADVERBS	1314.56
NOUNS	1209.48	COPULATIVE	1091.27	NOUNS	1219.52
STEM	596.12	VERBS	1054.38	VERBS	1191.18
REFLEXIVES	562.76	COPULATIVES	798.12	DEM	1068.77
IDEOPHONES	450.20	PLURAL	734.68	ADJ	981.19
PLURAL	436.76	SINGULAR	609.03	SUFFIX	980.06
PREFIX	387.42	PREFIX	549.19	ADJECTIVES	912.92
PREFIXES	287.91	STEM	487.98	ADJECTIVE	870.45
		PLURALS	440.21	CL	830.74
				PRON	820.91
				POS	748.93
				PREFIX	665.35
				COPULATIVE	655.69
				VERBAL	651.68

3.4.2 N-lingual lexicography

Monolingual, bilingual, trilingual and multilingual lexicography are all being discussed in *Lexikos*, but not in equal proportions, and those proportions change with time, as summarized in Figure 8.

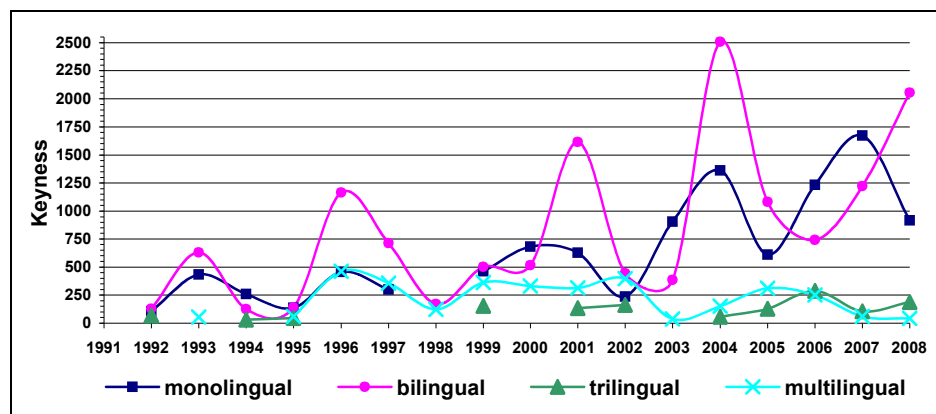


Figure 8: Keyness values for n-lingual in each *Lexikos* issue

Overall, bilingual lexicography is more popular than monolingual and multilingual lexicography in lexicographic discussions. While trilingual lexicography only appears sporadically, multilingual topics are actually slightly decreasing. Both bilingual and monolingual lexicography are on the rise, with the acceleration of the former greater than that of the latter.

The core task in monolingual lexicography is defining, so one could investigate whether the occurrence of key terms like *defining*, *definition(s)* or *explanatory* have changed during the past eighteen years. Apart from one peak for *definitions* in 2007, all these key terms remained stable, as seen in Figure 9.

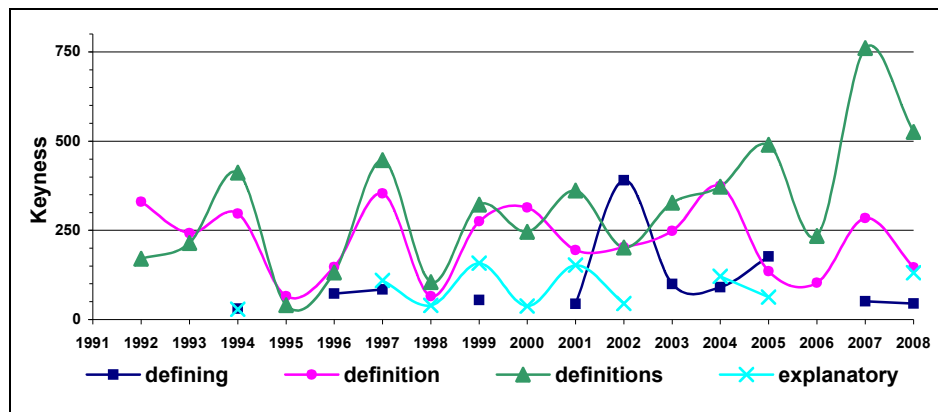


Figure 9: Keyness values for *defining* etc. in each *Lexikos* issue

Likewise, the core task in bilingual lexicography is to provide translation equivalents, so Figure 10 shows whether or not changes have occurred in this regard, looking at the key terms *equivalence*, *equivalent(s)*, *translating*, and *translation(s)*. Here one notices an increased activity for the period 2002–2004.

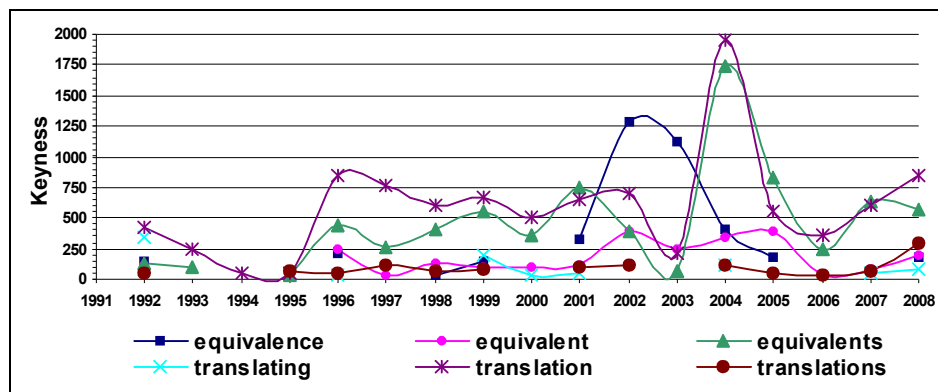


Figure 10: Keyness values for *translating* etc. in each *Lexikos* issue

3.4.3 Terminology and terminography

With regard to the sub-disciplines of terminology and terminography, Figure 11 clearly indicates that there has been a flurry of activity in the period 1997–2004, especially in the years 2000 and 2002.

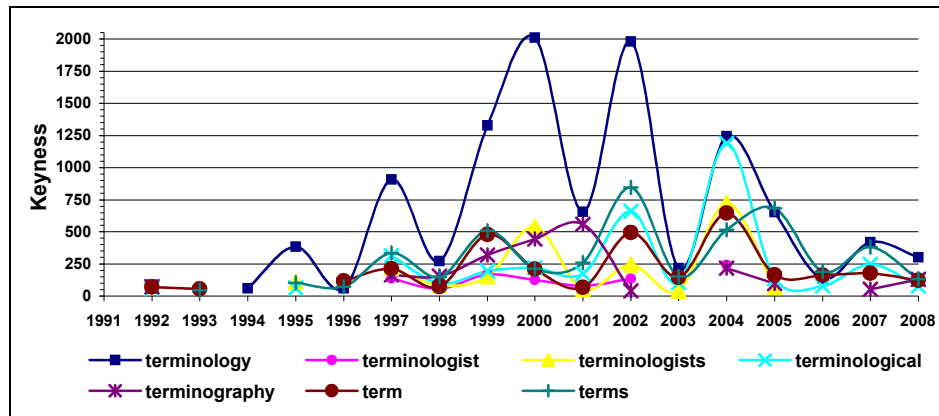


Figure 11: Keyness values for *terminology etc.* in each *Lexikos* issue

3.4.4 Metalexigraphy

In Section 3.2 above, we noted *Lexikos's* openness towards metalexigraphy. We are now in a position to analyse this aspect in more depth. Figure 12 plots the distribution of the general keywords *concept(s)*, *principle(s)*, *scientific* and *theoretical*. From the graph it is clear that the period 1999–2004 should be regarded as the most active. Focusing on lexicography itself, this is confirmed in Figures 13 through 15.

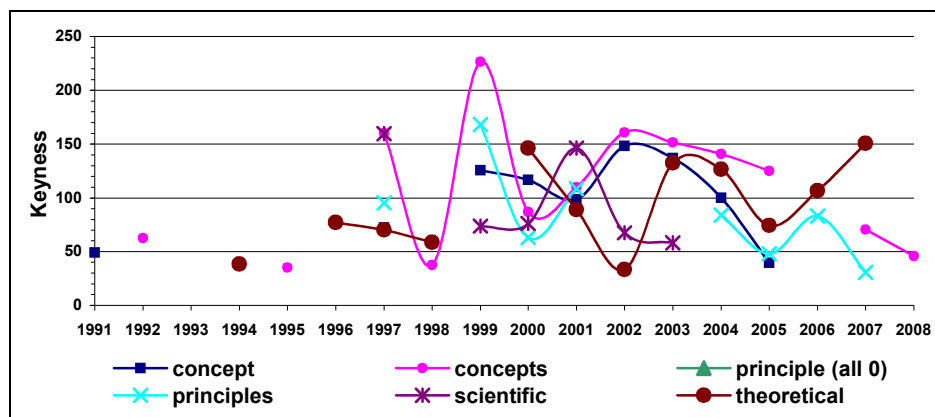


Figure 12: Keyness values for *theoretical etc.* in each *Lexikos* issue

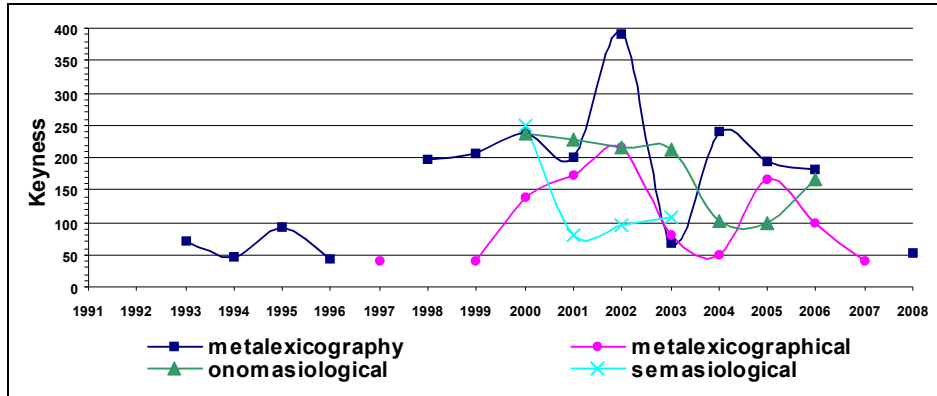


Figure 13: Keyness values for *metalexigraphy* etc. in each *Lexikos* issue

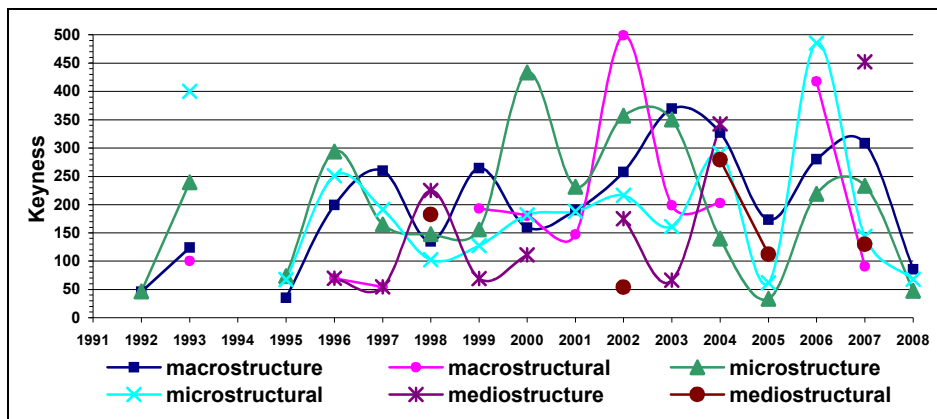


Figure 14: Keyness values for *x-structure* in each *Lexikos* issue

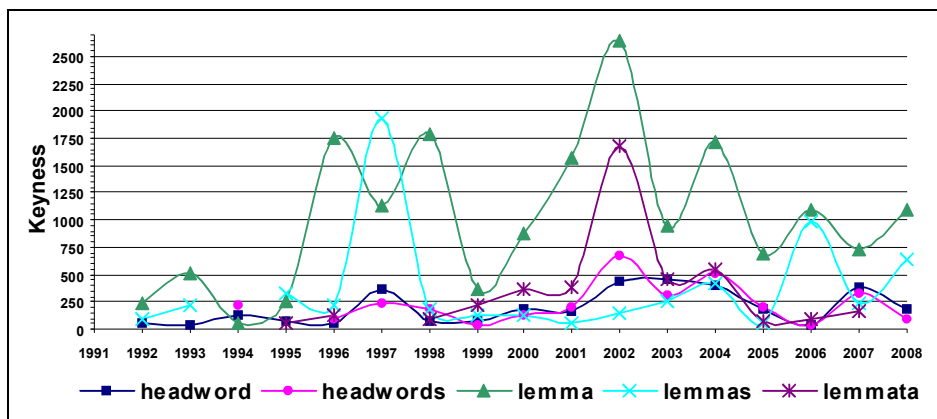


Figure 15: Keyness values for *lemma* etc. in each *Lexikos* issue

In Figure 13, the terms *metalexicography* and *metalexicographical* are especially key for the period 2000–2005, while highly technical terms like *onomasiological* vs. *semasiological* also become key during that period. Figure 14 looks at where terms like *macro-*, *micro-* and *mediostucture* peak, which is roughly 2000–2004. In Figure 15, lastly, the common *headword(s)* is contrasted with the scientific *lemma(s)/lemmata*, which clearly indicates that the vast majority of the contributors to *Lexikos* indeed use field-specific terms.⁷ These results stand in sharp contrast to those found for *IJL*, where far fewer theoretical terms are found in the keyword lists.

3.4.5 Polysemy and synonymy

Figures 16 and 17 summarize the use of respectively polysemous and synonymous terms in *Lexikos*.

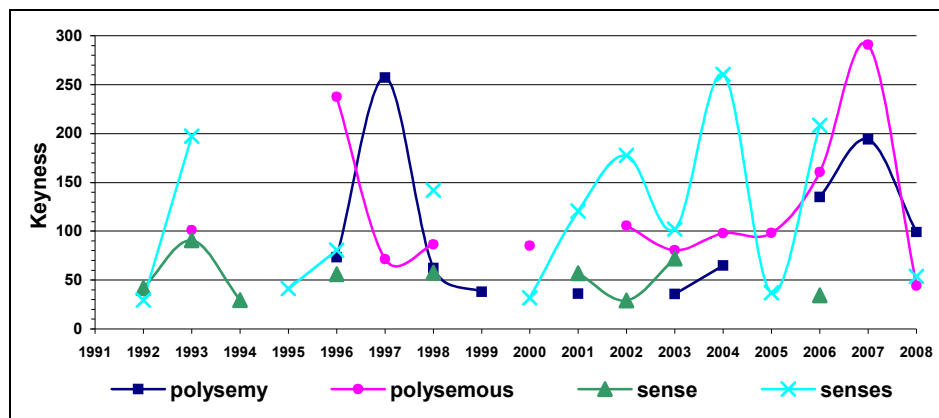


Figure 16: Keyness values for *polysemy* etc. in each *Lexikos* issue

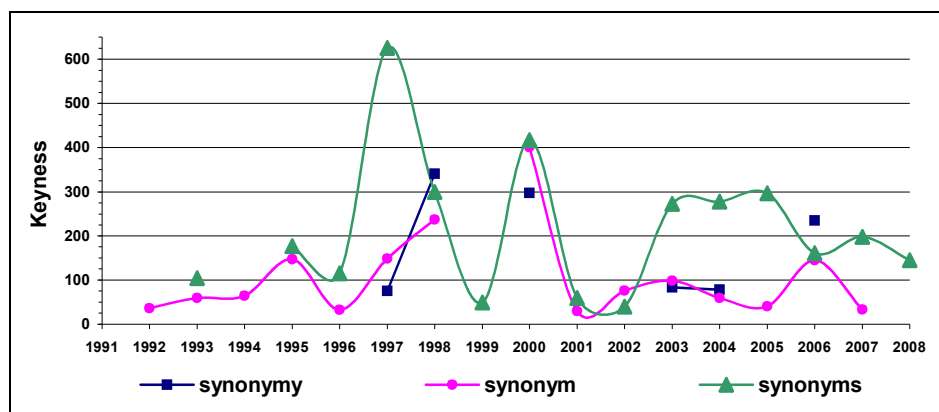


Figure 17: Keyness values for *synonymy* etc. in each *Lexikos* issue

From Figures 16 and 17 one may conclude that polysemy and synonymy are (a) not typically treated together in metalexicographical discussions, and (b) not always at the centre of such discussions. Both these findings run against what one would have expected.

3.4.6 Meaning, collocations, and examples vs. idioms/proverbs, and dialects/slang

While there are lexicographic topics for which one expects some kind of continuity, such as *meaning*, *collocations* or *examples*, others only appear sporadically, such as *idioms* and *proverbs* or *dialects* and *slang*. The difference can clearly be seen from Figures 18 through 22. Such graphs thus visualize where which topics are being discussed, and with which weighted frequency. In each of those graphs, peaks nonetheless pinpoint heightened activity.

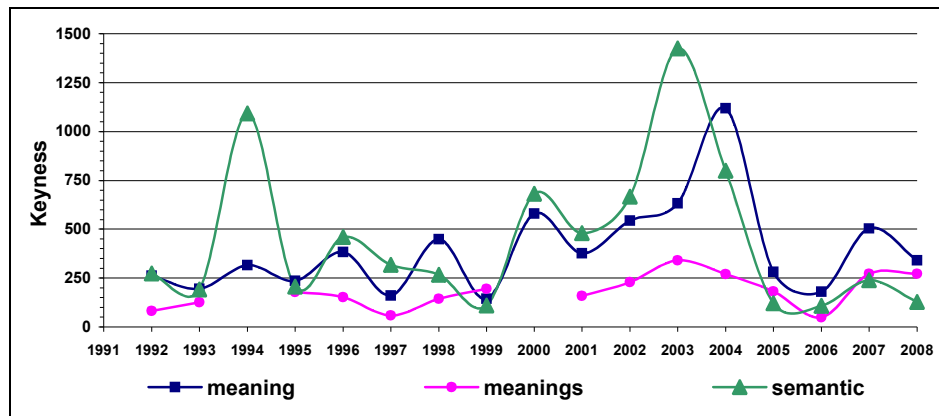


Figure 18: Keyness values for *meaning(s)* and *semantic* in each *Lexikos* issue

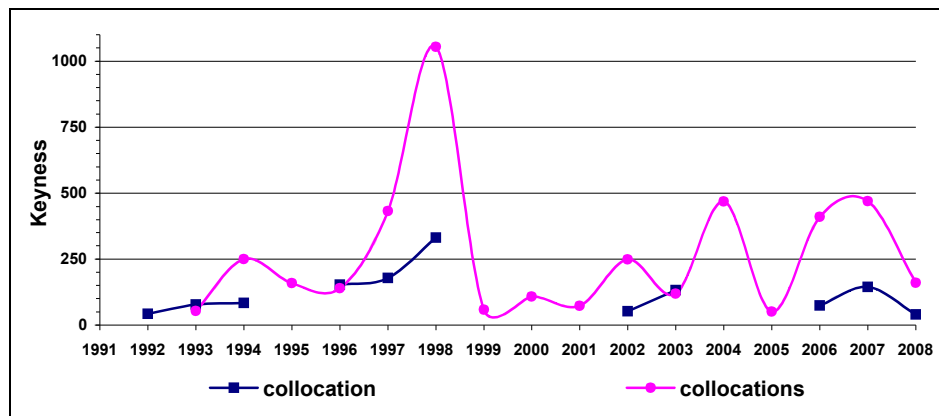


Figure 19: Keyness values for *collocation(s)* in each *Lexikos* issue

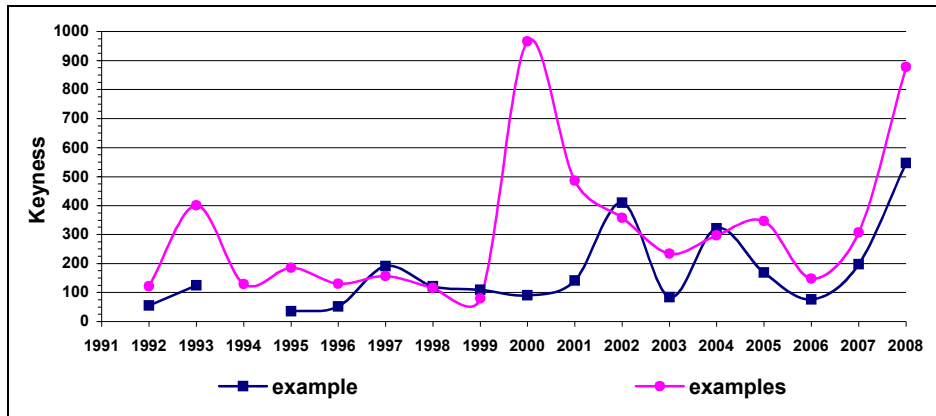


Figure 20: Keyness values for *example(s)* in each *Lexikos* issue

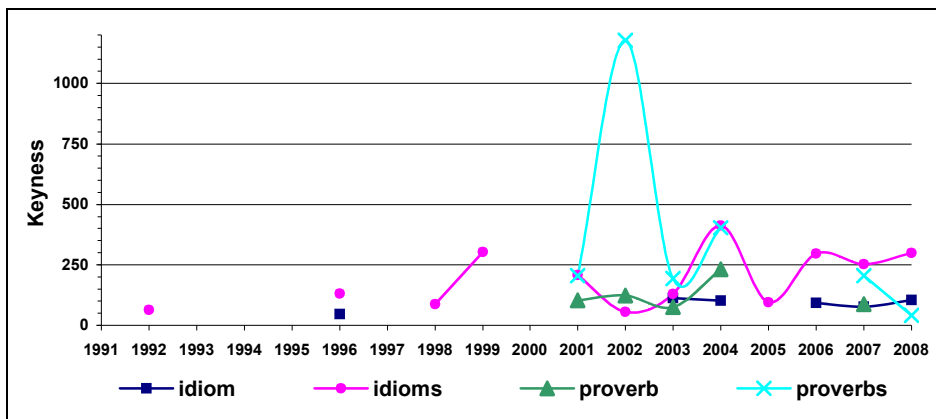


Figure 21: Keyness values for *idiom(s)* and *proverb(s)* in each *Lexikos* issue

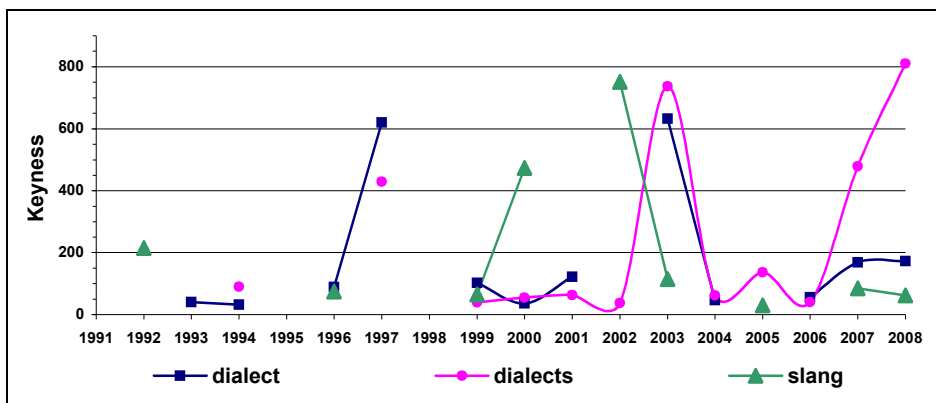


Figure 22: Keyness values for *dialect(s)* and *slang* in each *Lexikos* issue

3.4.7 The user perspective

To conclude this section, one may also briefly look at the user perspective. While the high day of the user perspective was found to be the period 1997–2004 in IJL, Figures 23 and 24 indicate that that period is narrowed down in *Lexikos* to basically two years: 2001 and 2004 (recall the special section on 'Learners' Dictionaries' that was added in Volume 14, cf. Section 2.1 above).

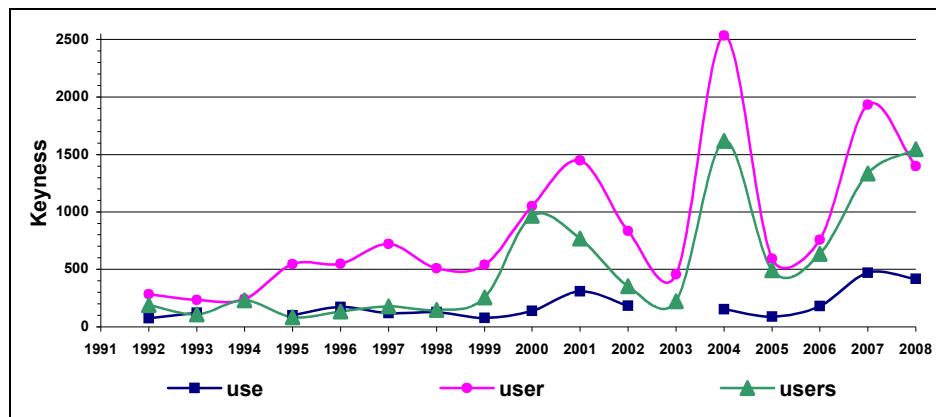


Figure 23: Keyness values for *use* and *user(s)* in each *Lexikos* issues

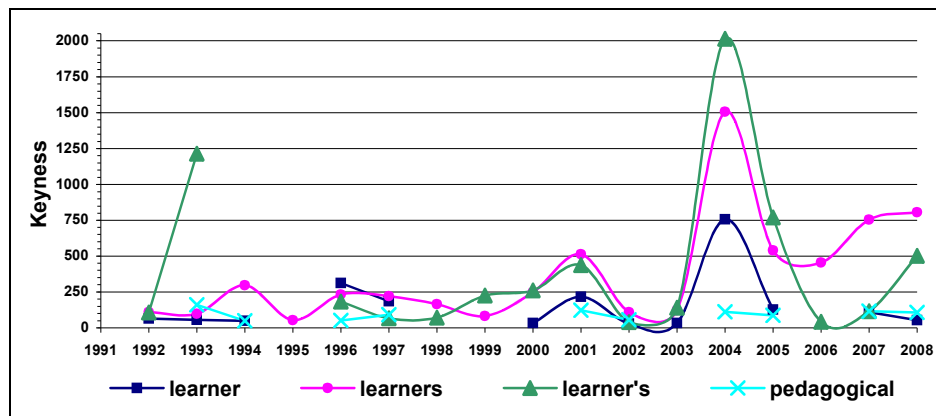


Figure 24: Keyness values for *learner(s) etc.* in each *Lexikos* issue

4. Summary and discussion

In this article an in-depth analysis was presented of both the publication details and the characteristic lexical items for the journal *Lexikos* from the first volume to the present. In Section 1 it was pointed out that *Lexikos* has become increasingly more professional with each new volume. Section 2 was devoted to an

analysis of the *Lexikos* Metadata Database. It was shown that the number of pages as well as the number of contributions — averaging 366 pages and 30 contributions per volume — steadily rose over the years. English has become the main discourse language (currently stabilized at 70%, on average 65%), followed by Afrikaans (currently stabilized at 20%, on average 26%). Other languages used are German, French, and Dutch. About 44% of the contributions are articles, 4% are review articles, and 20% are book reviews. As many as 83% of the contributions are single-authored, 12% are joint efforts. A total of 260 scholars contributed to *Lexikos* during the first eighteen years, one third of them (34%) more than once. Those repeat contributors are also 'faithful' in that their input is regular. Most repeat contributors moreover started publishing in *Lexikos* over the past few years, which bodes well for the future of *Lexikos*. About three quarters of the contributors (74%) are based in Africa, 86% of those in South Africa. Within South Africa, 30% are based in Pretoria and as many as 46% in Stellenbosch. In Pretoria, the most productive centre is the Department of African Languages at the University of Pretoria (33%), while in Stellenbosch it is the Department of Afrikaans and Dutch at the University of Stellenbosch (39%) and especially the Bureau of the WAT (45%).

Some caution in interpreting the last set of results is in order. While it is undeniable that a very large percentage of the material published in *Lexikos* emanates from South Africa, the African reach and especially the interaction with the rest of Africa is larger than these statistics suggest. Firstly, through various cooperation agreements, students from the rest of Africa are enrolled at South African universities, where they obtain degrees in lexicography. The prime example here is the Department of Afrikaans and Dutch at the University of Stellenbosch, where several Gabonese students have obtained their PhDs over the past few years. Many of them published their results in *Lexikos*, using their South African affiliations while in South Africa. This is for example the case for Mavoungou, nr. 12 in Table 6. The same is true for PhD students from Namibia (for example Beyer, nr. 16 in Table 6). Secondly, success breeds success. Both *Lexikos* and AFRILEX can definitely be seen as South African lexicographic success stories, and as a result of their reputation, scholars now continue to visit South Africa, wishing to do their research in South Africa rather, sometimes even on South African languages (as is the case for the current writer).

Section 3 was devoted to an analysis of the *Lexikos* Corpus, more in particular that part containing all 355 English-language contributions which appeared in *Lexikos* to date, totalling 1 744 536 tokens and 65 485 types. That corpus was partitioned in three sub-corpora (of 6 years each) as well as 18 sub-corpora (i.e. per year), in order to study various trends across time (based on the keyness measure as compared with the BNC). With regard to dictionaries discussed in *Lexikos*, a rather stable picture was found, both in terms of number of keywords and keyness values. In contrast, the number of lexicographers competing for keyword status increases with time. Looking at the languages as

objects of investigation, those were found to be overwhelmingly English and Afrikaans first, and biased towards the official South African languages second. Within the latter group, the order of attention is (a) Sotho languages (with especially Northern Sotho), (b) Nguni languages (with Zulu the most recent focus), (c) Venda, and lastly (d) Tsonga.

Seven key topic areas were also studied, the first being parts of speech, for which it was found that the list of word classes studied from a lexicographic point of view most definitely grows with the years in *Lexikos*. The second area was *n*-lingual lexicography, for which it was shown that bilingual lexicography is more popular than monolingual and multilingual lexicography in scholarly discussions. Thirdly, when it comes to terminology and terminography, a flurry of activity for the period 1997–2004 could be pinpointed. In a fourth section *Lexikos* was found to be highly theoretically oriented, certainly more than the *International Journal of Lexicography*, with the vast majority of the contributors using field-specific terms. Fifthly, for polysemy and synonymy the conclusions were that they are (a) not typically treated together in metalexicographical discussions, and (b) not always at the centre of such discussions. In a sixth section continuity was found for topics such as meaning, collocations or examples, while topics such as idioms and proverbs or dialects and slang are more sporadic. Lastly, attention to the user perspective in *Lexikos* was found to peak in the years 2001 and 2004.

Looking back, then, one cannot but be impressed by the achievements of *Lexikos*. One recognizes a healthy research culture, to which both lexicographers and metalexicographers contribute, on a variety of languages, in a variety of sub-disciplines. There is still a strong South African bias, however, which *Lexikos* as an adult should now shed.

Acknowledgements

Many thanks are due to Ms Riette Ruthven, who helped in assembling the *Lexikos* 1–10 material underlying the *Lexikos* Corpus. She managed to extract most data from legacy Word files, re-scanned the missing parts, and helped restoring diacritics. Volumes 11–17 were downloaded via the Sabinet Online service. *Lexikos* 18 was received directly as a single PDF file from the Bureau of the WAT.

Endnotes

1. Only very rarely is an article a reworked version of a published conference proceedings paper, such as Gouws (2001) vs. Gouws (2001a), or Mafela (2004) vs. Mafela (2005).
2. The percentage of editorial material is relatively high. This is so because all editorial material — written by the *Lexikos* Editor and the AFRILEX President — always appears in both Afrikaans and English.

3. In calculating this, some of the names were collapsed: 'Potgieter, Liezl' with (née) 'Gouws, Liezl', and 'Nong, Salmina' with (née) 'Maepa, Salmina'.
4. See endnote 2.
5. For the calculation of keywords and associated keyness values, the log-likelihood statistic of WordSmith 4.0 was used, with the minimum keyword frequency set at 3, and the maximum p value = 0.0000001. See for more details on the extraction of keywords for lexicographic purposes, Taljard and De Schryver (2002).
6. As keyness values are sensitive to (sub-)corpus size, figures for six-year windows are not directly comparable to per-year figures.
7. As one anonymous adjudicator correctly pointed out, this statement assumes (since these are keyness values, not frequencies) that the lemmas [headword] and [lemma] are of equal frequency in general English / the BNC. This is indeed a risky assumption, but as it turns out, is true: In the lemmatized BNC, the occurrences are respectively 65 vs. 60.

References

- BNC World Edition*. 2009. A 100 million word collection of samples of written and spoken language from a wide range of sources, designed to represent a wide cross-section of British English from the later part of the 20th century [online]. <http://www.natcorp.ox.ac.uk/>
- 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.** 2009. A Few Words from AFRILEX. *Lexikos* 19: xiv-xv.
- De Schryver, G.-M.** 2009a. Bibliometrics in Lexicography. *International Journal of Lexicography* 22(4).
- Du Plessis, J.C.M.D.** 2008. Foreword. *Lexikos* 18: xi.
- Gouws, R.H.** 2001. Enhancing the Access Structure of Translation Dictionaries. Lee, S. (Ed.). 2001. *ASIALEX 2001 Proceedings, Asian Bilingualism and the Dictionary: 77-83*. Seoul: Yonsei University, Center for Linguistic Informatics Development.
- Gouws, R.H.** 2001a. The Use of an Improved Access Structure in Dictionaries. *Lexikos* 11: 101-111.
- Harteveld, P.** 1991. Foreword. *Lexikos* 1: xii-xiv.
- Mafela, M.J.** 2004. Bilingual Dictionary and Meaning Discrimination in Venda. Williams, G. and S. Vessier (Eds.). 2004. *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004, Lorient, France, July 6-10, 2004: 501-509*. Lorient: Université de Bretagne Sud, Faculté des Lettres et des Sciences Humaines.
- Mafela, M.J.** 2005. Meaning Discrimination in Bilingual Venda Dictionaries. *Lexikos* 15: 276-285.
- Scott, M.** 2009. *WordSmith Tools* [online]. <http://www.lexically.net/wordsmith/>
- Taljard, E. and G.-M. de Schryver.** 2002. Semi-automatic Term Extraction for the African Languages, with Special Reference to Northern Sotho. *Lexikos* 12: 44-74.
- Web of Science*. 2009. Provides researchers, administrators, faculty, and students with quick, powerful access to the world's leading citation databases. Authoritative, multidisciplinary content covers over 10 000 of the highest impact journals worldwide [online]. <http://apps.isiknowledge.com/> [subscription required]
- Zgusta, L.** 1993. Book Review: Harteveld, P., Ed. 1991. *Lexikos*. *International Journal of Lexicography* 6(4): 305-306.

Addendum 1: Keywords and keyness values for *Lexikos* sub-corpus 1: 1991–1996

#	Keyword	Keyness	#	Keyword	Keyness	#	Keyword	Keyness	#	Keyword	Keyness
1	DICTIONARY	20062.88	76	Y	874.20	151	BANTU	441.47	226	STANDARD	298.10
2	AFRIKAANS	11942.42	77	X'S	856.70	152	SEMANTICS	440.44	227	ARTICLES	295.97
3	DICTIONARIES	9584.93	78	ENTRY	846.61	153	METAFORMS	438.95	228	MALAYO	292.63
4	ENGLISH	8755.20	79	ITEM	832.12	154	WÖRTERBUCH	438.95	229	WOORDENBOEK	292.63
5	LANGUAGE	6035.01	80	BENG	826.54	155	PLURAL	436.76	231	LIT	292.02
6	LEXICOGRAPHY	5827.79	81	TEXTUAL	824.81	156	CITATIONS	433.80	230	DEROGATORY	289.50
7	LEXICAL	4967.50	82	TEXT	823.28	157	ET	430.71	232	MORPHEME	288.87
8	AFRICAN	4861.51	83	MALAYU	810.37	158	ALLEX	427.69	233	OFFENSIVE	287.93
9	LANGUAGES	4281.88	84	NIE	808.10	159	NEUTRAL	425.71	234	PREFIXES	287.91
10	DUTCH	3783.88	85	EDS	781.55	160	TL	423.08	235	TERM	287.78
11	LEXEMES	3415.93	86	MAGN	780.80	161	REFERENCES	420.90	236	YA	287.69
12	WORD	3198.02	87	HAUSMANN	766.13	162	STELLENBOSCH	417.49	237	MORPHEMES	286.96
13	WAT	3039.17	88	ARTICLE	752.27	163	NATIVE	416.80	238	STRUCTURE	286.71
14	SOUTH	2817.94	89	LEARNERS	744.21	164	OXFORD	411.96	239	AFRIKA	286.63
15	VERB	2440.16	90	LEMMAS	734.55	166	ENCYCLOPEDIA	409.48	240	TRANSLATING	284.49
16	USAGE	2435.08	91	USERS	721.04	165	SYNONYMS	405.81	241	FURCHT	281.38
17	LEXICOGRAPHICAL	2427.54	92	NACC	694.09	168	KLA	405.18	242	GROLL	281.38
18	LEMMA	2413.09	93	AMERICAN	692.96	169	LEMATIC	405.18	243	KAMUSI	281.38
19	LEXICOGRAPHER	2357.89	94	BLOUNT	686.81	167	ÜBERWINDEN	405.18	244	VERFLIEGEN	281.38
20	SWAHILI	2202.69	95	AFK	686.57	170	PRONUNCIATION	400.25	245	EQUIVALENCE	281.21
21	RACIST	2112.91	96	AFRILEX	686.57	171	GRAPHS	395.77	246	MORPHOLOGICAL	280.59
22	SEMANTIC	2043.33	97	REFERENCE	661.17	172	BÖSCHUNG	393.93	247	ABBREVIATIONS	280.26
23	NON	2002.94	98	FR	648.36	173	IPA	390.51	248	SLANG	280.11
24	BILINGUAL	1998.65	99	EQUIVALENTS	640.55	174	IN	390.49	249	PHONETIC	277.58
25	WORDS	1919.72	100	LINGUISTS	635.97	175	DB	389.23	250	EXAMPLE	277.41
26	USER	1799.97	101	OCCURRENCE	627.08	176	TEXTS	389.17	251	NIGGER	274.58
27	LEXICOGRAPHERS	1791.22	102	CONDENSATION	624.27	177	PREFIX	387.42	252	DISCUSSION	273.04
28	VERBS	1782.64	103	EDITORIAL	611.66	178	ASHKENAZIC	384.80	253	DSYNTA	270.12
29	LINGUISTIC	1769.81	104	USAGES	600.86	179	LEXICOGRAPHICA	382.67	254	LEXIKOGRAFIE	270.12
30	YIDDISH	1664.82	105	IS	600.30	180	WUT	382.67	255	OSSELTON	270.12
31	ENTRIES	1644.69	106	STEM	596.12	181	FORM	378.27	256	SCHÜREN	270.12
32	LEXICOGRAPHIC	1638.07	107	MICROSTRUCTURAL	590.54	182	EQUIVALENT	378.20	257	DICTIONNAIRE	270.04
33	CF	1633.09	108	MICROSTRUCTURE	577.50	183	ORTHOGRAPHIC	374.30	258	TONO	269.36
34	AFRICA	1628.05	109	ANGST	564.87	184	AL	367.56	259	OED	267.87
35	ETYMOLOGY	1525.77	110	REFLEXIVES	562.76	185	ILLUSTRATIVE	367.24	260	CATEGORIES	267.04
36	TRANSLATION	1489.92	111	BRANFORD	559.64	186	WNT	360.16	261	ELKE	266.36
37	ITEMS	1452.08	112	UNIVERSITY	550.47	187	SHONA	358.27	262	HABEN	265.28
38	WOORDEBOEK	1384.41	113	LFS	549.51	188	HOFFNUNG	357.09	263	PRESS	264.49
39	ZULU	1365.02	114	TERMINOLOGY	547.88	189	MACROSTRUCTURE	352.84	264	AUSTRALIAN	261.75
40	MEANING	1359.95	115	CORPORA	547.11	190	WÖRTERBÜCHER	348.91	265	PEDAGOGICAL	260.47
41	CAPE	1336.03	116	MULTILINGUAL	537.95	191	ALPHABET	344.36	266	INDIGENOUS	260.43
42	JEWISH	1310.92	117	MEANINGS	537.46	192	POLYSEMIOUS	343.40	267	ONESELF	260.39
43	NOUN	1305.44	118	LF	529.46	193	CONCISE	342.38	268	ÄRGER	258.87
44	HARTMANN	1301.68	119	ORTHOGRAPHIES	523.47	194	LABEL	340.78	269	DICTIONNAIRES	258.87
45	CORPUS	1266.75	120	COLLOCATIONS	523.05	195	KISI	337.65	270	EMPFINDEN	258.87
46	LEARNER'S	1258.91	121	SPELLING	513.43	196	GRAMMAR	336.41	271	ENTZÜCKEN	258.87
47	MONOLINGUAL	1229.53	122	ORIGIN	513.31	197	EXPRESSIONS	333.49	272	ERREGEN	258.87
48	NOUNS	1209.48	123	USED	509.56	198	FORMS	331.60	273	ILSON	258.87
49	TAAL	1188.01	124	OF	506.45	199	NORTHEASTERN	331.40	274	JOOD	258.87
50	AFRIKAANSE	1181.81	125	PRETORIA	498.75	200	COBUILD	330.31	275	MANIFESTABLE	258.87
51	LINGUISTICS	1173.62	126	ZGUSTA	495.23	201	NIEMEYER	328.64	276	TWEETALIGE	258.87
52	COMPILERS	1169.18	127	LEXICOGRAPHER'S	495.19	202	HEADWORD	328.39	277	UNTERDRÜCKEN	258.87
53	LEXEME	1165.18	128	EDITION	491.29	203	BEGEISTERUNG	326.40	278	VERSETZEN	258.87
54	EMOTION	1162.40	129	SYNTACTIC	483.07	204	PANIK	326.40	279	WECKEN	258.87
55	GRAMMATICAL	1154.34	130	KISWAHILI	480.59	205	COLLOCATION	325.44	280	VOWEL	257.64
56	ARABIC	1151.52	131	ETC	479.51	206	SENSES	323.63	281	SUCH	256.82
57	BUREAU	1149.35	132	HOTTENTOT	478.79	207	NASAL	323.29	282	DATA	255.94
58	XHOSA	1144.14	133	NADEN	478.79	208	INCLUSION	321.69	283	TERMS	252.47
59	ACC	1090.04	134	MUSLIM	478.49	209	FEM	317.87	284	VS	250.87
60	LEXIKOS	1080.51	135	LEXICON	473.80	210	EIFERSUCHT	315.14	285	LEGEN	247.61
61	SPEAKERS	1071.30	136	GOUWS	472.71	211	FÜHLEN	315.14	286	MEL'CUK	247.61
62	NETHERLAND	1057.53	137	ISIXHOSA	472.71	212	GIVEN	313.79	287	NACHLASSEN	247.61
63	DEFINITION	1035.40	138	USE	471.46	213	SICH	312.65	288	OTJHERERO	247.61
64	GEFÜHL	1012.97	139	ACHTUNG	471.20	214	SYNONYM	311.47	289	SCHAM	247.61
65	ORTHOGRAPHY	985.06	140	FREUDE	467.72	215	LOB	310.39	290	ÜBERKOMMEN	247.61
66	WIEGAND	984.95	141	BL	464.54	216	THUS	310.28	291	LIEBE	244.88
67	ED	952.68	142	TÜBINGEN	461.46	217	TARGET	306.93	292	HEADWORDS	243.54
68	VOCABULARY	947.04	143	NTS	459.10	218	BEKOMMEN	303.89	293	CONDENSED	242.32
69	EXAMPLES	946.51	144	LINGUIST	454.15	219	EURALEX	303.89	294	BRUIN	240.94
70	SOTHO	945.81	145	LEARNER	453.59	220	LEIDENSCHAFT	303.89	295	GHANA	239.69
71	INFORMATION	944.23	146	IDEOPHONES	450.20	221	OR	303.39	296	NGUNI	239.39
72	DEFINITIONS	922.19	148	INL	450.20	222	COMPILATION	302.90	297	CATEGORY	237.46
73	KAFFIR	921.47	147	KAFFER	450.20	223	ECD	300.46	298	LABELS	236.93
74	NDAT	900.42	149	NAMES	443.92	224	ISO	300.29	299	BEDAVERN	236.36
75	PHRASAL	896.81	150	EXETER	443.51	225	AS	299.12	300	MITLEID	236.36

Addendum 2: Keywords and keyness values for *Lexikos* sub-corpus 2: 1997–2002

#	Keyword	Keyness	#	Keyword	Keyness	#	Keyword	Keyness	#	Keyword	Keyness
1	DICTIONARY	3722.13	76	ITEMS	1536.58	151	ELECTRONIC	783.27	226	DEFINING	538.12
2	DICTIONARIES	19532.26	77	BOHLALE	1525.54	152	MICROSTRUCTURAL	779.39	227	CONCEPT	537.93
3	LANGUAGE	13562.42	78	EQUIVALENCE	1521.71	153	MEANINGS	764.71	228	LIST	535.53
4	LEXICOGRAPHY	11015.13	79	AFRIKAANSE	1515.44	154	AL	757.99	229	PANSALB	535.45
5	LANGUAGES	8885.86	80	NDEBELE	1485.13	155	REFERENCES	756.41	231	GABONESE	530.05
6	DIE	8761.30	81	GRAMMATICAL	1455.08	156	KNOWLEDGE	750.00	230	SEMI	528.63
7	AFRICAN	7485.60	82	TERM	1410.82	157	TONGUE	749.67	232	SMIT	525.56
8	CORPUS	7412.14	83	WOORDEBOEKE	1384.10	158	CONCEPTS	748.61	233	PHRASES	525.15
9	WIEGAND	7095.03	84	TERMINOLOGICAL	1374.35	159	LOANWORDS	739.18	234	DRC	521.89
10	ENGLISH	6987.90	85	WIEGAND'S	1353.79	160	GABON	738.47	235	TEXTUAL	521.63
11	VAN	6791.82	86	DEFINITION	1347.76	161	SUID	737.50	236	DUTCH	519.55
12	LEMMA	6682.36	87	UNIVERSITY	1344.69	162	PLURAL	734.68	237	SETSWANA	518.29
13	TERMINOLOGY	6400.04	88	XHOSA	1335.53	163	FIELD	725.42	238	NOORD	515.43
14	WORD	6174.76	89	NOUNS	1327.65	164	DIALECT	723.68	239	RECHISHONA	515.24
15	LEXICAL	5329.38	90	MPŤA	1293.17	166	STANDARD	721.04	240	TERMINOLOGIST	515.24
16	AFRIKAANS	5313.14	91	SWAHILI	1289.27	165	GRUYTER	716.85	241	PARAPHRASE	514.66
17	LEXICOGRAPHIC	5216.82	92	SYNONYMS	1287.77	168	MOSADI	707.19	242	CATEGORIES	510.05
18	USER	4914.39	93	SCRHYVER	1261.29	169	DICTIONNAIRE	706.04	243	EXPLANATORY	509.26
19	WORDS	4630.34	94	LEARNERS	1247.80	167	FORMS	701.07	244	LEKSIKOGRAFIESE	505.14
20	SOTHO	3946.56	95	MICROSTRUCTURE	1242.85	170	CB	699.17	245	ONOMASIOLOGICAL	505.14
21	TRANSLATION	3693.35	96	TERMINOGRAPHY	1242.65	171	ENCARTA	697.09	246	LINGUISTS	498.46
22	BILINGUAL	3542.47	97	ET	1242.16	172	LEMATIZATION	697.09	247	SPELLING	497.15
23	HAUSA	3424.94	98	ALPHABETICAL	1237.81	173	TÜBINGEN	697.09	248	NAMELY	495.55
24	WAT	3318.27	99	ORTHOGRAPHY	1208.73	174	INTERNET	686.18	249	WOORDE	495.03
25	SEPEDI	3212.81	100	HEADWORDS	1199.94	175	ALBERTS	682.21	250	WÖRTERBÜCHER	495.03
26	LINGUISTIC	3160.48	101	NIE	1164.97	176	NA	680.52	251	MACQUARIE	492.82
27	ARTICLE	3079.72	102	BA	1140.54	177	CONSIDERATE	677.78	252	EDITION	492.50
28	SHONA	3058.44	103	TAAL	1123.86	178	COMPILED	676.60	253	SCIENTIFIC	490.98
29	TEXT	2955.75	104	SESOTHO	1122.08	179	SOURCE	673.48	254	FUNCTIONAL	490.14
30	CF	2710.90	105	PROVERBS	1106.09	180	DIPUKU	666.78	255	STEM	487.98
31	WOORDEBOEK	2667.21	106	ED	1096.87	181	GILLES	653.54	256	SEARCH	487.81
32	LEXICOGRAPHICAL	2641.68	107	COPULATIVE	1091.27	182	AS	652.44	257	SENSES	486.42
33	EQUIVALENTS	2488.19	108	SUBLEXICAL	1089.83	183	WÖRTERBUCH	646.58	258	LA	485.61
34	USERS	2484.20	109	HEADWORD	1087.33	184	ENTRY	644.56	259	COMPONENT	483.39
35	SOUTH	2450.55	110	SPECIFIC	1087.08	185	BERGENHOLTZ	636.47	260	OUTER	483.31
36	INFORMATION	2409.82	111	ACCESS	1078.63	186	EDITORS	634.70	261	GRAMMAR	481.12
37	LEXICOGRAPHERS	2374.74	112	ITEM	1072.15	187	ZGUSTA	626.37	262	SAGER	474.41
38	SEMANTIC	2366.90	113	SA	1060.79	188	KEYWORDS	623.40	263	VARIOUS	473.96
39	TEXTS	2346.14	114	VERBS	1054.38	189	ISIXHOSA	616.27	264	SEMANTICALLY	473.43
40	CORPORA	2335.47	115	SLANG	1042.87	190	REGARD	609.50	265	GE	472.78
41	VERB	2335.04	116	EXAMPLE	1011.57	191	SINGULAR	609.03	266	DESCRIPTIVE	470.10
42	LEXICOGRAPHER	2270.21	117	STELLENBOSCH	1005.08	192	ISIZULU	606.17	267	TERMINOLOGIE	465.12
43	NOUN	2218.57	118	MACROSTRUCTURE	1000.16	193	LEXEMES	603.77	268	ALLEX	464.73
44	LEXIKOS	2212.56	119	FRAÇAIS	990.08	194	AFRIKA	602.14	269	MEDIOSTRUCTURE	464.73
45	TERMS	2187.85	120	COPYRIGHT	972.04	195	NIEMEYER	602.10	270	METALEXICOGRAPHICAL	464.73
46	MONOLINGUAL	2158.57	121	COMPILERS	959.51	196	KRIEL	599.05	271	COLLOCATION	462.68
47	MEANING	2142.08	122	STRUCTURE	956.05	197	FREQUENCY	595.98	272	PRESENTED	459.78
48	KA	2138.15	123	LEARNER'S	955.96	198	HARTMANN	592.92	273	GROOT	458.70
49	GOUWS	2121.63	124	CONSONANT	949.92	199	TYPES	591.36	274	TECHNICAL	458.32
50	PRETORIA	2114.45	125	TERMINOLOGISTS	949.67	200	IDIOMS	589.38	275	NORTHERN	456.66
51	LEMMATA	2107.06	126	ENCYCLOPEDIA	933.89	201	CROSS	588.67	276	IN	455.25
52	EDS	2058.02	127	METALEXICOGRAPHY	929.46	202	MANTŠU	585.96	277	TRANSLATIONS	454.35
53	TARGET	2057.42	128	USE	929.07	203	TARP	585.90	278	PRINCIPLES	454.32
54	LEXICON	2010.31	129	LEXIKOGRAFIE	919.36	204	NON	585.67	279	VENDA	454.28
55	LEMMAS	2004.32	130	PROJECT	915.25	205	SYNONYMY	583.33	280	SENTENCES	452.24
56	AFRICA	1954.78	131	DSAE	909.25	206	SEMANTICS	581.43	281	ENCYCLOPAEDIC	451.08
57	EXAMPLES	1954.06	132	YA	897.54	207	CULTURAL	578.34	282	ETYMOLOGY	451.08
58	ARTICLES	1828.75	133	INDIGENOUS	889.38	208	ABBREVIATIONS	576.14	283	FEEDBACK	450.86
59	REFERENCE	1806.51	134	ARTIKEL	889.05	209	AFRIKATALE	575.86	284	THOUGHTFUL	449.75
60	USAGE	1777.37	135	OF	880.04	210	CLUVER	575.86	285	NICHE	445.59
61	VOCABULARY	1774.09	136	ARABIC	879.79	211	DURAMAZWI	575.86	286	KONERDING	444.52
62	COMPILATION	1714.60	137	LSP	869.89	212	ORDERING	574.86	287	ISN	444.37
63	SPEAKERS	1688.41	138	COBUILD	850.31	213	MULTI	567.96	288	ENCYCLOPEDIA	442.46
64	NICHING	1677.09	139	DATA	836.42	214	ALSO	562.00	289	SWEDISH	442.12
65	PRINSLOO	1665.49	140	BANTU	834.20	215	RESEARCH	559.41	290	ACCORDING	441.38
66	ZULU	1648.46	141	LEKSIKOGRAFIE	808.22	216	LOUW	557.01	291	BUSINESS	440.43
67	MULTILINGUAL	1641.82	142	MACROSTRUCTURAL	808.22	217	DIALECTS	555.02	292	PLURALS	440.21
68	COLLOCATIONS	1631.31	143	EQUIVALENT	808.09	218	PREFIX	549.19	293	THEREFORE	439.82
69	DEFINITIONS	1606.27	144	COPULATIVES	798.12	219	SL	545.88	294	NEEDS	439.36
70	KE	1597.31	145	SYNONYM	796.57	220	INL	545.55	295	REFERENCING	438.69
71	AFRILEX	1596.26	146	USED	795.13	221	LOUWRENS	545.55	296	LEMATISATION	436.90
72	MORUTIŠI	1586.16	148	NASAL	792.15	222	WYK	541.63	297	TREATMENT	436.57
73	ENTRIES	1579.21	147	IS	791.62	223	PRONUNCIATION	538.62	298	KAF	434.94
74	LINGUISTICS	1558.84	149	ETC	789.05	224	DGC	538.35	299	SEGMENT	434.49
75	HAUSMANN	1543.78	150	CILUBÄ	788.02	225	ROM	538.23	300	VERKLARENDE	434.42

Addendum 3: Keywords and keyness values for *Lexikos* sub-corpus 3: 2003–2008

#	Keyword	Keyness	#	Keyword	Keyness	#	Keyword	Keyness	#	Keyword	Keyness
1	DICTIONARY	60677.92	76	SETSWANA	1778.25	151	SCHOLTZ	983.35	226	HARARE	697.04
2	DICTIONARIES	34268.95	77	FRANÇAIS	1771.51	152	ISIXHOSA	982.02	227	MEDIOSTRUCTURE	693.19
3	LANGUAGE	22342.33	78	KA	1759.05	153	ADJ	981.19	228	SPOKEN	691.46
4	LEXICOGRAPHY	16604.31	79	USAGE	1737.13	154	SUFFIX	980.06	229	DEUTSCH	690.33
5	LANGUAGES	11389.05	80	WOORDEBOEKE	1723.37	155	DEFINITIONAL	973.51	231	HAUSA	685.63
6	ENGLISH	9489.79	81	ITEM	1678.11	156	LEKSIKOGRAFIE	972.39	230	EVALUATION	684.54
7	LEXICOGRAPHIC	8408.29	82	INTERNET	1649.28	157	COMPILERS	971.13	232	BOSMAN	683.80
8	SHONA	8319.34	83	STELLENBOSCH	1641.89	158	GRAMMAR	968.81	233	SVEN	683.58
9	CORPUS	8222.76	84	GABON	1631.54	159	FREQUENCY	962.22	234	EENTALIGE	683.56
10	NDEBELE	7895.39	85	ORTHOGRAPHY	1628.97	160	LEXIKOGRAFIE	953.14	235	MACROSTRUCTURAL	683.56
11	USER	7256.95	86	EQUIVALENCE	1626.88	161	SHANGANI	953.14	236	CROSS	680.86
12	BILINGUAL	6939.91	87	ISN	1624.42	162	SPELLING	949.29	237	CAPE	680.60
13	LEXICAL	6218.81	88	TÜBINGEN	1617.46	163	YA	944.02	238	FIELD	674.58
14	AFRIKAANS	6101.93	89	NON	1614.95	164	TSHIVENDA	943.51	239	TERMINOLOGISTS	673.93
15	WORD	6093.01	90	AFRICA	1570.08	166	CULTURE	934.00	240	KISWAHLI	669.92
16	WORDS	5703.51	91	SA	1569.55	165	DIALECT	926.54	241	PREFIX	665.35
17	AFRICAN	5608.01	92	ET	1569.43	168	GRUYTER	913.13	242	COPULATIVE	655.69
18	USERS	5540.67	93	LIST	1545.90	169	ADJECTIVES	912.92	243	NIENABER	655.69
19	MONOLINGUAL	5325.01	94	AFRIKAANSE	1530.81	167	INDIGENOUS	906.86	244	HÜLLEN	654.68
20	LEXICOGRAPHERS	5118.73	95	NIEMEYER	1529.75	170	LEARNER	905.73	245	VERBAL	651.68
21	LEMMA	5052.34	96	HARTMANN	1518.58	171	SEPEDI	905.00	246	TREATMENT	649.34
22	TEXT	4614.96	97	TSHWANELEX	1482.67	172	MICROSTRUCTURAL	904.76	247	NA	642.26
23	SOTHO	4607.16	98	TERMINOLOGICAL	1481.61	173	LEMMATISATION	900.28	248	COMMUNICATION	638.75
24	TRANSLATION	4161.60	99	NOUN	1481.39	174	ETYMOLOGICAL	898.83	249	IS	638.30
25	ARTICLE	4062.62	100	TAAAL	1462.62	175	HTTP	895.37	250	PRESCRIPTIVE	634.44
26	LEXICOGRAPHICAL	3968.11	101	OUTER	1459.17	176	YILUMBU	895.37	251	RESEARCH	628.99
27	LEXICOGRAPHER	3895.24	102	COLLOCATIONS	1441.60	177	PRONUNCIATION	892.31	252	POLYSEMOUS	628.43
28	LEARNERS	3806.67	103	PHAROS	1415.66	178	LEKSIKOGRAFIESE	885.74	253	MAIOR	627.43
29	LINGUISTIC	3735.64	104	VENDA	1409.96	179	EXPRESSIONS	884.43	254	LA	627.17
30	TEXTS	3717.04	105	DURAMAZWI	1396.02	180	ENTRY	883.62	255	ISICHAZAMAZWI	625.79
31	EQUIVALENTS	3552.15	106	HEADWORD	1383.63	181	ADJECTIVE	870.45	256	INTERLINGUAL	625.04
32	REFERENCE	3425.34	107	NIE	1380.57	182	RECHISHONA	866.49	257	SUBLEXICAL	625.04
33	GOUWS	3321.64	108	IDIOMS	1332.29	183	LSP	866.42	258	SENSES	622.08
34	ONLINE	3174.96	109	GRAMMATICAL	1328.31	184	INCLUSION	862.09	259	ELECTRONIC	621.15
35	SCHRYVER	3163.63	110	TERM	1327.90	185	THESE	859.80	260	ZIMBABWEAN	618.98
36	WIEGAND	3051.72	111	MEANINGS	1321.60	186	SEMANTICS	855.74	261	INFLECTIONAL	618.57
37	SWAHILI	3051.10	112	EXAMPLE	1314.65	187	ZGUSTA	847.23	262	X	616.54
38	SESOTHO	3029.74	113	ADVERBS	1314.56	188	TYPES	845.42	263	YIPUNU	616.17
39	WOORDEBOEK	2994.28	114	USE	1297.21	189	MICROSTRUCTURE	844.61	264	CRITERIA	611.04
40	LEXIKOS	2965.39	115	TARGET	1286.53	190	OF	839.19	265	TRILINGUAL	608.26
41	MEANING	2920.72	116	SPECIFIC	1270.07	191	COMPILING	835.00	266	MAROLE	606.54
42	LEARNER'S	2912.65	117	DEFINITION	1250.81	192	ALPHABETICAL	834.15	267	WÖRTERBÜCHER	606.54
43	WAT	2895.51	118	SOUTH	1239.26	193	CL	830.74	268	EDITORS	606.11
44	EDS	2797.59	119	SYNONYMS	1234.83	194	LONGMAN	829.72	269	NGUNI	605.85
45	ITEMS	2712.20	120	ARTIKEL	1232.34	195	PRESENTED	826.68	270	ETYMOLOGY	603.59
46	TERMINOLOGY	2686.55	121	HESELENG	1232.34	196	PHONETIC	824.18	271	BASED	601.95
47	COMPILATION	2645.66	122	HEADWORDS	1221.03	197	PRON	820.91	272	PRESENTATION	601.39
48	VOCABULARY	2605.56	123	NOUNS	1219.52	198	EURALEX	818.35	273	TOIT	599.28
49	DATA	2571.41	124	REFERENCES	1217.97	199	CULTURAL	809.36	274	AS	598.02
50	DEFINITIONS	2558.94	125	PHILIPPSON	1213.09	200	MAVOUNGOU	808.72	275	GA	596.57
51	LEBOA	2512.87	126	FORMS	1192.56	201	KEYWORDS	808.70	276	FOREIGN	593.24
52	SEMANTIC	2478.52	127	VERBS	1191.18	202	REGARD	793.71	277	DANISH	591.96
53	GABONESE	2472.23	128	MACROSTRUCTURE	1189.81	203	VARIANTS	791.83	278	ACCESS	587.32
54	BERGENHOLTZ	2464.73	129	LEXICON	1188.27	204	EDITION	787.19	279	SHOULD	585.16
55	ZULU	2401.48	130	EQUIVALENT	1180.09	205	KHWE	779.84	280	LINGUISTS	585.06
56	INFORMATION	2372.10	131	USED	1175.08	206	LEMMATIZATION	760.58	281	ZEZURU	577.66
57	ENTRIES	2360.37	132	AL	1175.01	207	TWEETALIGE	760.58	282	ALLEX	568.03
58	ED	2282.52	133	BANTU	1169.46	208	LEXSM	750.95	283	LEMMATIZED	568.03
59	CF	2236.65	134	DICTIONNAIRE	1164.80	209	POS	748.93	284	BIBLIOGRAPHY	565.35
60	ARTICLES	2217.55	135	OXFORD	1120.67	210	STRUCTURE	748.79	285	THEORETICAL	561.91
61	LINGUISTICS	2198.63	136	TLS	1113.15	211	DIFFERENT	746.43	286	FORM	555.74
62	SPEAKERS	2153.77	137	GILLES	1105.95	212	CANADIAN	743.56	287	ORTHOGRAPHIC	553.87
63	PRETORIA	2145.64	138	LE	1096.98	213	KNOWLEDGE	741.92	288	METALANGUAGE	551.73
64	TARP	2134.20	139	FUNCTIONS	1090.11	214	WA	741.80	289	LEXIKOGRAFIE	548.77
65	PRINSLOO	2133.77	140	LEXICOGRAPHICA	1087.93	215	ETC	734.21	290	RAPONDA	548.77
66	EXAMPLES	2094.41	141	COBUILD	1083.54	216	DESCRIPTION	730.45	291	MAURICE	547.90
67	CORPORA	2061.21	142	MORPHOLOGICAL	1072.77	217	MULTILINGUAL	719.94	292	HENNING	545.84
68	UNIVERSITY	2043.17	143	DEM	1068.77	218	SPEECH	717.52	293	JOFFE	544.51
69	LEMMAS	2010.79	144	LEMMATA	1059.53	219	PROVERBS	715.76	294	NEEDS	539.75
70	DIALECTS	1970.74	145	COMPILED	1057.86	220	HABE	711.44	295	INTRALINGUAL	539.15
71	AFRILEX	1944.81	146	WÖRTERBUCH	1049.42	221	ADDENDUM	711.21	296	METALEXICOGRAPHY	539.15
72	TERMS	1884.03	148	HAUSMANN	1047.65	222	SEARCH	709.00	297	MULTILEXICAL	539.15
73	ZIMBABWE	1841.84	147	KE	1029.89	223	ALSO	708.81	298	CATEGORIES	535.78
74	ALRI	1819.67	149	EFL	1011.61	224	LABELS	705.25	299	JOURNAL	532.78
75	VERB	1791.75	150	CHIMHUNDU	991.65	225	FUNCTION	700.96	300	SUCH	532.59

Keeping in Touch. A Survey of Lexicography Periodicals

R.R.K. Hartmann, *Exeter, United Kingdom* (*r.r.k.hartmann@exeter.ac.uk*)

Abstract: Journals and other serial publications are important to the work of lexicographers, so it is necessary to know what is available. On the basis of a list of 40 periodicals of relevance to lexicography, a comparative summary is attempted to give an impression of the major topics treated and formats used, including some desiderata for future improvements.

Keywords: LEXICOGRAPHY PERIODICALS, READERS' NEEDS, BIBLIOGRAPHIC TREATMENT

Opsomming: Om in voeling te bly. 'n Oorsig van leksikografietydskrifte. Tydskrifte en ander periodieke publikasies is belangrik vir leksikograwe se werk, dus is dit noodsaaklik om te weet wat beskikbaar is. Aan die hand van 'n lys van 40 tydskrifte wat tersaaklik is vir die leksikografie, is 'n kort vergelykende oorsig onderneem om 'n indruk te kry van die hoofonderwerpe wat behandel en die formate wat gebruik word, insluitende 'n aantal tekortkominge vir toekomstige verbetering.

Sleutelwoorde: LEKSIKOGRAFIETYDSKRIFTE, LESERSBEHOEFTE, BIBLIOGRAFIESE BEHANDELING

1. Introduction

What is the use of periodicals? This is an important question, as it can determine our attitude to progress in the field of lexicography. I experienced this recently while reviewing a book on phraseology (Hartmann 2009), a set of conference proceedings concerned with the linguistic study of multiword expressions, which did not give me a complete picture of how this might affect their treatment in dictionaries. Then I happened to come across a journal with a thematic section on 'Collocations in European lexicography and dictionary research' (more on this in Section 2 below), which helped me to place the whole subject in its proper context. What this example shows us is that periodicals can be extremely useful in providing information, although we still have to know what to expect, where to look for it, and how to decide what is relevant.

So what are the needs of readers, how do such publications meet them, and how do their editors and publishers know (or learn) to locate and present the required information? One representative case to examine is the journal *Lexicographica* (more on this in Section 2 below), which was started by Herbert Ernst Wiegand (Heidelberg) in the wake of the foundation of the European Association for Lexicography and in parallel with the book series *Lexicographica Series Maior*. With informal links to EURALEX and DSNA since 1985, it has produced 24 annual volumes of international repute.

I shall attempt to provide a survey here of relevant periodicals, some similar to and some different from *Lexicographica*. There are no precedents for this (I could not find any other easily accessible guide like this anywhere), so I had to undertake the task with the help of the appended list, which provides data on a selection of 40 serial publications, i.e. journals, magazines and newsletters for the field of lexicography in the narrow sense as well as some for neighbouring disciplines, differentiating where it seems appropriate between international and national organs, with a deliberate intention to look beyond English-language media. The Appendix gives details on their titles, sponsors, editors, ISSNs, publishers and numbers of volumes (and issues p.a.) published.¹

2. A case study: *Lexicographica*

Published once a year, *Lexicographica* has managed to follow (and often lead) major trends in the field, promoting a European and global outlook. The 13 papers on the theme of collocation in the most recent volume already mentioned (No. 24, 2008) are symptomatic of the journal's bridge-building approach, bringing together people from different countries, languages and specialisations under a capable editor — in this case Franz Josef Hausmann (Erlangen) — to reflect on recent developments in the subject under discussion: Henning Bergenholtz (Aarhus) on Danish, Peter Blumenthal (Cologne) on French, Sibilla Cantarini (Ferrara) on Italian, František Čermák (Prague) on Czech, Peter Ďurčo (Bratislava) on German and Slovakian, Martin Everaert (Utrecht) on Dutch, Thomas Herbst and Brigitta Mittmann (Erlangen) on English, Zita Hollós (Budapest) on Hungarian, Vida Jesenšek (Maribor) on Slovenian, Sven-Göran Malmgren (Gothenburg) on Swedish, Tadeusz Piotrowski (Opole) on Polish, Bernhard Pöll (Salzburg) on Spanish and Portuguese, and Kathrin Steyer (IdS Mannheim) on German. Although such overviews can help to clarify the current position, I must admit that I am still somewhat puzzled by the conflicting interpretations and definitions of even the basic notion of 'collocation' (do such constructions consist of two or more lexical items, do they involve modified nouns (like *stock-taking exercise*), verbs (*keep in touch*), and/or other word classes (*note down, on a humorous note*), are they 'fixed' like idioms or more 'free', and is statistical frequency a factor?) as well as several tensions that remain between linguistic theory and lexicographic practice.

Other topics covered in thematic issues of *Lexicographica* guest-edited by appropriate experts have included dictionary criticism, dictionary history, dictionary structure (such as the treatment of meaning, examples and idioms), dictionary typology (such as monolingual vs. bilingual lexicography, cultural, thesaurus and LSP lexicography, authors' and learners' dictionaries), and dictionary IT (such as computer applications and corpus linguistics). In addition to the regular feature of a thematic part, *Lexicographica* contains papers on other topics, reports on events such as conferences and dictionary projects, book reviews and occasional lists of bibliographical references to new publications.

3. Taking stock

Most bibliographies do not provide separate lists of periodicals. One exception is the international *Bibliographie Linguistique*, which started in 1949 and has appeared in over 50 volumes, although its list of excerpted lexicography journals covers only 17 of the periodicals discussed here; another is the *International Bibliography* of the Modern Language Association (MLA) which includes a Master List of Periodicals (now part of OCLC — see below); a third is the *Bibliographie Thématique* by Boccuzzi et al. (2007), which cites over 250 titles, although only a small proportion of these are devoted to the field of lexicography.

Many lists of bibliographic references to journals are incomplete or out of date (even the one on the EURALEX website), and there are very few attempts to extract the periodical literature in a systematic way (the journal *Linguistics and Language Behavior Abstracts* (LLBA) has recently agreed to increase its coverage of the periodicals discussed here to more than half of the 40 listed), a problem which is aggravated by the fact that journals themselves do not always keep track of their own papers in the form of regular indexes, which might be made worse by the reluctance on the part of editors and publishers to acknowledge the achievements of their competitors. Motivated by such deficiencies, I started a panoramic 'Zeitschriftenschau' many years ago (Hartmann 1964–1970) in an effort to make readers of a German journal of translation studies more aware of the literature in other journals of relevance to the subject. Similar compilations have been provided by Wiegand et al. in the above-mentioned annual journal *Lexicographica*, at regular intervals from Vol. 1 (1985) to Vol. 17 (2001), and also under the heading 'Zeitschriftenschau' in the journal *Zeitschrift für germanistische Linguistik* (since 2004).

Radical improvements have been made recently in the form of digital services such as the American *Genamics-Openly* and *OCLC-JournalSeek* database (part of the *Online Computer Library Center* at Dublin, Ohio), the *Thomson-Reuters Journal Citation Index*, EBSCO's *Electronic Journals Service*, and the Spanish *Dialnet* library search system at the Universidad de la Rioja. In spite of their impressive coverage (MLA and OCLC claim to list over 17 000 titles, and EBSCO over 19 000), not all lexicography periodicals are cited there; thus, *Lexicographica* is missing from LLBA, OCLC, *Thomson-Reuters* and EBSCO, although it can be found on *Dialnet*.

4. Lexicography and dictionary research

4.1 International journals

Our comparative survey can start with a description of the *International Journal of Lexicography* (IJL), which began a few years after *Lexicographica* as the journal of EURALEX with a full programme of essays, reports, review articles as well as lists and reviews of new publications. There have been three active editors

since 1988, Robert Ilson (London), Tony Cowie (Leeds) and Paul Bogaards (Leyden), supported by two associate editors, a review editor and an international editorial board, ensuring that all aspects of lexicography are covered, particularly through the medium of thematic issues, most recently on the legacy of John Sinclair (Birmingham) for the study of collocations and the use of corpus linguistics, guest-edited by Patrick Hanks (Brno) in Vol. 21 No. 3 (2008), on planning bilingual dictionaries, edited by Willy Martin (Amsterdam) in Vol. 20 No. 3 (2007), on corpus-based studies of German idioms, edited by Christiane Fellbaum (Berlin/Princeton) in Vol. 19 No. 4 (2006), on the 250th anniversary of Samuel Johnson's *Dictionary of the English Language*, edited by Anne McDermott and Rosamund Moon (Birmingham) in Vol. 18 No. 2 (2005), on lexicography in Italy, edited by Carla Marengo (Torino) in Vol. 17 No. 4 (2004), on multilingual databases for the benefit of bilingual lexicography, edited by Paul Bogaards in Vol. 17 No. 2 (2004) and on the FrameNet approach to lexicography, edited by Thierry Fontenelle (Microsoft, Redmond WA) in Vol. 16 No. 3 (2003).

There are not many truly international or continental or regional journals available for the field of lexicography. In addition to the *IJL*, the following three deserve to be singled out: *Dictionaries*, *Lexikos* and *LexicoNordica*, all with respectable histories and ambitious agendas. Several years before the foundation of EURALEX, the Dictionary Society of North America had been established (with the help of Edward Gates at Terre Haute IN), and by 1979 its journal had begun, entitled *Dictionaries*, with a series of editors from Richard W. Bailey (Michigan) and William Chisholm (Cleveland State) to Michael Adams (on leaving based at Indiana) and William Frawley (Chevy Chase MD). One of its innovations was a ten-year Index (in 1990 and 2000), although nowadays it has become easier to search for the contents of published issues, especially in journals with online connections. Another notable feature have been thematic issues dedicated in the last few years to the treatment of neologisms in lexicography and a tribute to Ladislav Zgusta (1924–2007) in Vol. 28 (2007), to 'hard problems' in dictionary-making in Vol. 27 (2006), to the memory of Allen Walker Read in Vol. 24 (2003), and to the completion of the *Middle English Dictionary* at the University of Michigan in Vol. 23 (2002).

Lexikos is the journal most familiar to its readers, so I need not go into details here except to acknowledge the regular and full coverage of most aspects of lexicography and its affiliation to the ever-growing African Association for Lexicography, although it had actually been initiated (by the *Woordeboek van die Afrikaanse Taal* at a meeting in Stellenbosch) some five years before the foundation of AFRILEX in 1995. The last volume, No. 18 (2008), offered 15 articles on a wide range of topics, dictionary genres and languages, a paper on lexicosoftware, a survey of dictionary use, a note on a pioneer lexicographer, a bibliography of material relevant to an etymological dictionary of Afrikaans, several reports on dictionary projects and other events, and a major review article on a new textbook.

No journals have yet emerged from the areas of AUSTRALEX and ASIALEX, but there is one run by the Nordic Lexicography Society (NFL) — *LexicoNordica* — which has become known particularly for its systematic treatment of thematic issues linked to an active programme of biennial conferences and annual symposia. Under the overall editorship of Henning Bergenholtz from Denmark and Sven-Göran Malmgren from Sweden, the last few volumes have covered topics like the user perspective (Vol. 15, 2008), Nordic encyclopedias and lexica (Vol. 14, 2007), historical dictionaries (Vol. 13, 2006) and electronic dictionaries (Vol. 12, 2005).

4.2 National journals

The journal *Lexicographica* has already been introduced, but in spite of its international links over the years it remains a national product, especially since the recent takeover of the publishing company Niemeyer by De Gruyter has brought about a change in the editorial team, concentrating on the interdisciplinary interests of five German academics.

The French journal *Cahiers de lexicologie* has been around for over half a century, undergoing several editorial and publishing changes. Topics covered include the French approach to dictionary research, reports on dictionary projects, 'hommages' to famous lexicographers, and comparative and historical accounts of important dictionaries as well as many issues in lexicological studies. To pick out just one sample issue, No. 88 (2006), it consists of 13 articles — one by Jean Pruvost (Cergy-Pontoise) on metalexigraphic theory, 11 by other authors on various features of specific dictionaries, and two on Josette Rey-Debove (1929–2005) — and three reviews of Le Robert and Larousse dictionaries. The new editors are keen to broaden the interdisciplinary coverage and international appeal of the journal.

The Chinese bimonthly journal *Cishu Yanjiu* (*Lexicographical Studies*) has been active for three decades. Sponsored by the Lexicographical Society of China (which has 8 committees on the various types and traditions of dictionary-making, and since the early 1990s has held numerous national conferences on general lexicography and symposia on bilingual dictionaries) and produced by the Shanghai Lexicographical Publishing House, *Cishu Yanjiu* has covered the practice and theory of lexicography with an impressively wide range of shortish articles, reviews, reports and occasional lists of references (multiplying the average number of papers per issue (10) by the number of issues p.a. (6) and the number of volumes published so far (30) results in the staggering figure of 1 800). It is entirely in Mandarin Chinese, but always supplies an English translation of its table of contents.

Another leading journal devoted to a particular country and the language(s) used there is the *Revista de Lexicografía*. Promoted as a national organ not by an association, but by an academic body, the Department of Spanish and Latin Philology at the University of Coruña, under the editorship of Mar Cam-

pos Souto and José Ignacio Pérez Pascual, it has been published in the form of annual volumes of about 6 to 9 papers each on mixed topics, and more recently it has been accompanied by 9 additional *Anexos*, which fulfil a function similar to that of thematic issues in *Lexicographica*, on such topics as the Spanish Royal Academy Dictionary (17 papers coordinated by the two journal editors, in Vol. 1, 2006), the 'Coruña' Spanish Dictionary (description and extracts, directed by José-Álvaro Porto Dapena, in Vol. 9, 2007), bilingual and learners' dictionaries (13 papers edited by Mar Campos Souto et al., in Vol. 8, 2008); interestingly, Vol. 3 (2006) of the *Anexos* offers 16 papers, edited by Margarita Alonso Ramos, on the dictionary treatment of collocations, idioms and other phraseological expressions, with references to such pioneers as John Sinclair and Igor Mel'čuk. The *Revista's* format is simpler than that of *Lexicographica*, with all papers being in Spanish (from now on a wider range of languages are being encouraged) and a gradually rising number of reviews. In contrast to *Lexicographica*, there is an international committee advising the editors of both the journal and the annexes.

Other nationally based journals include *Studi di Lessicografia Italiana* in Italy,² *Onomasiology Online* in Germany, *Lexicon* in Japan, and the journal of KOREALEX.

Onomasiology Online is special in that it exemplifies the growing tendency towards publications on the internet with relation to a rather rare alternative approach in lexicography. Most traditional dictionaries have been 'semasiological' in that they start from given words and attempt to explain their meaning(s), while 'onomasiology' starts with a particular thing or notion and tries to find the various ways in which it can be designated, an approach that has been associated mainly with thesaurus and synonym lexicography, and also with onomastics and technical terminology (see Sections 5.2 and 5.3 below). This German e-journal explores the ways cognitive linguistics, knowledge management and corpus linguistics can be brought to bear on a better understanding of lexical and communicative processes. The three editors have the help of an international board of consultants both for questions of substance and for the use of English by authors of individual contributions.

Hanguk-sajeonhak, the journal of the Korean Lexicography Association, has been published in the form of 2 volumes per year since 2003, under the editorship of Sang-kyu Seo (Yonsei, Seoul) with the support of a 13-member national editorial committee, providing articles and reports on all aspects of lexicography, including the newsletter of KOREALEX.

Lexicon is the journal published annually by the Iwasaki Linguistic Circle, an informal research network chaired by Shigeru Takebayashi (Tokyo), concerned with progress in English lexicography for over three decades, particularly in terms of developing objective criteria to apply to dictionary criticism and the study of dictionary history.

There is also a small group of local journals which started as newsletters (see Section 6 below), but have grown into full-fledged bulletins, magazines or journals, some even with international connections, such as the *Barnhart Dic-*

tionary Companion, which specialises in (American) English neologisms and features the Editor's Page on aspects of word-formation and lexicography, and the *Kernerman Dictionary News* which publishes reports on old and new dictionary projects from around the world, most recently (No. 16, 2008) on English, Japanese and Estonian, and on the future of Merriam-Webster.

We need to acknowledge that occasionally brilliant papers on lexicographic topics may be published in non-lexicographic journals, just as non-lexicographic topics may appear in lexicography journals. Two examples of the former are the thematic issues on the application of corpus linguistics to dictionaries and translation in a Polish journal (Kaszubski 2006) and on various aspects of dictionary use around the world in one published in Brazil (Welker 2007), but I cannot find any examples of the latter.

5. Wider perspectives

We have been concentrating on lexicography in the narrow sense of the word, but now we need to broaden the scope beyond it to other fields, beginning with *Reference Reviews*, one of the few journals which has managed to transcend many of the artificial boundaries that still exist today, e.g. between general and specialised dictionaries, between lexical and encyclopedic reference works, and between lexicography and its surrounding disciplines, thus contributing one answer to the intriguing question whether the whole of lexicography, together with other subjects mentioned below, might be moving towards what has been called 'reference science' (McArthur 1998) or 'informology' (Tarp 2007).

Started in the late 1980s by M.C.B University Press and now run by the Emerald Group in Northern England, *Reference Reviews* is unique in offering an open-minded but critical panorama 'of new reference materials prepared by librarians for librarians'. Thus, the final two issues of Vol. 22, Nos. 7 and 8 (2008), contain articles reviewing reference works, in part this time with special attention to 'output from lesser-known publishers', under the following subject headings: 'general' works (such as the *New Walford Guide to Reference Sources* reviewed by Stuart James, past editor of *Reference Reviews*, commenting on the irony that printed guides of this kind are still essential today in spite of the ever more widely available information on the internet), philosophy and religion (such as the *Encyclopedic Sourcebook of Satanism*), social sciences (such as the *Oxford Handbook of the United Nations*), business and management (such as the online *Business Plans and Profiles Index*), languages and literature (such as *The Greenwood Encyclopedia of Folk Tales and Fairy Tales*), science and technology (such as the online *U.S. Patent and Trademark Office*), arts (such as the *Cambridge Companion to Ballet*), geography, biography and history (such as *The 'Daily Telegraph' Dictionary of Tommies' Songs and Slang 1914–18*), area studies (such as *A Companion to Latin American History*), plus editorials and indexes.

Now we can branch out to neighbouring disciplines, such as descriptive linguistics (from phonetics and syntax to semantics and textology, with or

without etymology and dialectology — see Section 5.1 below), the study and codification of names (see Section 5.2 below), the study and codification of language(s) for specific purposes (LSP) or technical terminology (see Section 5.3 below), and various other aspects of applied linguistics (from language teaching to translation, with or without such fields as library science and semiotics — see Section 5.4 below), all of them supported by information technology (see Section 5.5 below). Again, for some of these, although perhaps not for all, it may make sense to distinguish international periodicals from those with a more regional, national or local orientation.

5.1 Language academies

One sub-genre of periodicals that deserves to be mentioned is the one specialising in the linguistic study of the world's languages, such as (in order of their numbers of native speakers) Mandarin Chinese, English, Spanish, Hindi, Arabic, Portuguese, Bengali, Russian, Japanese, German, and French — without forgetting 'smaller' languages such as Albanian, Basque, Cherokee, Hebrew or Nama. Many of these have academies or similar bodies devoted to their detailed linguistic description, which often involves corpus archives and historical, general and dialect dictionary projects. Some of these institutions publish journals and newsletters (such as *Trefwoord* of the Frisian Academy in the Netherlands and *Leksikografski Pregled*³ of the Language Institute of the Bulgarian Academy of Sciences), although it is worth pointing out that even for some of these bodies and publications, lexicographic issues are not often considered of more than secondary importance.

We will now widen our view to a type of periodical which concentrates on the description of particular languages, but only brief summaries of five such publications can be given: *American Speech*, *English Language and Linguistics*, *Language in India*, *Southern African Linguistics* and *Applied Language Studies*, and *Lexique*.

American Speech is one of the oldest authorities on dialectology, but as any website search would confirm, it does not neglect lexicographic codification as an important topic; indeed, one of the original goals of its sponsor, the American Dialect Society, was to publish a dictionary of American English. The journal carries the typical articles and reviews, but also has regular features (such as the over 60-year-old column 'Among the New Words') and special issues, such as the Winter 2002 issue on Frederick Cassidy (1907–2000). A series of supplements and monographs which accompanies the journal, *Publication(s) of the American Dialect Society*, has had several volumes dedicated to lexicographic issues and achievements, such as dialect dictionaries, usage labelling and Allen Walker Read (1906–2002).

English Language and Linguistics is linked to a number of departments at British universities, namely Manchester, Edinburgh and University College London. The English Department at UCL is the home of the Survey of English

Usage which hosted the Third Conference on the Linguistics of Contemporary English in July 2009 (together with colleagues from Queen Mary, University of London). Lexical studies form part of the editorial emphasis of ELL, such as the thematic issue No. 2 of Vol. 12 (2008) on intensifiers, guest-edited by Belén Mendez-Naya (Santiago de Compostela).

Language in India is a genuinely online journal attempting to build bridges between Indian scholars and linguists around the world. Its website and its pages are full of information on Indian languages such as Hindi, Punjabi, Gujarati, Kannada, Tamil, Telugu, Rajasthani, Urdu and Malayalam, including conferences, publications and research degrees, although occasionally the editors' rushing into e-text can detract from serious academic judgement and idiomatic fluency in the predominantly English discourse.

Southern African Linguistics and Applied Language Studies was established as a combination of the two journals of the Linguistic Society of Southern Africa and the Southern African Applied Linguistics Association with the purpose of reporting on research into all the languages of Southern Africa from both the theoretical-descriptive and the applied-practical points of view. It has a national editorial committee and an international advisory committee and publishes articles on a wide range of topics, research reports and book reviews, predominantly through the medium of English, occasionally on subjects of relevance to lexicography.

Lexique is the title of the Lille-based revue, which has appeared in 19 volumes since 1982, at least 6 of which have been dedicated to lexicological and lexicographic themes: Vol. 2 (1982) edited by Bernard Al and Jaap Spa on 'le dictionnaire', Vol. 4 (1986) edited by Claude Buridant on 'lexicographie au moyen âge', Vol. 9 (1990) edited by Michel Glatigny on 'les marques d'usage dans les dictionnaires (XVIIe–XVIIIe siècles)', Vol. 12–13 (1995) edited by Pierre Corbin and Jean-Pierre Guillermin on 'dictionnaires et littérature/littérature et dictionnaires', and Vol. 19 (2008) edited by Pierre Corbin and Nathalie Gasiglia on 'changer les dictionnaires?'.

5.2 Onomastics

Onomastics is concerned with the study of personal names and place names as well as designations for other objects, events and products, a fascinating but still rather underdeveloped discipline. *Onoma*, the journal of the International Council of Onomastic Sciences, was founded in 1950 and has recently been making an effort to catch up with its sequence of annual thematic volumes since No. 39 (2004), guest-edited by Botolv Helleland (Oslo) with 20 papers on the teaching of the subject in 12 different countries, in the United Nations and on the web, and No. 40 (2005), edited by Grant Smith (Eastern Washington) and Friedhelm Debus (Kiel), concentrating on literary onomastics. Other annual issues due to be published soon are Vol. 41 edited by Richard Coates (Bristol) on name theory, Vol. 42 edited by Staffan Nyström on urban topony-

my, and Vol. 43 edited by Antje Zilg et al. on commercial names; other topics to be treated include African onomastics, the history of onomastics, and names and minorities.

Of other onomastic organs on offer, I shall just mention five: *Names*, the journal of the American Name Society, *Nomina*, the journal of the Society for Names Studies in Britain and Ireland, the French *Nouvelle Revue d'Onomastique* and the *Rivista Italiana di Onomastica*. Some are supported by international editorial advisory boards; most present a combination of articles, reviews and notices, often with references to relevant conferences, dictionary projects and published glossaries.

5.3 Terminology

In this section, we turn our attention to an area that is often neglected in lexicography, viz. technical terminology, a field that hovers awkwardly between the study of specialised discourse ('LSP') and various efforts to standardise the lexis used in scientific communication ('terminography'). Although there are many national institutions and even international conferences dedicated to this kind of specialisation, there seems to be a regrettable lack of cooperation between them, and teaching and research are still rather underdeveloped.

Ibérica is the journal of AELFE, the European Association of Languages for Specific Purposes, which contains articles (e.g. on the vocabulary and discourse features of particular disciplines, notably in Spanish and English), research notes (e.g. on various text genres and their teaching and translation) and book reviews (e.g. of descriptions of and textbooks on LSP and occasionally of terminological dictionaries).

Terminology has the ambitious sub-title *International Journal of Theoretical and Applied Issues in Specialized Communication*. Its two editors, Kyo Kageura (Tokyo) and Marie-Claude L'Homme (Montréal), take its 15th anniversary as an occasion for describing its raison d'être in Vol. 14 No. 2 (2008) by highlighting three sets of tensions: terminology studies vs. subject-domain disciplines, terminological theory vs. applied issues, and local terminology vs. global terminology. The journal attempts to present various aspects of these not only in the form of conventional articles, reports and reviews, but occasionally by means of 'special issues', such as the five devoted to term extraction in Japanese, recent trends in computational terminology, application-driven terminology engineering, pattern-based approaches to semantic relation extraction, and processing of terms in specialised dictionaries and terminological databases.

There are several other national and international journals, of which I will mention only two more. The *Cahiers du Rifal* (RIFAL being a research network linking the Organisation Internationale de la Francophonie and the Communauté Française de Belgique) has appeared in approximately annual volumes since 1989, initially under the title *Terminologie Nouvelle*, with the aim of promoting interdisciplinary connections between terminology studies and transla-

tion, e.g. in Vol. 26 (2007), which has the theme 'Terminology, culture and society' and draws on papers from two conferences at Dakar (Bangladesh) and Gatineau (Canada).

Fachsprache, the *International Journal of Specialized Communication*, based at an Austrian publisher, has been quite influential in presenting and linking some of the recurring topics and recent trends in LSP, celebrating its 30th anniversary with a new editorial team and an international advisory board, adding Spanish to its main languages English, French and German, and developing a newsletter as part of its website.

5.4 Other aspects of applied linguistics

The distinction made in earlier sections between international and national periodicals seems less relevant here in a territory where interdisciplinary contact and collaboration are difficult enough as an objective in such fields as applied linguistics, language teaching, translation, semiotics, library science and genealogy; and it is only possible for me to give a brief impression of them.

Applied Linguistics (AL) is a truly global periodical, supported by the International Association for Applied Linguistics (famous for the powerful impact of its 15 triennial congresses since 1966), the American Association for Applied Linguistics and the British Association for Applied Linguistics. Among the applications of linguistic knowledge to the solution of practical problems mentioned on the AL website are language learning, teaching and testing, translation, bilingual education, discourse analysis and stylistics, language planning and lexicography. Periodicals for applied linguistics, particularly those specialising in language learning and teaching (either of a mother tongue/first language, LT1, or a foreign/second language, LT2) are too numerous to mention, especially those which concentrate on English, so AL will have to be their only representative here.

Babel is the journal of the International Federation of Translators, with an active tradition since 1955, when it was founded by Pierre-François Caillé, with financial support from UNESCO. The current editor is René Haeseryn, who has the backing of both an editorial board and an advisory committee. There are 4 issues per year, with papers in French, English and a number of other languages. One representative issue, Vol. 52 No. 1 (2006), contains 4 papers on different aspects of translation, an update on 'The Life of FIT' (and some of its publications), a report on a UNESCO series on literary translation, and bibliographical and lexicographic references (including two reviews).

Founded by Alexander Lane and led by a series of editors, including the translator and lexicographer Günther Haensch (and other academics at Augsburg and Leipzig), the journal *Lebende Sprachen* has been run by the dictionary publisher Langenscheidt since 1958 and recently been taken over by Niemeyer/De Gruyter, with the support of the German Federation of Translators. It often carries innovative articles on all aspects of literary and technical trans-

lating and interpreting, (bilingual and multilingual) glossaries of specialised vocabulary, bibliographies and reviews of relevant publications and reports on relevant events.

Founded in 1969 as the organ of the International Association for Semiotic Studies, the journal *Semiotica* reports on research and reviews the literature, occasionally with thematic issues, on the role of letters, sounds, pictures and other signs and symbols in the graphic, phonic, visual and acoustic media and their effectiveness in specific communicative contexts and genres. The interpretation and codification of vocabulary and terminology come up for treatment now and again, e.g. in the fields of medicine, law and computing in issue No. 172 (2008).

Having met the journal *Reference Reviews* at the start of Section 5, we can acknowledge other periodicals connected with library science and book publishing. *Linguistics and Language Behavior Abstracts*, already mentioned in Section 3 above, has done much for over 40 years to catalogue the literature contained in linguistic periodicals. *The Indexer*, published in print as well as in electronic format by the (British) Society of Indexers on behalf of indexing societies around the world, including those in Australia/New Zealand, Canada, China, Germany, the Netherlands and the U.S., covers a wide range of issues from different conventions across languages, subjects and ages to an ever widening range of appropriate tools. *The Genealogists' Magazine* addresses a range of topics in its most recent issues, from the tracing of ancestors by various means and through specific archives, records and registers to the role of emigration and immigration and the ancestry of famous people, plus obituaries and library news.

5.5 Information technology and corpus linguistics

As early as 1988, Laurence Urdang (1927–2008) was the pioneering guest-editor of a thematic issue of *Lexicographica* on what by now has become a fairly commonplace feature, the use of computers in dictionary-making. Other aspects of information technology which are occasionally relevant to lexicography include database development, archive and corpus design, artificial intelligence, natural language processing, and machine translation.

Computational Linguistics, published by the Association for Computational Linguistics well-known for its international conferences, has been addressing some of these topics since 1974. In the four issues of 2008, they include statistical parsing, semantic role labelling, machine learning and the construction of corpora. On the latter, the *International Journal of Corpus Linguistics* has been particularly relevant in two recent thematic issues, Vol. 11 No. 3 (2006) focusing on lexical cohesion, guest-edited by John Flowerdew and Michaela Mahlberg, and Vol. 12 No. 2 (2007) on the work of John Sinclair, guest-edited by Rosamund Moon.

More journals are, of course, available for English and other languages

and countries, but it is impossible to cover every aspect of this ever more dominating field. Meanwhile, it is easy to observe a trend for more computerisation, which not only helps all aspects of lexicography, but affects the way more and more periodical publications are developing electronic versions (documented, among others, by *IngentaComment*, a service provided by the British-American company Publishing Technology).

6. Newsletters

Related to journals are newsletters, typically issued by lexicography associations, dictionary publishers or dictionary projects, such as the *EURALEX Newsletter* (which originally appeared as a paperback *Bulletin* after the LEXeter'83 conference and was then published in *IJL*, first as a *Bulletin* every other issue and since 1990 as a regular 4-issue feature). The *DSNA Newsletter* has been published by the Dictionary Society of North America since 1977 to report on biennial meetings, other conferences, important projects and outstanding lexicographers.

There is a tendency that bulletin-like publications can gradually grow into journals, as has happened to the KDN (of Kernerman Dictionaries Ltd.) and the BDC (of Barnhart/LexikHouse). At the same time, websites have taken over from printed newsletters for the purpose of publicising events, annual reports and lists of people and references, as has happened in the case of many national and international associations (e.g. the AFRILEX website offers cross-references to its nine newsletters published since 1995) as well as publishing companies (e.g. Oxford University Press's *OED News* since 1995 have become part of the regular updates on the online pages of the *Oxford English Dictionary*).

7. Conclusion

I have tried to single out some of the representative types of journals, magazines and newsletters which have a bearing on lexicographic specialisation(s), both in printed and in online format (as in *Onomasiology Online* and *Language in India*), although it is of course impossible to cover all relevant institutions, countries, and languages. As we have seen, the topics covered in these periodicals range widely, from practice to theory, and from one genre of reference work to another (cf. Hartmann 2007); in their entirety they make a significant contribution to defining and supporting lexicography.

There are inherent limitations in the editing and publication of all periodicals, owing to academic-personal, commercial-public and other problems, some of which lead to difficulties such as reduced sales, irregular delivery, changes of title or topics treated. Examples of these are the closing down of *Lexicology*, *the International Journal on the Structure of Vocabulary* after only three years, the slow start of a number of journals (such as *Ibérica* between 1969 and 1999),

changes in the sequence of volumes or annual numbers of issues published (e.g. *Semiotica* is one of those journals which numbers issues cumulatively and volumes by year rather than annual number), or even a complete change of editors/sponsors/publishers (as in the *Cahiers de lexicologie* after 40 years or in *Fachsprache* after 30 years).

A number of desiderata seem to be in order, then, to improve the periodical scene further. One of these is international standardisation of publishing practices (such as the now almost globally uniform ISSN, cited for all periodicals listed in the Appendix). A second desideratum is better bibliographic treatment (e.g. by abstracting and indexing — we noted that abstracting journals like LLBA do not list all important titles from the lexicographic field, and several of the world's indexing systems like *Dialnet*, *OCLC*, *Thomson-Reuters* or *Ingenta* do not include the full range of lexicography periodicals either). A third desideratum is wider academic recognition by universally respected bodies (such as the European Science Foundation which includes 19 of the 40 journals discussed above and lists them in its *European Reference Index for the Humanities* classified into one of three grades, from A to C). A fourth desideratum is better awareness and provision of more critical reviews of periodicals (such as the one of *Lexique* by Dupont and Bonnet 1999). A fifth desideratum is more accurate, up-to-date and reliable websites for the institutions and publishers responsible for journals.

It is impossible to cover every single periodical in a survey of this kind. Not all linguistic, regional and (inter)disciplinary serials could be listed — some may even have escaped my searches altogether, in which case I apologise to those concerned. However, I hope that my attempt at sharing some of my experiences as a compiler, researcher, teacher and author has at least given readers an impression of the wide range of periodical publications that are out there, produced for us all to benefit from.⁴ Keep in touch and stay connected!

Notes

1. I have benefited from many lists (e.g. by ESF and OCLC, see below) as well as from the websites of publishers and sponsors of periodicals, and I have had some help from over 90% of all the publishing companies and journal editors involved. (A few whose details I could not verify had to be omitted.) I am particularly grateful to a number of friends and personal contacts for sending me relevant information (e.g. Kaoru Akasu, Gilles-Maurice de Schryver, Young-kuk Jeong, Robert Lew, Rosamund Moon, Jean Pruvost, Adam Smith, Kathrin Steyer and Herbert Welker).
2. This is one of the journals I had to omit as I could not verify the details by reference to its website or its editor/publisher.
3. This is another periodical whose publication details were difficult to verify, so I have omitted it from the list in the Appendix.
4. If you would like to make a contribution to my project of an *International Directory of Lexicography Institutions* (IDLI), of which the subject of Periodicals forms one of 12 sections, you can contact me by email: r.r.k.hartmann@exeter.ac.uk

References

- Bibliographie Linguistique/Linguistic Bibliography* ed. by the Permanent International Committee of Linguists. Vol. 1 (1949) to Vol. 49 (1999) published by Spectrum at Utrecht and Antwerpen; from Vol. 50 (2000) published online, eds. S. Tol and H. Olbertz. Leyden: Brill.
- Boccuzzi, C. et al.** 2007. *Bibliographie thématique et chronologique de métalexicographie 1950–2006*. Bibliotheca della Ricerca. Bibliographica 10. Fasano: Schena Editore.
- Dupont, N. and V. Bonnet.** 1999. La Revue *Lexique*. *International Journal of Lexicography* 12(2): 99-106.
- Hartmann, R.R.K.** 1964–1970. Zeitschriftenschau [I to XV]. *Lebende Sprachen* (München) Vol. 9 (1964) to 15 (1970).
- Hartmann, R.R.K.** 2007. Pure or Hybrid? The Development of Mixed Dictionary Genres. *Facta Universitatis, Series Linguistics and Literature* (University of Niš) 3.2: 193-208.
- Hartmann, R.R.K.** 2009. Review of Phraseology. An Interdisciplinary Perspective ed. by Sylviane Granger and Fanny Meunier. *International Journal of Lexicography* 22(1): 87-89.
- Kaszubski, P. (Ed.)**. 2006. Assessing the Potential of Corpora. *Poznań Studies in Contemporary Linguistics* 41: 3-274.
- McArthur, T.** 1998. What then is Reference Science? McArthur, T. *Living Words. Language, Lexicography, and the Knowledge Revolution*: 215-222. Exeter: University of Exeter Press.
- Tarp, S.** 2007. Lexicography in the Information Age. *Lexikos* 17: 170-179.
- Welker, H.A. (Ed.)**. 2007. O uso de dicionários. *Horizontes de Lingüística Aplicada* (Brasília) 6(2): 5-294.
- Wiegand, H.E. et al.** 1985–2001. Continuous Bibliography. *Lexicographica*, from Vol. 1 (1985) to 17 (2001).
- Wiegand, H.E. et al.** 2004–2009. Zeitschriftenschau. *Zeitschrift für germanistische Linguistik*, from Vol. 31 (2003–04) to 36 (2008–09).

Appendix: List of cited periodicals

Title of serial (Sponsoring body)	Editor(s) ISSN [E for online]	Publisher	Number of volumes (and issues p.a.) published
<i>American Speech</i> (AS). <i>A Quarterly of Linguistic Usage</i> (American Dialect Society)	Michael Adams 0003-1283 E: 1527-2133	Duke University Press, Durham NC, USA	83 (x 4 issues p.a.) since 1925; available online
<i>Applied Linguistics</i> (AL) (Association Internationale de Linguistique Appliquée, American Association for Applied Linguistics, British Association for Applied Linguistics)	Jane Zuengler, Guy Cook 0142-6001 E: 1477-450X	Oxford University Press, Oxford, GB	29 (x 4 issues p.a.) since 1980; available online
<i>Babel. Revue Internationale de la Traduction / International Journal of Translation</i> (Fédération Internationale des Traducteurs)	René Haeseryn 0521-9744 E: 1569-9668	J. Benjamins Publishing Co., Amsterdam, NL	54 (x 4 issues p.a.) since 1955; available online
<i>(The) Barnhart Dictionary Companion</i> (BDC). <i>A Quarterly of New Words</i>	David K. Barnhart 0736-1122	Lexik House, Hyde Park NY, USA	13 (x 4 issues p.a.) since 1982; online planned
<i>Cahiers de lexicologie. Revue internationale de lexicologie et lexicographie</i> (L.D.I. @ Univ. Paris 13 and Univ. Cergy Pontoise)	Gaston Gross, Gérard Petit 0007-9871	Éditions Garnier, Paris, FR	48 (x 2 issues p.a.) since 1957; online planned
<i>(Les) Cahiers du Rifal</i> (Réseau International Francophone d'Aménagement Linguistique)	Martine Garsou 1015-5716	RIFAL c/o TERMISTI, Institut Supérieur de Traducteurs et Interprètes, Bruxelles, BE	27 (x 1 issue p.a.) since 1989
<i>Cishu Yanjiu / Lexicographical Studies</i> (Zhongguo Cishu Xuehui / Lexicographical Society of China)	CHAO Feng, XU Xiamin 1000-6125	Cishu Chubanshe / Lexicographical Publishing House, Shanghai, CN	30 (x 6 issues p.a.) since 1979
<i>Computational Linguistics</i> (CL) (Association for Computational Linguistics)	Robert Dale 0891-2017 E: 1530-9312	MIT Press, Cambridge MA, USA	34 (x 4 issues p.a.) since 1974; available online
<i>Dictionaries. Journal of the Dictionary Society of North America</i>	William Frawley 0197-6745	DSNA c/o Lisa Berglund, Department of English, Buffalo State College, Buffalo NY, USA	29 (x 1 issue p.a.) since 1979
<i>DSNA Newsletter</i>	Katherine Isaacs no ISSN	DSNA → <i>Dictionaries</i>	32 (x 2 issues p.a.) since 1977; available online
<i>English Language and Linguistics</i> (ELL)	Bas Aarts, David Denison, April McMahon 1360-6743 E: 1469-4379	Cambridge University Press, Cambridge, GB	12 (x 3 issues p.a.) since 1997 available online

EURALEX Newsletter	Paul Bogaards	→ <i>Intern. Journal of Lexicography</i>	25 (x 4 issues p.a.) since 1984
Fachsprache/International Journal of Specialized Communication	Jan Engberg, Susanne Göpferich, Nina Janich 1017-3285	W. Braumüller, Vienna, AT	30 (x 4 issues p.a.) since 1979; online planned
(The) Genealogists' Magazine. Journal of the Society of Genealogists	Michael Gandy 0016-6391	SoG, London, GB	30 (x 4 issues p.a.) since 1925
Hanguk-sajeonhak/Journal of the Korean Association for Lexicography	Sang-kyu Seo 1598-8694	KOREALEX, Seoul, KR	6 (x 2 issues p.a.) since 2003
Ibérica. Journal of the European Association of Languages for Specific Purposes	Ana Bocanegra-Valle 0153-0364	AELFE c/o Departament d'Estudis Anglesos, Universitat Jaume I de Castelló, Castelló de la Plana, ES	10 (x 2 issues p.a.) since 1999
(The) Indexer. The International Journal of Indexing (Society of Indexers)	Maureen MacGlashan 0019-4131 E: 1756-0632	SI, Sheffield, GB	26 (x 4 issues p.a.) since 1958; available online
International Journal of Corpus Linguistics (IJCL)	Michaela Mahlberg 1384-6655 E: 1569-9811	J. Benjamins Publishing Co., Amsterdam, NL	13 (x 4 issues p.a.) since 1996; available online
International Journal of Lexicography (IJL) (European Association for Lexicography)	Paul Bogaards 0950-3846 E: 1477-4577	Oxford University Press, Oxford, GB	21 (x 4 issues p.a.) since 1988; available online
Kernerman Dictionary News (KDN)	Ilan Kernerman 1565-4745	K. Dictionaries, Tel Aviv, IL	16 (x 1 issue p.a.) since 1994
Language in India	Madasamy Thirumalai et al. 1930-2940	M.S. Thirumalai, Bloomington MN, USA	9 (x 12 issues p.a.) since 2001; online only
Lebende Sprachen (LES). Zeitschrift für fremde Sprachen in Wissenschaft und Praxis (Bundesverband der Dolmetscher und Übersetzer)	Peter A. Schmitt, Reinhold Werner 0023-9909 E: 1868-0267	Max Niemeyer Verlag, Tübingen and W. de Gruyter, Berlin, DE	53 (x 4 issues p.a.) since 1956; available online
Lexicographica. International Annual for Lexicography/Revue Internationale de Lexicographie/Internationales Jahrbuch für Lexikographie (Dictionary Society of North America, European Association for Lexicography)	Ulrich Heid, Stefan Schierholz, Wolfgang Schweickard, Herbert Ernst Wiegand, Werner Wolski 0175-6206	Max Niemeyer Verlag, Tübingen and W. de Gruyter, Berlin, DE	24 (x 1 issue p.a.) since 1985; available online
Lexicon (Iwasaki Linguistic Circle)	Hideo Masuda 0385-566X	ILC c/o Kenkyusha Co. Ltd., Tokyo, JP	38 (x 1 issue p.a.) since 1972
LexicoNordica (Nordisk Forening for Leksikografi)	Henning Bergenholtz, Sven-Göran Malmgren 0805-2735	NFL c/o Språkrådet, Oslo, NO	15 (x 1 issue p.a.) since 1994

Lexikos (African Association for Lexicography)	J.C.M.D. du Plessis 1684-4904	Buro van die WAT, Stel- lenbosch, ZA	18 (x 1 issue p.a.) since 1991
Lexique (Université Lille 3 Labora- toire Savoirs Textes Lan- gages UMR 8163)	Georgette Dal (et al.) 0756-7138	Presses Universitaires du Septentrion, Lille, FR	19 (x 1 issue p.a.) since 1982
Linguistics and Language Behavior Abstracts (LLBA)	Lynette Hunter 0888-8027	Cambridge Scientific Abstracts, Bethesda MD, USA	42 (x 5 issues p.a.) since 1967; available online
Names. A Journal of Onomas- tics (American Name Society)	Frank Nuessel 0027-7738 E: 1756-2279	Maney Publishing, Cam- bridge MA, USA	56 (x 4 issues p.a.) since 1953; available online
Nomina. Journal of the Society for Name Studies in Britain and Ireland	Maggie Scott 0141-6340	SNSBI c/o University of Salford, Salford, GB	31 (x 1 issue p.a.) since 1978
(La) Nouvelle Revue d'Ono- mastique (NRO) (Société Française d'Ono- mastique)	Martina Pitz 0755-7752	SFO c/o Université Jean Moulin Lyon 3, FR	51 (x 1 issue p.a.) since 1949; online planned
Onoma. Journal of the Inter- national Council of Onomastic Sciences	Doreen Gerritzen 0078-463X E: 1783-1644	Editions Peeters, Leuven, BE	40 (x 1 issue p.a.) since 1950; available online
Onomasiology Online (OnOn)	Alfred Bammes- berger, Joachim Grzega, Marion Schöner 1616-9692	Katholische Universität Eichstätt, Eichstätt-Ingol- stadt, DE	8 (x 1 issue p.a.) since 2000; online only
Reference Reviews (RefRev)	Tony Chalcraft 0950-4125	Emerald Group, Bingley, GB	22 (x 8 issues p.a.) since 1987; available online
Revista de Lexicografía (with thematic <i>Anexos</i> since 2006)	José Ignacio Pérez Pascual, Mar Campos Souto 1134-4539	Departamento de Filología Española e Latina, Universidade da Coruña, A Coruña, ES	13 (x 1 issue p.a.) since 1995; online planned
Rivista Italiana di Onomas- tica (RION) (Laboratorio Internazionale di Onomastica, Roma)	Enzo Caffarelli 1124-8890	Società Editrice Romana, Roma, IT	13 (x 2 issues p.a.) since 1995
Semiotica. Journal of the International Association for Semiotic Studies	Marcel Danesi 0037-1998 E: 1613-3692	W. de Gruyter, Berlin, DE	39 (x 5 issues p.a.) since 1969; available online
Southern African Linguistics and Applied Language Studies (SALALS) (Linguistic Society of South- ern Africa and Southern African Applied Linguistics Association)	Jacobus A. Naudé 1607-3614 E: 1727-9461	National Inquiry Services Centre, Grahamstown, ZA	27 (x 4 issues p.a.) since 1983; available online

<i>Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication</i>	Kyo Kageura, Marie-Claude L'Homme 0929-9971 E: 1569-9994	J. Benjamins Publishing Co., Amsterdam, NL	14 (x 2 issues p.a.) since 1994; available online
<i>Trefwoord, elektronisch lexicografisch tijdschrift (Fryske Akademy)</i>	Anne Dykstra et al. no ISSN	Fryske Akademy, Leeu- waarden, NL	9 (x 1 issue p.a.) since 1999; online only

South Africa's New African Language Dictionaries and their Use for the African Speech Communities*

Juliane Klein, *University of Leipzig, Leipzig, Germany*
(*julianeklein.trier@googlemail.com*)

Abstract: During the last 15 years, the lexicographic scene in South Africa changed drastically as many new dictionaries for the African languages were compiled. The different dictionary types and publication modes discussed in this article are: general dictionaries, restricted dictionaries, printed dictionaries, electronic dictionaries, online and cell phone dictionaries. Although there are different dictionary types, they all have three major uses for the speech communities. Dictionaries are useful tools for language documentation and standardization, as they try to cover and document the general vocabulary (general dictionaries) or the specialized vocabulary (technical dictionaries). They empower the language users because they help to improve communication by providing users with the necessary vocabulary they need. In addition, dictionaries have a high symbolic value for a language. Having dictionaries, and especially technical, online or cell phone dictionaries, is the visible proof that a language is standardized and modern, and can be used in all domains of life.

Keywords: LEXICOGRAPHY, GENERAL DICTIONARIES, RESTRICTED DICTIONARIES, ONLINE DICTIONARIES, CELL PHONE DICTIONARIES, LANGUAGE DOCUMENTATION, LANGUAGE STANDARDIZATION, EMPOWERMENT, COMMUNICATION, PSYCHOLOGICAL FACTOR, SOCIOLINGUISTICS, LANGUAGE PLANNING

Opsomming: Suid-Afrika se nuwe Afrikataalwoordeboeke en hul gebruik vir die Afrikataalgemeenskappe. Gedurende die afgelope 15 jaar het die leksikografiese toneel in Suid-Afrika ingrypend verander deurdat baie nuwe woordeboeke vir die Afrikatale saamgestel is. Die verskillende woordeboeksoorte en publikasievorme wat in hierdie artikel bespreek word, is: algemene woordeboeke, beperkte woordeboeke, gedrukte woordeboeke, aanlyn- en selfoonwoordeboeke. Alhoewel daar verskillende woordeboeksoorte is, het hulle almal drie hoofgebruike vir die taalgemeenskappe. Woordeboeke is nuttige werktuie vir taaldokumentasie en -standaardisering deurdat hulle probeer om die algemene woordeskat (algemene woordeboeke) en die gespesialiseerde woordeskat (tegniese woordeboeke) te dek en te dokumenteer. Hulle bemagtig die taalgebruikers omdat hulle kommunikasie verbeter deur die gebruikers van die nodige woor-

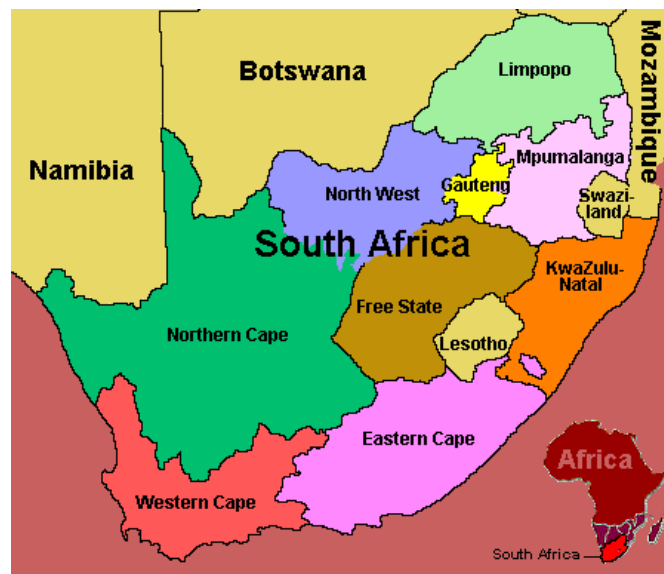
* This article is a revised version of a paper presented at the Thirteenth International Conference of the African Association for Lexicography, organized by the Bureau of the *Woordeboek van die Afrikaanse Taal*, Stellenbosch, Republic of South Africa, 30 June–3 July 2008.

deskat wat hulle benodig, te voorsien. Daarbenewens het woordeboeke 'n hoë simboliese waarde vir 'n taal. Om woordeboeke te besit, en veral tegniese, aanlyn- en selfoonwoordeboeke, is 'n sigbare bewys dat 'n taal gestandaardiseer en modern is, en op alle terreine van die lewe gebruik word.

Sleutelwoorde: LEKSIKOGRAFIE, ALGEMENE WOORDEBOEKE, BEPERKTE WOORDEBOEKE, AANLYNWOORDEBOEKE, SELFOONWOORDEBOEKE, TAALDOKUMENTASIE, TAALSTANDAARDISERING, BEMAGTIGING, KOMMUNIKASIE, SIELKUNDIGE FAKTOR, SOSIOLINGUISTIEK, TAALBEPLANNING

1. Introduction and sociolinguistic background

The lexicographic scene in South Africa changed markedly after the 1996 Constitution and the establishment of the eleven National Lexicographic Units. The last decade saw the compilation of many African language dictionaries. This article discusses these new dictionaries and their use for the African language speech communities. In addition, the need to establish a dictionary culture will be discussed, as a functional dictionary culture can help to empower people.



This map shows South Africa with its nine provinces (www.afrilux.co.za/quickness/South_Africa.htm).

South Africa is a sociolinguistically complex country, which is "characterized by: a multiplicity of languages and cultures; an overlapping demographical and geographical distribution of languages; the politicisation of its languages and cultures due to the colonial past and the policy of apartheid" (Webb 1996:

143). It is not possible to give the exact number of languages spoken in South Africa. Webb and Kembo-Sure (2000: 50) indicate that there are between 25 and 80 languages. In contrast to this high number, *Ethnologue* lists 31 languages of which "25 are living languages, 3 are second languages without mother-tongue speakers, and 3 are extinct". It has eleven official languages, of which nine are African languages, i.e. Ndebele, Sesotho, Sesotho sa Leboa, Setswana, SiSwati, Tsonga, Venda, Zulu and Xhosa, and two are Germanic languages, i.e. Afrikaans and English.

Table 1: Home languages in the nine provinces (Census 2001)

Home language	Eastern Cape	Free State	Gauteng	KwaZulu Natal	Limpopo	Mpumalanga	Northern Cape	North West	Western Cape	South Africa
Afrikaans	9.3	11.9	14.4	1.5	2.3	6.2	68.0	7.5	55.3	13.3
English	3.6	1.2	12.5	13.6	0.5	1.7	2.5	1.2	19.3	8.2
IsiNdebele	0.1	0.4	1.9	0.2	1.5	12.1	0.1	1.3	0.0	1.6
IsiXhosa	83.4	9.1	7.6	2.3	0.3	1.5	6.2	5.8	23.7	17.6
IsiZulu	0.8	5.1	21.5	80.9	0.7	26.4	0.3	2.5	0.2	23.8
Sepedi	0.0	0.3	10.7	0.1	52.1	10.8	0.1	4.2	0.0	9.4
Sesotho	2.4	64.4	13.1	0.7	1.3	3.7	1.1	5.7	0.7	7.9
Setswana	0.0	6.8	8.4	0.1	1.6	2.7	20.8	65.4	0.1	8.2
SiSwati	0.1	0.3	1.4	0.1	1.1	30.8	0.1	0.6	0.0	2.7
Tshivenda	0.0	0.1	1.7	0.0	15.9	0.2	0.0	0.5	0.0	2.3
Xitsonga	0.0	0.3	5.7	0.0	22.4	3.8	0.0	4.7	0.0	4.4
Other	0.2	0.2	1.0	0.4	0.3	0.3	0.7	0.6	0.4	0.5
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 1 shows that most people speak an African language as their home language. Although most provinces have at least one dominant home language, the language distribution does not follow provincial borders. This heterogeneous occurrence explains the distribution of the official provincial languages and is a reason why all provinces have at least three official languages and have chosen at least one African language.

Table 2: Provincial languages (Geyser online resource)

Province	Provincial languages
Eastern Cape	Xhosa, Afrikaans, English, Sign language
Free State	English, Afrikaans, Sesotho, Xhosa
Gauteng	Afrikaans, English, Sepedi, Zulu
KwaZulu-Natal	Zulu, Afrikaans, English
Limpopo	Sepedi, Afrikaans, Venda, Xitsonga, Ndebele, English
Mpumalanga	Ndebele, SiSwati, Afrikaans, English
Northern Cape	Setswana, Afrikaans, Xhosa, English
North West province	Setswana, Afrikaans, English, Sign language
Western Cape	Afrikaans, English, Xhosa

South Africa does not only have eleven official languages, but it also has multilingualism and the development of the African languages enshrined in the constitution. This is an important aspect, as most multilingual African countries chose only the former colonial language as official language. In most countries,

the African languages are not recognized as official languages; they either have no status or are called "national languages" without having the status and legal backing official languages have. In contrast to this, South Africa is one of the few countries providing a legal basis for the development of the indigenous languages spoken by the majority of people in South Africa.

2. Lexicographic scene in South Africa

The National Lexicographic Units Bill from 1996 paved the way for most of the new dictionaries in South Africa. According to the Bill, Lexicographic Units for each official language had to be established. The existing Units were not dissolved but reorganized so that they met the requirements stated in the Bill. Having such a legal backing was probably the main reason why the lexicographic scene in South Africa developed and expanded quite fast during the last decade. At present, general dictionaries are available for all official languages. These include printed dictionaries, online and cell phone dictionaries.

Because of the need felt for them, technical dictionaries and word lists are being compiled by many individual researchers and by the National Language Service. Although some technical dictionaries already exist, they are not yet widely distributed among the speech communities.

3. Interdependent dictionary types and publication modes and their use for the people

Although different dictionary types and publication modes can be distinguished, the boundaries between these groups are blurred as most dictionaries belong to more than one group.

3.1 General dictionaries

General dictionaries aim to cover the general vocabulary of a language. They can be either monolingual, bilingual or, in the case of South Africa, multilingual. The major difference between monolingual and bilingual or multilingual dictionaries is that in monolingual dictionaries, the headwords are defined in the same language as the dictionary entries, whereas bi- or multilingual dictionaries provide translations of terms or define terms in a target language if a direct translation is not possible. Bi- or multilingual dictionaries can be monodirectional, i.e. from source into target language only, or bidirectional, i.e. from source language into target language as well as from target language into source language.

General dictionaries cover the biggest user group, as they are meant for mother-tongue speakers and language learners as well. This implies that a single dictionary is used by a broad range of users, e.g. of different age groups and for a variety of tasks. An important sub-group of general dictionaries are

school and learner's dictionaries, as they are often the first dictionaries students come in contact with. A dictionary from this category is frequently seen as *the* dictionary by its users. In countries where people live in a pre-dictionary culture, general dictionaries are often the only dictionaries used because they are often the only dictionaries known. In addition, specialized dictionaries mostly do not exist in a pre-dictionary culture or are not known to the speech communities. In South Africa, the general dictionary is the best-covered dictionary type, as all official languages have at least one general dictionary. Although all of these dictionaries fall into the same category, they are, depending on the compiler(s), quite different with regard to their coverage of the language. Some are very small and basic with only a few thousand entries, often just providing the lemmata and their translations, sometimes even without any part of speech information. In contrast to such word lists, other dictionaries cover more of the general vocabulary and provide much more information, for example, parts of speech, pronunciation, frequency, and usage examples.

General dictionaries are often made available as paperback or hardback editions. Other publishing modes are online dictionaries, electronic dictionaries or cell phone dictionaries.

3.2 Restricted dictionaries

Gouws and Prinsloo (2005: 47) define restricted dictionaries as dictionaries that "focus on one field, e.g. a specific semantic field or a specific subject domain ... or one type of lexical item ... or they limit their treatment to one data category". Some dictionaries belonging to this category are: technical dictionaries, idiom dictionaries, dictionaries of abbreviations, pronunciation dictionaries, etymological dictionaries or dictionaries of synonyms. The most important type in the South African context is technical dictionaries. English and Afrikaans are well covered, having a broad range of restricted dictionaries. The African languages, needing many more restricted and especially technical dictionaries, are still at the beginning. Nevertheless, some small online technical dictionaries exist and can be consulted by everyone who has access to the internet. Examples include a terminology list of linguistic terms in Sesotho sa Leboa and English (<http://africanlanguages.com/sdp/linguistics/>), and a birds and trees list in English and Sesotho sa Leboa (<http://africanlanguages.com/sdp/ff/index.php?l=en>). There is also a webpage containing some basic medical terms and phrases in English and Zulu (<http://www.wolfescape.com/WebPages/ZuluDict.htm>). In addition, the National Language Service has published some technical dictionaries, for example: the *Multilingual Natural Sciences & Technology Dictionary for Grade 4 to 6* (English–Afrikaans–Zulu–Xhosa–SiSwati–Ndebele; English–Afrikaans–Setswana–Sepedi–Sesotho; English–Afrikaans–Tshivenda–Xitsonga) and the *Multilingual Mathematics Dictionary Grade R–6* (English into all official languages), both available in printed form or accessible online (http://www.dac.gov.za/chief_directorates/NLS/list.htm).

3.3 Different publication modes

3.3.1 Traditional paper dictionaries

The most commonly known publication mode for dictionaries is still paper, as borne out by the fact that a dictionary is often defined as a book containing words and their translations into another language or as a book explaining the meaning of words.

An advantage of this traditional publishing mode is that the user just needs the book without requiring additional technologies, such as electricity and a computer, to be able to use it. Another advantage is that the user, by seeing the whole book, gets an idea of the size of the stock of words of the language. Nevertheless, this can be a disadvantage as well, if the dictionary contains only the basic vocabulary of a language, as this could imply that the language lacks enough words to be a "good" language. Another disadvantage is that the space in every printed dictionary is restricted so that not all words of a language can be included, and the number of examples has to be limited. Large, multivolume dictionaries which are usually expensive can also be quite intimidating for an inexperienced user because they contain too many words, making it difficult for the user to find the correct entry, especially if there is more than one volume.

3.3.2 Electronic dictionaries

Usually electronic dictionaries are the electronic versions of some already existing printed dictionary. They are static versions which can either be downloaded to a computer or come on a CD or as pocket organizers, palm pilots and personal digital assistants. Electronic dictionaries are as restricted as their paper versions, because they are static, needing an electronic medium to be used. Nevertheless, they are useful for users of computers for, as the dictionary is available on the computer, no other medium (book) is needed. In addition, they can be easier to use for inexperienced users who only have to type in the search word without having to browse through a book to find a word.

3.3.3 Online dictionaries

Online dictionaries, which can be general or restricted dictionaries, are usually dynamic dictionaries available on the internet and accessed online instead of being downloaded to the user's computer. One of the greatest advantages of online dictionaries is that they virtually have no space restrictions, so that they can contain more entries and more examples than a printed dictionary. Sound files for each entry can be added, which means that the user can listen to the correct pronunciation of a word, a very useful feature for language learners. Having no space restrictions entails that they can be compiled in a more cost-

effective way, making the final product cheaper for the user, an important factor in economically disadvantaged societies. Another advantage is that they are dynamic, which makes it possible to easily add changes in the language or additional words and examples as well as user-feedback. Nevertheless, the great disadvantage of an online dictionary is that users either need a computer with an internet connection or a cell phone with an internet connection.

3.3.4 Cell phone dictionaries

Cell phone dictionaries, which is one of the newest publication modes used in South Africa, are a subcategory of electronic dictionaries. The two most common types are either downloadable dictionaries which stay on the user's cell phone or text-message based dictionaries where the user texts a query and receives a text message back containing the translation. This means that the first type is a static electronic dictionary that is downloaded to the cell phone (instead of the computer) and that can be used as long as the cell phone is operative. Pharos is compiling a set of four cell phone dictionaries, developed from *Pharos's Afrikaans-English-Xhosa-Zulu Dictionary*. This publisher also hopes to launch the cell phone version of *Pharos's Afrikaans-English-Northern Sotho-Tswana Dictionary* soon (W. Smith-Muller 2009, personal communication). The advantage here is that, because the dictionary stays on the cell phone, it is always available. A disadvantage is that such dictionaries are even more restricted in space than printed dictionaries because of the small cell phone memory card. Nevertheless, such a publication mode can be used by language planners to distribute dictionaries easily, as they could, like ring tones or games that come with the cell phone, be downloaded before the cell phones are sold. The second type is a translation service rather than a dictionary, as the user cannot access the whole dictionary but only single words.

4. Different dictionaries and their use for the speech communities

There are three main possibilities for the speech community: (a) Dictionaries as tools for language documentation and standardization; (b) Empowerment through communication; and (c) Psychological value of the dictionary as a status symbol for the language and its users.

4.1.1 Documentation and standardization: general dictionaries

A standard language is defined by Haugen (1966) as a language having minimal variation and stability in form and maximal variation in function. This means that a key feature of standardization is the creation of a uniform written variety that replaces a variety of forms with one invariable form (Milroy 1999). Such a variety can be achieved through the process of codification, i.e. through "giving explicit definition to the norm, principally through the production of

authoritative grammars, dictionaries, spellers and the like" (Ferguson 2006: 21). Although dictionary compilers often claim that their dictionaries are descriptive and not prescriptive, dictionaries are mainly perceived as prescriptive by their users. They usually see the dictionary as an authority when it comes to spelling, i.e. the form given in the dictionary is considered to be the correct form. This is one reason why dictionaries document the standard form of a language. South African examples of dictionaries documenting the official languages are for example the *Pukuntšutlhaloši ya Sesotho sa Leboa* (comprehensive monolingual Sesotho sa Leboa dictionary), *The Greater Dictionary of IsiXhosa* and the *Woordeboek van die Afrikaanse Taal* (WAT).

4.1.2 Documentation and standardization: restricted dictionaries

Not only general dictionaries but also restricted dictionaries are important tools to document and to standardize a language. Restricted dictionaries are the visible outcome of the process of lexical modernization. According to Cooper (1989: 149), "modernization ... refers to the process whereby a language becomes an appropriate medium of communication for modern topics and forms of discourse." It is mainly effected at the lexical level by enlarging the vocabulary of a language. Typical ways of enlarging the vocabulary are, for example, coining new terms, and introducing loan-words or translations into the language.

Another area of standardization is unification. Here technical dictionaries play a vital role because they provide a standardized term for a concept, i.e. there are not several different terms for one concept but only one term considered to be the standard one. Having a standard form makes communication much easier as the same term is used for the same concept. In addition, technical dictionaries pave the way for other projects, for example, HLT applications such as spell checkers, telephone interpreting systems, information systems, reservation systems or multilingual ATMs. This makes them important tools for modernizing a language.

4.1.3 Documentation and standardization: online dictionaries

As mentioned earlier, the advantage of online dictionaries is that they have virtually no space restrictions, which allows a thorough documentation of a language. This includes all forms of a word, not just the lemmatized stem. Online dictionaries are ideal tools for standardizing a language because new forms and words can easily be added to the dictionary — an important feature to keep it up-to-date as documentation tool during the process of standardization whenever new terms are created. A change in language use can be documented and included if the online dictionary has a user-feedback form where users can inform the compilers how they use a word or can enquire why certain words are omitted.

4.2.1 Empowerment through communication: general dictionaries

General dictionaries are important tools of empowerment because they make it possible for users to speak about a broad range of topics. Bi- and multilingual dictionaries which include at least one African language, enable communication between users of different first languages, providing them with the necessary words to make themselves understandable.

School and learners' dictionaries are important subcategories of general dictionaries. They help pupils to understand the teacher and teaching materials if the medium of instruction is not their first language. An example for this would be the *Oxford Bilingual School Dictionary: Northern Sotho and English*. Another advantage is that they familiarize pupils with dictionaries, i.e. they learn how to use dictionaries and can profit from this knowledge when using other kinds of dictionaries. "Beginners" dictionaries introducing students to the world of dictionaries include, for example, the *South African Oxford Multilingual Primary Dictionaries* series (Sotho and Nguni languages), *Oxford's First Bilingual Dictionary* series (an African language and English) or the *Pharos Multilingual Illustrated Dictionary*. Although these dictionaries contain only a basic vocabulary, they are pupil-oriented, acquainting learners with the concept of dictionaries and extenuating their fear of contact with dictionaries.

General dictionaries help users to improve communication in the areas of text production, text reception, spelling, pronunciation (if this is included in the dictionary) and the usage of words (corpus-based dictionaries with usage examples).

4.2.2 Empowerment through communication: online dictionaries

Instead of the traditional lemmatized forms a user finds in paper dictionaries all forms of a word can be included as headwords in an online dictionary. Including all forms instead of just the stem is important for agglutinative languages, such as, for example, the Nguni languages, because this allows the inexperienced user to find a word in a dictionary. A language learner who probably does not know the stem of an unknown word found in a text can simply enter the actual word form from the text and will be lead to the correct entry and be provided with a translation.

Online dictionaries give quick and easy help, especially during text production and reception. They might become useful tools for the younger generation, as there is a growing use of the internet. As most online dictionaries are free of charge, they can be used by tertiary education students who usually have access to computers at their institutions.

4.2.3 Empowerment through communication: restricted dictionaries

Restricted dictionaries and especially technical dictionaries allow users to communicate on a broad range of subjects, such as, for example, medicine, law,

languages and economy. They are useful tools for everybody who wants to reach people through languages. Journalists or radio and TV presenters of programmes broadcast in an African language need a standardized vocabulary if they want to avoid continuous code-switching, sounding incompetent for not knowing the correct words in their own language. The government and governmental organizations need specialized dictionaries in order to distribute important information in all official languages. Technical dictionaries are a prerequisite to give users access to new concepts, especially in science. If they do not possess the vocabulary, they cannot speak or write about technical topics.

4.2.4 Empowerment through communication: cell phone dictionaries

Because they are available virtually everywhere, cell phone dictionaries help users to communicate by providing a quick solution, i.e. a translation, for personal communications as well as when reading or writing a text. During a face-to-face communication, they might enable understanding when participants speak different languages. An important example here is the doctor-patient communications, where a misunderstanding can have tragic consequences. Cell phone dictionaries can provide instant help in such cases because they can at least provide a basic vocabulary which can help the doctor to understand the patient and the patient to comprehend what the doctor says, for example what dosage of medication to take.

5. Symbolic value of dictionaries

Dictionaries are valuable means to implement measures not only for corpus planning but also for status planning. They can convey a high symbolic value if they are introduced in such a way that the target users will accept and value them. In short, a dictionary can be a status symbol for a language. A dictionary is the visible proof that a language is developed and has a standard form, i.e. the language has progressed from a dialect to a "proper" language. Dictionaries can conduce first language speakers of less prestigious languages to take pride in their languages by showing them that they can be used in many domains of life. Although there are, from a linguistic point of view, no worthless languages or better languages, people perceive different languages differently. In South Africa, languages have often been used as means to suppress people, especially during the apartheid years. This led to a negative image of the African languages, especially among their mother-tongue speakers. English was seen as the language of the struggle and liberation, while the African languages were used to separate the people and to exclude them from important jobs. The implementation of the language stipulations in the new constitution and the language bill are necessary to change such negative connotations into positive ones. Dictionaries are useful tools to do so, as they are not only a visible proof of corpus planning measures but can also be used in status planning by contributing to enhance the status of a language.

5.1 General dictionaries as status symbols

A comprehensive dictionary is the visible proof that the language has actually reached a similar high level as for example Afrikaans or English. It shows that the language is more than a "worthless" dialect which has no written standardized form and which lacks the vocabulary for its use in more formal domains of life. Having a multi-volume dictionary for a language shows the speakers that their language actually has an extensive vocabulary. Many of the African languages are still underrated by their mother-tongue speakers and still induce negative connotations. They are often considered worthless, because their speakers believe that they cannot be used in the workplace with their supervisors, in academic situations or when consulting doctors or lawyers. Therefore, improving the status of these languages is very important. Demonstrating that the language is developed is a difficult and protracted task but making the language visible is a first step towards enhancing its status. Compiling attractive dictionaries, i.e. dictionaries having an interesting cover and printed on good quality paper is another step towards user acceptance. Even if the first use of such an attractive-looking dictionary is just to impress visitors and show them that the possessor is proud of his culture and language, having acceptable dictionaries is important as only such products can serve as status symbols.

5.2 Restricted dictionaries as status symbol

For smaller languages and languages which are still not used regularly in all domains of life, it is important to have restricted and especially technical dictionaries. They show that the language actually has the vocabulary to discourse on a broad range of topics. The dictionary is a visible proof that there are words, that they have a standardized form, and that they can be used. Being able to use such standardized forms makes users to appear more educated because they can speak and write properly about a particular subject which enhances their self-esteem.

5.3 Online and cell phone dictionaries as status symbols

Modern publications such as online and cell phone dictionaries carry a highly emotional value as they symbolize a link between a traditional language that is still not used in all domains of life and modern technologies, as for example the internet and cell phones. In other words, they show that the language has enough status and is well enough developed to be used on the internet and on cell phones. Having a dictionary on their cell phones has another psychological advantage, because it gives the users the reassurance that they are able to communicate with the help of the dictionary which furnishes them with the

necessary words whenever and wherever they need them.

Online dictionaries are accessible from everywhere in the world. Having online dictionaries makes a language visible and effects access to the World Wide Web. Examples of successful online dictionaries are the *isiZulu.net Zulu–English Dictionary* (<http://www.isizulu.net/>) and the *Sesotho sa Leboa (Northern Sotho)–English Dictionary* (<http://www.africanlanguages.com/sdp/index.php?1=en>). Such dictionaries may also be used by people from abroad who will recognize that the language is used on the internet. Although this will mainly be an unconscious realization, it helps to change the status of a language. The language is there, it has an online dictionary, so it has progressed from a vernacular language to a modern language.

6. Conclusion

The lexicographic landscape in South Africa has changed during the last decade. By now all official languages have general dictionaries and some technical dictionaries have also been compiled. New publication modes have emerged; there are online and cell phone dictionaries available for the African languages. This has been made possible because the languages are supported by the constitution, and because some language planning policies have been successfully implemented. These new African language dictionaries are very important for South Africa. They can empower their speakers and can elevate the status of the languages. But to do so, they must be freely available to the speech community. As many people in South Africa still live in a pre-dictionary culture, they must be user-friendly and have a clear user-guide in order to be useful for the inexperienced user.

Acknowledgements

I am grateful to Ms A. Barfs for a very helpful and constructive critique of an earlier version of this article. However, she obviously carries no responsibility for any part of it.

References

- Cooper, R. 1989. *Language Planning and Social Change*. Cambridge: Cambridge University Press.
- Ferguson, G. 2006. *Language Planning and Education*. Edinburgh: Edinburgh University Press.
- Geyser, R. 2005. *Language Distribution and the Situation in the Other Provinces* [online]. <http://www.fs.gov.za/.../3presentation%20roelof%20geyser.pp> [21 February 2009].
- Gordon, R.J., Jr. (Ed.). 2005. *Ethnologue: Languages of the World*. Fifteenth Edition. Dallas: SIL International. http://www.ethnologue.com/show_country.asp?name=ZA [23 March 2009].
- Gouws, R. and D. Prinsloo. 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS

- Haugen, E.** 1966. Linguistics and Language Planning. Bright, W. (Ed.). 1966. *Sociolinguistics: Proceedings of the UCLA Sociolinguistics Conference, 1964*: 50-70. The Hague: Mouton.
http://www.afrilux.co.za/quickies/South_Africa.htm [23 March 2009].
- Joffe, D.** 2003. *Linguistics Terminology Sesotho sa Leboa–English* [online]. <http://africanlanguages.com/sdp/linguistics/> [23 March 2009].
- Milroy, J.** 1999. The Consequences of Standardisation in Descriptive Linguistics. Bex, T and R. Watts (Eds.). 1999. *Standard English: The Widening Debate*: 16-39. London: Routledge.
- National Language Service.** 2003. *Multilingual Mathematics Dictionary Grade R–6* [online]. http://www.deac.gov.za/chief_directorates/NLS/list.htm [23 March 2009].
- National Language Service.** 2005. *Multilingual Natural Sciences & Technology Dictionary for Grade 4 to 6* [online]. http://www.deac.gov.za/chief_directorates/NLS/list.htm [23 March 2009].
- TshwaneLex.** 2008. *Birds and Trees List: English–Sesotho sa Leboa* [online]. <http://africanlanguages.com/sdp/ff/index.php?l=en> [23 March 2009].
- Webb, V.N.** 1996. Language Planning and Politics in South Africa. *International Journal of the Sociology of Language* 118: 139-162.
- Webb, V.N. and E. Kembo-Sure.** 2000. The Languages of Africa. Webb, V.N. and E. Kembo-Sure (Eds.) 2000. *African Voices. An Introduction to the Languages and Linguistics of Africa*: 26-54. Oxford: Oxford University Press.
- Wolfe, J.** 1993. *Important Zulu Medical Words — Their Pronunciation and Translation* [online]. <http://www.wolfescape.com/WebPages/ZuluDict.htm> [23 March 2009].

Natural Science and Technology Terminology in the Sesotho sa Leboa Monolingual Dictionary*

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

Abstract: The main purpose of this article is to make a comparative analysis of the structural and the lexical differences between coinage and transliteration, as methods for the acquisition of foreign natural science and technology lexical items, which are meant for inclusion in the Sesotho sa Leboa comprehensive monolingual dictionary, *Pukuntšutlhaloši ya Sesotho sa Leboa*. With the exception of the multilingual glossaries ('dictionaries' as they are called), which are compiled by the National Language Service of the Department of Arts and Culture, the Sesotho sa Leboa National Lexicography Unit does not at present have any monolingual dictionary specializing in specific subjects like natural science and technology. As a start, the Lexicography Unit decided to include this specialized terminology in the second edition of its comprehensive monolingual dictionary. The main purpose for lemmatizing the natural science and technology terminology is to give these lexical items comprehensive definitions in Sesotho sa Leboa, instead of 'a one word translation definition' as is the case in the bilingual and multilingual dictionaries.

Keywords: TRANSLITERATION, COINAGE, COINED LEXICAL ITEM, LOAN WORD, ADOPTIVE, FOREIGN WORD, LEXICOGRAPHY, SOTHOIZED WORD, INDIGENOUS WORD, NATURAL SCIENCE, TECHNOLOGY, COMPOUND WORD, AMBIGUITY

Opsomming: **Natuurwetenskaplike en tegnologiese terminologie in die Sesotho sa Leboa eentalige woordeboek.** Die hoofdoel van hierdie artikel is om 'n vergelykende ontleding te maak van die strukturele en die leksikale verskille tussen nuutskepping en transliterasie as metodes vir die verwerking van vreemde natuurwetenskaplike en tegnologiese leksikale items wat bedoel is vir insluiting in die Sesotho sa Leboa omvattende eentalige woordeboek *Pukuntšutlhaloši ya Sesotho sa Leboa*. Met die uitsondering van die meertalige woordelyste ("woordeboeke" soos hulle genoem word) wat deur die Nasionale Taaldiens van die Departement Kuns en Kultuur saamgestel word, het die Sesotho sa Leboa Nasionale Leksikografie-eenheid op die oomblik geen eentalige woordeboek wat spesialiseer in spesifieke onderwerpe soos die natuurwetenskap en tegnologie nie. As 'n begin het die Leksikografie-eenheid besluit om hierdie gespesialiseerde terminologie in die tweede uitgawe van sy omvattende eentalige woordeboek in te sluit. Die hoofdoel vir die lemmatisering van hierdie natuurwetenskaplike en tegnologiese ter-

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

minologie is om hierdie leksikale items omvattende definisies in Sesotho sa Leboa te gee in plaas van " 'n eenwoordvertalingdefinisie" soos die geval is in die tweetalige en meertalige woordeboeke.

Sleutelwoorde: TRANSLITERASIE, NUUTSKEPPING, NUUTGESKEPTE LEKSIKALE ITEM, LEENWOORD, ONTLENING, VREEMDE WOORD, LEKSIKOGRAFIE, VERSOTHODE WOORD, INHEEMSE WOORD, NATUURWETENSKAP, TEGNOLOGIE, SAMESTELLING, DUBBELSINNIGHEID

1. Introduction

The main objective of this article is to analyze the structural and the lexical differences between the natural science and technology terms adopted through coinage and transliteration in Sesotho sa Leboa or Northern Sotho. Most of the examples of the science and technology terms used for comparative analysis in this article are derived from the publication of the National Language Services of the Department of Arts and Culture, i.e. the *Multilingual Natural Science and Technology Dictionary for Grades 4 & 6* (2005). The lemmatization of this terminology in the Sesotho sa Leboa Monolingual Dictionary is aimed at giving the terms full and detailed monolingual definitions, instead of single translation definitions, as those appearing in this word list. The Sesotho sa Leboa natural science and technology terms are mostly derived from foreign languages, especially Latin, Greek, English and Afrikaans, and usually enter the Sesotho sa Leboa lexicon through coinage or transliteration. Coinage is possible where the equivalent indigenous lexical items exist in Sesotho sa Leboa and where lexical items, including loan lexical items, are used as compound words to refer to new concepts. Most natural science and technology terms do not have equivalents in Sesotho sa Leboa, and in these cases, transliteration is used to form new lexical items for the new concepts.

2. Coinage and transliteration

Most Sesotho sa Leboa linguistic scholars and theorists prefer the coinage system in formal situations, especially in formal writings and publications. This preference is based on the conviction that the indigenous languages are capable of providing indigenous equivalents for the newly acquired science and technology concepts. For example, the research by Nong, De Schryver and Prinsloo (2002: 17) reveals the following:

By studying the respondents' comments as a whole, it is clear that they prefer the (more) indigenous words to be treated in dictionaries, and that loan words should only be used if there is no good alternative in Northern Sotho. Quite a number of respondents even suggest that words should be coined in order to have a Northern Sotho word instead of an adoptive from other languages. Where offered a direct borrowing and a (Sothoised) loan as only options, the

(Sothoised) loan is preferred to the direct borrowing. This thus suggests the following preference hierarchy: indigenous word > Sothoised loan word > direct borrowing.

According to this point of view, the best way to develop the vocabulary of the language is to promote or to elevate the indigenous lexical item by widening its scope of reference to refer also to the newly developed science and technology concept.

The *Longman Modern English Dictionary* defines 'coinage' as the inventing of new words, phrases or expressions. This type of linguistic borrowing is sometimes referred to as neologism, i.e. where only the meaning of the word is adopted and not the word itself. The meaning of the foreign term (the science and technology term) is associated with an indigenous word, especially a word referring to a more or less related sense. The meaning of the indigenous word is extended to include, also, the meaning of the foreign term. The following are examples of lexical items which originated through coinage, and which are listed in the *Multilingual Natural Science and Technology Dictionary*:

lefaseterekhomphutha 'desktop computer'
legonggodišaboleng 'veneer'
lephelotswadišo 'propagation cycle'
sešepisameno 'toothpaste'

According to the *Longman Modern English Dictionary*, transliteration is the replacement of the letters of one alphabet by the letters of another alphabet with the same phonetic sounds. The *Encarta Concise English Dictionary* defines 'transliterate' as:

To represent letters or words written in one alphabet using the corresponding letters of another.

The following are examples of transliterated lexical items taken from the *Multilingual Natural Science and Technology Dictionary*:

silibere 'silver'
lelampi 'lamp'
galase 'glass'
softwere 'software'
ketlele 'kettle'
faunthene 'fountain'
sepikara 'speaker'
kuka 'cook'
karatšhe 'garage'

The lexical and structural comparison between the coined and the transliterated lexical items is based on the following factors: ambiguity, increase in vocabulary, compounding, and user-friendliness.

3. Ambiguity

Ambiguity is one of the major factors underlying the differences between the coined and the transliterated lexical items. It is defined by many linguists, lexicographers and semanticists. According to the *Longman Modern English Dictionary*, ambiguity is:

The quality of having more than one meaning: an idea, statement or expression capable of being understood in more than one sense.

Kooij (1971: 7) points out that a sentence is inherently ambiguous when it sounds the same but has more than one meaning. With regard to ambiguity, Hurford and Heasley (1983: 122) say:

A word or phrase is said to be ambiguous if it has two, or more, synonyms that are not themselves synonyms of each other.

Mojela (1991: 26) mentions that:

Ambiguity is found where a loan word is used synonymously with an indigenous Northern Sotho word to refer to one and the same referent.

The major challenge facing lexicographers in the lemmatization of coined lexical items is their definition. When compared to the lexical items created through transliteration, the definitions of the coined science and technology terms in the Sesotho sa Leboa monolingual dictionary are more complicated. These complications are due to the fact that almost all the coined lexical items are basically ambiguous, i.e. 'have two, or more, synonyms that are not themselves synonyms of each other' (Hurford and Heasley 1983: 122). Adoption of science and technology terms into the Northern Sotho vocabulary through coinage occurs through:

- (a) Extending or widening the meaning of existing lexical items to refer also to the meaning of the new terms and, therefore, causing them to become ambiguous, e.g.
mogala: Basic meaning: 'a rope (used to tame cattle or heifers)'
Extended meaning: 'telephone'
- (b) Converting the action, function or meaning of the foreign term into a newly coined lexical item, usually in the form of a compound word, to name the new concept, e.g.
sellathekeng: Literally: 'that which cries at the waist'
Meaning: 'cellular telephone'
seoketšamodumo: Literally: 'that which increases sound'
Meaning: 'amplifier'

In contrast to the preceding coined lexical items which are characterized by ambiguity, i.e. having more than one meaning, the transliterated lexical items are formed by adaptation of the foreign words themselves to the morphological and the phonological system of Sesotho sa Leboa. The transliterated lexical items are usually characterized by having the same meaning as their foreign counterparts, as shown in the following examples:

rayone 'rayon'
khomphutha 'computer'
kerafu 'graph'
bithamine 'vitamin'

4. Increase in vocabulary

The second major difference between coinage and transliteration as important methods of foreign acquisition in Sesotho sa Leboa, is that coinage does not directly lead to an increase in the quantity of lexical items as is the case with transliteration. The reason for this is that in coinage one word is repeatedly used polysemously to refer to several concepts instead of adding new lexical items to the vocabulary. Transliteration leads to an increase in the quantity of lexical items in Sesotho sa Leboa, while coinage only increases the meanings of existing lexical items in the language. The lexicographer will enter the coined lexical item in the monolingual dictionary as a polysemous lexical item, i.e. 'as an instance where a lexical item has a set of more than one, different but related meanings' (Mojela 1991: 31). But the loan lexical item which originated through transliteration in Sesotho sa Leboa is entered in the monolingual dictionary as a new and independent lexical item, therefore increasing the vocabulary of the language instead of only increasing the semantic system of the existing lexical items. The following are examples of differences in the dictionary entries of the coined and the transliterated lexical items in the Sesotho sa Leboa monolingual dictionary:

Coinage:

mosebe ('arrow')

1. *Sebetša goba sedirišwa sa go dirwa ka tshipi goba motato seo se šomišwago go hlaba goba go bolaya, kudukudu se šomišwa ke batsomi ge ba tsoma diphoofole goba madira ge a le ntweng.*

('A weapon or an instrument, made from steel or wire, which is used to stab or to kill, especially used by hunters when hunting animals or by soldiers during war.')

2. *Misaele goba mosebe wa mollo wo o šomišwago ke madira dintweng tša selehono.*

('A missile or firearm which is used by soldiers in modern warfare.')

Transliteration:*misaele* ('missile')

Sebetša sa go bopega bjalo ka mosebe se se tlabaketšwego ka dithuthupiši seo se šomišwago ke madira dintweng tša sehono. Sebetša se se rethwa go tšwa sekepeng sa ntwa, sefofane goba go tšwa go setlabakelo sa go thuntšha se se theilwego ka nepo ya go thuntšhetša sebetša se fao se tlogo thuthupa se le gona. ('An arrow-like explosive weapon used by soldiers in modern warfare. This weapon is shot from a naval boat, military aircraft or a special instrument designed to fire and direct this weapon where it is required to explode.')

The preceding examples show the following differences between *mosebe* 'an arrow' (coinage) and *misaele* 'missile' (transliteration), i.e.:

mosebe: One polysemous lexical item referring to two senses, i.e. (1) an indigenous weapon and (2) a modern military weapon:

mosebe (1) 'arrow' (2) 'missile'

misaele: One transliterated lexical item which bears the basic meaning of its original English source term:

misaele 'missile'

These examples also show that when compared to transliteration, word coinage whose most important feature is polysemy, does not contribute much to the increase of the vocabulary of Sesotho sa Leboa but only of the meaning of existing lexical items in the language. The lexical item, *mosebe* 'arrow' is entered in the dictionary as one polysemous lexical item even though its reference has been extended to include the newly acquired scientific meaning of a weapon, i.e. 'a missile', while the transliterated lexical item, *misaele* 'missile', is a new addition to the Sesotho sa Leboa lexicon because the term has never existed before in its vocabulary.

5. Compounding

The use of compound lexical items is one of the distinctive features of coined lexical items. Most coined lexical items referring to science and technology concepts are usually descriptive compound words, since the lexical item is derived from the meaning of the foreign compound words, and not from the term itself, as shown in the following examples:

segodišamodumo 'amplifier' Literally: 'that which increases sound'

sefatanagantefe 'taxi' Literally: 'motor-car pay me'

sehlapišameno 'toothbrush' Literally: 'that which washes teeth'

tankaserwaladiela 'tanker' Literally: 'tank which carries liquid'

phelakadingwe 'parasite' Literally: 'live on/by others'

seelakgokaganya mohlagase 'electrolyte' Literally: 'electricity linkage current'
moriti wa senepe 'film negative' Literally: 'shadow of photograph'

Here compounding is usually the result of a combination of verbs, substantives and qualificatives to form new compound lexical items which describe the actions or the meanings of the newly acquired or adopted science and technology terms. Most of these newly coined terms are usually long and ambiguous.

The compound lexical item, *segodišamodumo*, referring to an amplifier, which is an audio device used to increase the volume of sound, consists of a deverbative noun *segodiša-* 'that which causes to increase', which is derived from the verb stem *-godiša-* 'cause to increase' and a noun *-modumo* 'sound'. This is the case also with the coinage *sefatanagantefe* 'taxi', which is a compound word consisting of the noun *sefatanaga-* 'car' and *-ntefe* 'pay me'. The word *sefatanaga* 'car' is furthermore a compound adopted and standardized earlier, consisting of a deverbative substantive *sefata-* 'that which ploughs the ground' (derived from *-fata-* 'plough'), and a noun *-naga* 'field'.

The compound lexical item *sehlapisameno* 'toothbrush' literally means 'that which washes the teeth', consisting of a deverbative noun *sehlapisa-*, derived from the verb stem *-hlapa-* 'wash', and the noun *-meno* 'teeth'. The compound *tankaserwaladiela* 'tanker' literally means 'the tank which carries liquids/fluids', consisting of the noun *tanka-*, a loan word from English 'tank', a deverbative noun *-serwala-* from the verb stem *-rwala-* 'carry' and the deverbative noun *-diela* 'liquids' or 'fluids' from the verb stem *-ela* 'flow'. In this case, *-serwala-* serves as a conjunctive between the two nouns *tanka-* and *-diela*.

The lexical item *phelakadingwe*, which literally means 'live by/on others', refers to a parasite. It is a compound consisting of a verb stem *phela-*, 'live', an instrumental prefix, *-ka-* 'with', and a substantive, *-dingwe*, 'others'. Despite the fact that this compound lexical item is descriptive, its reference can only be understood when used in context. As such, it is too ambiguous when compared to the transliterated loan lexical item, *pharasaete* 'parasite', as demonstrated in the following sentences:

Ke phelakadingwe:

1. 'It is that which lives on/by others/other things' (which may not necessarily be a parasite)
2. 'It is a parasite'

Ke pharasaete: 'It is a parasite' (which has a direct meaning)

Seelakgokaganya mohlagase 'electrolyte' and *moriti wa senepe* 'film negative' are multiword lexical items, each one serving as a single grammatical unit. *Seelakgokaganya mohlagase* which literally means 'electricity linkage current', consists of a deverbative noun *seela-* 'current' from the verb stem *-ela* 'flow', a verb stem *-kgokaganya-* 'link' and the noun *-mohlagase* 'electricity'. *Moriti wa senepe* refers to 'film negative', literally meaning 'shadow of a snap/photograph'. It consists of a noun *moriti-* 'shadow', a possessive concord *-wa-* 'of' and a noun *-senepe*

'photograph' which is a loan word from English 'snap'.

In order to name the newly acquired science and technology concepts, the coinage system is also characterized by the compounding of transliterated lexical items or a combination of loan lexical items and the indigenous lexical items to form new compounds. This is contrary to the preference of most linguistic and indigenous scholars who favour coinage instead of transliteration as a means of foreign acquisition, i.e. the preference of using indigenous terms or indigenous meanings instead of converting the foreign terms to name the newly acquired science and technology concepts. Lexical items such as the following are examples in this regard:

motšheneatšhelete 'ATM (automatic teller machine)'
thelebišenitherei 'plasma television'
lefaseterekhomphutha 'desktop computer'

The coined word *motšheneatšhelete* referring to an ATM (automatic teller machine) is a compound consisting of two loan words, which are both derived directly from foreign terms, i.e. *motšhene-* 'machine' (noun), *-a-* possessive concord (conjunction), and *-tšhelete* 'money' (noun, from Afrikaans *geld*). Instead of coining a new, long and ambiguous lexical item by combining foreign lexical items to refer to the ATM, it would have been convenient to consider creating a new loan lexical item from this acronym through transliteration, e.g. *eithieme* (from the pronunciation of the term). This new creation will not only be an addition to the vocabulary of the language, but will also have a direct meaning when compared to the coined lexical item *motšheneatšhelete* which can be interpreted or translated to mean 'any machine associated with money', such as a till, a cash, a minting, or a money-counting machine. The suggested transliterated lexical item, *eithieme*, will refer only to an ATM.

This is also the case with the coined lexical item *thelebišenitherei* 'plasma television' which consists of two foreign lexical items which were adopted into Sesotho sa Leboa through transliteration, i.e. *thelebišeni-* from 'television' and *-therei* from 'tray'. This compound expresses the sense that 'a plasma television' is a television which is shaped like a tray.

The lexical item *lefaseterekhomphutha*, referring to a desktop computer, is a combination of the following loan lexical items: *lefasetere-* 'window' (from Afrikaans *venster*) and *-khomphutha* 'computer'. Both the lexical items, i.e. *lefasetere* and *khomphutha*, are loan lexical items, and another loan lexical item can still be formed through the transliteration of 'desktop', e.g. *deskethopho*, which will not only be an addition to the Sesotho sa Leboa lexicon, but will also have the direct and unambiguous meaning 'desktop computer'.

6. User-friendliness

In comparison with the transliterated lexical items, the coined lexical items are not as user-friendly. Improvements in the knowledge and pronunciation of

foreign languages, especially English and Afrikaans, the two former official languages of the Republic of South Africa, led to improvements in the pronunciation of foreign lexical items among the Basotho ba Leboa communities. As a result, the transliterated lexical items came to be preferred above the coined lexical items since users opt for words which are closer to the source language and which have meanings equivalent to the original foreign terms. Ambiguity and compounding are among the major factors causing user-unfriendliness as far as coined lexical items are concerned. Ambiguity means that speakers will always have to explain the coined terms in sentences for listeners to understand their meaning. For instance

- (1) *Ke hlapa meno ka kholkeiti* 'I clean my teeth with colgate', instead of:
Ke hlapa meno ka sešepisameno 'I clean my teeth with toothpaste/soap for the teeth'.

The reason for this preference is that the lexical item *kholkeiti* is a direct equivalent to its original English counterpart, the trade name 'Colgate', when compared to the long, compound and ambiguous coined lexical item, *sešepisameno*, which consists of the following components: *sešepi-* 'soap' (noun, loan word from Afrikaans *seep*), *-sa-* 'of' (possessive concord), and *-meno* 'teeth' (noun).

Besides referring to toothpaste, this lexical item can refer to any soap-like cleansing agent for the teeth. However, *kholkeiti* will always mean 'toothpaste'.

- (2) *Ke butše deskthopho ya khomphutha ya gagwe* 'I opened his/her desktop computer', instead of:
Ke butše lefaseterekhomphutha ya khomphutha ya gagwe 'I opened his/her desktop computer'

The transliterated word *deskthopo* is not only short and direct to the point, but has an equivalent meaning with the original word 'desktop', while the coined word *lefaseterekhomphutha* is very long and ambiguous, as explained under section 5 above.

7. Conclusion

This article has compared the user-friendliness of the coined and the transliterated natural science and technology terminology in Sesotho sa Leboa. The example terms have been derived from the *Multilingual Natural Science and Technology Dictionary*. This comparative analysis has revealed the following:

- The majority of scholars and theorists prefer to use the coined lexical items in formal publications and written language, even though they never use most of these lexical items in practical situations.
- The coined lexical items are not user-friendly, due to ambiguity and compounding.

- In comparison to the transliterated lexical items, the lemmatization of the coined lexical items is usually a challenge due to the ambiguity and compounding, as in the following:
 - (a) In a compound word, the lexicographer will have to be conversant with the etymology for all the components in a compound lexical item in order to give a coined lexical item a complete and accurate lexicographic definition.
 - (b) Most coined lexical items are entered into the dictionary as polysemous words, i.e. as one word referring to more than one related meaning (therefore ambiguous), while transliterated lexical items always refer to the same meaning as that of the source term in the foreign language (therefore unambiguous).
 - (c) Some of the coined compound lexical items are very long, and to enter them as lemmata into the dictionary is less convenient than lemmatizing a transliterated lexical item.

References

- Department of Arts and Culture.** 2005. *Multilingual Natural Science and Technology Dictionary for Grades 4 to 6*. National Language Service.
- Hurford, J.R. and B. Heasley.** 1983. *Semantics: A Course Book*. Cambridge: Cambridge University Press.
- Kooij, J.G.** 1971. *Ambiguity in Natural Languages*. Amsterdam: North-Holland Publishing Company.
- Mojela, V.M.** 1991. *Semantic Changes Accompanying Loan-words in the Northern Sotho Lexicon*. Unpublished M.A. Dissertation. Pretoria: Vista University.
- Nong, S., G.-M. de Schryver and D.J. Prinsloo.** 2002. Loan Words versus Indigenous Words in Northern Sotho — A Lexicographic Perspective. *Lexikos* 12: 1-20.
- Rooney, K. (Ed.-in-Chief).** 2001. *Encarta Concise English Dictionary*. London: Bloomsbury.
- Watson, O. (Ed.).** 1976. *Longman Modern English Dictionary*. Harlow: Longman.

Populating Sub-entries in Dictionaries with Multi-word Units from Concordance Lines

Thapelo J. Otlogetswe, *Department of English, University of Botswana, Gaborone, Botswana (otlogets@mopipi.ub.bw)*

Abstract: Lexicography is primarily concerned with the representation of words and their senses in dictionaries. By *words* most dictionary users and lexicographers refer to a combination of characters delineated by spaces on both sides. This article discusses the weakness of this approach in the selection of dictionary entries. Through an inspection of concordance lines generated from a multi-million Setswana corpus, it is argued and demonstrated how multi-word units (MWUs), also known as multi-word expressions (MWEs), may be extracted from concordance lines to supplement dictionary entries. It is illustrated how both monolingual and bilingual Setswana dictionaries may be enhanced by the addition of MWEs as sub-entries.

Keywords: SETSWANA, LEXICOGRAPHY, MULTI-WORD UNIT, CORPUS, CONCORDANCE, MULTI-WORD EXPRESSION, COLLOCATION, WORD, SUB-ENTRIES, DICTIONARY

Opsomming: Die aanvulling van subinskrywings in woordeboeke met meerwoordige eenhede uit konkordansiereëls. Leksikografie is hoofsaaklik gemoeid met die weergawe van woorde en hul betekenis in woordeboeke. Met *woorde* verwys die meeste woordeboekgebruikers en leksikograwe na 'n kombinasie van lettertekens afgegrens deur spasies aan beide kante. Hierdie artikel bespreek die swakheid van hierdie benadering by die keuse van woordeboekinskrywings. Deur 'n ondersoek van konkordansiereëls gegenereer uit 'n multimiljoen-Setswanakorpus, word daar geredeneer en verduidelik hoe meerwoordige eenhede (MWE's), ook bekend as meerwoordige uitdrukkinge (MWU's), uit konkordansiereëls onttrek kan word om woordeboekinskrywings aan te vul. Daar word aangetoon hoe sowel eentalige as meertalige Setswanawoordeboeke uitgebrei kan word deur die toevoeging van MWU's as subinskrywings.

Sleutelwoorde: SETSWANA, LEKSIKOGRAFIE, MEERWOORDIGE EENHEID, KORPUS, KONKORDANSIE, MEERWOORDIGE UITDRUKKING, KOLLOKASIE, WOORD, SUBINSKRYWINGS, WOORDEBOEK

1. Introduction

At the centre of lexicography lies the problem of what constitutes a word. The problem is not only a lexicographic one. It is also a linguistic one. McArthur (1998: 45-47) identifies eight types of words: orthographic, phonological, morphological, lexical, grammatical, onomastic, lexicographical and statistical words.

What constitutes words is critical in corpus linguistics, since it translates into the problem of what gets counted by the computer. Lexicographically, those ones considered as words are listed in the dictionary.

2. The word problem

In frequency analysis, there is therefore a need to clarify what constitutes a word in a language and how words get counted. In linguistic literature, the term *word* is defined in a variety of ways. Some of these definitions, while useful for theoretical linguistics, are useless for computational word counts. Finch (2000: 132) defines a word as "a unit of expression which native speakers intuitively recognize in both spoken and written language" and adds that "there is a certain indeterminacy about the definition of a word". Finch's definition is unhelpful in that "a unit of expression" could be anything from a word, a phrase, a clause or a sentence. His definition also leaves the determination of what a word is to a speaker's intuition which may vary from one speaker to another. Aitchison (1992: 49) points out that "the best-known definition of a word is the one proposed by the American linguist Bloomfield who defined it as a minimum free form, that is, the smallest form that can occur by itself". She further argues that distinctions must be made between lexical items, syntactic words and phonological words. If we consider lexical items, a form such as *fly* represents at least two words:

fly [noun]: an insect with two wings.

fly [verb]: to move through the air in a controlled manner.

The two lexical items have different syntactic forms associated with them. The noun could either be singular (*fly*) or plural (*flies*). The verb on the other hand could occur as *fly*, *flying*, *flies*, *flew* and *flown*. This therefore raises problems for the Bloomfieldian approach.

Leech et al. (1982: 27) consider a word as "delimited, for most purposes by a space (or punctuation mark other than a hyphen or apostrophe) on each side". This is known in linguistic literature as an orthographic word. However they also acknowledge that "the boundaries of words ... are not always clear; e.g. we can write the sequence *piggy + bank* in three ways: *piggy bank*, *piggy-bank*, or *piggybank*".

In most computational processes, a word is treated as a "minimal free form, the smallest unit that can exist on its own" (Dash and Chaudhuri 2000: 189) and "delimited by a space ... on each side" (Leech et al. 1982: 27). This approach is helpful if one is studying forms delineated by spaces. However, in this article, larger units which have spaces within them are studied. Moon (1998) calls these fixed expressions and idioms. In other literature they are called multi-word units or MWUs (Schone and Jurafsky 2001) or multi-word expressions or MWEs (Sharroff 2004; Oflazer and Çetinoğlu 2004; Villavicencio

et al. 2004; Fazly and Stevenson 2007). Bannard (2007: 1) gives the following definition:

A multi-word unit is usually taken to be any word combination (adjacent or otherwise) that has some feature (syntactic, semantic or purely statistical) that cannot be predicted on the basis of its component words and/or the combinatorial processes of the language. Such units need to be included in any language description that hopes to account for actual usage.

Sag et al. (2002: 2) characterize MWEs as "idiosyncratic interpretations that cross word boundaries (or spaces)". And if Jackendoff's (1997: 156) estimate that the number of MWEs in a speaker's lexicon is of the same order of magnitude as the number of single words, then MWEs deserve focus and will significantly enhance dictionary entries.

MWEs therefore include idioms, phrasal verbs, proverbs, compound words, etc. English examples are *by and large*, *kick the bucket*, *in step*, *take up*, *take off*, *shake up*, *telephone booth*, *pull strings*, *fresh air*, *fish and chips*, *salt and pepper*, etc. Setswana examples are *solegela molemo* (benefit), *kukega maikutlo* (be upset), *iphaga dikoro* (involve oneself in other people's business), *tsholetsa maoto/dinao* (walk faster), *opisa tlhogo* (cause trouble), *tsaya karolo* (participate), *tsaya tsia* (pay attention), *nna le seabe* (take part), *ja monate* (enjoy), etc. The immediate problem arises with their identification, since they can be written in diverse and inconsistent ways. Take for example the following different spellings which are acceptable in both English and Setswana as written in Botswana.

<i>houseboat</i>	<i>house-boat</i>	<i>house boat</i>
<i>tradeoff</i>	<i>trade-off</i>	<i>trade off</i>
<i>khuduthamaga</i>	<i>khudu-thamaga</i>	<i>khudu thamaga</i>
<i>pelotshetlha</i>	<i>pelo-tshetlha</i>	<i>pelo tshetlha</i>
<i>rampatshetlha</i>	<i>rampa-tshetlha</i>	<i>rampa tshetlha</i>
<i>motshwaradiphala</i>	<i>motshwara-diphala</i>	<i>motshwara diphala</i>
<i>kgakalakangakala</i>	<i>kgakala-kgakala</i>	<i>kgakala kgakala</i>

The examples *houseboat* and *kgakalakangakala* will each constitute a single token, while *house boat* and *kgakala kgakala* will form two tokens each. Words joined by a hyphen can either be recognized as single words or as two separate words depending on the tokenizing program. The difference is not trivial in statistical linguistics, since the number of tokens will vary significantly depending on what is counted.

3. Methodology and experiments

For our experiments, we follow Brunner and Steyner (2008) and use corpus data. By a corpus is meant, according to Renouf (1987: 1), "a collection of texts,

of written or spoken words, which is stored and processed on computer for the purpose of linguistic research". The Setswana corpus used for the experiments is just over 13 million tokens. The software employed is *Oxford Wordsmith Tools Version 4* (Scott 2004–2006). It is applied to study a specific word in context in some detail in terms of co-texts to its left and to its right. This is achieved by generating a key word in context (KWIC), often referred to as concordance lines. Dash and Chaudhuri (2000: 190) give the following definition:

A concordance is an index of the surface word forms in a text. It is a collection of the occurrences of a word form, each in its own textual environment.

A concordance reveals the context of a word, its collocates, and thereby reveals meanings and usages which are hard to recover through mental recall (Otlogetswe 2007: 56). We illustrate this below with the example of the word *pele* (heart).

Figure 1: Concordance results for the word *pele*

o ka kgopolo ya gore Morwadi o tlaa wela pele. A mo gaupanya. ka legofl fa ga re ngwatiaka, O se tshoge bono wa ka, Wela pele ga o seitaodi re Use rotlhe, O sek gang, o tla e rola morago o sena go wela pele. ' / ' r ~ a emelela, o b-ua a le esi) a ka seatla. "O sale sentle, moratiwa wa pele ya me. Ga ke itse gore ke~ tla go se o lela jalo? Ke a go rata moratiwa wa pele ya me." Fa a sa ntse a e phimola, ne a ithuta ona. "Gomotsega, moratiwa wa pele, ya me." Mosele a didimala, mme go rebe la Mokwena, a buledisa moratiwa wa pele ya gagwe. O ne a tsamaya ka bonya, a. : Ke go reile ka re o seka wa utlwisa pele botlhoko tlhe rra! O a itse gore b g mo matlhong a gago. O se ka wa utlwisa pele ya gago botlhoko ka nna, ke swetse gatlhisa thata., Ba, utlwil ba mo tswela pele tota.. ,I, Mmaago Molebi a mo roma be, a di phailela kwa, a re ba mo tswela pele. Le ene Pule tota tsala ya gagwe y hegela ke moratiwa, e seng go mo tswela pele kgotsa go mo tlhoafalela. Seno se sadi yo montle, mme phokojwe a mo tswela pele. Phokojwe a leka maano a le mantis wana wa kgosi, noga ya bona peba ya tswa pele ka ene e ntse lobaka lo lo leele e wana wa kgosi, noga ya bona peba ya tswa pele ka ene e ntse lobaka lo lo leele e ro ya gagwe. NtsVwa Mosetsanyana ya tswa pele, ya metsa mathe a keletso. Saitsan swe ke marago a tanka e nngwe. A tshwara pele monnamogolo wa batho mme a ntse a gotla-tshekelo. Mmaagwe Sereri a tshwara pele ya gagwe, a bua ka tidimalo le bad e motlha mongwe lokwalo lo tla tshikinya pele ya ga Mmatheebe. A kwalela kgarebe ologeletsweng morwadi wa we la tshikinya pele ya gagwe gore a bale ka mabogo a a nala kwa Naledi a tla a ratile go tlola pele. Mogoma e re ntlhomane a feta fa M o ngwega Uncle Boot O ne a rata go tlola pele. Mmaagwe ene, a ipega fa a ne a le r ." Bikibiki a re, "Ngwana yo o tlhomola pele. Lefatshe le mo itaya ka ntlha ya re ruri rre Rapitso wa batho o tlhomola pele ka tshenyo e e leng ka mo supamake le bobotlana. Motho wa tsona o tlhomola pele. Keikepetse o ntse jalo mo teramen ka go tena Ontefile. Ketshedile a tlhapa pele ka o ne a sa ntse a senka leano la a. Bofelo a mo tsepela leitlho. A tlhapa pele gonne fa go rata Bofelo, tsotlhe d ng. jaanong Morwadi a nametsega a tlhapa pele. A tsaya tlhogo a e latsa mo sehub otlhapelo a ngwana wa mpa. "Nnake, tiisa pele. O sa ntse o na le mogomotsi e bon bosigo. Tlogela go nna legatlapa. Tiisa pele. Ga o sa tlhola o le mosimane. Ka a gagamatsa thamo ya gagwe, a thatafatsa pele ya gagwe, mo a bileng a se ka a bo 2Mme le ka sebaka seo Farao a thatafatsa pele ya gagwe gape, a se ka a naya mora wa gee." "Ke mang?" A botsa a swegaswega pele. Ngwananyana a bolela fa ene a bon a ke matlhagatlhaga, e bile a swegaswega pele. O ne a batla go balela kwa pele.

In the above concordance lines, *pele* together with its collocates, is rarely used to convey the meaning of the physical heart, "a hollow muscular organ that pumps the blood through the circulatory system by rhythmic contraction and

dilation" (Pearsall 1998: 847). In the first lines, *wela pelo*, which literally translates as "have your heart fall down", means "be at peace or be settled". In the next lines, *moratiwa wa pelo* (the loved one of the heart) is equivalent to "sweet-heart" or "beloved". Further on, *tshwara pelo* (handle or hold the heart) means "be in control of your emotions".

It is by inspecting collocates that we can uncover different MWEs such as proverbs, compounds, idioms, sayings, phrasal verbs, etc. Such structures can then be entered into dictionaries as sub-entries. Through the use of computer programs or concordance software, it is relatively easy to obtain a list of all the co-occurrences of a particular word in context and see all the meanings associated with the word (Biber et al. 1998: 27). The concordance lines above reveal the different subtle meanings associated with the word *pelo*. From such a study of concordance lines, a possible 84 sub-entries of the headword *pelo* have been extracted:

<i>ama pelo</i>	<i>mabetwa-e-pelo</i>	<i>pelo yotlhe</i>
<i>balabala ka pelo</i>	<i>masetla pelo</i>	<i>pelo-e-thata</i>
<i>baya pelo</i>	<i>matlhomola pelo</i>	<i>pelo-kgale</i>
<i>beta pelo</i>	<i>matlhotlha-pelo</i>	<i>pelo-telele</i>
<i>betwa ke pelo</i>	<i>nametsa pelo</i>	<i>pelo-tlhomogi</i>
<i>bofa pelo</i>	<i>ngomola pelo</i>	<i>pelo-tshetlha</i>
<i>bolawa ke pelo</i>	<i>ngona pelo</i>	<i>phatlola pelo</i>
<i>bolwetse jwa pelo</i>	<i>nna pelo</i>	<i>ritibatsa pelo</i>
<i>bona pelo</i>	<i>nona pelo ka mathe</i>	<i>sephiri sa pelo</i>
<i>bongwefela jwa pelo</i>	<i>ntsha pelo</i>	<i>sera pelo</i>
<i>bonosi jwa pelo</i>	<i>ntsha pelo pelaelo</i>	<i>sethunya sa pelo</i>
<i>boteng jwa pelo</i>	<i>pateletsa pelo</i>	<i>sisa pelo</i>
<i>bua ka pelo</i>	<i>pelo e boela mannong</i>	<i>sulafatsa pelo</i>
<i>bula pelo</i>	<i>pelo e e botlhoko</i>	<i>swa pelo</i>
<i>busa pelo</i>	<i>pelo e e letlapa</i>	<i>swegaswega pelo</i>
<i>fela pelo</i>	<i>pelo e ja serati</i>	<i>thiba maroba a pelo</i>
<i>feretlha pelo</i>	<i>pelo e khibidu</i>	<i>thuba pelo</i>
<i>fetola pelo</i>	<i>pelo e rotha madi</i>	<i>tlala pelo</i>
<i>gapa pelo</i>	<i>pelo e rutha</i>	<i>tlalelana pelo</i>
<i>garoga pelo</i>	<i>pelo e setlhogo</i>	<i>tlhomola pelo</i>
<i>kgaoga pelo</i>	<i>pelo e thata</i>	<i>tlola pelo</i>
<i>go sena letsapa le fisang pelo</i>	<i>pelo khutshwane</i>	<i>tshwara ka pelo</i>
<i>gonolwa ke pelo</i>	<i>pelo namagadi</i>	<i>tshwara pelo</i>
<i>isa pelo mafisa</i>	<i>pelo ntsho</i>	<i>tswa pelo</i>
<i>itaya pelo</i>	<i>pelo pedi</i>	<i>tswela pelo</i>
<i>itse pelo</i>	<i>pelo pholwana e a golegwa</i>	<i>uba pelo</i>
<i>kgwaralatsa pelo</i>	<i>pelo potsane e a golegwa</i>	<i>wa pelo</i>
<i>lala ka pelo e rotha madi</i>	<i>pelo tshweu</i>	<i>wela pelo</i>

In Table 1 below, only 10 of these are explained.

Table 1: Corpus-derived possible sub-entries for the entry *pele*

Collocates	Literal translation	Meaning
<i>ama pele</i>	touch the heart	hurt someone
<i>balabala ka pele</i>	speak too much by the heart	talk aloud to yourself; be absent-minded
<i>baya pele</i>	put the heart	relax; lay back
<i>beta pele</i>	suffocate the heart	persevere
<i>betwa ke pele</i>	be choked by the heart	be very angry
<i>bofa pele</i>	tie the heart	restrain yourself
<i>bolawa ke pele</i>	be killed by the heart	desire something but be unable to acquire it
<i>bolwetse jwa pele</i>	the disease of the heart	heart attack
<i>bona pele</i>	see the heart	see somebody's intentions or thoughts
<i>bua ka pele</i>	speak with the heart	be troubled to the extent that you talk to yourself

In Setswana, the phenomenon of idiomaticity when considering a word and its collocates is not unique to *pele*. Words like *molomo* (mouth), *mpa* (stomach), *nkô* (nose), *monwana* (finger), *kgomo* (cow), and many others display similar characteristics. Such idiomatic expressions can enrich dictionary entries as sub-entries. Tables 2–5 present the idiomatic expressions for *molomo* (mouth), *mpa* (stomach), *lonaô* (foot) and *matlhô* (eyes) respectively which have been extracted through studying concordance lines.

Table 2: Corpus-derived possible sub-entries of the entry *molomo*

Collocates	Literal translation	Meaning
<i>bolwetsi jwa tlhako le molomo</i>	disease of hoof and mouth	foot and mouth disease
<i>itoma molomo wa tlase</i>	bite the lower mouth	be determined
<i>itshwara molomo</i>	hold/touch the mouth	be shocked
<i>ntsha ka molomo</i>	release with the mouth	speak
<i>pula molomo</i>	that which opens the mouth	money paid before someone speaks in lobola negotiations
<i>pipa-molomo</i>	that which covers the mouth	a bribe
<i>rwala molomo</i>	carry the mouth on your head	be angry and tight-lipped
<i>roka molomo</i>	sew the mouth	remain quiet
<i>tswa molomo</i>	grow a mouth	speak
<i>tlhoka molomo</i>	lack a mouth	have nothing to say

Table 3: Corpus-derived possible sub-entries for the entry *mpa*

Collocates	Literal translation	Meaning
<i>bana ba mpa</i>	children of a stomach	relatives
<i>bipa mpa ka mabele</i>	cover the stomach with the breasts	withhold bad information to protect a relative or friend
<i>gare ga mpa ya bosigo</i>	in the centre of the belly of the night	in the middle of the night
<i>gare ga mpa ya lefatshe</i>	in the centre of the stomach of the world	in the middle of nowhere
<i>gare ga mpa ya naga</i>	in the centre of the belly of the wilderness	in the middle of nowhere
<i>mpa ya sebetse</i>	the belly of the liver	flat on the stomach
<i>mpa e tuka molelo</i>	a belly burning fire	filled stomach
<i>go ja ka mpa tsoopedi</i>	eat with two stomachs	eat until the stomach is full
<i>ntsha mpa</i>	take out a stomach	commit abortion
<i>imelwa ke mpa</i>	be overladen with a belly	have a full stomach

Table 4: Corpus-derived possible sub-entries of the entry *lonaô/dinaô*

Collocates	Literal translation	Meaning
<i>apaya ka lonaô</i>	cook with a foot	avoid cooking and eat at other people's homes instead
<i>goga dinaô</i>	drag the feet	move slowly
<i>fodisa dinaô</i>	cool the feet	have a rest
<i>motsamaya ka dinaô</i>	one who walks with the feet	a pedestrian
<i>ngotla dinaô</i>	reduce the feet	walk slower
<i>tlhatlosa dinaô</i>	raise the feet	walk faster
<i>baya lonaô</i>	put a foot	be in a place
<i>tsholetsa dinaô</i>	lift the feet	walk faster
<i>kgwele ya dinaô</i>	ball of the feet	football
<i>tsosa dinaô</i>	wake up the feet	walk faster; hurry up
<i>tiisa dinaô</i>	strengthen the feet	walk faster

Table 5: Corpus-derived possible sub-entries of the entry *matlhô*

Collocates	Literal translation	Meaning
<i>bula matlhô</i>	open the eyes	educate; make aware; open the eyes
<i>diga matlhô</i>	drop the eyes	look down
<i>digalase tsa matlhô</i>	glasses of the eyes	spectacles; sunglasses

<i>latlhela matlhô</i>	throw the eyes	look briefly
<i>matlhô a phage a lebane</i>	the eyes of a wild cat face to face	face to face
<i>kala matlhô</i>	measure the eyes	confuse
<i>tlodisa matlhô</i>	make the eyes jump	overlook someone or something
<i>kgarakgaratsha matlhô</i>	make the eyes move from one place to another	look from one place to another
<i>tlhatlosa matlhô</i>	raise the eyes	look up
<i>tlhaetsa matlhô</i>	shorten the eyes from	despise someone

4. Treatment of multi-word units in Setswana dictionaries

When idiomatic collocates are treated as sub-entries in dictionaries, it is important that the type of dictionary should be kept in mind. Normally general dictionaries, which have a more inclusive nature can accommodate more sub-entries than standard or school dictionaries, which, because of their smaller nature, have to exclude many sub-entries. In the case of very economical, restrictive and selective dictionaries, all sub-entries will have to be omitted. When, in the following discussion, we therefore indicate how the sub-entries in some Setswana dictionaries may be increased, it does not necessarily mean that all these sub-entries should be included. It merely shows what are available. When a choice has to be made, which sub-entries have to be included in accordance with a specific type of dictionary, corpus evidence will be helpful to indicate which idiomatic collocates are the most commonly and generally used.

Setswana dictionaries have attempted to include sub-entries based on the idiomaticity of collocates. However, some of these have been few because of a lack of sufficient corpus evidence. Above we have shown that 84 sub-entries for *pelo* could be extracted from a corpus. When the entry *pelo* in Matumo (1993: 306-3007) is referred to, we can see that he lists only 20 sub-entries. Presented below are examples of how the entry *molomo* has been treated in Setswana dictionaries to illustrate the nature and extent of this.

Brown (1925: 210) identifies only two sub-entries *kgwedi ya molomo* and *go cwa molomo*:

Molomo, n., pl. *melomo*, A mouth (outside); a beak of a bird; a foreskin. *Kgwedi ea molomo*, the first month of the Sechuana year; the month of eating first-fruits. *Go cwa molomo*, to open the mouth, in speaking.

Kgasa (1976: 71) does not list any sub-entry for *molomo*. It may be that Kgasa's dictionary, which was aimed at primary schools was simplified for this reason; he might have seen no need to complicate entries with sub-entries:

molomo(me) kgôrô e dijô di yang mo 'ganong ka yônê.

Kgasa and Tsonope (1998: 171) list only a single sub-entry: *molomo o tlola noka e tletse* (a claim is easy to make):

mo•lomo TTT *ln./3.* me-. phatlha e e tswalwang ke dipounama tse pedi e go tsenngwang dijô ka yônê go ya ko mpeng le go bua. ♠ *molomo o tlola noka e tletse* = *motho o kgôna go bua dilô tse di ntsi tse a ka di dirang mme ntswa a se ka ke a kgôna*

While Snyman et al. (1990) do not enter *molomo* in their dictionary at all, Matumo (1993: 260) lists only two sub-entries, *go tswa molomo* and *sejô sennyê ga se fete molomo*:

molomo, N. CL, 3 *mo-*. SING. OF *melomo*, a mouth; lip; a beak of a bird; an opening, as a tube, piping or tunnel; a foreskin. ID. EXPR., *go tswa molomo*, to open the mouth in speaking. PROV., *sejô sennyê ga se fete molomo*.

All the above dictionary treatments of the entry *molomo* are deficient and will benefit greatly from the use of corpus evidence. For instance, the definition from Matumo (1993) may be revised in the following way, ■ being used to mark a sub-entry. This shows how the study of collocations can enrich dictionary entries.

molomo, *n.* 1. mouth 2. a lip 3. a beak 4. an object opening, as that of a bottle ■ **bolwetsi jwa tlhako le molomo**: foot and mouth disease ■ **itoma molomo wa tlase**: be determined ■ **itshwara molomo**: be shocked ■ **ntsha ka molomo**: speak; express an opinion; express a view ■ **pula molomo**: money paid before someone speaks during lobola negotiations ■ **pipa molomo**: a bribe ■ **rwala molomo**: be angry and tight-lipped ■ **roka molomo**: remain quiet ■ **tswa molomo**: speak; say something; contribute; express an opinion ■ **tlhoka molomo**: have nothing to say; be dumbstruck; be rendered speechless ■ **molomo o tlola noka e tletse**: it is easy for someone to claim that they can achieve what they cannot do

We conclude this section by illustrating how dictionary entries for *mpa*, *lonaô* and *matlhô* could be enriched by means of information in Tables 2–5 derived from a corpus. The proposed entries in each case are compared with entries from Matumo (1993).

Matumo (1993: 276):

mpa N. CL, 90-, SING. OF *dimpa*, a belly; a stomach. ID. EXPR. *mpa ya lentswê*, the middle of a hill; *mpa ya lonaô*, the sole of a foot. PROV., *seboba re bata sa mokwatla sa mpa re a mpampetsa*.

Matumo's entry of *mpa* with only three sub-entries may be improved in the following manner with the addition of nine sub-entries:

mpa *n.* a belly; a stomach ■ **bana ba mpa**: relatives ■ **bipa mpa ka mabele**: withhold bad information to protect a relative or friend ■ **gare ga mpa ya bosigo**:

in the middle of the night ■ **gare ga mpa ya lefatshe/naga:** in the middle of nowhere ■ **mpa ya sebetse:** flat on the stomach ■ **mpa e tuka molelo:** with a full stomach ■ **go ja ka mpa tsoopedi:** eat until the stomach is full ■ **ntsha (senya) mpa:** commit abortion ■ **imelwa ke mpa:** have a full stomach

Matumo (1993: 212):

lonaô N. CL. 11 *lo-*, SING OF *dinaô*, a foot. ID EXPR, *go baba lonaô*.

Matumo's entry of *lonaô* with a single sub-entry may be improved with the addition of eleven sub-entries as follows:

lonaô *n.* a foot ■ **apaya ka lonaô:** avoid cooking and eat at other people's homes instead ■ **goga dinaô:** move slowly ■ **fodisa dinaô:** take a rest ■ **motsamaya ka dinaô:** a pedestrian ■ **ngotla dinaô:** walk slower ■ **tlhatlosa dinaô:** walk faster ■ **baya lonaô:** set foot in a place ■ **tsholetsa dinaô:** walk faster ■ **kgwele ya dinaô:** football ■ **tsosa dinaô:** walk faster ■ **tiisa dinaô:** walk faster

Matumo (1993: 232):

matlhô N. CL. 6 *ma-*, PL OF CL. *leithô*; *maitlhô* is still used in a few areas, eyes.

Matumo's entry of *matlhô* which lacks any sub-entry, may be improved by the addition of nine sub-entries:

matlhô *n.* eyes. ■ **bula matlhô:** educate, make aware, enlighten ■ **diga matlhô:** look down ■ **digalase tsa matlhô:** spectacles, sunglasses ■ **latlhêla matlhô:** look briefly ■ **matlhô a phagê a lebane:** face to face ■ **kala matlhô:** confuse ■ **tlodisa matlhô:** overlook someone or something ■ **kgarakgaratsha matlhô:** look from one place to another ■ **tlhatlosa matlhô:** look up

The updating does not only apply to the bilingual dictionaries. Monolingual Setswana dictionaries could be enhanced in a similar manner, as the following example of *tsaya* (take) from Kgasa and Tsonope (1998: 303):

tsaya GT|tseile *tpt.* -ile. 1. tlosa sengwe fa se ntseng se le teng ka go se tsenya mo diatleng tsa gago 2. inêla ka molaô ga monna go tshela le mosadi; nyala ♠ *go tsaya seditse = go dumêlwa ke ba bangwe mo go se o se buileng*

Kgasa and Tsonope's treatment of *tsaya* with a single sub-entry may be revised in the following comprehensive manner with the aid of concordance lines to add 28 sub-entries:

tsaya *ld.* 1. amogela mo diatleng 2. sutisa sengwe fa se neng se le teng 3. tsamaya ka tselana; ya ntlheng nngwe 4. nyala 5. nna le sengwe; tshola ■ **tsaya botshe-lo:** bolaya ■ **tsaya dinopolo:** utswa diphiri ■ **tsaya ditaelo:** sala morago melawana ■ **tsaya ka motlhala:** sala morago ■ **tsaya dipilisi:** metsa dipilisi ■ **tsaya karolo:** nna le seabe ■ **tsaya ka letsogo la molema:** sotla; nyatsa; kgetholola ■ **tsaya kgakololo:** amogela kgakololo **tsaya kgato:** dira sengwe ■ **tsaya lobaka:** go diragala mo nakong e telele ■ **tsaya mongwe/sengwe**

motlhofo: nyatsa ■ **tsaya puso:** simolola go etelela mmuso ■ **tsaya phekelo e sele:** go senyegela pele ■ **tsaya tshwetso:** dira mogopolo ■ **tsaya nako:** iketle ■ **tsaya motlhala:** kopa sengwe se se siameng ■ **tsaya mosadi:** nyala ■ **tsaya mogote:** tlhola selekanyo sa mogote mo mongweng ■ **tsaya matsapa:** dira sengwe mo nakong e telele ■ **tsaya tsia:** tlotla; tlhokomela ■ **tsaya malatsi:** ikhutsa; nna o ye tirong ■ **tsaya malebela:** kopisa sengwe se se ntle ■ **tsaya maikarabelo:** nna wena o tshwereng sengwe ■ **tsaya loeto:** eta ■ **tsaya maemo:** simolola maemo ■ **tsaya setshwantsho:** dirisa khamera go tshwantsha ■ **tsaya sekgele:** fenyha ■ **tsaya sebaka:** sengwe sa nako e telele

5. Conclusion

In this article, we have attempted to illustrate what could be achieved by a simple study of concordance lines to extract MWEs for the significant improvement of dictionary entries. Considering only single words as candidates for dictionary entry impoverishes a dictionary and betrays a rudimentary understanding of what constitutes a word in language. If Jackendoff's estimate that the number of MWEs in a speaker's lexicon is of the same order of magnitude as the number of single words is accurate, then MWEs in African languages deserve intensive study, which they have hitherto not received. To generate concordance lines is inexpensive, and free concordance programs are available online to aid researchers explore the complexity of texts. Dictionaries of African languages would therefore benefit greatly from populating sub-entries with MWEs harvested from concordance lines.

References

- Aitchison, J. 1992. *Teach Yourself Linguistics*. London: Hodder & Stoughton.
- Bannard, C. 2007. A Measure of Syntactic Flexibility for Automatically Identifying Multi-word Expressions in Corpora. *Proceedings of the ACL Workshop on a Broader Perspective on Multi-word Expressions, Prague, Czech Republic, June 2007*: 1-8.
- Biber, D., S. Conrad and R. Reppen. 1998. *Corpus Linguistics: Investigating Language Structure and Usage*. Cambridge: Cambridge University Press.
- Brown, T.J. 1925. *English Dictionary*. Johannesburg: Pula Press.
- Brunner, A. and K. Steyner. 2008. *Corpus-Driven Study of Multi-Word Expressions Based on Collocations from a Very Large Corpus*. Paper presented at the Fourth Inter-Varietal Applied Corpus Studies (IVACS) Conference, University of Limerick, Ireland, 13-14 June 2008.
- Dash, N.S. and B.B. Chaudhuri. 2000. The Process of Designing a Multidisciplinary Monolingual Sample Corpus. *International Journal of Corpus Linguistics* 5(2): 179-197.
- Fazly, A. and S. Stevenson. 2007. Distinguishing Subtypes of Multiword Expressions Using Linguistically-motivated Statistical Measures. *Proceedings of the ACL Workshop on a Broader Perspective on Multi-word Expressions, Prague, Czech Republic, June 2007*: 9-16.
- Finch, G. 2000. *Linguistic Terms and Concepts*. Basingstoke: Macmillan Press.
- Jackendoff, R. 1997. *The Architecture of the Language Faculty*. Cambridge, MA: MIT Press.

- Kgasa, M.L.A.** 1976. *Thanodi ya Setswana ya Dikole*. Cape Town: Longman.
- Kgasa, M.L.A. and J. Tsonope.** 1998. *Thanodi ya Setswana*. Gaborone: Longman.
- Leech, G., M. Deuchar and R. Hoogenraad.** 1982. *English Grammar for Today: A New Introduction*. Basingstoke: Macmillan Press.
- Matumo, Z.I.** 1993. *Setswana–English–Setswana Dictionary*. Gaborone: Macmillan.
- McArthur, T.** 1998. *Living Words: Language, Lexicography and the Knowledge Revolution*. Exeter: University of Exeter.
- Moon, R.** 1998. *Fixed Expressions and Idioms in English: A Corpus-based Approach*. Oxford: Oxford University Press.
- Oflazer, K. and Ö. Çetinoğlu.** 2004. Integrating Morphology with Multi-word Expression Processing in Turkish. *Second ACL Workshop on Multi-word Expressions: Integrating Processing, Barcelona, Spain, July 2004*: 64-71.
- Otlogetswe, T.J.** 2007. *Corpus Design for Setswana Lexicography*. Unpublished Ph.D. Thesis. Pretoria: University of Pretoria.
- Pearsall, J.** 1998. *The New Oxford Dictionary of English*. Oxford: Oxford University Press.
- Renouf, A.** 1987. Corpus Development. Sinclair, J.M. (Ed.). 1987. *Looking Up. An Account of the COBUILD Project in Lexical Computing and the Development of the Collins COBUILD English Language Dictionary*. London/Glasgow: COBUILD.
- Sag, I.A., T. Baldwin, F. Bond, A. Copestake and D. Flickinger.** 2002. Multiword Expressions: A Pain in the Neck for NLP. *Proceedings of the Third International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-2002), Mexico City, Mexico, March 2002*: 1-15.
- Schone, P. and D. Jurafsky.** 2001. Is Knowledge-free Induction of Multi-word Unit Dictionary Headwords a Solved Problem? *Proceedings of the 2001 Conference on Empirical Methods in Natural Language Processing, Pittsburgh, PA*: 100-108.
- Scott, M.** 2004-2006. *Oxford WordSmith Tools Version 4*. Oxford: Oxford University Press.
- Sharoff, S.** 2004. What is at Stake: A Case Study of Russian Expressions Starting with a Preposition. *Second ACL Workshop on Multi-word Expressions: Integrating Processing, Barcelona, Spain, July 2004*: 17-23.
- Snyman, J.W., J.S. Shole and J.C. le Roux.** 1990. *Dikisinare ya Setswana–English–Afrikaans Dictionary/Woordeboek*. Pretoria: Via Afrika.
- Villavicencio, A., A. Copestake, B. Waldron and F. Lambeau.** 2004. Lexical Encoding of MWEs. *Second ACL Workshop on Multi-word Expressions: Integrating Processing, Barcelona, Spain, July 2004*: 80-87.

An Analysis of *Practical Lexicography: A Reader* (Ed. Fontenelle 2008)

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

Abstract: Intended as a companion volume to *The Oxford Guide to Practical Lexicography* (Atkins and Rundell 2008), Fontenelle's book aims to bring together the most relevant papers in practical lexicography. This review article presents a critical analysis of the success thereof, both in quantitative and qualitative terms.

Keywords: PRACTICAL LEXICOGRAPHY, DICTIONARY, MONOLINGUAL, BILINGUAL, QUANTITATIVE EVALUATION, QUALITATIVE EVALUATION, CITATION PATTERNS, CORPUS TOOLS, MEANING, DEFINITIONS, EXAMPLES, EQUIVALENCE, DICTIONARY USE, ENGLISH, FRENCH

Samenvatting: Een analyse van *Practical Lexicography: A Reader* (Fontenelle 2008). Bedoeld als achtergrondlektuur bij *The Oxford Guide to Practical Lexicography* (Atkins en Rundell 2008), is het doel van Fontenelle's boek om de meest relevante papers m.b.t. praktische lexicografie samen te brengen. Dit recensieartikel is een kritische analyse van het succes daarvan, zowel in kwantitatieve als kwalitatieve termen.

Sleutelwoorden: PRAKTISCHE LEXICOGRAFIE, WOORDENBOEK, EENTALIG, TWEETALIG, KWANTITATIEVE EVALUATIE, KWALITATIEVE EVALUATIE, CITATIEPATRONEN, CORPUS SOFTWARE, BETEKENIS, DEFINITIE, VOORBEELDEN, EQUIVALENTIE, WOORDENBOEKGEBRUIK, ENGELS, FRANS

1. The Editor and the Reader

Think of Microsoft and computational lexicography, and a couple of names spring to mind. Ken Church is one of them, Thierry Fontenelle the other. The first will return below, the second is the editor of the book under review. In addition to his work as a Senior Program Manager with Microsoft's Natural Language Group (where he was responsible for the French lexical database used in a variety of natural language applications), Fontenelle is well known as an Associate Editor of the *International Journal of Lexicography* (IJL), as a Past President of Euralex, for the projects he led at the European Commission Translation Service, for his innovative research into collocations and semantic

networks, and for his contribution to bilingual dictionaries published by CUP. Fontenelle is currently the Head of the General Affairs Department at the Translation Centre for the Bodies of the European Union.

Invited by Sue Atkins to compile a companion volume to *The Oxford Guide to Practical Lexicography* (Atkins and Rundell 2008), henceforth OGPL, he proceeded with such speed that his 'accompanying' collection, *Practical Lexicography: A Reader* (Fontenelle 2008), henceforth PLR, was completed and published before OGPL. We opted to analyse these two works in their logical order, however, with a review of OGPL appearing in last year's *Lexikos* (De Schryver 2008), and a review of PLR herewith.

PLR is a collection of 23 papers. All of these papers have been published elsewhere, with the exception of the Introduction by Fontenelle, which provides an insightful overview of the whole topic, as well as a brief summary of each paper. Some of the papers are not easily accessible elsewhere and, in the opinion of this reviewer, well over half of them are required reading for any student with an interest in words, meanings, and dictionaries. The references for each paper have all been brought together into one single section at the end of the book,¹ and a combined author/subject index concludes PLR.

The core facts about each of the 22 papers as well as the Introduction are enumerated in the Addendum, to which reference is made for the details of various claims below. As may be seen from the Addendum, there are, altogether, 25 authors for the 23 contributions: Atkins and Kilgarriff have each authored or co-authored three; while Fontenelle, Grefenstette, Hanks, and Rundell are each author or co-author of two. The number of pages varies widely, from 4 pages (for Bolinger) to 39 pages (for Atkins and Varantola); the average being 14.8 pages. Following the Introduction, the 22 papers have been grouped into twelve parts, with either one, two or three papers per group.

2. Reader statistics

This is not the first reader in lexicography/lexicology. Two of the better known ones are Hartmann's (2003) *Lexicography: Critical Concepts* (reviewed in De Schryver 2005), and Hanks's (2008) *Lexicology: Critical Concepts* (reviewed in De Schryver 2008a). Both are actually multi-volume anthologies (respectively three and six volumes), in contrast to the single-volume PLR. Their aim is also different: While both anthologies intend to provide an overview of the whole field (of lexicography/lexicology), the current reader sets out to support a specific textbook (OGPL). Cross-comparing various statistics for these three collections will nonetheless prove fruitful.

A first aspect of interest is to look at the dates of publication for each of the 23 texts in PLR. This is done in Figure 1, from which one sees that Fontenelle especially selected texts produced during the first half of the 1990s, and to a lesser extent from the following decade. In comparison, in Hartmann's lexicography collection, "an increasing number of influential texts were seemingly

written during the 1980s" (De Schryver 2005: 94). For Fontenelle, then, the lexicographic contributions that best support OGPL are written about a decade later. Compared to Hanks's anthology, which shows that lexicology has "continued to pick up momentum ever since [the 1950s]" (De Schryver 2008a: 421), an inverse trend may be noticed in the papers from the past two decades in PLR.

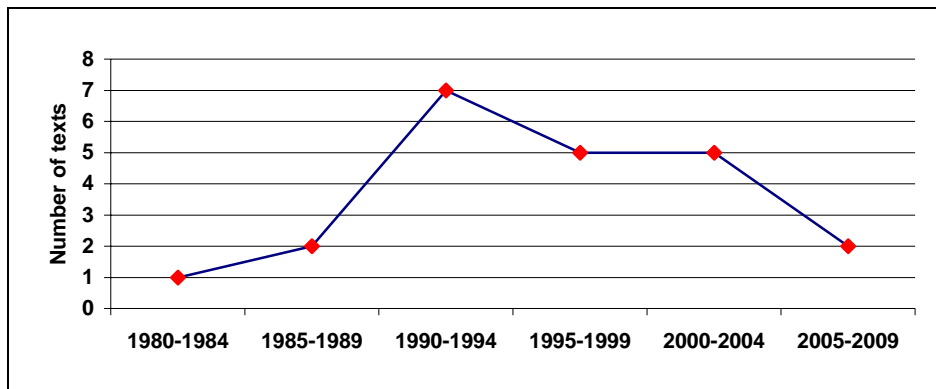


Figure 1: Distribution across time of the texts in *Practical Lexicography: A Reader* [not showing one mid-eighteenth-century text]

A second aspect of interest is a study of the sources of each of the texts in PLR. The distribution is shown in Figure 2, from which one may see that nearly 40% of the texts were initially published in journals (four, or nearly half, of them in IJL), and about 35% in conference proceedings (with as many as six, or three quarters, in Euralex proceedings). If anything, the large amount of material from IJL and the Euralex proceedings immediately delineates the field of study as Anglo-Saxon. In this it indeed supports OGPL, which "is fully embedded into the English and European cultural world" (De Schryver 2008: 431).

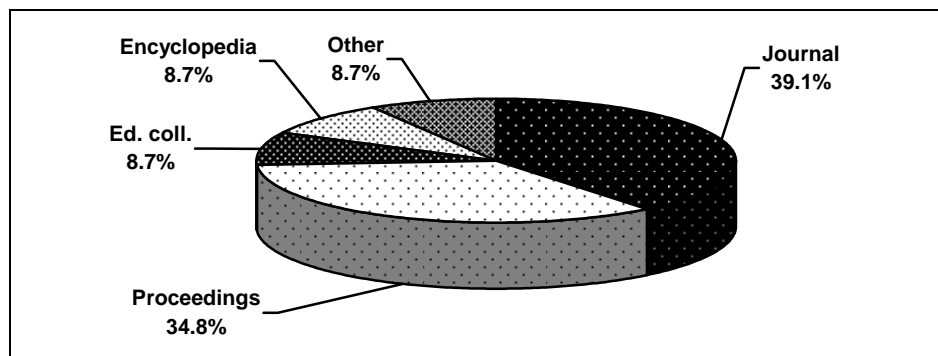


Figure 2: Sources of the texts in *Practical Lexicography: A Reader*

Comparing the distribution of the texts in PLR with the distributions of the texts in the two anthologies, one obtains the data listed in Table 1.

Table 1: Source distribution in three lexicography/lexicology collections

	PLR (Fontenelle 2008)		Lexicology: CC (Hanks 2008)		Lexicography: CC (Hartmann 2003)	
	N	%	N	%	N	%
Journal	9	39.1	41	41.0	12	17.1
Proceedings	8	34.8	15	15.0	17	24.3
Edited collection	2	8.7	26	26.0	12	17.1
Dictionary / Encyclopedia	2	8.7	3	3.0	1	1.4
Book	–	–	13	13.0	18	25.7
Textbook	–	–	–	–	4	5.7
PhD	–	–	–	–	1	1.4
Other	2	8.7	2	2.0	5	7.1
	23	100.0	100	100.0	70	100.0

Clearly PLR's distribution is closer to Hanks's than to Hartmann's, with the book category entirely absent (where it was still 26% in Hartmann, and half of that, 13%, in Hanks). With an increased focus on journals and conference proceedings, rather than on books (and textbooks), this points to a lively and healthy research environment in practical lexicography.

A third aspect of interest is the answer to the question: "At what age does one write material worthy of inclusion in a reader of practical lexicography?" That age seems to be 49, compared to 47 in the lexicology anthology, and 51 in the lexicography anthology. The PLR average thus falls right in-between those of general lexicology and general lexicography — in line with what one would have expected, given the relatively high percentage of computational lexicographers (who tend to be younger than their colleague-lexicographers) in PLR.

Lastly, one can also compare the three collections with reference to the specific texts selected, as well as their contributing authors. Here the differences in overlap are striking. While only one text (Johnson 1747) and four authors (Cowie, Hanks, Johnson, and Varantola) are shared between Hartmann's collection and PLR, two texts (Church and Hanks 1989, and Kilgarriff et al. 2004) and as many as thirteen authors (Apresjan, Atkins, Bolinger, Church, Fellbaum, Fillmore, Fontenelle, Hanks, Kilgarriff, Miller G.A., Rychlý, Smrž, and Tugwell) are shared between Hanks's collection and PLR. With more than half of the PLR authors also represented in the lexicology collection, one is tempted to accord more value to the selections in these two, which depict a more unified field, rather than to the more esoteric selection found in the lexicography collection, which shows "a strong bias towards especially Asian authors and colleagues working in Asia" (De Schryver 2005: 94).

3. Quantitative evaluation

This section presents a quantitative evaluation of PLR. The evaluation proceeds in two steps. Firstly, one can assume that the main purpose of a reader is not to present esoteric texts that the editor may happen to know; rather, the true core texts of a field ought to be brought together. From this it follows that quite a number of these texts should have attracted a substantial number of citations over the years. Bibliometrics, especially in lexicography, is still in its infancy (cf. De Schryver 2009a), but one rather relevant and freely accessible source in this regard is Google Scholar. With it, the number of citations for each of the 22 texts in PLR may be checked — the results of which are shown in Figure 3.

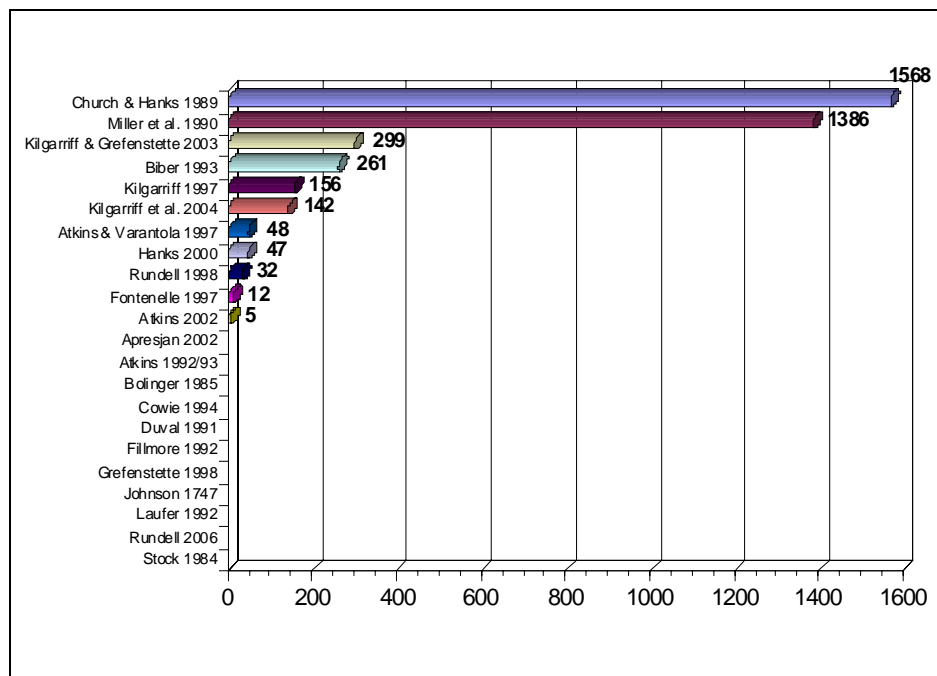


Figure 3: Number of citations for the texts in *Practical Lexicography: A Reader* (according to Google Scholar statistics on 5 August 2009)

As may be derived from Figure 3 (and thus according to the Google-crawlable material only!) half of PLR's contributions have indeed been cited (11 out of 22). The other half, however, has not, or at least much less (knowing that Google does not (yet) see 'everything').² Furthermore, the top six most-frequently cited are unsurprisingly contributions in the sub-field of computational lexicography — 'unsurprisingly', as the NLP community is (a) much larger than the lexicographic community, (b) prominently present on the Web, and (c) characterized by a much greater speed in communicating (and citing) research results. This

leaves us with just five truly (traditional) lexicographic papers that have been cited: Atkins and Varantola (1997), Hanks (2000), Rundell (1998), Fontenelle (1997), and Atkins (2002). If one merely looks at the author names, these papers from PLR simply *must* support the arguments in OGPL; although, of course, one cannot help but notice some circularity in the undertaking.

For the second step in the quantitative evaluation, one can take one step back to OGPL. In OGPL, each of the twelve chapters is concluded with a reading section, divided into 'Recommended reading' (R), 'Further reading on related topics' (F), and 'Websites' (W). One could assume that each recommended source would at least have been mentioned once in the chapter to which it belongs, but this is not the case, as may be seen from the statistics presented in Table 2.

Table 2: Referred vs. non-referred reading material in OGPL (per chapter, 'Ch.')

Ch.	Recommended				Further reading				Websites			
	Referred		Not ref.		Referred		Not ref.		Referred		Not ref.	
	N	%	N	%	N	%	N	%	N	%	N	%
1	2	40.0	3	60.0	1	25.0	3	75.0	5	71.4	2	28.6
2	1	20.0	4	80.0	5	23.8	16	76.2	0	0.0	1	100.0
3	6	75.0	2	25.0	10	55.6	8	44.4	1	11.1	8	88.9
4	2	66.7	1	33.3	1	14.3	6	85.7	0	0.0	6	100.0
5	6	66.7	3	33.3	12	32.4	25	67.6	1	25.0	3	75.0
6	2	50.0	2	50.0	2	6.9	27	93.1	0	0.0	1	100.0
7	0	0.0	5	100.0	0	0.0	51	100.0	0	0.0	2	100.0
8	6	50.0	6	50.0	17	35.4	31	64.6	–	–	–	–
9	0	0.0	5	100.0	7	10.0	63	90.0	–	–	–	–
10	7	87.5	1	12.5	17	23.6	55	76.4	–	–	–	–
11	0	0.0	4	100.0	0	0.0	19	100.0	1	50.0	1	50.0
12	0	0.0	4	100.0	0	0.0	37	100.0	–	–	–	–
	32	44.4	40	55.6	72	17.4	341	82.6	8	25.0	24	75.0

Of the five sources recommended at the end of Chapter 2, for example, only one was actually referred to in Chapter 2. Overall, there are 72 recommended references (including doubles), 44.4% of which have been referred to in the respective chapters. The number of sources referred to from the further-reading section is even lower: 17.4% overall (for a total of 413 references); with the website section falling in-between: 25.0% referred (for a total of 32 references).

While this mismatch was obviously a design feature by the authors of OGPL, one in which they departed from the academic practice of providing all relevant references throughout the text (cf. De Schryver 2008: 434-435), this leaves the present reviewer with the task to find an alternative way to check whether the texts in PLR are a good representation of the material referred to in OGPL. The only way to do so remains to check *all* references in OGPL, throughout the *entire* text. While reading through OGPL, a citation database was kept, recording not only the data summarized in Table 2, but also the

occurrence of each and every citation (C). Table 3 lists those references that were mentioned at least five times overall.

Table 3: Top section of the citation database built for OGPL, contrasted with the texts and authors in PLR

#	Author(s)	Year	R	F	C	Sum	Text in PLR?	Author(s) in PLR?
1	<i>Landau</i>	2001	4	2	6	12	–	–
2	Atkins	1992/93	3	2	1	6	✓	✓
3	Atkins, Rundell & Sato	2003	2	2	2	6	–	✓✓–
4	Apresjan	1973	2	1	3	6	–	✓
5	<i>Cruse</i>	2004	2	15	17	17	–	–
6	Fillmore & Atkins	2000	1	3	1	5	–	✓✓
7	<i>Cruse</i>	1986	1	2	9	12	–	–
8	Geeraerts	1990	1	2	4	7	–	–
9	Bolinger	1965	1	1	4	6	–	✓
10	Hanks	1987	1	1	4	6	–	✓
11	<i>Aitchison</i>	2003	1	1	3	5	–	–
12	Johnson	1755	1	12	13	13	–	✓
13	Johnson	1747	1	7	8	8	✓	✓
14	Sinclair	2003	1	6	7	7	–	–
15	Kilgarriff, Rundell & Dhonnchadla	2007	1	5	6	6	–	✓✓–
16	Moon	1987a	1	5	6	6	–	–
17	Stock	1984	1	5	6	6	✓	✓
18	Atkins, Clear & Ostler	1992	1	4	5	5	–	✓--
19	Bogaards	1990	4	1	5	5	–	–
20	<i>Hoey</i>	2005	3	7	10	10	–	–
21	Cowie	1999	3	2	5	5	–	✓
22	<i>Lakoff & Johnson</i>	1980	2	3	5	5	–	--
23	Rundell	1998	2	3	5	5	✓	✓
24	<i>Taylor</i>	1995	2	3	5	5	–	–
25	Moon	1987b	1	4	5	5	–	–
26	Fontenelle	1996	5	5	5	5	–	✓
27	Hanks	2004	5	5	5	5	–	✓

When relative weight ($R > F > C$) is combined with number of citations, the most important source referred to in OGPL is (the second edition of) Landau's well-known textbook, with 12 references overall (4 x R, 2 x F, and 6 x C). This is followed by Atkins's (1992/93) DSN paper (3 x R, 2 x F and 1 x C). Of course, Fontenelle could not have been expected to include entire textbooks or general linguistic monographs in his reader,³ which immediately excludes numbers 1, 5, 7, 11, 20, 22 and 24 (all displayed in italics) from Table 3. Of the 20 remaining sources, four have been included in PLR (these are highlighted for ease of reference). While just four is not a very high number, the author representation is better: 11 are in (Atkins, Apresjan, Bolinger, Cowie, Fillmore, Fontenelle, Hanks, Johnson, Kilgarriff, Rundell, and Stock), versus 14 that are out (Aitchi-

son, Bogaards, Clear, Cruse, Dhonnchadla, Geeraerts, Hoey, Lakoff, Landau, Moon, Ostler, Sato, Sinclair, and Taylor).

One may now also proceed to analyse the reverse, namely, how does Fontenelle's selection compare to the citations in OGPL? The result of this comparison is shown in Table 4.

Table 4: Texts in PLR, contrasted with the citations in OGPL

#	Author(s)	Year	R	F	C	Sum
1	Atkins	1992/93	3	2	1	6
2	Johnson	1747	1		7	8
3	Stock	1984	1		5	6
4	Rundell	1998		2	3	5
5	Apresjan	2002	2		1	3
6	Cowie	1994	1	2	1	4
7	Atkins & Varantola	1997	1	2		3
8	Biber	1993	1	2		3
9	Fillmore	1992	1	2		3
10	Kilgarriff	1997	1	1		2
11	Hanks	2000	1		3	4
12	Kilgarriff & Grefenstette	2003	1		2	3
13	Bolinger	1985	1		1	2
14	Grefenstette	1998	1		1	2
15	Kilgarriff, Rychlý, Smrž & Tugwell	2004	1		1	2
16	Atkins	2002	1			1
17	Duval	1991		2		2
18	Rundell	2006		1	3	4
19	Laufer	1992		1	1	2
20	Miller GA, Beckwith, Fellbaum, Gross & Miller KJ	1990		1	1	2
21	Church & Hanks	1989		1		1
22	Fontenelle	1997		1		1

The following observations can be made. Firstly, each of the texts from PLR is mentioned at least once in OGPL. Secondly, all texts from PLR have either been recommended (R) and/or suggested for further reading (F) in OGPL. Thirdly, three texts from PLR are only listed once (Atkins 2002, Church and Hanks 1989, and Fontenelle 1997), which is meagre indeed. (After all, within the roughly fifty other items from the recommended lists, there are many with far more mentions in OGPL.) Including Atkins (2002) was not a good idea: A full rework and update of her paper was included as part of OGPL itself (as §9.2.5.5 in there). In the light of the citation patterns seen in Figure 3, it is truly astonishing that the paper by Church and Hanks (1989) only gets one mention in OGPL, so it is good news that it was 'rescued' for PLR. When Table 3 is compared with 4, it is also clear that item 26 in Table 3 would have been a better choice than Fontenelle (1997), but then it is the prerogative of the editor of a reader to select what they want to be remembered for.

All in all, then, the various aspects of the quantitative evaluation presented in this section indicate that about half the selection is warranted, both in terms of citation patterns and from the point of view of its aim to be a companion volume to the textbook it is to support.

4. Qualitative evaluation

With the quantitative evaluation behind us, we are now in a position to present a qualitative evaluation. While each of the 23 contributions will be given attention, no attempt was made at balancing the length of each evaluation. This is not because of the varying length of the papers themselves (some of the shorter ones actually contain the most insightful thoughts), but simply because not all papers warrant the same attention from the point of view of being included in a reader on practical lexicography.

4.1 Fontenelle's *Introduction* (2008)

In his *Introduction*, Fontenelle both explains the genesis of his book and provides succinct summaries of each of the 22 texts he selected, successfully contextualizing each of them in the process, and providing additional references for further reading. Re-summarizing this part will only serve to perpetuate summaries of summaries, which cannot be the idea. Rather, two observations will be made. Firstly, Fontenelle claims that the structure of PLR reflects the structure of OGPL (p. 2).⁴ While it is true that both OGPL and PLR contain the same number of chapters, twelve, that is where the similarity in structure ends, as there is hardly any correlation between the two series of twelve. What Fontenelle probably meant to say is that the *topics* selected for PLR parallel some of the *topics* developed further in OGPL.

Secondly, Fontenelle claims that, with reference to excellent papers in lexicography, "it must be acknowledged that many of these papers are often published in hard-to-access conference proceedings", and points out that his collection "attempts to meet the need for a coherent and easily accessible compilation of papers" (p. 2). It is not immediately clear whether or not the Euralex proceedings are really so hard to find, but what is verifiable is to check what can already be found in digital form. From the details listed in the Addendum, one sees that as many as 12 of the 22 texts are currently available online: 6 on the Internet, 1 through Google Books, and another 5 by subscription. Given that the Euralex proceedings are soon "to be made available free of charge through the Euralex website under a Creative Common[s] Licence agreement" (Bogaards 2009: 354), another 5 papers will easily become accessible at that point, bringing the total to 17 out of 22. As was pointed out in reviewing Hanks's lexicology collection (De Schryver 2008: 429-430), the future of readers and anthologies is online.⁵

4.2 Johnson's *Plan of a Dictionary of the English Language* (1747)

Fontenelle's collection, quite rightly, starts with Samuel Johnson's Plan. It is the first of three texts in Part I, *Metalexigraphy, Macrostructure, Microstructure, and the Contribution of Linguistic Theory*. Even though Johnson's Plan was written over 260 years ago, Johnson remains the only major literary figure in any language ever to have undertaken the burdensome task of compiling a dictionary — which he did alongside a prodigious outpouring of journalism, reviews, literary criticism, short biographies, and poetry. Hanks (2005: 265) remarks that, in the Preface to his dictionary (1755), "Johnson crisply addresses theoretical issues which were subsequently neglected for some two hundred years." Fontenelle has selected Johnson's less well-known "Plan of a Dictionary of the English Language", written eight years before the dictionary was published. A comparison of the two documents reveals some fascinating differences. In the Plan, Johnson announces that his chief intent is "to preserve the purity and ascertain the meaning of our English idiom" (p. 20). By the time the dictionary was published, Johnson had realized that natural languages cannot be fixed, but are governed by irresistible forces of change — an insight that to this day eludes certain conservative members of the academies of France, Italy, and Spain, as well as many other people who dislike linguistic change.

According to Johnson, a dictionary should provide guidance on at least orthography, pronunciation (he gives examples such as *tear* rhyming with *dare* vs. *tear* rhyming with *peer*), inflections, and etymology (a science which was in its infancy in Johnson's day), as well as the meaning of words. He has interesting comments on analogy ("speech was not formed by an analogy sent from heaven. It did not descend to us in a state of uniformity and perfection, but was produced by necessity and enlarged by accident" (p. 24)), syntax, phraseology, and several other matters. The Plan announces an intention that the dictionary should 'instruct the learner' as well as 'delight the critic' (p. 20). Both these objectives were to be amply fulfilled. The following quotation provides helpful guidance to metalexigraphers, even a quarter of a millennium after it was written: "The unlearned much oftener consult their dictionaries, for the meaning of words, than for their structures or formations" (p. 20).

Johnson's Plan recognizes the problem of coverage of technical terms, but is rather ambivalent about how many should be included. In the event, the dictionary was to include many technical terms from 18th-century science, sometimes with a quotation from an authoritative source in place of a definition. Collectively, these terms provide an amazing insight into how much science has changed in 250 years: 18th-century scientific terms and explanations of terms are very often incomprehensible to modern readers without special training.

Concerning the order of senses of polysemous words, Johnson was well aware of the tendency of words to develop new senses through analogical or

metaphorical change. He offers (p. 26) the first known account of what came to be known as 'historical principles'. For example for *arrive*:

- (1) to reach the shore in a voyage (the "natural and primitive signification" [because *riva* in Latin means 'shore']).
- (2) to reach any place whether by land or sea (the "consequential meaning").
- (3) to obtain any thing desired, as in: he *arrived* at a peerage (a "metaphorical sense").

Furthermore, Johnson anticipates Sinclairian notions of phraseology and semantic prosody: "the word *arrive* [...] cannot be properly applied but to words signifying something desirable; thus we say, a man *arrived* at happiness, but cannot say, without a mixture of irony, he *arrived* at misery" (p. 26).⁶

4.3 Atkins's *Theoretical Lexicography and its Relation to Dictionary-Making* (1992/93)

Much of what is said in this article is covered (often in a revised and improved form) in OGPL, so including it here may seem like unnecessary duplication of effort. Nevertheless, it does make a valuable contribution, especially for readers who only read PLR and not OGPL.

Atkins contrasts the single-volume "trade dictionary, a product created to be sold in the marketplace" (p. 31), with scholarly historical dictionaries. According to Atkins (p. 33), there are two steps to dictionary-making: analysis and synthesis. They may even be carried out by different groups of people. Analysis consists of "trying to discover as many relevant linguistic facts [about each word] as possible" (p. 33), by means of studying a large corpus and/or a collection of citations and/or pre-existing dictionaries and/or the lexicographer's own intuitions. Fortunately, the days in which lexicographers had to rely on their intuitions for evidence of linguistic facts about a word are mostly behind us, although for some languages (especially the undocumented ones), evidence based on the lexicographer's own intuitions still plays a large part. Gradually, large corpora of texts are being built for all languages, although in some cases — such as American English (see e.g. Hanks 2009) — dictionary publishers have not yet realized the need for corpus evidence. It should be borne in mind that publishers do not really care what the dictionary entries say, or whether what is said is true or false, as long as deadlines are met and the product meets its sales targets.

In the synthesis stage, the lexicographers write up the entries. Dictionaries are sometimes written by large teams, rather than by a single author, so a style guide, drawn up by senior editors, is needed to ensure consistency of treatment.

Atkins might have added a third stage: refinement. The wording of definitions or translations, as well as the choice of examples, is checked in dic-

tionaries far more often than in most other kinds of literary product, and changes — sometimes quite radical changes — are introduced at quite a late stage. The success of a dictionary in the market place may be quite radically affected by changes at the refinement stage, for the initial rough-hewn synthesis may not be clear or convincing to a general reader.

Atkins says (p. 34): "One clear, albeit indirect, contribution that the theoretical linguist may make to the synthesis process is to give the would-be lexicographer language skills and language awareness." However, as Atkins herself goes on to show, the criticisms levelled by academic linguists at dictionaries are often irrelevant and in other cases take no account of practical constraints such as space or the needs of users. Twenty years on, it is now clear that only some linguists can give the would-be lexicographer language skills and language awareness. Others merely confuse the issue and miss the point. Linguists such as Apresjan, Fillmore, and Halliday certainly have useful things to say to lexicographers, but others do not.

In a personal communication, Patrick Hanks mentions that, during the heyday of Chomskyan syntactocentrism (in the 1960s, 70s, and 80s), he regarded a degree in linguistics as an active disadvantage for a would-be lexicographer. Graduates with degrees in botany, classics, or literature made better lexicographers, for they had less to unlearn and more skill in discrimination. We may conclude from such attitudes by a well-known project leader in lexicography that even today would-be lexicographers need guidance on what sort of linguistics might improve their relevant language skills. On p. 45 Atkins seems to acknowledge this, for, after alleging that "lexicography will improve if more lexicographers read theoretical papers", she goes on to say: "the most helpful and common-sense papers on defining that I know are those written by practicing and practical lexicographers." PLR provides precisely the sort of selection of articles that Atkins in 1992/93 was pleading for.

On pp. 43-48 Atkins distinguishes 'internal facts' (facts about a word's spelling, inflections, and meaning) from 'external facts': "the editor's greatest problems with external facts lie in SYNTAGMATIC relationships" (p. 47). The truth of this important comment is now widely recognized and the problems that Atkins alludes to are beginning to be addressed in corpus-based syntagmatic research.

4.4 Apresjan's *Principles of Systematic Lexicography* (2002)

The Russian tradition in linguistics (Ščerba, Apresjan, Mel'čuk, amongst others) has always enjoyed an integral relationship between the study of lexis, pragmatics, semantics, syntax, and other aspects of language. It has never been bedevilled by a division between 'commercial dictionaries' and 'academic research'. Instead, dictionaries and research have gone hand in hand. A prime example is Apresjan, who is both a leading academician and practical dictionary maker. He was working without the benefit of corpus evidence, so he

sometimes got the details wrong, but his five general principles, outlined in this contribution, repay study by every would-be lexicographer. In brief, they may be paraphrased as follows.⁷

According to the first principle each language 'forces' its speakers to express specific meanings, such as in certain Russian sentences where one is forced to specify the manner of locomotion through verbs like *walking*, *flying* or *crawling*, as the use of the (more) general *leaving* results in doubtful constructions. The second principle insists on a perfect 'coordination' of dictionaries and grammars. The current discrepancy is convincingly exemplified with a discussion of the labelling of numerals as either 'nouns' or 'adjectives' in English dictionaries vs. their characterization as 'numerals' in their own right in grammars. In the third principle it is advocated that 'lexical classes' ought to be treated in full and described uniformly in dictionaries. Under the heading 'lexicographic types', factive and putative predicates are looked into as an example, leading to highly interesting oppositions (in English) such as 'knowledge has a source, but not a reason' vs. 'opinions have a reason, but never a source'. The converse verbs *buy*, *sell*, *pay* and *cost* are discussed as an instance of a 'lexical-semantic paradigm' under a second heading. Here, Apresjan anticipates Fillmore's theory of Frame Semantics. Moving from the macro- to the microcosm, the fourth principle stresses the importance of an exhaustive linguistic description of lexemes, while the fifth points to the need to pay attention to meaning interaction across language units.

Apresjan's text is a highly entertaining one, and the five principles of systematic lexicography that are defined and illustrated are indeed of great importance to practical dictionary making.

4.5 Biber's *Representativeness in Corpus Design* (1993)

Biber's article is the first of two that make up Part II, *On Corpus Design*. This text has attracted a considerable number of citations (cf. Figure 3), mainly from the computational linguistics community, it has been referred to from OGPL (cf. Table 4), and it is useful for corpus builders, *but* it is only indirectly relevant to lexicography, and in fact irrelevant to practical lexicography. While lexicographers of course need to know that the corpus evidence they are using contains a reasonably wide variety of different texts and text types, they do not need the sort of details on sampling procedures and corpus design that are to be found in this article.

The conclusion, too, is somewhat of a downturn: "the parameters of a fully representative corpus cannot be determined at the outset. Rather, corpus work proceeds in a cyclical fashion" (p. 86). Given the ease with which corpora are built these days, and especially the ease with which any number of sub-corpora may be combined and recombined at any point in time, Biber's text offers less than is intimated.

4.6 Kilgarriff and Grefenstette's *Introduction to the Special Issue on the Web as Corpus* (2003)

Kilgarriff and Grefenstette start their article with the important question "Is the Web a corpus?" Their answer is "Yes." Our answer is "Not really." The Web is a vast collection of texts, constantly growing and changing. A corpus is also a collection of texts, but it is a collection with a purpose. Insofar as that purpose is linguistic analysis, a corpus must be stable, for corpus analysis can only be effective if it compares like with like and is able to measure linguistic facts statistically. Thus, the Web is not really a corpus, the reason being that it is unstable. It is a resource from which many corpora can be built (i.e. the Web 'for' corpus building rather than the Web 'as' a corpus, a distinction introduced in De Schryver 2002) — just as it is a resource for all sorts of other purposes. For example, by using the Web, armchair linguists have the best chances of finding authentic evidence to support their pet theories, especially if they are too bizarre to be reflected in a small sample and therefore are not part of conventional usage. In contrast to this, the advantage of a real corpus — a stable corpus — is that it can be analysed statistically, enabling lexicographers and linguists to distinguish normal usage from abnormal (even if authentic) usage, or as Hanks (2008a: 228) put it: "authenticity alone is not enough: evidence of conventionality is also needed."

Kilgarriff and Grefenstette then present a very idiosyncratic account of the history of corpora, point out that the BNC is not big enough (no corpus linguist would disagree; but then, no one is arguing a case for reducing the size of corpora), with the subsequent sections of little relevance to practical lexicography, and in places rather tedious. So, this article (after all, merely an introduction to a special issue of the journal *Computational Linguistics*) is really disappointing, at least as far as lexicography is concerned. Here is why. Lexicography is concerned, among other things, with discovering and representing the conventional, normal use and meaning of each word in a language. The Web as corpus is a source of masses of evidence for word use of all kinds, but it is not a reliable source of evidence for conventional word use or significant collocations. It is therefore surprising that this paper has been included in a collection claiming to provide readings in practical lexicography.

In the opinion of the present reviewer, then, neither of the papers in Part II belong in this reader. They are not really relevant to lexicography, and they do not give a good overview of the *relevance* of corpus data to lexicography. In that sense, one could have hoped that the next contribution (cf. Section 4.7) would have obviated this problem, but unfortunately, it does not.

4.7 Fillmore's '*Corpus Linguistics*' or '*Computer-aided Armchair Linguistics*' (1992)

Part III, *On Lexicographical Evidence*, consists of just one paper, Fillmore's. The paper opens with an amusing caricature of the mutual incomprehension be-

tween theoretical (generative) linguists and corpus linguists, and concludes by acknowledging that he "refuses to give up his old ways [involving introspection including the invention of evidence] but [nevertheless] finds profit in being a consumer of some of the resources that corpus linguists have created" (p. 105). The paper includes a full rehash of the well-known investigation by Fillmore and Atkins (1992) of the semantic frame for *risk*. This work was a precursor of the FrameNet database, in which the frame elements for all words participating in a given frame are identified. In the case of *risk*, the main frame elements are:

- (1) the person taking the risk — the **Protagonist**.
- (2) the **Harm** that the Protagonist might suffer.
- (3) a **Valued Possession** which the Protagonist puts at risk.
- (4) an **Act** performed by the Protagonist.
- (5) the **Goal** of the Protagonist.

Syntactically, (1) is the subject of the active verb *risk*. (2), (3), and (4) compete for the direct object slot, so only one of them can be explicitly present in any given sentence. (5), if present at all, is governed by the preposition *for*. Very often, the Goal is implicit rather than expressed. Nevertheless, semantically, all five arguments are implicitly present.

There is also (pp. 113-121) an extended and very interesting discussion of the word *home*. Both these case studies (*risk* and *home*) are highly thought-provoking. They are full of insights which contribute greatly to the understanding of a wide range of lexical items, and they may result (we hope) in due course in significant improvements in monolingual lexicography. They should be required reading for every novice lexicographer.

Two aspects of Fillmore's work may be criticized. One has already been mentioned, namely his view that corpus evidence is supplementary to evidence invented by introspection: He is not a supporter of *systematic* corpus analysis. The second is that he fails to distinguish evidence from interpretation: "Should it ever come about that linguistics can be carried out without the intervention and suffering of a native-speaker analyst, I will probably lose interest in the enterprise" (p. 122). Here, Fillmore seems to imply that corpus linguists want to abolish the use of intuitions for purposes of linguistic analysis. It is true that some extremists in the corpus-linguistics world believe that the right thing to do is to let language learners loose on concordances with WordSmith (Scott 2009) or some similar toolkit, and leave them to draw their own conclusions. An answer to such people is implicit in the paper by Laufer in this collection (cf. Section 4.15).

4.8 Hanks's *Do Word Meanings Exist?* (2000)

Hanks's article is the first of three texts in Part IV, *On Word Senses and Polysemy*. In it, Hanks is concerned with the question "Do Word Meanings Exist?" His answer is "Yes, but ..." (p 133):

- Yes, but traditional dictionaries give a false impression. What dictionaries contain are (more or less inaccurate) statements of *meaning potentials*, not meanings.
- Yes, but only *in context*.
- Yes, but the meaning potential of a word consists of a cluster of *semantic components*, only *some of which are activated* when the word is used in context.

The article contains a corpus-based contrastive analysis of the two English nouns *bank*, in terms of their semantic components and semantic types, which Hanks associates with Pustejovsky's notion of the 'lexical conceptual paradigm'. It shows very clearly the systematic variability that characterizes word use in natural language such as English, and (by implication) the need for a new kind of lexicography that will link the prototypical phraseology of each word with a prototypical meaning.

4.9 Kilgarriff's *I Don't Believe in Word Senses* (1997)

Although the title of Kilgarriff's article may seem like a response to Hanks's question, this is misleading, for (a) it was published before Hanks's study, and (b) it is actually a statement (with which Kilgarriff agrees) by Sue Atkins dating from October 1994. While Hanks looked at meanings from the point of view of a seasoned writer and editor of monolingual dictionary definitions, Kilgarriff takes the NLP view, and is interested in automated word sense disambiguation (WSD). More particularly, Kilgarriff wants to find out whether the word sense division as seen in dictionaries or thesauri is of any use in NLP applications. His answer, unsurprisingly, is in the negative: "The set of senses defined by a dictionary may or may not match the set that is relevant for an NLP application" (p. 151). This outcome led to much further research, most of it undertaken under the umbrella of the Senseval project (see the Senseval website for more information, and Kilgarriff's Home Page for the numerous WSD publications).

Nonetheless, as was the case with the two contributions in Part II, although relevant to the NLP community, it is not immediately clear how practicing lexicographers may benefit from this study, apart from knowing that the results of their labour cannot be put to good use for a purpose for which they were not intended.

4.10 Stock's *Polysemy* (1984)

Although rarely quoted (cf. Figure 3), Stock's paper is indeed a must-read for all practicing lexicographers. Published in 1984 and thus well before the COBUILD project heralded a revolution in lexicography — in the form of a very new type of learner's dictionary (Sinclair and Hanks 1987), and the first

book-length guide to making a dictionary from a corpus (Sinclair 1987) — this is the very earliest paper to draw attention to the importance of collocational analysis in actual corpus-driven lexicography.

Stock starts her paper by pointing out her dissatisfaction with Ayto's working method for sense division: "Distinct superordinate or genus words suggest distinct senses" (p. 153). While working on the COBUILD project, she noticed that it is "possible to disambiguate meanings from written material with minimal, and purely linguistic, context" (p. 156). With reference to an even earlier work by Jones and Sinclair (1974), she continues with: "It is clear that, in a large number of cases when working from concordanced citational material, an examination, sometimes even a fairly cursory examination, of the syntactic and collocational patterns in the environment of the node word (the word under analysis), clarifies which meaning is being used" (p. 156).

With ample corpus evidence at hand, she then proceeds to offer her own instructive method to determine isolable meanings of polysemous words. That method consists of three procedures: An analysis of the syntactic behaviour, a study of the collocational patterns, and a final check of each possible reading.

4.11 Cowie's *Phraseology* (1994)

Cowie's text is the first of two that belong to Part V, *On Collocations, Idioms, and Dictionaries*. Cowie is well known as the co-author of the two-volume *Oxford Dictionary of Current Idiomatic English* (Cowie and Mackin 1975; Cowie, Mackin and McCaig 1983), henceforth ODCIE, which is basically a dictionary of idiomatic phrases, based on an OED-style collection of citations, accumulated at vast and admirable expenditure of time and effort, rather than a corpus. A new edition of this classic work in lexicography, thoroughly revised in the light of corpus evidence, is overdue.⁸

The problem with ODCIE highlights the problem with the selected article: Cowie refers to an "accumulation of descriptive studies throughout the 1970s and 1980s" (p. 163), but these (meritorious as they may be) are based on introspection and citations, not on a corpus-based search for regularities. Much corpus-based work has been done in phraseology since this short article was written and it now appears rather dated.

4.12 Fontenelle's *Using a Bilingual Dictionary to Create Semantic Networks* (1997)

Fontenelle used the 1st edition of the Collins–Robert French–English bilingual dictionary to create a semantic network. He augmented it with lexical functions based on the lexicographic work of Igor Mel'čuk, more in particular the Meaning-Text Theory (MTT). The result is a bilingual lexical-semantic database that can be used and/or integrated with WSD programs, translation systems or corpus query tools. From an end user's point of view: "One of the ultimate

goals of lexical knowledge acquisition is to make it possible for a user to navigate within a lexical knowledge base through concepts and lexical relations" (p. 170).

Although this article is very representative of Fontenelle's own (earlier) research interests and work, one notices that (a) the resulting database is not publicly available (which calls into question its true research and exploitation potential), and (b) the relevance to practicing lexicographers can only be said to be indirect. Fontenelle himself suggests that his work "should be seen as a contribution to the study of lexical-semantic relations and, more specifically, of collocational knowledge" (p. 188).

4.13 Bolinger's *Defining the Indefinable* (1985)

If a prize were to be awarded to the highest quality/page ratio in PLR, it would definitely go to Bolinger's four-page squib, the first of two texts in Part VI, *On Definitions*. The following quote from the opening page (p. 193) should suffice to send everyone straight to Bolinger's text:

Lexicography is an unnatural occupation. It consists in tearing words from their mother context and setting them in rows — carrots and onions and beetroot and salsify next to one another — with roots shorn like those of celery to make them fit side by side, in an order determined not by nature but by some obscure Phoenician sailors who traded with Greeks in the long ago. Half of the lexicographer's labor is spent repairing this damage to an infinitude of natural connections that every word in any language contracts with every other word, in a complex neural web knit densely at the center but ever more diffusely as it spreads outward. A bit of context, a synonym, a grammatical category, an etymology for remembrance's sake, and a cross-reference or two — these are the additives that accomplish the repair. But the fact that it is a repair always shows, and explains why no two dictionaries agree in their patchwork, unless they copy each other.

Brilliantly written, it is full of insight. In order to illustrate how much lexicographers destroy when they define, Bolinger then uses the suffix *-less* as his case study.

4.14 Rundell's *More Than One Way to Skin a Cat: Why Full-Sentence Definitions Have Not Been Universally Adopted* (2006)

Rundell's paper evaluates COBUILD-style full-sentence definitions; his point is that full-sentence definitions are appropriate only in certain circumstances, for example, the phrasal verb *to be laid up* in COBUILD-3 (Sinclair and Fox 1995):

If someone is **laid up** with an illness, the illness makes it necessary for them to stay in bed.

COBUILD definitions are always longer than their conventional equivalents. Rundell's charge that sometimes they are often also unnecessarily verbose is

well justified. Rundell quotes the noun *retreat* from COBUILD-1 (Sinclair and Hanks 1987):

A **retreat** is a change in your position when you have decided that you do not want to do what you have agreed or promised to do, usually because it has become too difficult, too expensive, or too embarrassing.

Rundell says with some justice: "longer definitions mean a heavier reading load (for readers whose linguistic resources are limited), and generally entail increased complexity. Thus the abandonment of traditional conciseness can bring new problems for users, who may go from the frying pan of unpacking a dense, formulaic definition to the fire of processing something two or three times longer" (p. 201). It was unfair of Rundell, however, to quote the 1st edition, as in COBUILD-3, the definition for this particular noun sense was dropped completely. The nearest thing to it is shown as a derivative of a verb sense, thus:

2 ... When an army **retreats**, it moves away from enemy forces in order to avoid fighting them. *The French, suddenly outnumbered, were forced to retreat.* | *Retreating soldiers were dousing homes and shops with petrol and setting them on fire.* [...] Also a noun. *In June 1942, the British 8th Army was in full retreat.*

Furthermore, for balance, Hanks's seminal 'Definitions and Explanations' (1987), in which the rationale for the COBUILD defining style is explained, should have been included in PLR. As it stands, the massive fifty-page section on definitions in OGPL (pp. 405-452 in there) is far more successful in presenting pros and cons.

4.15 Laufer's *Corpus-based versus Lexicographer Examples in Comprehension and Production of New Words* (1992)

Laufer's text is the only one in Part VII, *On Examples*. In it, Laufer attacks one of the well-established practices of COBUILD lexicography, viz. the use of authentic examples only in the dictionary. She conducted research showing the following results:

- (1) Learners perform much better in acquiring understanding and learning to use new words if they are given a definition *and* examples, than if they are given examples alone (as in some corpus-based classroom work) — even if the examples are carefully preselected and sorted.
- (2) Examples invented by lexicographers tend to be more useful to learners than authentic examples taken from a corpus.

She argues that possible slight loss of naturalness is a small price to pay for improved comprehensibility. She has a point, and it is well supported by her research.

We might add, however, that there is a danger of confusing academic research with practical tool creation. A dictionary is a practical tool for learners, who do not want to be bothered with the niceties of academic disputes. On the other hand, Sinclair (1984) showed that the practice of invention of examples can seriously distort the patterns of conventional usage associated with every word. This latter point may be more important in an academic research context than in a language learning context, but of course as more and more data are made available and corpora grow larger and larger, it becomes easier and easier to obtain the best of both worlds, by selecting examples from corpus data that combine brevity and clarity with authenticity.

As with Rundell's contribution on definitions, which required a voice from the other side, Fox's (1987) 'The Case for Examples', in which she sheds light on the COBUILD approach to examples, should have been included in PLR for balance.

4.16 Rundell's *Recent Trends in English Pedagogical Lexicography* (1998)

Rundell's article is the only one in Part VIII, *On Grammar and Usage in Dictionaries*. As background reading to OGPL, it fulfils its role admirably well. In short, OGPL deals with the monolingual learner's dictionary (MLD), which is (a) a commercial product, and (b) for human consumption. OGPL is thus a textbook dealing with the production of real dictionaries, for use in real situations, by users with real needs. Those needs are both of the decoding as well as the encoding type. Providing dictionary users with encoding skills (an aspect formerly only found in teaching material of the non-dictionary type) is the first truly 'hard part' of modern MLD compilation. The use of corpus evidence since the mid-1980s has further led to the increased realization that the mere description of words in isolation is simply not sufficient: In order to convey meanings one needs to know more about the typical company words keep and the contexts in which they are used. Hence the greater attention to, among others, lexical collocations, multi-word expressions, syntactic environments and usage labelling. Getting all of this right is the second 'hard part' of modern MLD compilation. In evaluating PLR as a companion volume to OGPL, one thus wants to see contributions that inform these issues further. This selection by Rundell does so.

Rundell's text starts with the observation that "the pace of change has been rapid — driven by a combination of theoretically-informed innovation, astonishing technological advances, and the creativity of dictionary-publishers in response to the known and perceived needs of users" (p. 221). He then proceeds with an analysis of how much MLDs have changed since A.S. Hornby, focusing on their descriptive and presentational improvements. Although the overview ends with the state-of-the-art of a decade ago (naturally, as it was published in 1998), all that is said (including the predictions made) remain valid.

4.17 Atkins's *Then and Now: Competence and Performance in 35 Years of Lexicography* (2002)

Atkins's paper is the first of two in Part IX, *On Bilingual Lexicography*. It is a good paper, but even though longer than its embedded and reworked version in OGPL itself (pp. 349-359 in there), it does not have the admirable quality of the OGPL version (cf. De Schryver 2008: 429). As such, including it in PLR may indeed provide the extra bit of data and insight, but it is mainly of interest to forensic text criticism. The 'Then' part (missing from OGPL) is moreover very short (less than two pages). Comparing the 'Then' with the 'Now', the text concludes with: "At last we are in a position to begin to reflect performance, and not our own competence, in our 21st century dictionary entries" (p. 271).

4.18 Duval's *Equivalence in Bilingual Dictionaries* (1991)

Initially published as an encyclopedia article in French, and translated into English by the author himself for PLR (an excellent translation at that), Duval's contribution is a welcome one indeed. Duval's topic is 'equivalence', which he extends to monolingual lexicography, where "there is equivalence between the entry word in the headword list from which the search starts and the body of the entry" (p. 273). From that angle, the title of his article is actually a misnomer. With reference to equivalence in bilingual lexicography — where the term is typically applied — Duval first points out that even so-called 'full equivalence' (between lemma sign and translation equivalent) does not always mean exact correspondence. He then proceeds by contrasting denotation vs. connotation, extension vs. comprehension, and language events vs. speech events. While the first dichotomy is also covered in OGPL (pp. 468-469 in there), the next two are not (and as such, constitute an informative extra).

Duval drew all his examples from the language pair French–English, two languages with very similar grammars. As a result, a whole range of additional thorny equivalence problems were avoided. A bilingual lexicographer working between a Bantu language and English, for example, is constantly faced with the problem that parts of speech in the one language do not correspond with those in the other (effectively turning, say, verbs into nouns, nouns into adjectives, etc.), or even with entire word classes (such as ideophones in Bantu) with no corresponding word classes nor translation equivalents (cf. e.g. De Schryver 2009). Bridging these various mismatches, in addition to the standard ones listed by Duval, is the real challenge of bilingual dictionary makers.

4.19 Church and Hanks's *Word Association Norms, Mutual Information, and Lexicography* (1989)

Church and Hanks's paper is the first of three in Part X, *On Tools for Lexicographers*. This paper had a galvanizing effect on the computational linguistics

world in 1989, when it was presented at the 27th annual meeting of the *Association for Computational Linguistics* (ACL) in Vancouver. There, it was the only paper to discuss statistical methods in computational linguistics, while at most (if not all) previous meetings of ACL there were none. Nowadays, such papers at ACL are in the majority.

The paper has attracted occasional hostile criticism, but seemingly only by people who feel threatened by it. For example, some people have proposed log-likelihood measures as a means of compensating for the so-called 'sparse data problem'. Computer scientists seem to like log-likelihood very much — it is elegant. But it has not been used in lexicography because it typically produces results that are less useful, practically speaking, than MI score (the statistical measure used by Church and Hanks in this paper) or t-score, which, as they were to comment in a later paper (Church et al. 1994) favours collocating function words, whereas MI favours collocations of pairs of content words.

The argument in the Church and Hanks paper is that collocations have a large role to play in decoding meaning, and that normal collocations are frequently recurrent in actual usage, so their relative importance can be measured by analysis of a large body of texts. What is more, Church and Hanks found (and published) a methodology for discovering the most significant collocates of any selected target word. The importance of this cannot be underestimated. Previous studies measured relations between two pre-selected target words, so they did not give us a *discovery* procedure. Church and Hanks then continue to show how collocates can be grouped to decide meaning.

When the proceedings of recent corpus linguistics conferences are read, it is surprising and saddening to note that there are many corpus linguists who have still, twenty years on, not yet adjusted their thinking to the most fundamental theoretical implication of this paper, namely that natural languages are analogical systems built around prototypes of many different sorts, and that corpora make it possible to identify these prototypes and measure agreement and variance statistically. If Church and Hanks are right about this (and their implication is hard to refute), it means that all linguistic categorization is a statistical procedure, a point of fundamental importance for lexicography of many different genres, as well as for theoretical and corpus linguistics.

MI is not really a "tool for lexicographers" (cf. the heading of Part X) — but it is the foundation on which one of the best corpus tools for lexicographers (the Sketch Engine) is based. This is discussed in the next section.

4.20 Kilgarriff, Rychlý, Smrž and Tugwell's *Sketch Engine* (2004)

The Sketch Engine is basically Manatee/Bonito to which word-sketch functionality was added. Manatee is a corpus query system (CQS), and Bonito its graphical user interface (GUI), both developed at Masaryk University, in Brno, by Pavel Rychlý (cf. e.g. Rychlý 2007). A word sketch is an automatically produced, corpus-based summary (i.e. 'sketch') of a word's grammatical and collo-

cational behaviour, first introduced by Adam Kilgariff and David Tugwell (2001). The Sketch Engine is arguably a magnificent tool for lexicographers (and corpus linguists in general), as it can be seen as a collocationally-annotated menu or index directly into the corpus. Figure 4, for instance, shows the word sketch for the noun *upset* in the Collins WordbanksOnline.

Home		Concordance		Word List		Word Sketch		Thesaurus		Sketch-Diff	
Turn on clustering		More data		Less data		Save					
upset WordBanks Online: English freq = 3866											change options
object of 1219 3.2		subject of 364 1.5		a modifier 1312 3.3		n modifier 505 2.6		pp over-p 22 6.0			
stomach 6 6.68		jumble 14 9.06		digestive 28 8.15		tummy 28 9.09		No. 4 2.05			
cause 289 6.36		ruin 20 6.26		gastric 13 7.8		stomach 125 8.33					
spring 16 6.22		highlight 12 4.66		first-round 28 7.42		decathlon 7 8.17		pp in-p 183 3.2			
bar 13 5.6		occur 5 2.11		stunning 58 7.18		cup 11 3.65		history 39 3.4			
score 48 4.86		cause 10 1.51		gastrointestinal 6 6.43		bit 40 3.14		tournament 8 2.84			
pull 38 4.39		happen 6 0.92		third-round 5 6.05		surprise 4 1.89		round 14 1.91			
avoid 22 3.77		continue 4 0.53		monumental 6 5.99		election 10 1.11		competition 4 0.96			
stage 5 3.62				second-round 6 5.91		nothing 8 0.57		race 5 0.6			
post 6 3.4		adj subject of 56 1.9		hormonal 5 5.87							
threaten 12 3.35		special 9 1.63		shocking 14 5.81		modifies 164 0.3		pp of-p 171 1.6			
suffer 16 2.84		possible 5 0.57		improbable 5 5.61		'Zona 24 11.99		No. 20 4.37			
predict 5 2.67				emotional 45 5.5		No. 67 6.11		tournament 8 2.84			
produce 19 2.55				major 150 5.3		tradition 4 1.82		century 9 1.67			
create 17 2.13				big 243 4.87				season 20 1.66			
record 4 2.11				electoral 9 4.7		and/or 385 1.3		decade 4 1.26			
provide 14 1.37				minor 16 4.68		diarrhoea 10 6.84		round 6 0.69			
get 72 1.36				mild 8 4.67		rash 5 5.35		day 17 0.2			
include 20 1.21				mighty 4 4.63		nausea 4 5.34					
consider 5 0.58				Greek 11 4.62		headache 6 4.19					
seem 8 0.58				sensational 4 4.4		cold 6 4.11					
believe 5 0.27				historic 9 4.27		streak 8 3.92					
see 17 0.11				least 11 4.17		controversy 6 3.43					
				huge 37 4.1		anger 5 2.85					
				memorable 4 4.0		pain 7 1.93					

Figure 4: Word sketch for *upset* (noun) in the Collins WordbanksOnline

For each grammatical relation (object of, subject of, etc.), lists are presented with the words that typically combine with the search word (1st columns), in order of statistical significance (3rd columns). The hyperlinked numbers (2nd columns) stand for the number of concordance lines, and clicking on them reveals all and only those instances of this particular combination.

In addition to word sketches, the Sketch Engine also offers a corpus-based thesaurus, and sketch differences. Computationally, 'all' that is required to obtain all of this is a corpus of a particular language, as well as a lemmatizer, a POS-tagger and a (regular expression) grammar for that language. At present, the Sketch Engine is available for about a dozen languages.

Note that, although extremely powerful, the Sketch Engine is not the most widely used corpus tool. That honour goes to WordSmith. One of the reasons is probably that the Sketch Engine "is designed for use over the web, with a server holding the data" (p. 304). In contrast, WordSmith is a lightweight stand-alone application, which enables users to use their corpus tool even when off-line and to keep their valuable corpora to themselves, rather than to have their data stored on a server in the Czech Republic. Moreover, lexicographers who compile their dictionaries with the dictionary production system TshwaneLex typically do not use any other corpus tools, as TshwaneLex contains a built-in and seamlessly integrated CQS. That system handles both raw and POS-tagged corpora (cf. De Schryver and De Pauw 2007), with which examples may be transferred directly from the corpus to the appropriate senses, and provides the core concordance functionality expected from a CQS. Word-sketch-like features are, at present, not included.

4.21 Grefenstette's *The Future of Linguistics and Lexicographers: Will there be Lexicographers in the Year 3000?* (1998)

Grefenstette starts by pointing out that generative linguistics is doctrinaire, like communism (with different, mutually disparaging factions) — deviance from the party line is severely punished, while corpus linguistics allows *laissez-faire*, like unbridled capitalism. He then argues that a middle course is needed, which he terms 'approximate linguistics'. Actually, today's techniques in computational linguistics are exactly that, "approximate linguistics, i.e. approximations to linguistic theory that are both incrementally perfectible as well as being robust and immediately useful" (p. 320-321). His paper describes the state-of-the-art of approximate linguistics a decade ago — walking the reader through tokenization, morphological analysis and lemmatization, POS-tagging, shallow parsing, and semantic tagging — and "gives a hint of what approximate linguistics can offer working lexicographers today in their task of describing some fundamental sense of a word" (p. 321). His vision of the future too is intriguing (cf. De Schryver 2003: 163 *ff.*), and where that future has already arrived, he was right (cf. De Schryver 2003: 180).

Finally, with regard to the question he poses in the title, he answers that there will always be aspects of the lexicographic enterprise where humans will be needed (compare also Section 4.7 above). These aspects are not really spelled out, however, and for this reason this text, despite being full of thought-provoking visions, is of only marginal relevance to a reader on practical lexicography.

4.22 Miller G.A., Beckwith, Fellbaum, Gross and Miller K.J.'s *Introduction to WordNet: An On-line Lexical Database* (1990)

Miller et al.'s text is the only one in Part XI, *On Semantic Networks and Wordnets*.

WordNet is an online thesaurus in which lexical items are arranged into sets of synonyms, called 'synsets', with senses across synsets related through lexical relations such as hyponymy/hyperonymy, meronymy/holonymy, troponymy, antonymy, pertainymy, etc. WordNet does not feature anywhere in OGPL (apart from merely being mentioned in two places, on pp. 82 and 159), has nothing to do with the production of commercial MLDs (compare also Section 4.16 above), and has little relevance for practical lexicography. It is used in NLP research though — simply because it happens to be available.

4.23 Atkins and Varantola's *Monitoring Dictionary Use* (1997)

Atkins and Varantola's text is the only one in the last part, Part XII, *On Dictionary Use*. It is a meticulously described account of a meticulously planned and meticulously executed experiment in monitoring how people actually use their dictionaries, monolingual and bilingual. A paper version of the think-aloud protocol is used, whereby a first subject, the 'dictionary user', performs a translation task using one or more paper dictionaries of their choice, and a second subject, the 'recorder', keeps track of all the look-ups needed in an attempt to solve a particular problem (i.e. a search). A total of 103 students performed 574 searches involving exactly 1 000 look-ups. The details of each of the 103 students were written down on a so-called 'cover sheet', and the recorder recorded the details of each look-up on a so-called pre-prepared 'recorder sheet', one per look-up. All the collected material was keyed into a database, and then analysed with the aim to provide answers to the following questions:

- How do people consult dictionaries?
- What kind of information are users looking for?
- Where do users find the information they need?
- Do users find what they look up?
- Are people satisfied with what they get?
- What do people do when frustrated?
- When do people use an L2 monolingual dictionary?

Answers are indeed obtained for each of those questions, with one of the most impressive outcomes, according to Atkins and Varantola, "the amount of reassurance sought from their dictionaries, particularly about L2 collocation, by even the most skilled of non-native L2 speakers, however experienced in translation they may be" (p. 371).

The aim of Atkins and Varantola "was to monitor the dictionary look-up process in as natural a situation as possible" (p. 338), and for the time (the experiments were carried out in 1991 and 1993, with paper dictionaries only) their approach was indeed the state-of-the-art. Their study rightfully attracted many citations, as seen in Figure 3 (where it is actually the first one following

the NLP section). These days, electronic dictionaries are studied using the computer as an invisible and unobtrusive 'recorder', making the dictionary look-up process even more natural. Modern experiments range from the monitoring of the use of real dictionaries consulted in a natural setting (see for an early example, De Schryver and Joffe 2004), to test lexica consulted in an experimental set-up (see for the most recent study, Lew and Doroszewska 2009).

5. How good is this Reader?

Bringing the quantitative and qualitative data together, one obtains Table 5.

Table 5: Contrasting the quantitative and qualitative evaluations of PLR

Ch.	Author(s)	Year	Citations	Refs from OGPL	Relevant to OGPL?
2	Johnson S	1747	–	***	***
3	Atkins BTS	1992/93	–	***	*
4	Apresjan JD	2002	–	**	**
5	Biber D	1993	**	**	–
6	Kilgarriff A & Grefenstette G	2003	**	**	–
7	Fillmore CJ	1992	–	**	**
8	Hanks P	2000	*	**	***
9	Kilgarriff A	1997	**	**	–
10	Stock PF	1984	–	***	***
11	Cowie AP	1994	–	**	–
12	Fontenelle T	1997	*	*	*
13	Bolinger D	1985	–	**	**
14	Rundell M	2006	–	*	**
15	Laufer B	1992	–	*	**
16	Rundell M	1998	*	***	***
17	Atkins BTS	2002	*	**	*
18	Duval A	1991	–	*	**
19	Church KW & Hanks P	1989	***	*	***
20	Kilgarriff A, Rychlý P, et al.	2004	**	**	**
21	Grefenstette G	1998	–	**	–
22	Miller GA, Beckwith R, et al.	1990	***	*	–
23	Atkins BTS & Varantola K	1997	*	**	***

In Table 5, the most important column is the last one, the qualitative evaluation, which is of course highly subjective. From it, and in the opinion of the present reviewer, one sees that 6 texts had better not been selected for PLR, while 3 are acceptable, 7 are good, and 6 are excellent choices. Contrasting this evaluation with the quantitative data, one notices a rather good overlap with the number of references from OGPL. A comparison of both of these with the number of citations is less convincing, however, but as was explained in Sec-

tion 3, (a) citation patterns show a bias towards papers in computational linguistics, and (b) in terms of authors, a much better fit is found.

Every reviewer of an anthology or reader of previously published papers no doubt has their own favourite authors and expectations, and the present reviewer is no exception. Among the selected papers are indeed some weak and/or less relevant ones — whose inclusion is surprising — while there are some even more surprising omissions. For example, it is quite astonishing that a modern reader on practical lexicography contains not a single paper by John Sinclair, the leading authority on collocations and the founder of statistical approaches to corpus linguistics, although papers by his colleagues (Stock, Hanks) are included. Collocations and phraseology are represented only by a single short encyclopedia article by Cowie, which reflects attitudes of the pre-corpus era and now seems rather out of date. This omission is perhaps not an oversight: It might to be the result of rivalry between John Sinclair and Sue Atkins (co-author of *OGPL*, which the volume under review is designed to support). It perhaps also explains why Hanks's important paper on 'Definitions and Explanations' is not included. The COBUILD full-sentence style of definition is represented only by a couple of less than enthusiastic accounts authored by Michael Rundell, the editor of a rival dictionary (*MEDAL*). The same reason may underlie the absence of Fox's 'The Case for Examples', another important product of the COBUILD project.

It is good to see the Russian tradition represented by Apresjan, but the equally important Mel'čuk appears only by proxy, in an article by Fontenelle himself. The generative tradition is not represented: One might have expected a paper each by Jackendoff, Pustejovsky, and Lakoff. Based on Table 3, Bogaards, Geraerts, Hoey, and Moon could and should also have been included. For good measure, a paper by Wierzbicka ought to have balanced out PLR.

In conclusion, this is a valuable reader of some very good papers and a handful of less good ones. However, its lack of balance means that supplementary reading is necessary for any would-be lexicographer.

Acknowledgements

I am grateful to Patrick Hanks and Robert Lew for comments on earlier drafts of this review article.

Endnotes

1. While elegant and space-saving, the disadvantages are that one cannot, from the references, quickly see what a particular paper is about or who is and is not quoted, while one will always need the entire book to keep things complete.
2. Given that the various papers which currently get zero hits in Google Scholar have most definitely been cited as well, one may be led to conclude that Google Scholar is particularly unreliable as an index of lexicographic scholarly activity. This is partly because it does not

survey all published sources, but only certain ones. If *all* publications were surveyed and if the time frame were 250 years, Johnson (1747), it would seem, would surely come out way ahead of Church and Hanks (1989).

3. Fontenelle could have considered selections, however, as Hartmann did for his anthology (cf. Table 1).
4. Unless otherwise noted, all page numbers preceded by 'p.' or 'pp.' are with reference to Fontenelle's PLR.
5. Actually, Fontenelle's PLR itself is now also and already available through Google Books, both in snippet view and as a limited preview.
6. Note that the collocational preferences of English *arrive* have changed since Johnson's time.
7. The following paragraph is taken from De Schryver (2005a: 422), where Apresjan's contribution was summarized.
8. Although both volumes have gone through a second edition (cf. the references), a thorough and truly corpus-driven revision is still outstanding. One of the selling points that accompanies the 2nd edition of volume 1 reads: "Examples of use taken from a wide range of contemporary sources, many drawn from the Oxford Corpus of the English Language" (<http://www.oup.com/elt/catalogue/isbn/6614?cc=global>), and for the 2nd edition of volume 2, tellingly shorter: "Example sentences taken from a wide range of contemporary sources" (<http://www.oup.com/elt/catalogue/isbn/6612?cc=global>). Selecting examples to support what was already there is very different from compiling such a dictionary from scratch driven by corpus evidence.

References

- Atkins, B.T.S. and M. Rundell. 2008. *The Oxford Guide to Practical Lexicography*. New York: Oxford University Press.
- Bogaards, P. 2009. EURALEX Newsletter — Autumn 2009. *International Journal of Lexicography* 22(3): 351-360.
- Church, K.W., W. Gale, P. Hanks, D. Hindle and R. Moon. 1994. Lexical Substitutability. Atkins, B.T.S. and A. Zampolli (Eds.). 1994. *Computational Approaches to the Lexicon*: 153-177. Oxford: Clarendon Press.
- Collins WordbanksOnline. 2009. An online service which gives subscribers access to language data based on the Collins corpora of modern written and spoken text [online]. <http://wordbanks.harpercollins.co.uk/Docs/Help/guide.html>
- Cowie, A.P. and R. Mackin (Eds.). 1975. *Oxford Dictionary of Current Idiomatic English. Volume 1: Verbs with Prepositions and Particles* (Retitled *Oxford Dictionary of Phrasal Verbs* for the second edition, 1993). Oxford: Oxford University Press.
- Cowie, A.P., R. Mackin and I.R. McCaig (Eds.). 1983. *Oxford Dictionary of Current Idiomatic English. Volume 2: Phrase, Clause and Sentence Idioms* (Retitled *Oxford Dictionary of English Idioms* for the second edition, 1993). Oxford: Oxford University Press.
- De Schryver, G.-M. 2002. Web for/as Corpus: A Perspective for the African Languages. *Nordic Journal of African Studies* 11(2): 266-282.
- De Schryver, G.-M. 2003. Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography* 16(2): 143-199.

- De Schryver, G.-M.** 2005. Book Review: R.R.K. Hartmann, ed. 2003. *Lexicography: Critical Concepts* (Routledge, three volumes, 1 327 pages). *International Journal of Lexicography* 18(1): 93-102.
- De Schryver, G.-M.** 2005a. Book Review: M.-H. Corréard, ed. 2002. *Lexicography and Natural Language Processing. A Festschrift in Honour of B.T.S. Atkins*. *Lexicographica: International Annual for Lexicography* 21: 420-425.
- De Schryver, G.-M.** 2008. An Analysis of *The Oxford Guide to Practical Lexicography* (Atkins and Rundell 2008). *Lexikos* 18: 423-445.
- De Schryver, G.-M.** 2008a. 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.** 2009. The Lexicographic Treatment of Ideophones in Zulu. *Lexikos* 19: 34-54.
- De Schryver, G.-M.** 2009a. Bibliometrics in Lexicography. *International Journal of Lexicography* 22(4).
- De Schryver, G.-M. and G. De Pauw.** 2007. Dictionary Writing System (DWS) + Corpus Query Package (CQP): The Case of *TshwaneLex*. *Lexikos* 17: 226-246.
- De Schryver, G.-M. and 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.
- Fillmore, C.J. and B.T.S. Atkins.** 1992. Toward a Frame-based Lexicon: The Semantics of RISK and its Neighbors. Lehrer, A. and E.F. Kittay (Eds.). 1992. *Frames, Fields and Contrasts: New Essays in Semantic and Lexical Organization*: 75-102. Hillsdale: Lawrence Erlbaum.
- Fontenelle, T. (Ed.).** 2008. *Practical Lexicography: A Reader*. New York: Oxford University Press.
- Fox, G.** 1987. The Case for Examples. Sinclair, J.M. (Ed.). 1987: 137-149.
- FrameNet*. 2009. An online lexical resource for English, based on Frame Semantics and supported by corpus evidence [online]. <http://framenet.icsi.berkeley.edu/>
- Google Books*. 2009. A search and preview of millions of books from libraries and publishers worldwide [online]. <http://books.google.com/>
- Google Scholar*. 2009. A search of scholarly literature across many disciplines and sources, including theses, books, abstracts and articles [online]. <http://scholar.google.com/>
- Hanks, P.** 1987. Definitions and Explanations. Sinclair, J.M. (Ed.). 1987: 116-136.
- Hanks, P.** 2005. Johnson and Modern Lexicography. *International Journal of Lexicography* 18(2): 243-266.
- Hanks, P. (Ed.).** 2008. *Lexicology: Critical Concepts* (6 volumes). Abingdon: Routledge.
- Hanks, P.** 2008a. The Lexicographical Legacy of John Sinclair. *International Journal of Lexicography* 21(3): 219-229.
- Hanks, P.** 2009. Book Review: S.J. Perrault, ed. 2008. *Merriam-Webster's Advanced Learner's English Dictionary*. *International Journal of Lexicography* 22(3): 301-315.
- Hartmann, R.R.K. (Ed.).** 2003. *Lexicography: Critical Concepts* (3 volumes). London: Routledge.
- Johnson, S.** 1755. *Johnson, Preface to the Dictionary*. Edited by Jack Lynch [online]. <http://andromeda.rutgers.edu/~jlynch/Texts/preface.html>
- Jones, S. and J.M. Sinclair.** 1974. English Lexical Collocations. A Study in Computational Linguistics. *Cahiers de Lexicologie* 24(1): 15-61.
- Kilgarriff, A.** 2009. *Adam Kilgarriff's Home Page* [online]. <http://www.kilgarriff.co.uk/>
- Kilgarriff, A. and D. Tugwell.** 2001. WORD SKETCH: Extraction and Display of Significant Collocations for Lexicography. *Proceedings of the ACL Workshop on 'COLLOCATION: Computational Extraction, Analysis and Exploitation'*: 32-38. Toulouse. [Reprinted as 'Sketching Words' in Cor-

- réard, M.-H. (Ed.). 2002. *Lexicography and Natural Language Processing. A Festschrift in Honour of B.T.S. Atkins*: 125-137. Euralex.]
- Lew, R. and J. Doroszewska.** 2009. Electronic Dictionary Entries with Animated Pictures: Lookup Preferences and Word Retention. *International Journal of Lexicography* 22(3): 239-257.
- Rundell, M.** 2002. *Macmillan English Dictionary for Advanced Learners*. Oxford: Macmillan. [MEDAL]
- Rychlý, P.** 2007. Manatee/Bonito — A Modular Corpus Manager. Sojka, P. and A. Horák (Eds.). 2007. *Proceedings of Recent Advances in Slavonic Natural Language Processing, RASLAN 2007*: 65-70. Brno: Masaryk University.
- Scott, M.** 2009. *WordSmith Tools* [online]. <http://www.lexically.net/wordsmith/>
- Senseval.** 2009. Evaluation Exercises for the Semantic Analysis of Text [online]. <http://www.senseval.org/>
- Sinclair, J.M.** 1984. Naturalness in Language. Aarts, J. and W. Meijs (Eds.). 1984. *Corpus Linguistics. Recent Developments in the Use of Computer Corpora in English Language Research*: 203-210. Amsterdam: Rodopi.
- 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: Collins ELT.
- Sinclair, J.M. and G. Fox.** 1995. *Collins COBUILD English Dictionary*. London: HarperCollins. [COBUILD-3]
- Sinclair, J.M. and P. Hanks.** 1987. *Collins COBUILD English Language Dictionary*. London: HarperCollins. [COBUILD-1]
- TshwaneLex.** 2009. TshwaneLex Dictionary Production System [online]. <http://tshwanedje.com/tshwanelex/>

Addendum: Fact sheet of the 23 texts in *Practical Lexicography: A Reader*

Author(s)	Born	Year	Title	Source	P.	Ch.	Pp.	Online	Cited
Aprésian JD	1930	2002	Principles of Systematic Lexicography	Fest - Atkins	I	4	10	-	-
Atkins BTS	1931	1992/93	Theoretical Lexicography and its Relation to Dictionary-Making	J - Ds: DSNA	I	3	20	-	-
Atkins BTS	1931	2002	Then and Now: Competence and Performance in 35 Years of Lexicography	Proc - Euralex '02	IX	17	26	-	5
Atkins BTS & Varantola K	1931, 1946	1997	Monitoring Dictionary Use	J - IJL	XII	23	39	\$	48
Biber D	1946	1993	Representativeness in Corpus Design	J - Ling & Lit Comp	II	5	25	\$	261
Bolinger D	1907	1985	Defining the Indefinable	Ed coll	VI	13	4	-	-
Church KW & Hanks P	(1957), 1940	1989	Word Association Norms, Mutual Information, and Lexicography	Proc - ACL	X	19	11	I	1568
Cowie AP	1931	1994	Phraseology	Encyclopedia	V	11	5	-	-
Duval A (Tr by the author)	(1949)	1991	Equivalence in Bilingual Dictionaries	Encyclopedia	IX	18	10	-	-
Fillmore CJ	1929	1992	'Corpus Linguistics' or 'Computer-aided Armchair Linguistics'	Proc - Nobel '82	III	7	18	G	-
Fontenelle T	1964	1997	Using a Bilingual Dictionary to Create Semantic Networks	J - IJL	V	12	21	\$	12
Fontenelle T	1964	2008	Editor's Introduction	N.A.	0	1	15	N.A.	N.A.
Grefenstette G	1956	1998	The Future of Linguistics and Lexicographers: Will there be Lexicographers in the Year 3000?	Proc - Euralex '98	X	21	17	-	-
Hanks P	1940	2000	Do Word Meanings Exist?	J - Comp & Hum	IV	8	10	(I)	47

Johnson S (Ed by Jack Lynch)	1709	1747	The Plan of a Dictionary of the English Language	Letter	I	2	12	I	-
Kilgarriff A	1960	1997	I Don't Believe in Word Senses	J - Comp & Hum	IV	9	17	I	156
Kilgarriff A & Grefenstette G	1960, 1956	2003	Introduction to the Special Issue on the Web as Corpus	J - Comp Ling	II	6	13	I	299
Kilgarriff A, Rychlý P, Smrží P & Tugwell D	1960, 1973, 1972, ?	2004	The Sketch Engine	Proc - Euralex '04	X	20	10	I	142
Laufer B	?	1992	Corpus-based versus Lexicographer Examples in Comprehension and Production of New Words	Proc - Euralex '92	VII	15	6	-	-
Miller GA, Beckwith R, Fellbaum C, Gross D & Miller KJ	1920, ?, (1950), 1968, ?	1990	Introduction to WordNet: An On-line Lexical Database	J - IJL	XI	22	8	\$	1386
Rundell M	1950	1998	Recent Trends in English Pedagogical Lexicography	J - IJL	VIII	16	23	\$	32
Rundell M	1950	2006	More Than One Way to Skin a Cat: Why Full-Sentence Definitions Have Not Been Universally Adopted	Proc - Euralex '06	VI	14	13	-	-
Stock PF	(1949)	1984	Polysemy	Proc - Euralex '83	IV	10	8	-	-

Abbreviations: Ed = edited, Tr = translated; P. = Part, Ch. = Chapter, Pp. = pages; \$ = by subscription, G = Google Books, I = Internet, (I) = idem but different version

Notes: One date was adapted here (Church & Hanks 1990 → 1989) compared to the one given in *Practical Lexicography: A Reader*, to better reflect the original year of publication. Years between brackets are informed extrapolations.

The Metalexicographical Contribution of Pedro A. Fuertes- Olivera and Ascensión Arribas- Baño's *Pedagogical Specialised Lexicography*: A Critical Review

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

Abstract: This article attempts to give a critical review of Pedro A. Fuertes-Olivera and Ascensión Arribas-Baño's *Pedagogical Specialised Lexicography*. It evaluates the book in view of the available metalexicographical literature as well as current trends in practical LSP lexicography in the wake of rapid technological and information developments. The article appreciates both the methodological and theoretical frameworks within which the book identifies, investigates and addresses lexicographical problems. The approaches make the book an important academic contribution with critical insights for metalexicographers, especially in the area of dictionary criticism. The theoretical insights provided in the book further possess the potential to accomplish exactly what metalexigraphy ought to accomplish, i.e. to facilitate the production of user-friendly dictionaries which meet both the reference needs and reference skills of the targeted users. A few concerns are, however, raised mainly regarding the technical aspects of the book. While these may have a negative impact on the reader, they are not severe enough to discredit the rigour with which the book was conceived.

Keywords: METALEXICOGRAPHY, PEDAGOGICAL SPECIALISED LEXICOGRAPHY, MONOLINGUAL LSP DICTIONARY, BILINGUAL LSP DICTIONARY, MEANING REPRESENTATION, DICTIONARY STRUCTURES, DEFINITIONS, EQUIVALENCE, EXAMPLES, DICTIONARY CRITICISM

Opsomming: Die metaleksikografiese bydrae van Pedro A. Fuertes-Olivera en Ascensión Arribas-Baño se *Pedagogical Specialised Lexicography*: 'n Kritiese beskouing. Hierdie artikel probeer om 'n kritiese beskouing van Pedro A. Fuertes-Olivera en Ascensión Arribas-Baño se *Pedagogical Specialised Lexicography* te gee. Dit beoordeel die boek met betrekking tot die beskikbare leksikografiese literatuur, sowel as huidige neigings in praktiese TSD leksikografie in opvolging van die vinnige tegnologiese en inligtingsontwikkelinge. Die artikel het 'n waardering vir sowel die metodologiese as teoretiese raamwerk waarbinne die boek die leksikografiese probleme bepaal, ondersoek en aanpak. Die benaderings maak die boek 'n belangrike akademiese bydrae met kritiese insigte vir metaleksikograwe, veral op die gebied van woordeboekkritiek. Die teoretiese insigte wat in die boek verskaf word, besit verder die moontlikheid om

dit te bewerkstellig wat metaleksikografie behoort te bewerkstellig, d.i. om die voortbrenging van gebruikersvriendelike woordeboeke te vergemaklik wat sowel die naslaanbehoefte as die naslaanvaardighede van die teikenbruikers bevredig. 'n Paar bedenkinge word egter geopper hoofsaaklik betreffende die tegniese aspekte van die boek. Alhoewel hulle 'n negatiewe uitwerking op die leser kan hê, is hulle nie ernstig genoeg om die noukeurigheid aan te tas waarmee die boek beplan is nie.

Slutelwoorde: METALEKSIKOGRAFIE, OPVOEDKUNDIGE GESPECIALISEERDE LEKSIKOGRAFIE, EENTALIGE TSD-WOORDEBOEK, TWEETALIGE TSD-WOORDEBOEK, BETENISAANBIEDING, WOORDEBOEKSTRUKTURE, DEFINISIES, EKWIVALENSIE, VOORBEELDE, WOORDEBOEKKRITIEK

1. Introduction

This article is a critical review of Pedro A. Fuertes-Olivera and Ascensión Arribas-Baño's (2008) book, *Pedagogical Specialised Lexicography*. Although lexicography has reached a stage where some of its practitioners and academics strongly claim its disciplinary status, literature about it is still limited when compared to other disciplines, even those closely related to it like linguistics. Of course, there are now lexicographical associations, accredited journals and courses in lexicography. However, book length publications such as this one are less common. This makes it important to evaluate each new publication in view of its predecessors in order to appreciate the formulation and application of methodological and theoretical insights. The discussion aims to put the book into the perspective of the challenges and visions of lexicography in the era of rapid knowledge development, the increased needs for access to information and efficient communication.

The authors of the book appear to be experts and experienced in specialised lexicography and related fields. Fuertes-Olivera, for example, is a professor of English for Specific Purposes, Applied Linguistics and Translation (<http://www.pedrofuertes.net>), which implies a direct academic interest in both pedagogical and specialised lexicography. The book already bears, as part of its blurb, recommendations by another two fine LSP lexicographers. Enrique Alcaraz Varó was a professor and the co-author of ten bilingual English–Spanish/Spanish–English LSP dictionaries (<http://www.aedean.org/>), the main objects of analysis in this book. Sven Tarp is the co-editor of the ground-breaking *Manual of Specialised Lexicography*, editor of seven LSP dictionaries (<http://www.aedean.org/>), author of numerous articles and book chapters on lexicography and also a professor of lexicography.

2. A Brief Synopsis of the Book

This section aims to provide an overview of the six chapters into which the book is structured. The chapterisation of the book has been respected in out-

lining its contents. However, caution has been taken to avoid reproducing the outline of the book given by the authors at the end of the first chapter.

Chapter 1, the "Introduction", clears the field for the analyses undertaken in the subsequent three chapters and raises issues conclusively addressed in the final chapter. Firstly, a background is provided for the linguistic concept of "Language for Specific Purposes" (LSP), which is the presupposition for LSP lexicography. By references to the works of well-known lexicographers, namely Atkins (1996), Worsch (1999) and Hartmann (2001), a clear indication is given that the subsequent chapters are conceived within a (meta)lexicographical framework. The authors delineate their scope of study, identifying the analysis of the representation of meaning in selected business English and Spanish LSP dictionaries as their main aim (p. 10). Also explained is the methodology of selecting the dictionaries for the study and the materials subjected to scrutiny in the book. The chapter ends by providing a brief overview of the remaining chapters. Although the outline offered in the book was used as the basis for this section, it should be borne in mind that additional information is given here for the benefit of the readers who are still unfamiliar with the book.

Chapter 2 is entitled "The macrostructure, mediostucture and access structures of business dictionaries". The authors show their awareness of various types of dictionary structures. They explain that their focus on the three selected types of dictionary structures is based on their view that meaning is distributed across these three. The chapter thus explores possible structural configurations of dictionaries resulting from the lexicographers' procedures of treating meaning information. The attention is mainly focused on the implications of homonymy and polysemy for the structure of dictionary articles and the user-friendliness of selected dictionaries.

Definitions are the subject of analysis in Chapter 3 entitled "Definitions in business dictionaries". The authors, concurring with many writers, show that the definition is the most widely used method of explaining meaning in dictionaries (p. 47). Accordingly, three main kinds of definitions are identified, namely terminological, encyclopaedic and semantic definitions. Before examining their employment and assessing their user-friendliness in the selected business dictionaries, the authors draw from other scholars to test the validity of the distinctions among the identified types of definitions. Having noted that the three types of definitions are distinguished according to the nature of dictionaries, the authors proceed to identify further types of definitions according to the contents of dictionaries and also according to users' needs. Consequently, the authors discuss the definition by intention, the definition by extension and the partitive definition before exploring the conceptual, the rational and the functional (also called encyclopaedic) definitions. The discussion of definitional styles brings the Cobuild definition to the fore.

Entitled "Equivalence in business dictionaries", Chapter 4 deals with the generally problematic phenomenon of equivalence in bilingual lexicography. The cultural factor and interlingual anisomorphism are identified as the pri-

mary challenges of bilingual lexicography. The chapter explores the approaches adopted to convey meaning information in their sample of bilingual LSP dictionaries. Inter alia, these are providing translation equivalents of lemmata, translating examples, collocations or idioms and sense discrimination. As will be noted in section 3.4, this chapter contributes not only to LSP lexicography but also to bilingual lexicography.

Chapter 5, with the title "Examples in business dictionaries", defines and categorises the notion of an example as a data category with a multi-functional role in conveying meaning information in dictionaries. The authors attempt a typology of lexicographical examples and explore the extent and effect to which meaning information is treated by means of this data category in the LSP dictionaries of the sample. In this respect, a triple distinction of invented examples, examples from a corpus without edition and examples from a corpus with edition is made. The criteria of arranging examples in articles in the same dictionaries are also explored.

In the sixth and final chapter, conclusions are drawn on the three areas of study which the authors identify. These are: (a) the representation of meaning in some monolingual business dictionaries; (b) the representation of meaning in some bilingual English–Spanish/Spanish–English business dictionaries; and (c) the construction of pedagogical LSP dictionaries (p. 133). The first two are covered in section 3.2 of this article while the third one is dealt with in section 3.1.

3. Arguments and Contributions Regarding Problems of LSP Lexicography

Although there has been a significant increase of metalexicographical literature pertaining to LSP lexicography, especially following the publication of Bergenholtz and Tarp (1995), LSP lexicography still faces challenges of a practical and theoretical nature (see, for example, Tarp 2000). Some of the challenges are not peculiar to LSP lexicography but affect lexicography in general. This makes it important to evaluate each new publication on this subject according to the extent to which at least one of these challenges is addressed. For the purpose of this article, focus will be firstly on the general theoretical insights provided for LSP lexicography and secondly on the insights the authors bring regarding their main objective, i.e. the analysis of the representation of meaning in LSP monolingual and bilingual dictionaries. Accordingly, this section is divided into two subsections.

3.1 On Problems Relating to the Theory of Lexicography

There seems yet to be no consensus whether lexicography is an independent scientific discipline or not. However, lexicography has over the years estab-

lished its identity which enables metalexigraphers to reach a consensus on what it is *not*. Most significantly, Wiegand (1984: 13) argues:

Lexicography is *not* (my italics) a branch of so-called applied linguistics. ... Lexicography is, at all events, more than the application of linguistic theories and philological findings. In a frequency dictionary, for example, the methods of statistics play the major role, and just imagine if linguistic knowledge alone were to be taken into account in a technical medical dictionary.

The argument advanced by Wiegand indicates that LSP lexicography would be in a nebulous position if lexicography is not granted its distinct identity from the other fields. LSP lexicography would be part of each and every field in which LSP dictionaries are compiled. Tarp's (2000) summation of the situation as absurd would have no effective equal in this regard. Tarp (2000) extracts an article from a dictionary of gene technology and convincingly argues that it could be written better by an expert in the field of gene technology without linguistic training than a trained linguist without the input of an expert. This effectively nullifies theoretical linguistics as an absolute qualification for lexicographical practice. At the same time, Tarp (2000) points out the limitation of the subject-field expert in practical LSP lexicography, indicating that some special lexicographical training is a prerequisite for producing a well-structured article. To explain the limitations of one of the dictionaries they review, Fuertes-Olivera and Arribas-Baño notably use the fact that the compiler is a subject-field expert lacking lexicographical training.

The main challenge of practical lexicography is to produce appropriate and user-friendly dictionaries, and the challenge of metalexigraphy is to provide practical lexicography with useful insights in this regard. The growth of metalexigraphy in the past half century or so has effectively established an identity for lexicography by indicating the extent of its relations with linguistics and many other diverse fields. In order to consolidate its identity, lexicography remains with the task of introspection, taking into account its own practical and theoretical elements. While issues of lexicographical qualifications which have already been explicated remain relevant, it is now a question of how established lexicographical institutions and trained lexicographers address specific problems in their quest to provide users with appropriate dictionaries. In the field of LSP lexicography, advances in science and technology, as well as globalisation imply an increased need for knowledge acquisition and dissemination, the challenges of which may be solved using LSP dictionaries. From this position, one may appreciate Fuertes-Olivera and Arribas-Baño's book as a significant theoretical contribution to LSP lexicography and lexicography in general.

As noted previously, the authors signal their adoption of a lexicographical approach in the first chapter. The reader is promptly reminded that "the content and design of a dictionary are determined by the needs of its users" (p. 2). The importance attached to the user-perspective is well-advised, given that it is

the utility value of dictionaries and the users' ability to use dictionaries whose concern distinguishes lexicography from other related fields. This is crucial for a student of lexicography who needs proper theoretical orientation. More significantly, the authors do this fully aware of the bigger picture of lexicography when they state that they "combine dictionary research and linguistics knowledge" because, although lexicography is an independent scientific discipline, it is connected to many related fields such as lexicology, syntax, corpus linguistics, contrastive linguistics and lexical semantics (p. 10). Thus one is reminded that while focus should be on the user needs many other external factors such as "the cost of material and staff; the selection of equipment and formats for corpus data, word-processing, printing and publishing; the planning and the designing of the layout; factual knowledge; dictionary research and knowledge of linguistics" need to be considered (p. 10). This is particularly important for aspiring lexicographers and many other subject-field experts who, noting the potential value of dictionaries within their fields, may aspire to venture into LSP lexicography. In South Africa, for example, the need to nurture multilingualism into the education curriculum currently sees a lot of activities of an LSP lexicographical nature in the form of the compilation of multilingual "specialised glossaries" which include the indigenous languages. This book may provide the necessary insights to practitioners involved in such work.

The theoretical position established in the first chapter largely prevails throughout the book, not disappointing the reader with a keen lexicographical interest. The analysis of the representation of meaning is firmly based on the user-perspective and evidently grounded in the so-called "modern" theory of lexicographical functions (Bergenholtz and Tarp 2003, Tarp 2008). This theory effectively enables the authors to analyse the representation of meaning information in the selected dictionaries in terms of the information needs of specific users (learners) in specific situations of learning business English LSP. As will be noted in the next section, the reader would realise that the criticism of some dictionaries and recommendations regarding them are validated within such a theoretical framework. They state that LSP dictionaries "should only include data on the basis of their respective functions" (p. 7). Given this, one could even suggest that reference to the works of Bergenholtz and Tarp among other proponents of the theory of lexicographical functions would have sufficed when the authors refer to "the new lexicography" and the pivotal determining role of users regarding the contents and design of dictionaries (p. 2). A reader familiar with the publications of the Danish metalexigraphers will even wonder whether translators or professional interpreters and LSP learners deserve mention as new additions which are carefully crafted to complete a statement of user characterisation within the theory of lexicographical functions, given that they have already been considered in publications such as Bergenholtz and Tarp (1995) and Tarp (2000, 2002, 2004). The book is certainly a good example of an excellent application of the theory of lexicographical functions which has mainly been popularised first through the publication of Bergenholtz and Tarp (1995).

Another theoretical issue which Fuertes-Olivera and Arribas-Baño consider and give insight into, regards the relationship between LSP lexicography and terminology. The authors content themselves with the position established by their predecessors such as Bergenholtz and Tarp (1995), Tarp (2000) and Bergenholtz and Nielsen (2006) who adopt a lexicographical approach. This approach "maintains that the often quoted distinction between LSP lexicography and terminology is of no practical use" (p. 8; cf. Bergenholtz and Nielsen 2006). Likewise, Fuertes-Olivera and Arribas-Baño point to a "methodological confluence between LSP lexicography and terminography" (p. 8) which has been embraced by scholars in the field of terminology, particularly Sager (1984, 1996). As indicated by the authors, the descriptive dimension of terminology now places a term rather than its represented concept at the centre of the general theory of terminology. Not only do they articulate this lexicographical position by making reference to scholars such as Temmerman (2000), they also proceed by indicating how the position determines the arrangement of lemmata in the macrostructure (see section 3.1) and even the formulation of definitions (see section 3.2). As will be seen in sections 3.1 and 3.2, the consolidation of the confluence between LSP lexicography and terminology offers more opportunities for LSP lexicographers and terminologists alike to provide users with user-friendly tools which are not based so much on the preservation of distinct identities of the two fields but mainly on the functional value of the products. It is significantly befitting that the book, which is a recent release in the series *Terminology and Lexicography Research and Practice*, pays attention to this debate in a conciliatory way, thereby contributing simultaneously to both LSP lexicography and terminology.

That lexicography has recourse to theories devised from other fields, to which Fuertes-Olivera and Arribas-Baño also concur, has been demonstrated, among others by Wiegand (1984), Tarp (2000) and Hartmann (2005). Among these, linguistic theories evidently dominate although it needs to be emphasised that they are not the only ones influencing lexicography. Accordingly, Fuertes-Olivera and Arribas-Baño consider their knowledge of linguistics as an important element of their theoretical framework. Meaning and its relations of polysemy, synonymy, antonymy, homonymy, hyponymy and hyperonymy are all linguistic phenomena, the analysis of whose representation in dictionaries would be impossible without a proper knowledge of linguistics. The Sapir-Whorfian hypothesis, a sociolinguistic theory, provides a framework for the analysis of meaning representation in the form of translation equivalents in dictionaries. Corpus linguistics also contributes to Chapter 5 in which corpus-based examples are considered. There is thus abundant evidence that the influence of linguistics needs not to be downplayed in lexicography (Hartmann 2001, Gouws and Prinsloo 2005), although it does not exhaustively address lexicographical issues. As will be shown in the next section, for Fuertes-Olivera and Arribas-Baño, a lexicographical approach which is also based on linguistic knowledge facilitates an effective analysis of the representation of meaning in

business dictionaries. While this enables the authors to fulfil their endeavour, it also contributes to the theory of LSP lexicography or lexicography in general by validating and elaborating on the theoretical pursuits of other scholars in the field.

The foregoing are the major highlights of how Fuertes-Olivera and Arribas-Baño engage with the main problems and theoretical issues in the area of lexicography. Within this theoretical framework, they sum up their arguments and proposals towards the compilation of pedagogical LSP dictionaries in the last chapter of their book. While this article has not given individual consideration to these proposals, the general framework within which they are made seems to be productive in such a way that readers may be rewarded when they put the proposals into practice, be it in the compilation or evaluation of dictionaries.

3.2 On the Representation of Meaning in LSP Dictionaries

The representation of meaning in English and Spanish business dictionaries is the main subject of Fuertes-Olivera and Arribas-Baño's book. The analysis is effected within the parameters of dictionary structures, definitions, equivalence and examples, each of which is reviewed from subsection 3.2.1 to 3.2.4. The objective of the review is to highlight the practical and theoretical contributions of the book regarding the treatment of meaning information in LSP dictionaries.

3.2.1 Dictionary Structures

In modern-day lexicography, practising lexicographers do not only have to consider the types of data they need to include in their dictionaries but also the structuring of the dictionary components and the spreading of the data across the various components. Elaborating on McArthur's (1986) view of dictionaries as containers of knowledge, Gouws (2004) sees such an approach as observing not only the knowledge contained in dictionaries but also dictionaries as containers — this thanks to Wiegand's metalexicography, especially his theory of lexicographical texts (Wiegand 1984, 1996).

Accordingly, Fuertes-Olivera and Arribas-Baño do not only single out meaning as the content of dictionaries but also consider dictionary structures as accommodating and conveying meaning. They focus on the macrostructure, microstructure, mediostructure and access structure, arguing that meaning is basically distributed within these structures. However, the authors' understanding of these types of dictionary structures seems to be somewhat confusing. Subsequently, their appreciation of the role of these dictionary structures regarding the representation of meaning will also be found inadequate.

The consideration and treatment of the linguistic phenomenon of homonymy as an element of meaning information in dictionaries mainly corresponds

to the macrostructure. Fuertes-Olivera and Arribas-Baño aim to determine whether the compilers of the selected business dictionaries took homonymy into account. This constitutes checking if some lemmata in the macrostructure are homonymous: "Dictionaries tend to treat homonyms in separate entries on the grounds that native speakers regard them as different words ..." (Hartmann and James 1998: 69). The authors of the book observe that only one out of their sample of six dictionaries deals with homonymy by entering homonyms as separate lemmata and indicating their grammatical categories. The example of *price* being both a noun and a verb, and hence having separate meanings, is given. This captures the lexical structure of the subject field by showing that, contrary to the dominant practice of lemmatising nouns only in LSP dictionaries (p. 135), other word categories such as verbs and adjectives have a special designation and are legitimate lemma candidates in LSP dictionaries (L'Homme 2003). For LSP learners, it assists in the conveyance of meaning information. Thus, the failure of the five evaluated dictionaries to provide information on homonymy is deservedly condemned. The minimalist approach of the *Oxford Dictionary of Business English* (1993) is recommended "because of the scarce relevance of etymological and/or semantic criteria for the student of business English" (pp. 21-22). This means that for some LSP dictionaries, etymological and semantic criteria should be used to complement the formal criterion in the treatment of homonymous lemmata.

Fuertes-Olivera and Arribas-Baño discuss the arrangement of lemmata under Section 2.4 which is entitled "The form of the lexicographical article in business dictionaries". As far as dictionary structures are concerned, this section is potentially confusing. The title suggests that the microstructure is the main subject of this section. On the contrary, the word *microstructure* does not appear anywhere in it. Apparently, it is the mediostructure which dominates the first two paragraphs of this section. Yet the reader is reminded that the section focuses on "the structuring criteria of the lemmas and sub-lemmas inside the dictionary article" (p. 22). Clearly, this concerns the macrostructure more than the mediostructure, but the criteria used in the former may determine the latter through what is referred to as "cross-reference conditions" (Nielsen 1999, Tarp 1999). The following statement of the authors of the book should be read the other way round: "the type of mediostructure to a considerabl[e] degree determines the synthetic or analytic character of the macrostructure" (p. 30). It is actually the macrostructure and the data distribution structure in general, which determines the relevance and employment of a mediostructure.

In spite of the above concerns, the reader will find that the authors' approach to the two main methods of arranging lemmata especially in LSP lexicography is consistent with the principles of modern lexicography on the subject (see McArthur 1986, Pederson 1995). The two main methods are an alphabetic and a thematic arrangement. The point of departure usually lies in acknowledging the strengths and weaknesses of each method before opting for one that would best assist the users with particular skills in consulting the dic-

tionary. This pertains to easy access to lemmata and easy establishment of conceptual, semantic, morphological and etymological relationships between lemmata. In their analysis, Fuertes-Olivera and Arribas-Baño note that all the dictionaries studied employ an alphabetic, but not a thematic macrostructure. Although they rightly argue that this is an easy method for the lexicographer who is also familiar with the users, they further demonstrate that it affects the inner cohesion of the subject field. This is particularly noted in the case of monolingual English or Spanish business dictionaries which adopt a straight alphabetical macrostructure. Niche-alphabetic arrangement and non-strict alphabetic arrangement principles observed mainly in the bilingual business dictionaries are to a certain extent bridges between the alphabetic macrostructure on the one hand and the thematic macrostructure on the other which the authors recommend: "Both of them allow space saving, and increase the pedagogical value of the dictionary, given that they permit the grouping of lexical units with semantic, morphological and etymological relationships" (p. 23). However, the authors suggest that the morphosyntactic attributes of a particular language should be considered in selecting criteria for arranging lemmata and their sublemmata:

The conceptual bond between them (lemmata and sublemmata) is sufficiently solid to legitimate the rejection of an alphabetical arrangement in the case of the English-Spanish part ('non strictly-alphabetical arrangement principle'), whereas, because of the morphosyntactic characteristics of the Spanish language, the arrangement of sublemmas in the Spanish-English part will have to be alphabetical.

By recommending a "half-way position" (p. 43) which consists in using both, even including a thematic introduction to the specific subject field and employing cross-references, the authors show their awareness of the need to restore lexicographical as well as terminological principles in LSP lexicography. This means facilitating easy access to lemmata and data retrieval necessary to perform both receptive and productive tasks with a clear view of the internal conceptual and terminological cohesion of the field in question.

Within the parameters of dictionary structures, the authors also deal with the linguistic phenomenon of polysemy and its treatment in the selected business dictionaries. They are quick to advise (p. 30) that:

Although scholars such as Svensén (1993) claim that polysemy should be analysed in relation to microstructure, we have opted for dealing with this linguistic issue in this chapter (Chapter 2).

This statement suggests that the treatment of polysemy is not analysed in relation to the microstructure. It also suggests that the microstructure does not constitute part of Chapter 2 of the book, although it has already been noted that the discussion on the "form of the lexicographical article" to some extent pertains to the microstructure. In a similar way, the discussion of the structure of

the entry (Subsection 2.5.1 of the book) also deals with some aspects of the microstructure, namely the arrangement of meaning and senses within articles. The same actually applies to the remainder of the chapter where sense differentiation and ordering in the articles corresponding to polysemous lemmata are microstructural aspects. It may be noted that what has been referred to as micro-architecture (Steyn 2004; Gouws and Prinsloo 2005) applies to some extent here. Therefore, for the authors to suggest that they analyse the treatment of polysemy in a different way is somewhat confusing. Of course, it is true that they do so within the parameters of the access structure, but the microstructure also falls within the domain of access structures, particularly the inner access structure.

Focusing on the treatment of polysemy as an element of meaning information in dictionaries, the authors identify sense discrimination and sense ordering as important lexicographical procedures which may increase the pedagogical value of LSP dictionaries. For sense discrimination, a distinction is made between old and modern lexicographical methods. The old method employs punctuation marks such as commas or slashes between the different senses in the semantic subcomment of the lemma. Because it does not provide any information on the usage of the different senses, this technique is not preferred. Instead, as modern method the authors recommend the use of numbers or letters for sense discrimination which clearly indicate the identified senses for each lemma. They are therefore called *polysemy indices* (p. 38). The use of letters or numbers would also indicate the order of the senses, which the authors regard as very important. Although there are various criteria for ordering the senses, the authors note that the limitations of each of these criteria necessitate that lexicographers combine them. The readers should be reminded that this would depend on the attributes of the lemmata being treated.

3.2.2 Definitions

As noted earlier, definitions are identified as the most widely used lexicographical data to explain meaning in dictionaries. Thus, it may be said that they are of great importance as far as meaning representation is concerned. Chapter 3 of the book deals with definitions. Three kinds, namely terminological, encyclopaedic and semantic definitions, are discussed. It would appear from this chapter that the definition types on their own are not of ultimate importance. Firstly, "there are no notable formal differences between the semantic definition and the purely terminological one", both being intentional (p. 54). Similar or even the same basic concepts are used in defining the same term in the selected dictionaries (p. 70). Secondly, some scholars dismiss the very existence of encyclopaedic definitions, as highlighted by the authors' reference (p. 57) to De Bessé (1990, 1997) who speaks of the fallacy of the encyclopaedic definition. What is otherwise more important is the content and style of the definition. In this regard, Fuertes-Olivera and Arribas-Baño (p. 57) criticise

the definition extracted from the terminological database, TERMIUM Plus, on the grounds that its selection of the hyperonym and conceptual characteristics is very poor. However, although they do not necessarily indicate preference for the other types of definitions, they mention the Cobuild definitional patterns for the "simplicity of their formulation, the high amount of information communicated, and the incorporation of the linguistic and situational environment into the definition" (p. 68). This should be read against the background of it being criticised for redundancy and lack of precision especially in highly technical terms.

From a metalexicographical point of view, it is necessary to comment on the theoretical orientation of the analyses and conclusions made in this chapter. There is a clear call for lexicographers to depart from the traditional defining principles. LSP lexicographers are also advised to escape from the trappings of the general theory of terminology which prefers the so-called terminological definition. The analysis of a definition from TERMIUM Plus also highlights that terminological products can be made more user-friendly if they adopt some lexicographical principles. Finally, lexicographers are encouraged to formulate definitions which satisfy the needs of the users for whom a particular dictionary is intended. This is clear when Fuertes-Olivera and Arribas-Baño (p. 70) recommend the Cobuild definitional formula for LSP dictionaries on the grounds that it "lends support to offering more and better syntactical and collocational information, and being of more help to students of specialised languages". For other LSP dictionaries, the needs may be different, but the analyses in *Pedagogical Specialised Lexicography* illustrate the appropriate procedures of meeting the specific needs.

3.2.3 Equivalence

As noted in section 2, Chapter 4 of the book deals with equivalence, which is central to meaning representation in bilingual dictionaries. Accordingly, it was stated that through this chapter, the book contributes to LSP lexicography as well as to bilingual lexicography. The dictionaries of a sample are not just analysed as LSP dictionaries, but also according to the principles of bilingual lexicography which are (bi-)directionality and bi-functionality. Thus, this chapter is somehow unique from the rest of the book because it may also be useful to readers who are more concerned with bilingual lexicography than with LSP lexicography.

The chapter is put into perspective by brief highlights of the interrelationships that constitute bilingual LSP lexicography. Section 4.2 deals with culture and lexicography while Section 4.3 deals with terminology and translation. It is in these sections that the authors conceptualise the challenges encountered in bilingual LSP lexicography. The implications of cultural disparities between two languages covered by a bilingual dictionary are crucial for bilingual LSP lexicography, especially in the so-called culture-dependent subject fields. Busi-

ness studies is one such field alongside others such as, for example, law, music and politics. In these fields, lexicographers usually encounter challenges in establishing equivalence owing to the differences between the systems underlying a particular subject field in two language communities. Specifically, these challenges relate to terminology and translation, whose activities are of great significance in bilingual LSP lexicography.

The main approach to the representation of meaning in bilingual dictionaries is through the provision of translation equivalents of lemmata. The other approaches analysed in the book are examples (translated and untranslated), translated collocations or idioms and meaning discrimination, whose relevance is underscored for the active part of the dictionary. While translation equivalents are essential, they are by no means sufficient in the realisation of all the functions of bilingual dictionaries. This also applies to the pedagogical function of assisting users with the LSP of a foreign language, which underlies the book.

To shed light on the problems associated with lexical equivalence, Fuertes-Olivera and Arribas-Baño (p. 79) refer critically to Neubert (1992) and Duval (1990) who respectively speak of the "triple fiction" and the "fallacy" of equivalence. As indicated earlier, the problems which stem from linguistic anisomorphism and cultural disparities clearly indicate that bilingual lexicography should transcend the notion of equivalence from the lexical level to the functional and pragmatic levels. This is reminiscent of the arguments by other metalexicographers that the mere inclusion of translation equivalents does not automatically make a dictionary bi- or multilingual (Gouws 2004a, Burkhanov 2004). Depending on the functions of a specific dictionary, other data types need to be considered. In the book, meaning discrimination receives the greatest attention, occupying half the space of the entire chapter. It is an important data category, especially on the active part of the dictionary on which the authors focus to evaluate the pedagogical potential of the dictionaries in assisting Spanish-speaking students of business English (p. 73). This also reminds the reader that two parts of a bilingual dictionary are of unequal importance to the respective speakers of the two languages covered by the dictionary. As already noted, this chapter is an important contribution to LSP lexicography as well as bilingual lexicography. As far as bilingual lexicography is concerned, the reader is advised to read yet another recent monograph in the series *Terminology and Lexicography Research and Practice*, namely Yong and Peng's (2007) *Bilingual Lexicography from a Communicative Perspective*.

3.2.4 Examples

Examples are the final type of data category studied in terms of meaning representation in *Pedagogical Specialised Lexicography*. In order to place the discussion into perspective, the authors begin by drawing a dichotomy between explicit and implicit information, already established by other scholars such as Bergenholtz and Tarp (1995). Examples are regarded as data categories which provide

implicit information. Explicit information would be provided by definitions and translation equivalents of lemmata. Significantly, the authors hasten to dismiss claims of superfluity attributed to implicit information, especially given that it exists alongside explicit information. They argue and demonstrate by means of articles extracted from the sample of business dictionaries that examples have a multi-functional role. The following functions are illustrated in the book:

- the meaning in the corresponding sense,
- grammar aspects, either morphological or syntactic,
- phraseology, particularly with regard to collocational information on the lexical items,
- stylistic and, more generally, pragmatic aspects,
- cultural and encyclopaedic aspects, the latter being of particular importance in the case of specialised dictionaries, and
- ideological elements, principally in the philological dictionary.

In addition to the above, Fuertes-Olivera and Arribas-Baño (p. 114) point out that examples in bilingual dictionaries may also facilitate translation. Apparently, the last section before the conclusion of the chapter deals with the role played by examples in bilingual dictionaries. It emerges that bilingual lexicography requires different considerations from monolingual lexicography regarding the use of examples. The compiler needs to be clear whether examples should be provided to illustrate lemmata, equivalents or both and whether they should be provided in both the active and the passive parts of a bilingual dictionary. In this regard, lexicographers are reminded (p. 114-115) that:

In the bilingual dictionary for active use only the equivalents need to be exemplified, since the person consulting the dictionary is a native speaker of the original language. On the other hand, users of the passive bilingual dictionary require no illustration of the equivalent, given their condition of native speaker of the target language. Nonetheless, what may be needed are contextual variants of the equivalent, and it is in the example where this information finds its most suitable mode of expression.

From the foregoing, it may be noted that Fuertes-Olivera and Arribas-Baño effectively underscore the importance of examples as a form of representing meaning in dictionaries. Importance is also attached to the principle which guides the types, sources, presentation and arrangement of examples. It is the general policy regarding examples which seems to be limited, with the authors observing that the compilers of the dictionaries analysed "adhere to very unsystematic lexicographical practices" (p. 133).

In the light of this, the authors make some recommendations regarding the use of examples to represent meaning information in LSP dictionaries. Significantly, the recommendations would seem valid for lexicography in general. It is proposed that a typology of lexicographical examples be established. First-

ly, it is noted that while a quotation and an example are both used for illustrative purposes, they, however, serve quite different purposes. Secondly, examples may be in the form of either sentences or phrases, with the authors preferring the former since it would convey more appropriately the meaning of the lemma. Thirdly, a distinction may be made between invented examples, unedited corpus examples and edited corpus examples. In this connection, the role of a corpus as a source of authentic examples for LSP lexicography is explored and the authors observe that its impact has thus far been very little. Probably this applies to the role of corpora in LSP lexicography in general. While corpora have proven to be the most legitimate sources of lexicographical evidence (Atkins and Rundell 2008), LSP corpora for lexicographical purposes have thus far had a limited appeal, with leading LSP lexicographers arguing that it is unremittingly labour intensive and costly (Bergenholtz 1995: 94). Finally, the authors recommend that lexicographers establish specific criteria for arranging and presenting examples in their dictionaries. This is important, for users have to identify them and relate them easily to the illustrated data categories. The notion of micro-architecture may also be applied here.

4. Technical Aspects of the Book

This discussion of the technical aspects of the book will focus on issues enhancing or inhibiting effective conceptualisation of the problems and presentation of the results, arguments and conclusions of the research presented. These mainly pertain to the scope and methodology of the research as well as the arrangement of the chapters and their sections.

Chapter 1 gives evidence that the book is the result of rigorous academic research. As an introduction, it situates the study within a metalexicographical framework which is cognisant of the latest developments in lexicography as well as other closely related fields. As noted earlier, a clear statement that the study is based on the theory of lexicographical functions is strangely absent, given that the works of Bergenholtz and Tarp, the proponents of this theory, populate the bibliography more than the other references. This indicates an unexpressed acknowledgement towards the theory of lexicographical functions which guides the arguments in the book.

The introduction also gives the reader the perspective of the book by outlining its aims. Although the reader may not be familiar with or even interested in Spanish–English dictionaries, an outline of the aims of the book is enough to gain some insights into the issues addressed. Apparently, practical LSP lexicographers and students of lexicography would find that the issues are of broader relevance to other fields and languages in which LSP dictionaries may be compiled. The representation of meaning is undoubtedly a topic of interest to anyone occupied with lexicography. Undertaking a book-length study of meaning representation is justified by the importance which dictionary consultation attaches to this information.

The methodology the authors employ seems to have been effective for their study. The choice of six dictionaries, all being business dictionaries, is enough to facilitate comparisons and allow for generally acceptable observations and recommendations. Dedicating a chapter to each of the relevant data categories yields more thorough analyses of the issues involved than elsewhere in the available literature. The arguments and conclusions are based on clearly presented data in the form of articles from the selected dictionaries. This methodology is the major technical strength of the book and although only business dictionaries are used as major references, the analyses and conclusions seem to apply to the entirety of LSP lexicography in a way that validates the title of the book.

Unfortunately there are some instances where data are presented to substantiate arguments, yet no elaboration is given to demonstrate how the data clarify arguments. Some sections and chapters, including the very last one, end with dictionary articles presented as examples. Some readers may have problems in relating such examples to the arguments, no matter how sound these are.

The ordering of sections and chapters of a book is also of great importance for the reader to follow all the arguments and relate issues raised in the different parts of the book. Illogical ordering may present readers with serious challenges which may lower the substantial impact of a book. As regards Fuertes-Olivera and Arribas-Baño's book, Chapters 4 and 5 may present problems for readers, but this would depend on how much readers need or require from the book or the chapter. Since the notion of equivalence is central in Chapter 4, it would perhaps have been better if the discussion in Section 4.5 was brought in much earlier. Sections 4.2 and 4.3 would have been deferred as their content provide the broader context of the challenges of dealing with the notion of equivalence, to which the reader is yet to be introduced at this point in the book. The same may be said of Section 5.6 in Chapter 5 which would be more helpful had it been presented earlier to acquaint the reader with the types of examples before discussing their functions, given that some functions may be relevant exclusively to certain types. A logical ordering of chapters and their sections in a book would allow for both selective reading and cover-to-cover reading. While the contents of the book warrant the reading of the book from the first to the last chapter, the reader is likely to have problems as illustrated above.

Finally, referencing is yet another weak technical point of the book. Firstly, in-text referencing reveals an error which needs to be corrected should there be a reprint or second edition of the book. Consider the following quotation (p. 23):

On many occasions the difference between non-strict alphabetical articles and "niche articles" is irrelevant. (Nielsen 1994: 200) Both of them allow space saving.

In this case, it is not at first glance clear which statement is attributed to Nielsen (1994). The bibliographical information is given outside the first sentence. A

second look will show that the reference to Nielsen is meant to refer to the first statement. There are numerous such instances, which undoubtedly are errors, trying the understanding and patience of the reader.

Another referencing issue pertains to the use of footnotes in the book. Footnote 1 is used on page 38. At this point, it does not matter whether it is the first one for Chapter 2 or for the entire book because Chapter 1 does not have any footnotes. When another footnote 1 occurs on page 63 (Chapter 3) and another one on page 84, followed by footnote 2 on the next page (Chapter 4), the reader may think that footnotes are used chapter by chapter. However, on page 144, the reader struggles to realise that footnotes 11 up to 22 belong to the same sequence as footnote 10 on page 103. Footnote 10 is used in Chapter 4 while footnotes 11 up to 22 are used in Chapter 6. On their own, the footnotes may not be a problem until the reader attempts the futile task of searching for note 11, not only in the list of footnotes but also in the illustrative dictionary articles which the footnotes seek to explain. It would also appear that the use of these footnotes does not help the reader very much as Example 60(b) does not have them, leaving the reader confused as to whether the article of the passive side (Example 60(a)) and the one of the active side of the bilingual dictionary should have exactly the same features. While the authors explain and illustrate the contrary in the book, following the principles of bilingual lexicography, they unfortunately miss it technically at this point. Regarding such technical weaknesses which may affect the reader, one could perhaps draw the attention of the authors to the importance of the lexicographical notion of user-friendliness, repeatedly emphasised in the book, but questioned by some (meta)lexicographers who prioritise functionality!

5. Conclusion

The length and detail of this article may have failed to unequivocally describe Fuertes-Olivera and Arribas-Baño's book better than Prof. Sven Tarp's blurb on the back cover of the book. In spite of the concerns raised in the article about some issues as well as the seemingly minor technical aspects, it remains "a scholarly theoretical rigorous account" of the representation of meaning in English and Spanish business dictionaries. It is certainly meant for every lexicographer, especially those working on LSP dictionaries, students of lexicography and all those involved in LSP pedagogy and translation studies. In particular, metalexicography is likely to draw theoretical and methodological insights from it for research on other lexicographical topics.

References

- Atkins, B.T.S. 1996. *Bilingual Dictionaries: Past, Present and Future*. Gellerstam, M., J. Järborg, S.-G. Malmgren, K. Norén, L. Rogström and C.R. Papmehl (Eds.). 1996. *Euralex '96 Proceed-*

- ings: I-II, Papers Submitted to the Seventh EURALEX International Congress on Lexicography in Göteborg, Sweden: 515-546. Gothenburg: Department of Swedish, Göteborg University.*
- Atkins, B.T.S. and M. Rundell.** 2008. *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Bergenholtz, H.** 1995. Material for the Dictionary. Bergenholtz, H. and S. Tarp. 1995: 90-93.
- Bergenholtz, H. and S. Nielsen.** 2006. Subject-field Components as Integrated Parts of LSP Dictionaries. *Terminology* 12 (2): 281-303.
- Bergenholtz, H. and S. Tarp (Eds.).** 1995. *Manual of Specialised Lexicography*. Amsterdam: John Benjamins.
- Bergenholtz H. and S. Tarp.** 2003. Two Opposing Theories: On H.E. Wiegand's Recent Discovery of Lexicographic Functions. *Hermes. Journal of Linguistics* 31: 171-196.
- Burkhanov, I.** 2004. Requirements for an "Ideal" Bilingual L1-L2 Translation Oriented Dictionary. *Lexikos* 14: 17-32.
- De Bessé, B.** 1991. Le contexte terminographique. *Meta* 36(1): 111-120.
- De Bessé, B.** 1997. Terminological Definitions. Wright, S.E. and G. Budin (Eds.). 1997. *Handbook of Terminology Management. Volume I: 63-74. Amsterdam/Philadelphia: John Benjamins.*
- Duval, A.** 1990. Nature et valeur de la traduction dans les dictionnaires bilingues. *Cahiers de Lexicologie* 56(1-2): 69-71.
- Fuertes-Olivera, P.A and A. Arribas-Baño.** 2008. *Pedagogical Specialised Lexicography. The Representation of Meaning in English and Spanish Business Dictionaries*. Amsterdam/Philadelphia: John Benjamins.
- Gouws, R.H.** 2004. Outer Texts in Bilingual Dictionaries. *Lexikos* 14: 67-88.
- Gouws, R.H.** 2004a. Monolingual and Bilingual Learner's Dictionaries. *Lexikos* 14: 264-274.
- Gouws, R.H. and D.J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.
- Hartmann, R.R.K.** 2001. *Teaching and Researching Lexicography*. Harlow: Pearson Education.
- Hartmann, R.R.K.** 2005. Lexicography and its Interdisciplinary Contacts with Special Reference to Onomasiology. *Lexikos* 15: 70-89.
- Hartmann, R.R.K. (Ed.).** 1984. *LEXeter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9-12 September 1983*. Lexicographica. Series Maior 1. Tübingen: Max Niemeyer.
- Hartmann, R.R.K. and G. James.** 1998. *Dictionary of Lexicography*. London/New York: Routledge.
- L'Homme, M.-C.** 2003. Capturing the Lexical Structure in Special Subject Fields with Verbs and Verbal Derivatives. A Model for Specialized Lexicography. *International Journal of Lexicography* 16(4): 403-421.
- McArthur, T.** 1986. *Worlds of Reference. Lexicography, Learning and Language from the Clay Tablets to the Computer*. Cambridge: Cambridge University Press.
- Neubert, A.** 1992. Fact and Fiction of the Bilingual Dictionary. Alvar Ezquerro, M. (Ed.). 1992. *EURALEX '90 Proceedings. IV International Congress: 29-42. Barcelona: VOX/Bibliograf.*
- Nielsen S.** 1994. *The Bilingual LSP Dictionary. Principles and Practice for Legal Language*. Tübingen: Gunter Narr.
- Nielsen, S.** 1999. Mediostuctures in Bilingual LSP Dictionaries. *Lexicographica. International Annual for Lexicography* 15: 91-113.
- Pederson, J.** 1995. Systematic Classification. Bergenholtz, H. and S. Tarp. 1995: 83-93.

- Sager, J.C.** 1984. Terminology and the Technical Dictionary. Hartmann, R.R.K. (Ed.). 1984: 315-326.
- Sager, J.C.** 1996. *A Practical Course in Terminology Processing*. Amsterdam: John Benjamins.
- Steyn, M.** 2004. The Access Structure in Learner's Dictionaries. *Lexikos* 14: 275-298.
- Svensén, B.** 1993. *Practical Lexicography. Principles and Methods of Dictionary Making*. Oxford: Oxford University Press.
- Tarp, S.** 1999. Theoretical Foundations of the So-called Cross-reference Structures. Wiegand, H.E. (Ed.). 1999. *Mediostrukturen in Fachwörterbüchern*. *Lexicographica* 15: 114-137.
- Tarp, S.** 2000. Theoretical Challenges to Practical Specialised Lexicography. *Lexikos* 10: 189-208.
- Tarp, S.** 2002. Basic Elements of Lexicographic Theory/Éléments de base de la théorie lexicographique. Emejulu, J. (Ed.). 2003. *Éléments de lexicographie gabonaise. Tome II*: 7-35. New York: Jimacs-Hillman.
- Tarp, S.** 2004. Basic Problems of Learners' Lexicography. *Lexikos* 14: 222-252.
- 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.
- Temmerman, R.** 2000. *Towards New Ways of Terminology Description. The Sociocognitive Approach*. Amsterdam/Philadelphia: John Benjamins.
- Tuck, A.** 1993. *Oxford Dictionary of Business English for Learners of English*. Oxford: Oxford University Press.
- Wiegand, H.E.** 1984. On the Structure and Contents of a General Theory of Lexicography. Hartmann, R.R.K. (Ed.). 1984: 13-30.
- Wiegand, H.E.** 1996. A Theory of Lexicographic Texts. An Overview. *South African Journal of Linguistics* 14 (4): 134-149.
- Worsch, W.** 1999. Recent Trends in Publishing Bilingual Learner's Dictionaries. Hartmann, R.R.K. (Ed.). 1999. *Recommendations, National Reports and Thematic Reports from the TNP Sub-Project 9: Dictionaries*: 99-107. Berlin: Freie Universität.
- Yong, H. and J. Peng.** 2007. *Bilingual Lexicography from a Communicative Perspective*. Amsterdam/Philadelphia: John Benjamins.

John Considine. *Dictionaries in Early Modern Europe: Lexicography and the Making of Heritage*. 2008, xiv + 393 pp. ISBN: 9780521886741. Cambridge: Cambridge University Press. Price: £55 / \$99.

Dictionary history or history of lexicography does not belong to one of the most studied metalexigraphic disciplines, although the International Society for Historical Lexicography and Lexicology regularly convenes conferences and publishes proceedings, and much literature (mainly in the Western world and mainly dealing with Western lexicography) has been published during the last five decades. Furthermore most of the work done deals with the subject quite specifically. General or versatile monographs are rather rare. Because of this, John Considine's *Dictionaries in Early Modern Europe* is a long-awaited and long-overdue work.

The overall subject demarcation of the book is partly geographically, partly linguistically motivated (cf. p. 17): Only Western European lexicography is dealt with, and also only partly, i.e. primarily English, German and French lexicography as well as Western European dictionaries with Latin (classical Latin and post-classical Latin) and Greek (ancient Greek and Byzantine Greek). Peripherally also Castilian (that is Spanish), Italian, Dutch, Danish and Swedish lexicography are treated and even more peripherally (Western) European dictionaries with Hebrew. The reason for this demarcation may be that these eight languages — i.e. English, German, French, Castilian, Italian, Dutch, Danish and Swedish — were the first eight vernaculars in the world to be the subject of lexicographical codification in printed dictionaries.

The specific time demarcation is the early modern period, i.e. from approximately 1500 to the French Revolution (1789–1799). Therefore it is difficult to understand why European lexicography from the 1700s is not discussed at all. The book actually deals only with lexicography from the 1500s and the 1600s.

Completeness is not pursued; many significant dictionaries are just mentioned in passing or not at all (cf. p. 17). The reader should not expect a general history of European dictionaries (p. 17):

This is not a general history of western European lexicography but a discussion of one group of themes in lexicographical thought, based on a selection of case-studies.

About the specific subject demarcation, i.e. the demarcation of the dictionary as a document type, the author states (p. 17):

I have not confined myself to the discussion of dictionaries in any narrow sense of the word, but have also considered a number of short wordlists and other studies of words.

But the argument also transgresses this demarcation when the author says (p. 18):

Lexicographical thought, which is the subject of this book, has not always been expressed in the writing of dictionaries; Guillaume Budé, who never published a dictionary but had major influence on lexicography, is a good example.

However, by far the major part (based on an estimate of about 90 per cent) of the works discussed (based on an estimate of between 150 and 300 titles) are regular dictionaries or glossaries.

The number of works discussed is relatively high, but it is only a very minor part of the actual number of dictionaries and glossaries produced in early modern Europe. In Europe, in the years from 1467 to 1600 (as far as is known), 858 German dictionaries and glossaries were published. And in the course of the 1600s, a further 1 150 German dictionaries appeared.

With the exception of the outside matter — consisting of the preliminaries, acknowledgements, information about conventions and abbreviations, conclusion, afterword, bibliography and index — the book contains eight chapters. Chapter one, having also the function of an introduction, deals with the association between dictionaries and ideas of heritage, i.e. cultural, intellectual, historical and linguistic heritage. Chapters two and three treat the heritages of the classical world and focus inter alia on the first dictionaries of the age of print; chapter two deals mostly with the work of Robert Estienne from the 1530s and the 1540s; chapter three is devoted solely to his son Henri Estienne and his work from the end of the 1540s to the beginning of the 1590s. Chapters four, five and six deal with the heritages of the early medieval Germanic world; chapter four discusses the rediscovery of the vernacular heritages and primarily treats Germany and the Netherlands in the sixteenth century and the first ten or twenty years of the seventeenth; chapter five deals with England from the early sixteenth century to around 1650; chapter six discusses England and Scandinavia and the learned Franciscus Junius from the second part of the seventeenth century. Chapter seven deals with Charles du Cange, his outside world and lexicographical activities from the 1670s and onwards and discusses the seventeenth century's relation to the heritages of the worlds of post-classical Latinism and Byzantine Grecism. Chapter eight comments on three ways in which seventeenth-century lexicographical thought built on and went beyond these heritages.

Dictionaries in Early Modern Europe is an erudite book written by a well-read scholar — a true scholar in the very positive sense of the word; a philologist interested in book history and historical bibliography. Therefore the reader will not mind the extensive use of footnotes (chapter five for instance, which is only less than 50 pages long, contains 223 footnotes, and chapter four, which is 55 pages long, contains 247 footnotes). But when the total of 1 271 footnotes in the book for a greater part contain — sometimes quite long — quotations mostly in Latin, moreover in classical Latin and in post-classical Latin, and further-

more in Italian, French, German, ancient Greek, Old High German, Old English as well as other Germanic languages from medieval times of which the greater part (based on an estimate of about 80–90 per cent) are not translated, the — obviously less learned — reader has the feeling that he is deprived of a whole book presented concurrently with the one he is reading. That quotations in modern German and French, and furthermore German, French and English from the 1500s and 1600s are left untranslated is to a certain extent acceptable, for they should mostly be understandable to a contemporary reader.

The book impresses with the extensive source materials on which it is based, as the 1 271 footnotes and the comprehensive 52 pages long and carefully written bibliography bear witness. Not only printed sources from earlier times are quoted but also non-printed material, manuscripts from the 1500s and 1600s, of which quite a few today only exist in one or a few copies. It is a highly praiseworthy accomplishment, that John Considine, who is an associate professor in the Department of English and Film Studies at the University of Alberta in Canada, has managed to gain access to these rare manuscripts at different European national libraries.

Considine's insight and understanding is impressive. For instance, in the chapter about Scandinavian dictionaries he eschews Jacoby (1990), whom he does not cite once, but instead resorts to Haugen (1984) and Molbech (1826). (About Jacoby (1990), who is highly untrustworthy and full of errors, cf. Gundersen (1994).) It is incorrect though, as is mentioned on page 236, that *Dictionarium Herlovianum* (1626) is the first known Danish–Latin dictionary in print — *Hortulus synonymorum* (1520) is in fact the first.

In a few cases, the book gives an uneven treatment of details: At one place several pages are devoted to an extensive exposition of the price of a given dictionary in the 1500s (parenthetically: which contemporary reader, who is not a historian, actually has sufficient knowledge of the monetary units and values at that time to interpret the given information?), while at another place elementary and important facts are lacking. When, for instance, did *Forensia* (a law dictionary from probably the 1500s) appear? Nothing in the book suggests that it is not known, rather the impression is created that the author has neglected to inform the reader.

It is regrettable, that the relations between textual criticism and lexicography are only dealt with very briefly and superficially (by and large only on pages 41 and 316). As is evident from page 316 of the conclusion, this aspect is very interesting and much more could have been presented on it:

The earliest forms or cognates of the European vernaculars — languages such as Old French, Old High German, Gothic and Old English — could not be registered with the same lexicographical techniques as the classical languages. Their texts were not as readily accessible, and they were not as well edited: indeed, the people who studied them encountered a circular problem, for the making of dictionaries depends on the availability of edited texts, and the editing of texts in obsolete language varieties depends on the availability of dictionaries. The early history of lexicography of these language varieties therefore looks like an anti-

climax after that of the great ordered registrations of the classical lexical heritages: a wordlist here, a glossary to a handful of freshly and imperfectly edited law-codes there, a few manuscript pages of onomastical notes, a series of dictionary projects that did not reach publication.

The most significant and recurring deficiency of the book is the general lack of metalexigraphic reflection. The reader has to reach page 85 before finding for the first time a metalexigraphic term (*macrostructure*); only very much later on page 231 he finds another term (*cross-references*), and eventually on page 315 *macrostructure* once again occurs. The use of metalexigraphic vocabulary is almost wholly absent. *Dictionaries in Early Modern Europe* is evidently written by a philologist with a profound interest in historical bibliography for other philologists with the same interest. It could also have been written by a learned librarian. The book will certainly be of great value and use primarily for philologists, librarians, book historians, historical bibliographers, and secondarily for historians and cultural historians. Dictionary historians, in so far as these regard themselves metalexigraphers, and metalexigraphers in general, will find the book lacking in metalexigraphic reflection.

It is a question of which fundamental conception one has (or does not have) of the document type called *dictionary* or *lexicographic reference work*. The author misses the fundamental realisation that a dictionary is a reference tool, a tool conceived for information retrieval to help specific users in specific user situations (regardless of whether it concerns the historic present, the Renaissance, or the Antiquity). The book might as well have dealt with any other type of document than the dictionary or the lexicographic reference work.

Now and then the reader finds a valuable quotation, for instance that of Robert Estienne from 1536 (translated by Considine, p. 47): 'What really earns the praise which is ours by right is, I say, our diligence in arranging our material, each word in its right order', which could have been used each time when lexicographic structures and, among these, access structures are mentioned — but it could not have been put into greater use, because the author does not reflect on access structures at all.

The overall aim of the book is given in the conclusion (p. 314-315):

I have tried to demonstrate in this book that an understanding of lexicography as shaped by ideas of heritage frees the history of dictionaries from too heavy a dependence on certain potentially arid kinds of narrative of the form '61 per cent of the entries in Y derive from entries in X', and engages it with broader and more humane questions about lexicology, the history of linguistics, the history of learned culture, indeed the history of culture in general.

The idea, which is repeated throughout the book, that dictionaries tell stories of personal and cultural heritage, is from a metalexigraphic point of view partly problematic, partly irrelevant. Regarding the former: Who has ever thought, that dictionary history or history of lexicography is about establishing whether '[so and so many] per cent of the entries in Y derive from entries in X'? Re-

garding the latter: It is self-evident that history of lexicography is history of culture in general (or on the whole, history in general, in so far as with history is meant the state of flux of all conditions of human life: institutions, norms, ideas, trends etc.). One can also argue for the fact, that the history of lexicography and the history of linguistics partly or at least in the 1500s, 1600s and 1700s went hand in hand. But whether dictionaries tell stories of personal and cultural heritage is from a metalexigraphic point of view unimportant. Most objects, if not all, created by human hand tell stories of either personal heritage, or cultural heritage or both at the same time. Therefore it is of no importance that dictionaries also have this quality or feature. It would have been interesting if this book on dictionaries in early modern Europe had concentrated on some of the core aspects of theoretical and practical lexicography: (1) user profiling or user adjustment; (2) data adjustment; (3) access adjustment; and (4) functionality. In other words: What ideas did lexicographers have in early modern Europe, when it came to user profiling, i.e. the question of how a given dictionary which is to be made for a given user group in accordance with optimising the genuine function of the dictionary as a tool for information retrieval should be adjusted in the best possible way? And how were these ideas put into practice? What ideas did lexicographers have in early modern Europe, when it came to data adjustment, i.e. the question of how the given data should be treated in the most optimal way in accordance with a user group? And how were these ideas put into practice? What ideas did lexicographers have in early modern Europe, when it came to access adjustment, i.e. the question of how the given data should most appropriately be made accessible in accordance with the subject or the language dealt with and the user group? And how were these ideas put into practice? What ideas, if any, did lexicographers have in early modern Europe about the functionality of dictionaries in accordance with the user needs that were to be accommodated? None of these questions are responded to in the book, because its aim is to describe the history of culture as manifested in dictionaries.

The book is provided with a meticulously written index, which is laid out in traditional flush-and-hang style with indented subentries. By page turns the main entry is repeated and marked with '(cont.)', so that the user can keep a sense of perspective; see for instance p. 392 'Swedish language (cont.)', which is an entry continued from the previous page. The index is six per cent (19 pages) against the body text (322 pages) and is therefore a so-called 'six per cent index', which actually is quite enough (cf. Mulvany 2005: 69-73). However, it displays a few unfortunate gaps: *Observationes in Ciceronem*, a dictionary from 1535, which is dealt with on page 63 and is important because it is a work of the so-called 'Ciceronianism' (a linguistic school from the Renaissance and early modern Europe, which maintained a rather conservative language view and reluctance towards post-classical Latinism), is not indexed either as main entry under its title or as subentry under the author's name (Mario Nizzoli, also known as Nizolius).

It is furthermore unfortunate, that double entries do not appear consistently, i.e.:

Celt'hellenisme ...

...

Trippault, Léon

Celt'hellenisme ...

Sometimes book titles are indexed having main entry status; sometimes they have only subentry status and are therefore only to be found under their respective authors. *Celt'hellenisme*, an etymological dictionary from 1582, is not indexed; it can only be found under the author's name (Trippault). The unfinished *Teütsche Sprach und Weißheit*, which was published by Georg Henisch in 1616, is not indexed either, not even under its Latin title *Thesaurus linguae et sapientiae germanicae* — only under the author's name. This is also the case with *Expositiones terminorum legum anglorum* (1520), a law dictionary written by John Rastell, and with many other works.

The importance of double entries is well-known and accepted amongst indexers — cf. for instance Wellisch (1995: 151-155), Booth (2001: 115) and Mulvany (2005: 83-84) — but can also be vindicated as follows: On page 127 one reads about a dictionary of plant names from the 1500s, that is a specialised dictionary, and one now remembers having read something previously in the book about another specialized dictionary, a law dictionary, with the title *Forensia*. One now wishes to reread the pages about *Forensia* in order to approach closer to some kind of an outline of specialised lexicography from the 1500s, but one cannot find the pages in question, because *Forensia* is not indexed. Therefore one has to skim the pages backwards one by one (from page 127 until one finds the book title *Forensia*, which happens on page 36, 34 and 32); here one can see, that a person named Guillaume Budé has written the book. In the index one indeed finds 'Budé, Guillaume' and as subentry here '*Forensia* 32, 34, 36'.

All in all, *Dictionaries in Early Modern Europe* is an exceptionally erudite, thorough and trustworthy book, written by a learned scholar, indisputably of great value and use for philologists, book historians and historical bibliographers. For the greater part, however, the book is of little value for metalexigraphers.

References

Booth, Pat F. 2001. *Indexing: The Manual of Good Practice*. München: K.G. Saur.

Gundersen, Dag. 1994. [Review] Michael Jacoby: *Historische Lexikologie zum nordgermanischen Raum: Lexika als Kultur- und Sprachdokumente zwischen Mittelalter und Neuzeit. Einflüsse von Toledo bis Paris, von London bis Berlin*. Wiesbaden: Otto Harrassowitz 1990. *LexicoNordica* 1: 253-261.

- Haugen, Einar.** 1984. *Lexicography in Scandinavia*. Haugen, Eva L. 1984. *A Bibliography of Scandinavian Dictionaries*: 1-61. White Plains/New York: Kraus International Publications.
- Jacoby, Michael.** 1990. *Historische Lexikologie zum nordgermanischen Raum: Lexika als Kultur- und Sprachdokumente zwischen Mittelalter und Neuzeit. Einflüsse von Toledo bis Paris, von London bis Berlin*. Wiesbaden: Otto Harrassowitz.
- Molbech, Christian.** 1826. Historisk Udsigt over de danske Ordbogsarbeider i det 17de og 18de Aarhundrede af Geheimeraad Matthias Moth, Conferentsraad Fr. Rostgaard og Statsraad J. Langebek. *Nye Danske Magazin* 5(4): 241-288.
- Mulvany, Nancy C.** 2005. *Indexing Books*. Second edition. Chicago/London: University of Chicago Press.
- Wellisch, Hans H.** 1995. *Indexing from A to Z*. Second edition, revised and enlarged. New York/Dublin: H.W. Wilson.

Loránd-Levente Pálfi
Centre for Lexicography
Aarhus School of Business
University of Aarhus
Denmark
(llp@asb.dk)

A.P. Cowie (Editor). *The Oxford History of English Lexicography. Volume I: General-purpose Dictionaries. Volume II: Specialized Dictionaries.* 2009. Volume I: xviii + 467 pp., Volume II: xix + 551 pp. ISBN Volume I-II: 978-0-19-928562-4. Volume I: 978-0-19-928560-0. Volume II: 978-0-19-928561-7. Oxford: Oxford University Press. Price: £140.

According to the blurb, *The Oxford History of English Lexicography* (volumes I-II) presents 'the fullest account yet published of the lexicography of English from its origins in medieval glosses, through its rapid development in the eighteenth century, to a fully-established high-tech industry that is as reliant as ever on learning and scholarship'. The term 'English lexicography' is interpreted broadly to embrace dictionaries not only of British English but also its national varieties such as Scots, American English, the varieties of English spoken in Australia, Canada, India, New Zealand, South Africa and the English-based Creoles of the Caribbean (cf. vol. I, p. 1). In addition, throughout the topical presentation numerous references are made to developments in Europe and elsewhere which have influenced the course of English lexicography.

In addition to other preliminaries and the 15 contributions (pp. 15-409), volume I contains the following nine components: (1) Contents (pp. v-vi), (2) Preface (pp. vii-viii), (3) Contents of Volume II (pp. ix-x), (4) Notes on Contributors (pp. xi-xiv), (5) List of Illustrations (p. xv), (6) List of Abbreviations (pp. xvi-xviii), (7) Introduction (pp. 1-14), (8) References (pp. 410-450), and (9) Index (pp. 451-467). The 15 contributions are grouped in two: 'Part I: Early Glossaries; Bilingual and Multilingual Dictionaries' (pp. 15-128), which contains five contributions, and 'Part II: The History of English Monolingual Dictionaries' (pp. 129-409), which comprises 10 contributions.

In addition to other preliminaries and the 17 contributions (pp. 23-478), volume II contains the following nine components: (1) Contents (pp. v-vi), (2) Preface (pp. vii-viii), (3) Contents of Volume I (pp. ix-x), (4) Notes on Contributors (pp. xi-xiv), (5) List of Illustrations (pp. xv-xvi), (6) List of Abbreviations (pp. xvii-xix), (7) Introduction (pp. 1-21), (8) References (pp. 479-533), and (9) Index (pp. 535-551). The 17 contributions are grouped in two: 'Part I: Dictionaries Specialized According to Ordering of Entries, Topical or Linguistic Content, or Speech Community' (pp. 23-336), which contains eleven contributions, and 'Part II: Dictionaries Specialized According to Uses and Users' (pp. 337-478), which contains six contributions.

Both volumes contain well-constructed and informative introductions by the editor, A.P. Cowie. In the course of the 14 and 21 pp. respectively, the author admirably succeeds in capturing and presenting the salient points of the two constituent volumes.

In volume I, part I, Hans Sauer explores the early development of glosses, glossaries and dictionaries in the Medieval Period; Janet Bately gives an outline of bilingual and multilingual dictionaries of the Renaissance and early seventeenth century; Monique C. Cormier writes about bilingual dictionaries of the

late seventeenth and eighteenth centuries; Carla Mareello deals with bilingual dictionaries of the nineteenth and twentieth centuries; and Donna M.T.Cr. Farina and George Durman contribute a chapter on bilingual dictionaries of English and Russian from the eighteenth to the twentieth centuries. In part II, N.E. Osselton explores the early development of the English monolingual dictionary (seventeenth and early eighteenth centuries); Allen Reddick writes about the work of Samuel Johnson and Charles Richardson; Sidney I. Landau gives an extensive outline of major American dictionaries (including sections on the work of Noah Webster and that of his successors in the USA); Lynda Mugglestone contributes a chapter on the *Oxford English Dictionary* (1879–1928), henceforth OED; Charlotte Brewer writes about the OED Supplements; Richard W. Bailey contributes a chapter on dictionaries of national and regional varieties of English; Margaret Dareau and Iseabail Macleod write about dictionaries of Scots; Michael Adams deals with period dictionaries (dictionaries of Old and Middle English); Jeannette Allsopp writes about dictionaries of Caribbean English; and Edmund Weiner contributes a chapter on the computerization of the OED.

In volume II, part I, Werner Hüllen gives an outline of dictionaries of synonyms and thesauri; Michael Rand Hoare contributes an extensive chapter on scientific and technical dictionaries; Carole Hough writes about dictionaries of place-names; Patrick Hanks writes about dictionaries of personal names; Joan C. Beal's contribution deals with pronouncing dictionaries in the eighteenth and early nineteenth centuries; Beverley Collins and Inger M. Mees contribute an extensive chapter on pronouncing dictionaries from the mid-nineteenth century to the present day; Thomas Herbst and Michael Klotz write about syntagmatic and phraseological dictionaries; Elizabeth Knowles contributes with a chapter on dictionaries of quotations; Anatoly Liberman deals with English etymological dictionaries; Robert Penhallurick deals with dialect dictionaries; and Julie Coleman contributes a chapter on dictionaries of slang and cant. In part II, Robert Allen writes about dictionaries of usage; Sidney I. Landau contributes a chapter on American collegiate dictionaries; A.P. Cowie writes about the earliest foreign learners' dictionaries; Thierry Fontenelle deals with linguistic research and learners' dictionaries; Rosamund Moon gives an account of the Cobuild project; and Hilary Nesi writes about dictionaries in electronic form.

It is regrettable that in volume I the first 1100 years, the period spanning ca. 700–1800, is allotted only approximately 100 pages (pp. 17–40, 41–64, 65–85 and 131–154), while the remaining 300 pages cover the nineteenth and twentieth centuries. Does this perhaps reflect a tendency that is prevalent in our day, especially in Western Europe, to focus too closely on current and recent developments? This trend is exemplified in the new Danish literary history, *Dansk Litteraturs Historie*, which is planned as a five-volume work, the first of which appeared in 2006 and covers the period 1100–1800, while the remaining four are to be devoted to the nineteenth and twentieth centuries (with all the vol-

umes to be of identical size). A similar bias is found in the recent 17-volume history of Denmark, *Gyldendal og Politikens Danmarkshistorie* (2002–2005), which devotes a total of five volumes (vols. 12–16) to the twentieth century, while the nineteenth is allotted two volumes (vols. 10–11), the eighteenth one volume (vol. 9), and the period 200–1250 is covered in three volumes (vols. 2–4), and, again, the volumes are identical in size.

Is this perhaps, on the other hand, to be explained by a lack of research in the history of lexicography in Europe, or in the world at large: Is there simply not known more, and therefore little to write about? Alternatively, was the editor unable to recruit a sufficient corps of researchers with relevant expertise in lexicographic history? Whatever the case, it is noticeable that the contributions on the oldest and early periods (ca. 700–1700) are of variable quality. Hans Sauer's contribution on medieval lexicography (vol. I, pp. 17–40), for example, and N.E. Osselton's piece on monolingual lexicography in the seventeenth and early eighteenth centuries (vol. I, pp. 131–154) reflect breadth, depth, thoroughness and critical analysis, while the chapter on the Renaissance and early seventeenth century (vol. I, pp. 41–64), for example, is superficial, disorganised and lacking in focus (a large number of dictionaries are mentioned, but little real insight is given into any of them).

It is also rather incomprehensible that a whole chapter (vol. I, pp. 105–128) is devoted to bilingual dictionaries of English and Russian — why English and Russian? What special role can Russian, as opposed to other world languages, have in a history of English lexicography? The contributions dealing with the period post-1700 are all, on the other hand, and with few exceptions, of a high standard (praiseworthy are, for example, Allen Reddick's chapter on Samuel Johnson's and Charles Richardson's work, Sidney I. Landau's contribution on American dictionaries, the chapters on the OED etc.).

The Oxford History of English Lexicography is an anthology, not a monograph. This presents both advantages and disadvantages. The principal advantage is that the work could be produced within a short time, which enables the most recent research to be incorporated. Had one or two academics been made responsible for the whole work, it might have had a gestation period of 10–15 years, with the result that the chapters written first could well be outdated by the time they reached publication. There is a further advantage to be gained since in the case of most specialist areas a high standard of writing and accuracy of detail is ensured by enlisting experts and experts alone, to produce the relevant chapters. Had only one or two metalexicographers written the entire work, it would hardly have reached the high standard it demonstrates, and many errors of detail would probably have passed unnoticed.

The disadvantages that an anthology entails are not easily disguised. The easiest to uncover is a lack of terminological consistency, even in the area of quite elementary, central lexicographic concepts or entities. Thus, for example, in the course of the first volume at least seven different expressions are used for 'lemma': (1) *lemma* (e.g. vol. I, p. 21), (2) *headword* (e.g. vol. I, p. 134), (3) *word-*

heading (e.g. vol. I, p. 172), (4) *entry* (e.g. vol. I, p. 184),¹ (5) *entry word* (e.g. vol. I, p. 190), (6) *word* (e.g. vol. I, p. 193), and (7) *term* (e.g. vol. I, p. 207). A second problem is that cohesion between the individual chapters is hardly greater than that to be found in a carefully-edited theme-volume ('History of Lexicography' as theme, for example) of an academic journal. Where is a *leitmotif* to be found? What are the basic assumptions or starting-points? What is the theoretical groundwork? What underlying approach is there to narrowing down the text genre or document type that lexicographic reference works represent? What is the overall view of lexicography — and how can one expect authors of individual contributions to achieve a consensus on such issues when they demonstrate no common ground on use when dealing with the lemma? It is generally accepted that there are diverging views as to what constitutes a lexicographic reference work — just as there are varying opinions on how lexicography itself is to be defined. Landau (2001), for example, regards lexicography as an art and a craft, and therefore apparently not as an academic discipline, whereas Michael Jacoby, who is responsible for a 1 000 page survey of Scandinavian dictionaries (Jacoby 1990), treats his overview as being an example of a *lexicological* work, despite the fact that others maintain that his book is a *lexicographic* work (cf. Gundersen 1994). Lindemann (1994: 2) distinguishes between *glossography*, knowledge and theorizing about word-lists, and *lexicography*, knowledge and theorizing about dictionaries and lexica. A widespread view, especially amongst linguists, is exemplified by Meier (2003: 307), who regards lexicography as a branch of *applied linguistics*. Wierzbicka (1985: 5) allows lexicography no scientific basis whatsoever, whereas Henriksen (1992) suggests a merger of *lexicography* and *encyclopaedism*, resulting in a synthesis to be called 'reference science' (Danish: *referenciologi*)! Finally there are Bergenholtz and Tarp (2002, 2003 *et passim*), Tarp (2008a, 2008b *et passim*), and Pálfi et al. (2008), who have yet other, but compatible, viewpoints. What is understood in *The Oxford History of English Lexicography* by 'lexicography' and what is the concept of the nature of the document type to which lexicographic reference works belong?

A second, and more important concern regarding lexicographic history, which is given no attention whatsoever, is the following central crux: To what form of historiography do the various contributions adhere? There are diverging approaches and assumptions on which the presentation of any form of historical development can be based, cf. Koerner (1978), Grotzsch (1982), Brekle (1985) and Dörner (1991). The range can be narrowed down to three:

- The historical-positivistic approach that seeks, monumentally, to collect as much data as possible and regards the described development as a constant progression towards a higher level.
- The selective approach which picks out the most important data and presents it without a (preconceived) evaluating frame; this is often associated with everyman's general interest in history: history is exciting.

- The selective approach in which especially significant data is critically selected and presented within an evaluative assessment frame which is often presented against a background of logic: one can learn from history.

Bergenholtz and Pálfi (2008) adopt the last of these three approaches but do not exclude elements of the second. In contrast to what emerges from some, but by no means all, of the chapters in the work under review, Bergenholtz and Pálfi (2008: 185; trsl. LLP) do not accept that lexicography has moved in curves constantly attaining higher levels:

On the contrary, we believe we are able to observe that lexicography in the period of the Enlightenment in many ways reached a level that subsequent dictionary writing was unable to attain [...]. When we take contemporary lexicography into consideration, we will not be motivated by a pattern of criticism or praise for the past, but rather by criticism of contemporary endeavours which have not always learned from earlier lexicographical mistakes or, indeed, predecessors' advantageous considerations.

On reading the chapters in these two volumes, it is clear that there is no shared view as to how history is to be written or approached.

It is also unfortunate that there are no theoretical generalisations. Neither the differences between Anglo-Saxon and Continental European lexicography (cf. Rothe 2001), nor the internal similarities within Anglo-Saxon lexicography (cf. Rothe 2001) are, for example, given any mention. It is questionable whether such elements can feature within the concept of an anthology, for which of the contributors should raise them? There should at least have been an editorial afterword or epilogue, as a supplement to the editor's introductions to the individual volumes — this could have highlighted and reviewed a number of theoretical issues.

The two indexes are laid out in traditional flush-and-hang style with indented sub-entries. At page turns the main entry is repeated and marked with '(cont.)', so that the user can retain a sense of perspective; see for instance 'proper names (cont.)' in vol. I, p. 464, which is an entry continued from the previous page. The index in vol. I is four per cent (17 pp.) of the body of the text (411 pp.) and is therefore a so-called four per cent index, whereas the index in vol. II is three and a half per cent (17 pp.) of the body of the text (480 pp.) and is therefore a three and a half per cent index. Neither of the indexes is sufficient in size and depth in relation to the density of information in the volumes (on estimating the size of an index cf. Mulvany 2005: 69-73). Both indexes display many unfortunate gaps; just to mention a few examples from the index in vol. I: *Libri de significatu verborum* (from the first century BC), which is mentioned on p. 29 and is important because it may be the very first dictionary from European Antiquity, is not indexed, not as a main entry with its title nor as sub-entry under the author's name (Verrius Flaccus); the author (Verrius

Flaccus), mentioned on p. 29, is not indexed either; *Etymologiae* (an encyclopedia from the sixth or seventh century), which is mentioned on p. 29 is not indexed, not as a main entry with its title nor as sub-entry under the author's name (Isidore of Seville); the author, mentioned on p. 29, is not indexed either; John Garland's *Synonyma* (a dictionary from 1483), mentioned on p. 37, is also missing, and so is its author, mentioned on p. 37; *Afrikaans* is not indexed, which means that the reader (who also receives no help from the list of contents) has no means of quickly establishing that bilingual dictionaries of English and Afrikaans are discussed in connection with bilingual dictionaries of English and Dutch (p. 101); *Patriot woordeboek/Patriot Dictionary* (1902–1904), mentioned on p. 101, is missing, as is its author, S. J. du Toit (p. 101); 'validation of words' (e.g. pp. 141–142, 147–148, 154) is missing, nor is it listed under 'words' (as 'words, validation of'). The number of missing entries in the two indexes is strikingly high, but responsibility for this probably lies with the publisher, rather than the indexer; the fact that the two indexes are identical in size (precisely 17 pp.), despite the fact that volume II is considerably larger than volume I, certainly points in this direction. The size of the indexes was probably dictated in advance, or they were trimmed by the publisher. This is regrettable because a sizeable amount of the mass of information presented by many experts is thereby not readily available, or, at worst, actually disappears (in the field of LIS (Library and Information Science) it is held that information that cannot be retraced or received can just as well be regarded as lost or non-existent). Especially in a work of the breadth of *The Oxford History of English Lexicography* it is very important that the text is fully indexed, since most readers, for lack of time, will probably never read through the whole work — and certainly not more than once to retrace non-indexed material which they believe they have encountered on the first read.

All in all, *The Oxford History of English Lexicography*, naturally enough, is a very useful publication and a rich source of information, which, despite the criticisms raised above, must be regarded as invaluable for metalexigraphers (primarily dictionary historians), first and foremost in Europe and the English-speaking parts of the world.²

Notes

1. The term *entry* is admittedly also used in the sense 'dictionary article' (e.g. p. 179) and it can well be argued that the word is employed in this sense on p. 184, and possibly elsewhere, and not in the sense of 'lemma'.
2. I am very grateful to Christopher Sanders for translating parts of this review from Danish.

References

- Bergenholtz, H. and L.-L. Pálfi. 2008. Danmarks første nationalordbog: Om Videnskabernes Selskabs Ordbog I–VIII (1793–1905). *Fund og Forskning* 47: 181–222.

- Bergenholtz, H. and S. Tarp.** 2002. Die moderne lexikographische Funktionslehre. Diskussionsbeitrag zu neuen und alten Paradigmen, die Wörterbücher als Gebrauchsgegenstände verstehen. *Lexicographica* 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.
- Brekle, H.E.** 1985. Was heißt und zu welchem Ende studiert man Sprachwissenschaftsgeschichte? Brekle, H.E. (Ed.). *Einführung in die Geschichte der Sprachwissenschaft*: 1-26. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Dörner, A.** 1991. Von der Gelehrtenrepublik zur Nationalbewegung. Kontexte der Sprachforschung in Deutschland an der Schwelle vom 18. zum 19. Jahrhundert. Bergenholtz, H., A. Dörner, R. Karatas, and G. Meder (Eds.). *Wegweiser durch die Grammatik von Heinrich Bauer. Verzeichnisse und Erläuterungen*: 9-60. Berlin/New York: Walter de Gruyter.
- Grotsch, K.** 1982. *Sprachwissenschaftsgeschichtsschreibung. Ein Beitrag zur Kritik und zur historischen und methodologischen Selbstvergewisserung der Disziplin*. Göppinger Arbeiten zur Germanistik, 352. Göppingen: Kümmerle.
- Gundersen, D.** 1994. Michael Jacoby: *Historische Lexikologie zum nordgermanischen Raum: Lexika als Kultur- und Sprachdokumente zwischen Mittelalter und Neuzeit. Einflüsse von Toledo bis Paris, von London bis Berlin*. In Kommission bei Otto Harrassowitz. Wiesbaden 1990. *LexicoNordica* 1: 253-261.
- Henriksen, P.** 1992. Forholdet mellem encyclopedologi og leksikografi. Fjeld, R.V. (Ed.). *Nordiske Studier i Leksikografi 1: Rapport fra Konferanse om Leksikografi i Norden 28.-31. mai 1991*: 219-227. Oslo: Nordisk Forening for Leksikografi.
- Jacoby, M.** 1990. *Historische Lexikologie zum nordgermanischen Raum: Lexika als Kultur- und Sprachdokumente zwischen Mittelalter und Neuzeit. Einflüsse von Toledo bis Paris, von London bis Berlin*. Wiesbaden: Otto Harrassowitz.
- Koerner, E.F.K.** 1978. *Toward a Historiography of Linguistics*. Amsterdam Studies in the Theory and History of Linguistic Science. Series 3. Studies in the History of Linguistics 19. Amsterdam: John Benjamins.
- Landau, S.** 2001. *Dictionaries: The Art and Craft of Lexicography*. Second Edition. Cambridge: Cambridge University Press.
- Lindemann, M.** 1994. *Die französischen Wörterbücher von den Anfängen bis 1600: Entstehung und typologische Beschreibung*. *Lexicographica*. Series Maior 54. Tübingen: Max Niemeyer.
- Meier, H.H.** 2003. Lexicography as Applied Linguistics. Hartmann, R.R.K. (Ed.). *Lexicography: Critical Concepts*: 307-318. London/New York: Routledge.
- Mulvany, N.C.** 2005. *Indexing Books*. Second edition. Chicago/London: University of Chicago Press.
- Pálfi, L.-L., P. Leroyer, A. Wagner and S.D. Vesterdahl.** 2008. Skomager, bliv ved din læst! Om politologiske leksika, politik i leksika og leksikografiske værktøjer. *LexicoNordica* 15: 197-218.
- Rothe, U.** 2001. *Das einsprachige Wörterbuch in seinem soziokulturellen Kontext: Gesellschaftliche und sprachwissenschaftliche Aspekte in der Lexikographie des Englischen und des Französischen*. *Lexicographica*. Series Maior 108. Tübingen: Max Niemeyer.
- 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.

Wierzbicka, A. 1985. *Lexicography and Conceptual Analysis*. Ann Arbor: Karoma.

Loránd-Levente Pálfi
Centre for Lexicography
Aarhus School of Business
University of Aarhus
Denmark
(llp@asb.dk)

ILALOK. *Dictionnaire vili-français/Mpisukulu bi kum' bi tshi vili ku tshi mputu.* 2008, 232 pp. ISBN 978-2-296-04939-0. Paris: L'Harmattan. Prix: €22.

Introduction

On connaît le vif intérêt de l'Institut des Langues Locales au Kouilou pour la promotion des langues maternelles et des cultures dont elles sont le véhicule irremplaçable. La parution aux Éditions L'Harmattan du *Dictionnaire vili-français* en 2008 vient couronner des années d'efforts acharnés et de privations. Ce dictionnaire, consacré au civil, poursuit ainsi la tradition des ouvrages lexicographiques commencés par les missionnaires. Pour le civil notamment on peut citer les contributions des révérends pères Derouet (n.d.) et Marichelle (1902) qui furent imprimés à la vieille imprimerie de Loango. Dans la préface du *Dictionnaire vili-français* (en abrégé DVF dans le reste du texte), le R.P. Marichelle précise que la parution de son dictionnaire vient compléter l'œuvre du R.P. Derouet auteur du *Dictionnaire français-vili*. Hormis les travaux lexicographiques sur le civil, deux autres ouvrages ont également été publiés. Il s'agit de *Proverbes et dictons du Loango en Afrique Centrale* de René Mavoungou Pambou (1997) et de *Parlons vili* de Gervais Loembé (2005). La parution du *Dictionnaire vili-français* complète parfaitement bien cette série.

Des prétextes du dictionnaire

Les prétextes (ensemble des textes figurant avant la section A-Z ou nomenclature) du DVF se limitent à la présentation de l'ouvrage faite par François Soumbou ainsi qu'à l'introduction signée de Marcel Poaty. Ces deux textes se complètent harmonieusement parce qu'ils présentent les acteurs de l'événement et exposent les principes théoriques et méthodologiques qui ont présidé au travail des auteurs. Par rapport à ces derniers, fait assez rare chez les Bantu, c'est un groupe de cadres congolais à la retraite désireux d'occuper leurs temps à la réflexion sur les langues du terroir ainsi qu'à leur revalorisation. Il s'agit de Joseph Tchiamas, François Soumbou, Gabriel Blaise Makosso, Alexandre Makosso, Joseph Siama, Simão Mamboma, et Jean Paul Serge Pangou.

À propos de la langue civil et de ses locuteurs

Le civil est une langue bantu parlée au Gabon, en République Démocratique du Congo, dans l'enclave du Cabinda (en Angola) ainsi qu'au nord de l'Angola, à la frontière avec la RDC. D'un point de vue linguistique, le civil fait partie du groupe Kongo (H10) de la branche linguistique bantu. Selon la classification de Guthrie (1948), le civil est désigné sous le sigle H12a. Le terme civil est la

dénomination ethnoscientifique de la langue que les Vili ou Bavili parlent. Il est transcrit phonétiquement tʃɪβɪli selon l'Alphabet Phonétique International (API). Ce glossonyme s'écrit également *tchivili* selon une orthographe calquée sur le français. Les missionnaires, administrateurs coloniaux et les locuteurs d'autres langues ont doté le civilo d'appellations comme *monvili*, *fioti*, *fiote*, *loango*, *lwangu*, *balwango*, etc. Ces missionnaires et administrateurs coloniaux ont produit une littérature sur le civilo touchant les domaines de la religion, de l'anthropologie, de la linguistique, etc. En ce qui concerne l'anthropologie, des études ont été consacrées aux Vili, qui avec d'autres groupes ethnolinguistiques viennent de l'ancien royaume de Loango (qui faisait lui-même partie du grand royaume de Kongo) formé au 14^{ème} siècle. Les Vili, principaux héritiers du royaume de Loango, s'identifient eux-mêmes comme *Fioti* ou *Fiote* (Hommes noirs). Au temps du royaume de Loango, le civilo était la langue standard de la cour du Maloango et celle de la classe moyenne. C'est pourquoi on va l'appeler tʃɪlɔwá:ngù qui signifie le parler de Loango (Ndinga-Koumba-Binza 2000: 10). Les parlers koci, yombe et lindji sont généralement présentés comme des dialectes du civilo. Mais en réalité, il s'agit de continuums linguistiques pour lesquels l'intercompréhension pourrait être facilitée par des facteurs socio-économiques (intégration économique, planification linguistique, etc.).

De la macro- et microstructure du dictionnaire

La macrostructure du DVF contient 4 007 articles classés par ordre alphabétique sous la lettre initiale du thème de chaque entrée. Les unités de traitement sont suivies de leurs préfixes respectifs entre parenthèses. Considerons l'exemple suivant:

<p>kobi (tshi):</p> <ol style="list-style-type: none"> 1. le talon 2. une danse semblable à la Mérengué, ainsi appelée parce que le tambour utilisé pour en rythmer la cadence se joue par frottement du talon sur la peau du tam tam, accompagné de l'action des mains.

Exemple textuel 1: Article **kobi** (DVF 2008: 78)

Cet ouvrage, comme tout dictionnaire, reflète l'état d'une langue à un moment donné. Le DVF inclut un inventaire de mots existants et expressions vili ainsi que des emprunts fortement intégrés. Ce fonds lexical renvoie aux mots et expressions vili de la vie de tous les jours. Dans la macrostructure, les substantifs prédominent, mais les autres classes de mots (verbes, adjectifs, interjections, etc.) n'en sont pas pour autant négligées. On note ça et là la présence de nombreux toponymes (*Bwäli*, *Diös'u*, *Hind'*, etc.), des anthroponymes (*Ngoy'u*, *Nzasi*, *Nzawu*, etc.), nombre d'emprunts (*balu*, *bwätu*, *fëlu*, etc.) et des expressions

idiomatiques (*tshi lwängu mwëgni, mwëndu land'n'*, etc.). Sur le plan microstructurel, ce dictionnaire fournit aux usagers une variété de catégories de données, à savoir: des traductions, des définitions, des exemples d'emploi, des notes encyclopédiques, etc. L'information rassemblée est utile, riche et diverse. Comparez les exemples suivants:

bayin':

Danse d'origine cabindaise, semblable à la valse. Un groupe de danses traditionnelles célèbre dans la région du Kouilou, li magni li tsi (l'assise du pays en perpétue la mémoire).

Exemple textuel 2: Article **bayin'** (DVF 2008: 25)

Bwäli:

Ancienne appellation du village Diosso, capitale du Royaume de Lwango. Ce village fut fondé par deux personnes: *bätu bwäli*, un homme et une femme, désireux de vivre loin de la communauté, d'où le nom de *bwäli* qui signifie: *deux*.

Exemple textuel 3: Article **Bwäli** (DVF 2008: 41)

Les lacunes du dictionnaire

Le travail de recensement du stock lexical vili par les auteurs aurait beaucoup été renforcé par une prise en compte de la catégorie grammaticale des différentes entrées. En effet, dans ce dictionnaire, l'unité de traitement n'est suivie ni d'une indication de la partie du discours à laquelle elle appartient ni d'une indication de son schème tonal.

On peut également regretter dans le corps du dictionnaire, la trop faible présence d'exemples d'emploi. En effet, leur emploi systématique aurait aidé le lecteur à se faire une meilleure idée des différentes traductions et définitions proposées pour les différentes entrées.

En guise de conclusion

Malgré ces quelques manquements le DVF reste un travail remarquable de dépouillement du fonds lexical vili qui, sans nul doute, servira de base à d'autres dictionnaires car, en matière de langue, le travail n'est jamais achevé. En conclusion, je voudrais adresser avant tout mes remerciements aux auteurs. Je les félicite pour leur excellent travail, ainsi que pour leur engagement en faveur du maintien de nos langues. Leur contribution ne s'arrête pas à l'étude du civilisé et à la réalisation de cet ouvrage, mais c'est aussi pour leur engagement vraiment actif que je les félicite en particulier.

References bibliographiques

- Derouet (R.P.)**. n.d. *Dictionnaire français–vili*. Loango: Imprimerie de la Mission.
- Guthrie, M.** 1948. *The Classification of the Bantu Languages*. Oxford: Oxford University Press.
- Loembé, G.** 2005. *Parlons vili: Langue et culture de Loango*. Paris: L'Harmattan.
- Marichelle, C.** 1902. *Dictionnaire français–vili*. Loango: Imprimerie de la Mission.
- Mavoungou Pambou, R.** 1997. *Proverbes et dictons du Loango en Afrique Centrale: Langue, culture et société*. Bajag-Meri.
- Ndinga-Koumba-Binza, H.S.** 2000. *Phonologie du civili de Mayumba: Langue bantu du Gabon (H12a)*. Mémoire de Maîtrise. Libreville: Université Omar Bongo.

P.A. Mavoungou
Groupe de Recherches en Langues et Cultures Orales (GRELACO)
Département des Sciences du Langage
Université Omar Bongo
Libreville
République Gabonaise
(moudika2@yahoo.fr)

Jean Pruvost (Éditeur). *Dictionnaires et mots voyageurs: les 40 ans du Petit Robert, de Paul Robert à Alain Rey.* 2008, 238 pages. ISBN 978-2-917268-02-5. Éragny: Éditions des Silves. Prix: €17.

Au moment où nous effectuons nos études en lexicographie avec le couple Corbin, Danièle et Pierre à Lille, Jean Pruvost lançait ce qui est devenue la «Journée des dictionnaires de Cergy» et devenait grâce à l'amitié qu'il vouait à Bernard Quemada, l'universitaire qu'il est devenu. Depuis qu'il l'organise, cette journée reçoit chaque année des visiteurs venus d'ailleurs et des conférenciers pour parler de ce vaste domaine d'étude qu'est la métalexigraphie. La sélection des conférenciers se fait à l'interne et ce sont souvent les mêmes personnes qui sont amenées à parler d'années en années ce qui enlève le caractère international que pourrait avoir ces journées si l'on y acceptait des personnes en dehors des réseaux d'amitiés interuniversitaires qui se suivent d'années en années. Les journées des dictionnaires que ce soit à Montréal, à Bari, en Allemagne procèdent toutes de la même manière.

Les actes de la Journée des dictionnaires de Cergy de 2007 sont publiés chez l'éditeur Éditions des Silves. L'ouvrage s'ouvre par une préface de Jean Pruvost, l'organisateur de cette journée, traitant de quelques lettres inédites de la Société du Nouveau Littré et de Paul Robert. Fidèle à son approche Pruvost s'intéresse à la vie des lexicographes Paul Robert, Alain Rey et Josette Rey-Debove, aux différentes préfaces des dictionnaires de langue française, etc. Il poursuit son approche dans le chapitre qu'il intitule «De Paul Robert à Alain Rey: deux personnalités marquantes du XXe et du XXIe siècle [...]». Dans son commentaire des biographies de Paul Robert que celui-ci insère dans le Robert 2, Pruvost reprend des propos qu'il a tenus ailleurs. Il évoque la vie de Paul Robert, l'implication de sa famille dans le dictionnaire, qui lui vaut, en octobre 1970, la formule d'Alain Rey, concernant la «famille Robert», son passé de chiffreur en Algérie. C'est à Georges Duhamel que l'on doit l'inspiration pour le nom de la société du Nouveau Littré qui est créée en 1951. Pruvost évoque aussi la vie d'Alain Rey, certes une des premières en milieu savant. Pruvost évoque également la proposition faite à Paul Robert de diriger le Trésor de la langue française et la collaboration de Rey à ce projet gaulliste. À lui seul Pruvost occupe un peu plus du quart de l'ouvrage pour 9 auteurs. Henriette Walter adopte un style vulgarisateur semblable à celui de Pruvost qui contraste avec le reste de l'ouvrage. Elle fait une sorte de résumé de l'ouvrage de Franck Resplandy sur les prêts de la langue française aux langues étrangères: *L'étonnant voyage des mots français dans les langues étrangères*. Cet article de Walter est fort instructif et nous apprend que 700 mots français sont récemment attestés dans la langue anglaise. Inès Sfar et Taoufik Massoussi s'intéressent à l'analyse et au traitement lexicographique des emprunts. Jean-Claude Boulanger et Anna Malkouwska s'intéressent aux emprunts dans le domaine alimentaire, particulièrement aux italianismes qui ont fait leur entrée récemment, comme par exemple *antipasti, cappelletti, penne, tiramisù, panini*, etc. L'italien domine avec

l'anglais au rang des langues prêteuses avec 16 mots. Il est fort intéressant de remarquer que la majorité des innovations lexicales en matière de cuisine italienne font d'abord leur entrée au Québec. La durée moyenne d'attente d'un mot entre son attestation dans les textes et son entrée dans le dictionnaire est de 16,9 ans (p. 114). Anna Anastassiadis-Syméonidis s'intéresse à l'emprunt du grec moderne au français analysant le phénomène d'une manière historique et plus générale que ne le font les autres contributeurs. Nathalie Gasiglia traite de l'emploi des emprunts dans les dictionnaires d'apprentissage français. C'est avec un sujet pour le moins original par rapport au reste que Frédéric Houbert laisse sa marque dans cet ouvrage. Il aborde les emprunts dans les dictionnaires juridiques se fondant sur les classiques en français (Gérard Cornu) et en anglais le *Black's Law Dictionary*. Il en vient à la conclusion (p. 219):

Dans la langue courante, l'emprunt peut permettre d'enrichir le lexique, servir à combler une lacune lexicale ou procéder d'un effet de mode. Dans le vocabulaire juridique, ses fonctions sont tout autres. Comme en témoignent les quelques exemples [...] l'emprunt y désigne le plus souvent une institution: copyright/droit d'auteur, trust, juge d'instruction, hypothèque.

John Humbley clôt l'ouvrage par un article sur les emprunts et leur remplacement, c'est-à-dire sur le traitement offert aux néologismes de langues étrangères suites aux propositions de commissions de terminologie par exemple.

Les journées des dictionnaires de Cergy sont donc toujours aussi actives, attirant chaque année de nombreux étudiants de Jean Pruvost ou de l'université Cergy-Pontoise mais aussi des enseignants, chômeurs, chercheurs, professeurs honorables ou honorés par Pruvost lui-même. La journée classique du mois de mars, tenue en 2009 le 18 mars, est aussi jumelée à la présentation d'ouvrages provenant de différentes maisons d'éditions traitant de lexicographie.

Jean-Nicolas De Surmont
Université de Leeds
Leeds
Royaume-Uni
(jdesurmont@yahoo.fr)

The Dictionary of Lexicography and Dictionary Research

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

Herbert Ernst Wiegand, *Germanistisches Seminar, Universität Heidelberg,
Heidelberg, Germany (herbert.ernst.wiegand@gs.uni-heidelberg.de)*

Abstract: This article presents a brief explanation of the current state of the work on the *Wörterbuch zur Lexikographie und Wörterbuchforschung*. It is done in a way that gives a general impression of the structure of this multivolume specialised dictionary.

Keywords: ALPHABETICAL EQUIVALENT INDEX, FUNCTIONAL CROSS-REFERENCE ARTICLE STRETCH, SPECIALISED DICTIONARY

Opsomming: Hierdie artikel bied 'n kort uiteensetting van die huidige stand van die werk aan die *Wörterbuch zur Lexikographie und Wörterbuchforschung*. Dit het ten doel om 'n algemene indruk te gee van die struktuur van hierdie meerdelige vakwoordeboek.

Sleutelwoorde: ALFABETIESE EKWIVALENTREGISTER, FUNKSIONELE KRUISVERWYSINGSARTIKELTRAJEK, VAKWOORDEBOEK

The full bilingual title of this specialised dictionary is: *Wörterbuch zur Lexikographie und Wörterbuchforschung / Dictionary of Lexicography and Dictionary Research. Mit englischen Übersetzungen der Umtexte und Definitionen sowie Äquivalenten in neun Sprachen / With English translations of the outer texts and definitions as well as equivalents in nine languages.*

This dictionary is edited and compiled by Michael Beißwenger (Dortmund), Rufus H. Gouws (Stellenbosch), Matthias Kammerer (Karlsruhe), Angelika Storrer (Dortmund), Herbert Ernst Wiegand (Heidelberg) and Werner Wolski (Paderborn). Including the lemma signs given in German, this is a dictionary that presents equivalents in ten languages. The following collaborators are responsible for translating the German lemma signs and finding appropriate translation equivalents:

- Afrikaans: Rufus H. Gouws (Stellenbosch)
- Bulgarian: Pavel Petkov (Sofia)
- English: Rufus H. Gouws
- French: Franz Josef Hausmann (Erlangen)
- Italian: Laura Giacomini/Giovanni Rovere (Heidelberg)
- Portuguese: Stefan J. Schierholz (Erlangen)/Rute Costa (Lissabon)

- Russian: Ekaterina Butina-Koller (München)
- Spanish: Teresa Fuentes Morán (Salamanca)
- Hungarian: Regina Hessky/Zita Hollós (Budapest)

All the outer texts of the dictionary are given in both German and English. Maria Smit (Stellenbosch) and Rufus H. Gouws are responsible for the English translations of these texts. The dictionary will contain more or less 13 000 lemmata, which includes more or less 7 400 cross-reference lemmata. The relative high number of cross-reference lemmata is due to the fact that all multiword terms are presented in a non-natural order, resulting in functional cross-reference article stretches like the following:

- Link, autogenerierter**
=> autogenerierter Link
- Link, dynamischer**
=> dynamischer Link
- Link, lexikalischer**
=> lexikalischer Link
- Link, navigatorischer**
=> navigatorischer Link
- Link, nutzgenerierter**
=> nutzgenerierter Link
- Link, systemgenerierter**
=> systemgenerierter Link
- Link, thematischer**
=> thematischer Link
- Link, toter**
=> toter Link
- Link, typisierter**
=> typisierter Link
- Link, unspezifizierter**
=> unspezifizierter Link
- Link, verweisbezogener**
=> verweisbezogener Link

The genuine function of a functional cross-reference article stretch consists therein that the user can recognise which and how many multiword terms of the same type, e.g. ADJECTIVE NOUN, can combine with the same noun (here *link*) to be presented as lemmata in this dictionary.

The dictionary will be published in four volumes by the publishing house Walter de Gruyter. The contents of the different volumes are as follows:

- Volume 1: Front matter texts, article stretches A–C
- Volume 2: D–K
- Volume 3: L–Z, back matter texts
- Volume 4: Index volume: nine alphabetical equivalent access indexes.

A brief description of the project with a demo-version of the dictionary articles can be found at <http://www.fabulex.de>. A more comprehensive presentation can be found at <http://woerterbuch-portal.de> as well as in the references cited below.

The current state of the project (July 2009) is as follows:

- (a) Volume 1 contains the following front matter texts, in both German and English:
- Two-toned graphical presentation of the consultation possibilities in the dictionary, given on the front inside cover (with the English version on the back inside cover)
 - Preface by the editors
 - Preface by the collaborators regarding the terminological equivalence
 - Guidelines for the user
 - Systematic introduction
 - Alphabetical access index for the *Systematische Einführung*
 - Alphabetical list A–Z of lemma signs from other subject fields

The German and English versions of these front matter texts have been completed, with some still in the final phase of editing. The article stretches A–C have been completed and Volume 1 is due in the fourth term of 2009.

- (b) Volume 2 contains the article stretches D–K. The dictionary articles of the article stretches D and E have been written and are already on the web. A part of the article stretches F–K have been completed and are on the web in a password-protected environment. The accompanying equivalents are being prepared in a separate equivalent table from where they are automatically integrated into the dictionary articles.
- (c) Volume 3 contains the article stretches L–Z and supplements A–Z. Many of these articles have already been written. This volume also contains the following back matter texts regarding the dictionary basis:
- the alphabetical list regarding Category I: Dictionaries and dictionary supplements
 - the alphabetical list regarding Category II: Metalexigraphic and other literature

The examples in the dictionary articles come from the references given in Category I. References in Category II are indicated in the reference position of the dictionary articles. The terms that receive a lexicographic treatment in the dictionary are employed in these references.

- (d) CD-ROMs. Volumes 1 and 2 have an accompanying CD-ROM. The CD-ROM in the first volume contains the volume-specific parts of the alpha-

betical lists of Category I and Category II as well as the nine partial equivalent indexes for the article stretches A–C. The CD-ROM in the second volume contains the respective partial lists and equivalent indexes of volumes 1 and 2. By means of the CD-ROMs, one can *firstly* find the references given in volumes 1 and 2 prior to the publication of the third volume and, *secondly*, the articles can be accessed via the partial equivalent indexes.

- (e) The fourth volume is an index volume, compiled by Matthias Kammerer. It contains nine alphabetical equivalent access indexes and is due shortly after the publication of the third volume.

The *Wörterbuch zur Lexikographie und Wörterbuchforschung* is a specialised learning and consultation dictionary for experts and semi-experts. Consequently it can be used effectively in situations of impeded specialised text reception with specific search questions regarding individual terms. It furthermore has a mediostructural arrangement that has been devised to ensure successful systematic dictionary-internal terminological consultation.

Further information

- Kammerer, Matthias.** 2001 [2002]. XML-getaggte Wörterbuchartikel. Ein Bericht aus der Praxis des Wörterbuchs zur Lexikographie und Wörterbuchforschung. *Lexicographica* 17: 249-301.
- Wiegand, Herbert Ernst.** 2003. Wörterbuch zur Lexikographie und Wörterbuchforschung/ Dictionary of Lexicography and Dictionary Research. Städtler, Thomas (Ed.). 2003. *Wissenschaftliche Lexikographie im deutschsprachigen Raum. Im Auftrag der Heidelberg Akademie der Wissenschaften*: 417-437. Heidelberg: Winter Verlag.
- Wiegand, Herbert Ernst.** 2003a. Wörterbuch zur Lexikographie und Wörterbuchforschung (WLWF). Dictionary of Lexicography and Dictionary Research. Eine kurze Projektbeschreibung. Botha, Willem (Ed.). 2003. *'n Man wat beur. Huldigungsbandel vir Dirk van Schalkwyk*: 368-384. Stellenbosch: Bureau of the WAT.
- Wiegand, Herbert Ernst.** 2004. Über die Unterschiede von Fachlexikographie und Terminographie. Am Beispiel des Wörterbuchs zur Lexikographie und Wörterbuchforschung. Wiegand, Herbert Ernst (Ed.). 2004. *Studien zur zweisprachigen Lexikographie mit Deutsch IX*: 135-152. *Germanistische Linguistik* 178. Hildesheim/Zürich/New York: Georg Olms.
- Wiegand, Herbert Ernst.** 2004a. Überlegungen zur Mediostruktur in Fachwörterbüchern. Auch am Beispiel des Wörterbuchs zur Lexikographie und Wörterbuchforschung. Brdar-Szabó, Rita and Elisabeth Knipf-Komlósi (Eds.). 2004. *Lexikalische Semantik, Phraseologie und Lexikographie. Abgründe und Brücken. Festgabe für Regina Hessky*: 339-365. Duisburger Arbeiten zur Sprach- und Kulturwissenschaft 57. Frankfurt a.M.: Peter Lang.
- Wiegand, Herbert Ernst.** 2004b. Reflections on the Mediostructure in Special-Field Dictionaries. Also According to the Example of the *Dictionary of Lexicography and Dictionary Research*. *Lexikos* 14: 195-221.
- Wiegand, Herbert Ernst.** 2007. Zum Stand der Arbeiten am Wörterbuch zur Lexikographie und Wörterbuchforschung. *Zeitschrift für germanistische Linguistik* 35: 182-184.

Publikasieaankondigings / Publication Announcements

Egbert Beijk, Lut Colman, Marianne Göbel, Frans Heyvaert, Tanneke Schoonheim, Rob Tempelaars en Vivien Waszink (Redactie). *Fons verborum: Feestbundel voor prof. dr. A.M.F.J. (Fons) Moerdijk, aangeboden door vrienden en collega's bij zijn afscheid van het Instituut voor Nederlandse Lexicologie*. 2009, 529 pp. ISBN 9789051796551. Leiden: Instituut voor Nederlandse Lexicologie/Amsterdam: Gopher.

Henning Bergenholtz, Sandro Nielsen and Sven Tarp (Editors). *Lexicography at a Crossroads: Dictionaries and Encyclopaedias Today, Lexicographical Tools Tomorrow*. 2009, 372 pp. ISBN 978-3-03911-799-4. Linguistic Insights. Studies in Language and Communication 90. Bern: Peter Lang. Price: SFR 93.

Julie Coleman. *A History of Cant and Slang Dictionaries. Volume III: 1859–1936*. 2009, xxiv + 489 pp. ISBN 978-0-19-954937-5. Oxford: Oxford University Press. Price: £75.

John Considine. *Dictionaries in Early Modern Europe: Lexicography and the Making of Heritage*. 2008, xiv + 393 pp. ISBN: 9780521886741. Cambridge: Cambridge University Press. Price: £55 / \$99. (Review in this issue.)

A.P. Cowie (Editor). *The Oxford History of English Lexicography. Volume I: General-purpose Dictionaries. Volume II: Specialized Dictionaries*. 2009. Volume I: xviii + 467 pp. Volume II: xix + 551 pp. ISBN Volume I–II: 978-0-19-928562-4. Volume I: 978-0-19-928560-0. Volume II: 978-0-19-928561-7. Oxford: Oxford University Press. Price: £140. (Review in this issue.)

Thierry Fontenelle (Editor). *Practical Lexicography: A Reader*. 2008, 405 pp. ISBN 978-0-19-929234-9. Oxford Linguistics. Oxford: Oxford University Press. Price: £27,50. (Review article in this issue.)

Pedro A. Fuertes-Olivera and Ascensión Arribas-Baño. *Pedagogical Specialised Lexicography: The Representation of Meaning in English and Spanish Business Dictionaries*. 2008, ix + 165 pp. ISBN 978 90 272 2335 7 (Hb). Terminology and Lexicography Research and Practice (TLRP) 11. Amsterdam/Philadelphia: John Benjamins. (Review article in this issue.)

ILALOK. *Dictionnaire vili-français/Mpisukulu bi kum' bi tshi vili ku tshi mputu*. 2008, 232 pp. ISBN 978-2-296-04939-0. Paris: L'Harmattan. Prix: €22. (Revue dans ce numéro.)

- Marijke Mooijaart and Marijke van der Wal (Editors). *Yesterday's Words: Contemporary, Current and Future Lexicography*. 2008, viii + 376 pp. ISBN (10): 1-84718-469-3. ISBN (13): 9781847184696. Newcastle: Cambridge Scholars Publishing. Price: £39.99.
- Minah Nabirye (Omwiluuzi), Kibuuka Balubuliza Kiingi, Deo Kawalya, John Kalema, Livingston Walusimbi, Manuel Muranga (Abawabuzi), Hassan Wasswa (Omuteesiteesi w'Ekitabo). *Eiwanika ly'Olusoga*. 2009, 636 pp. ISBN 978-9970-101-00-9. Kampala: Menha Publishers. (See article in this issue.)
- Vincent B.Y. Ooi, Anne Pakir, Ismail S. Talib and Peter K.W. Tan (Editors). *Perspectives in Lexicography: Asia and Beyond. New and Selected Papers from the Fourth ASIALEX International Conference, National University of Singapore*. 2009, 290 pp. ISBN 978-965-90207-1-3. Tel Aviv: K Dictionaries.
- Jan Posthumus: *Woordenboeken en hun lotgevallen. Studies over lexicografische praktijk (ca. 1850 tot heden)*. 2009, x + 331 pp. ISBN/EAN 978-90-8880-007-8 / ISBN/EAN 978-3-89323-534-6. Cahiers voor Taalkunde 28. Amsterdam: Stichting Neerlandistiek VU / Münster: Nodus Publikationen.
- Anton F. Prinsloo. *Annerlike Afrikaans*. 2009, 536 pp. ISBN 978-1-86919-265-5. Pretoria: Protea Boekhuis. Prys: R250.
- Salonee Priya (Editor). *Lexicography: The Changing Landscape*. 2009, VI + 252 pp. ISBN 978-81-314-2438-4. Hyderabad: The Icfai University Press.
- Jean Pruvost (Éditeur). *Dictionnaires et mots voyageurs: les 40 ans du Petit Robert, de Paul Robert à Alain Rey*. 2008, 238 pp. ISBN 978-2-917268-02-5. Éragny: Éditions des Silves. Prix: €17. (Revue dans ce numéro.)

VOORSKRIFTE AAN SKRYWERS

(Tree asseblief met die Buro van die WAT in verbinding (wat@sun.ac.za) vir 'n uitvoeriger weergawe van hierdie instruksies of besoek ons webblad: <http://www.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 kantlyne (ongeveer 2,5 cm) oorskry nie. Manuskrip moet verkieslik in elektroniese formaat as ASCII-teks, as volledig geformateerde Microsoft Word (DOS of Windows) lêers of as WordPerfect (DOS of Windows) lêers op rekenaarskyf (360 KB tot 1.44 MB) voorgelê word. 'n Rekenaardrukstuk van die artikel moet die skyf vergesel. Elke artikel moet voorsien wees van 'n Engelse opsomming van tussen 150 en 250 woorde, sowel as tussen 10 en 30 Engelse sleutelwoorde.

2. Grafika

Een stel duidelike oorspronklike illustrasies, tabelle, grafieke, diagramme, of kwaliteitsafdrukke daarvan, moet voorgelê word. Die plasing van grafika binne die teks moet duidelik aangedui word.

3. Bibliografiese gegewens en verwysings binne die teks

Kyk na onlangse nommers van *Lexikos* vir meer inligting.

INSTRUCTIONS TO AUTHORS

(For a more detailed version of these instructions, please contact the Bureau of the WAT (wat@sun.ac.za) or refer to our website: <http://www.wat.co.za>)

A. EDITORIAL POLICY

1. Type and content of articles

Articles may deal with pure lexicography or with the implications that related fields such as linguistics, general linguistics, computer science and management have for lexicography.

Contributions may be classified in any one of the following categories:

(1) **Articles:** Fundamentally original scientific research that has been done and the results that have been obtained, or reflecting existing research results and other facts in an original, synoptic, interpretative, comparative or critically evaluative manner.

(2) **Review articles:** Research articles presented in the form of a critical review of one or more published scientific sources.

Contributions in categories (1) and (2) are subjected to strict anonymous evaluation by independent academic peers in order to ensure the international research quality thereof.

(3) **Reviews:** An analysis and critical evaluation of published scientific sources and products, such as books and computer software.

(4) **Projects:** Discussions of lexicographical projects.

(5) **Lexiconotes:** Any article containing practice-oriented information, suggestions, problems, questions, commentary and solutions regarding lexicography.

(6) **Lexicovaria:** Any of a large variety of articles, announcements and press releases by lexicographic societies which are of particular value to the practising lexicographer.

(7) **Reports:** Reports on conferences and workshops.

Contributions in categories (3)-(7) must all meet the requirements of academic writing and are evaluated by the editors with this in mind.

2. Academic standard and evaluation procedure

The Department of Education of the South African Government has approved *Lexikos* as a subsidized, i.e. income-generating research journal.

Articles will be evaluated on the following aspects: language and style; conciseness and comprehensibility; problem formulation, reasoning and conclusion; references to the most important and most recent literature; substantial contribution to the specific discipline.

3. Language of contributions

Afrikaans, Dutch, English, French or German.

4. Copyright

Neither the Bureau of the WAT nor the African Association for Lexicography (AFRILEX) accepts any responsibility for claims which may arise from contributing authors' use of material from other sources.

Copyright of all material published in *Lexikos* will be vested in the Board of Directors of the Woordeboek van die Afrikaanse Taal. Authors are free however to use their material elsewhere provided that *Lexikos* (AFRILEX Series) is acknowledged as the original publication source.

5. Originality

Only original contributions will be considered for publication. Authors bear full responsibility for the originality and factual content of their contributions.

6. Free offprints and copies

Authors will receive five free offprints of each of their articles or review articles published, as well as one complimentary copy of the issue containing such article(s). Authors of purely evaluative reviews and of contributions to the categories Lexiconotes, Lexicovaria, Projects, and Reports receive five free offprints of their contributions. In the case of the latter four categories, the editors may, however, depending on the nature and scope of the contributions, decide to grant the author a copy of the issue concerned.

7. Invitation and editorial address

All interested authors are invited to submit contributions for publication in *Lexikos* to:

The Editor: LEXIKOS
Bureau of the WAT
P.O. Box 245
7599 STELLENBOSCH
Republic of South Africa

B. PREPARATION OF MANUSCRIPTS

Manuscripts of articles must meet the following editorial requirements:

1. Length and format

Contributions should not exceed more than 20 type-written A4 pages with double spacing and ample margins (about 2,5 cms). Manuscript should preferably be in electronic form on a (360 KB to 1.44 MB) floppy disk as either ASCII text, fully-formatted Microsoft Word (DOS or Windows) or WordPerfect (DOS or Windows) files. A computer printout of the article should accompany the disk. Each article must be accompanied by an English abstract of 150 to 250 words, and between 10 and 30 English keywords.

2. Graphics

One set of clear original drawings, tables, graphs, diagrams or quality prints thereof must be submitted. The locations of graphics must be clearly indicated in the text.

3. Bibliographical details and references in the text

Examine recent issues of *Lexikos* for details.

HINWEISE UND RICHTLINIEN FÜR AUTOREN

(Nehmen Sie bitte mit dem Büro des WAT Kontakt auf (wat@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 (wat@sun.ac.za)
ou consulter notre website: <http://www.wat.co.za>)

A. POLITIQUE ÉDITORIALE

1. Caractéristiques et contenu des articles

Les articles seront consacrés à la lexicographie pure, ou aux rapports entre la lexicographie et les disciplines voisines telles que la linguistique, la linguistique générale, l'informatique et le management.

Les contributions pourront appartenir à l'une des catégories suivantes:

- (1) **Articles:** Recherches scientifiques originales, avec leurs résultats; ou présentations originales, synoptiques, interprétatives, comparatives, évaluatives et critiques des résultats de recherches en cours;
- (2) **Articles bilans:** Articles de recherche présentés sous forme de bilan critique de travaux scientifiques déjà publiés.

Les contributions appartenant aux catégories (1) et (2) seront soumises de manière anonyme à des experts spécialistes indépendants afin d'en assurer la qualité scientifique au niveau international.

- (3) **Recensions:** Analyses et évaluations critiques de travaux de recherche et de productions scientifiques, telles que livres ou logiciels;
- (4) **Projets:** Présentations de projets lexicographiques;
- (5) **'Lexiconotes':** Textes contenant des informations pratiques, ou des suggestions, des problèmes, des questions, des commentaires et des solutions concernant des activités lexicographiques;
- (6) **'Lexicovaria':** Articles, annonces, communiqués de presse émanant de centres de lexicographie et qui revêtent un intérêt particulier pour les lexicographes;
- (7) **Rapports:** Rapports sur des colloques et ateliers.

Les contributions dans les catégories (3) à (7) devront répondre aux exigences de qualité des publications scientifiques et seront évaluées dans cette optique.

2. Critères et procédures d'évaluation

La revue *Lexikos* est reconnue et subventionnée par le Ministère de l'Éducation du gouvernement Sud-Africain comme revue devant générer des revenus.

Les articles seront évalués selon les critères suivants: langue et style, concision et clarté, formulation de la problématique, raisonnement et conclusion, référence aux travaux les plus importants et les plus récents, contribution substantielle à la discipline.

3. Langue des contributions

Afrikaans, allemand, anglais, français ou néerlandais.

4. Copyright

Le Bureau du WAT ou l'*African Association for Lexicography* (AFRILEX) décline toute responsabilité en cas de réclamations motivées par l'utilisation d'autres sources par les auteurs.

Les droits d'auteurs des documents publiés dans *Lexikos* appartiennent au Conseil d'administration du *Woordboek van die Afrikaanse Taal* (WAT). Cependant,

les auteurs sont libres d'utiliser leurs textes dans d'autres publications, à condition d'indiquer *Lexikos* (collection AFRILEX) comme source de la publication originale.

5. Originalité

Seules les contributions originales seront acceptées pour la publication. Les auteurs conservent l'entière responsabilité de l'originalité et du contenu factuel de leur texte.

6. Tirés-à-part et exemplaires gratuits

Les auteurs recevront gratuitement cinq (5) tirés-à-part de chaque article ou article de bilan, ainsi qu'un exemplaire gratuit de la publication contenant cet article.

Les auteurs des recensions et des publications dans les catégories 'Lexiconotes', 'Lexicovaria', Projets et Rapports recevront cinq (5) tirés-à-part de leur texte. Dans ces quatre dernières catégories, le responsable de la revue pourra néanmoins décider, en fonction de la nature et des dimensions des publications, d'accorder un exemplaire gratuit de la revue à leurs auteurs.

7. Adresse de la revue

Les auteurs intéressés sont invités à soumettre leurs propositions à:

L'éditeur: LEXIKOS
Bureau du WAT
Boîte postale 245
7599 STELLENBOSCH
République d'Afrique du Sud

B. PRÉSENTATION DES MANUSCRITS

Les manuscrits se conformeront aux exigences suivantes:

1. Longueur et format

Les contributions ne devraient pas excéder 20 pages dactylographiées, de format A4, avec double espacement et marges suffisantes (environ 2,5 cm), si possible en format électronique (disquette 360KB à 1,44MB), sous forme de fichier ASCII, complètement formaté sous Microsoft Word ou sous WordPerfect (DOS ou Windows). La disquette sera accompagnée d'un tirage papier. Chaque article sera pourvu d'un résumé en anglais de 150 à 200 mots, et de 10 à 30 mots-clés.

2. Tableaux et graphiques

Les dessins, tableaux, graphiques et diagrammes seront envoyés, soit sous leur forme originale soit sous forme d'une copie de bonne qualité. Leur place dans le texte devra être clairement indiquée.

3. Bibliographie et références dans le texte

Voir les exemplaires récents de *Lexikos*.