

Lexikos 16

Lexikos 16

Redakteur

Editor

J.C.M.D. du Plessis

Resensieredakteur

Review Editor

T. Harteveld



African Association for Lexicography

AFRILEX-REEKS 16:2006

AFRILEX SERIES 16:2006



BURO VAN DIE WAT

STELLENBOSCH

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

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

Uitgewer Publisher

BURO VAN DIE WAT
Posbus 245
7599 STELLENBOSCH

Kopiereg © 2006 deur die uitgewer
Alle regte streng voorbehou
Eerste uitgawe 2006

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 0-9584971-3-3
ISSN 1684-4904

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

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

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

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

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

Indekse Indexes

Arts & Humanities Citation Index, R.R.K. Hartmann Bibliography of Lexicography, Sabinet Online, SIL Bibliography Serials Index

Adviesraad / Advisory Board

Prof. H. Béjoint (Frankryk/France)
Prof. H. Chimhundu (Zimbabwe)
Prof. A. Delbridge (Australië/Australia)
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)
Prof. P.G.J. van Sterkenburg (Nederland/The Netherlands)
Prof. L.S. Vikør (Noorweë/Norway)
Prof. H.E. Wiegand (Duitsland/Germany)
Prof. L. Zgusta (VSA/USA)

Redaksiekomitee / Editorial Committee

Prof. H.M. Batibo (Botswana)
Dr. H.L. Beyer (Namibië/Namibia)
Prof. W.A.M. Carstens (RSA)
Mnr. E. Chabata (Zimbabwe)
Dr. A.R. Chuwa (Tanzanië/Tanzania)
Prof. C.J. Conradie (RSA)
Prof. J.D. Emejulu (Gaboen/Gabon)
Prof. A.E. Feinauer (RSA)
Prof. R. Finlayson (RSA)
Dr. S. Hadebe (Zimbabwe)
Prof. I.M. Kosch (RSA)
Prof. E.F. Kotzé (RSA)
Dr. P.A. Louw (RSA)
Prof. L.J. Louwrens (RSA)
Mnr. K.J. Mashamaite (RSA)
Dr. P.A. Mavoungou (Gaboen/Gabon)
Dr. J.S. Mdee (Tanzanië/Tanzania)
Prof. B.M. Mini (RSA)
Mnr. M.C. Mphahlele (RSA)
Mnr. M.H. Mpungose (RSA)
Prof. D.J. Prinsloo (RSA)
Prof. P.H. Swanepoel (RSA)

Inhoud / Contents

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

Artikels / Articles

The Treatment of Terms Relating to Islam and the Arab World in English Dictionaries	1
<i>Hashan Al-Ajmi and Abdullah Al-Otaibi</i>	
How to Do Language Policy with Dictionaries	13
<i>Henning Bergenholtz and Rufus H. Gouws</i>	
'n Voorlopige model vir die sistematiese beskrywing van gebruikersvriendelikheid in woordeboeke	46
<i>Herman L. Beyer</i>	
Do Dictionary Users Really Look Up Frequent Words? — On the Overestimation of the Value of Corpus-based Lexicography	67
<i>Gilles-Maurice de Schryver, David Joffe, Pitta Joffe and Sarah Hillewaert</i>	
Discipline, Dilemmas, Decisions and Data Distribution in the Planning and Compilation of Monolingual Dictionaries	84
<i>Rufus H. Gouws</i>	
Inclusion Strategies for Multi-word Units in Monolingual Dictionaries	95
<i>Phillip Louw</i>	

The <i>Dictionnaire français–yipounou/yipounou–français</i> of the Église Évangélique du Sud-Gabon (CMA) 1966 <i>Ludwine Mabika Mbokou</i>	104
A Trilingual Dictionary Yilumbu–French–English: An Ongoing Project <i>Paul Achille Mavoungou</i>	121
Challenges to Issues of Balance and Representativeness in African Lexicography <i>Thapelo Joseph Otlogetswe</i>	145
Die hantering van teenoorgesteldheid in enkele tipes woordeboeke, toegelig met Afrikaanse voorbeelde <i>Anna Nel Otto</i>	161
Die bewerking van idioeme in tweetalige woordeboeke <i>Liezl Potgieter</i>	180
Compiling a Bidirectional Dictionary Bridging English and the Sotho Languages: A Viability Study <i>D.J. Prinsloo</i>	193
Das Lern- und Konsultationswörterbuch. Ein neuer Fachwörterbuchtyp am Beispiel der <i>Wörterbücher zur Sprach- und Kommunikationswissenschaft (WSK)</i> <i>Herbert Ernst Wiegand</i>	205

Lexikoprogrammatuur / Lexicosoftware

Internationalisation, Localisation and Customisation Aspects of the Dictionary Application <i>TshwaneLex</i> <i>Gilles-Maurice de Schryver</i>	222
---	-----

Leksikonotas / Lexiconotes

Des considérations sur la lexicographie et la linguistique au Gabon <i>Guy-Modeste Ekwa Ebanéga et Fatima Tomba Moussavou</i>	239
Thoughts and Views on the Compilation of Monolingual Dictionaries in South Africa <i>N.C.P. Golele</i>	251

The Heart of the Matter: Methodological Challenges in Developing a Contemporary Reading Programme for Monolingual Lexicography, from the Perspective of the Dictionary Unit for South African English <i>Leela Pienaar</i>	257
De la problématique des articles synopsis dans la compilation des dictionnaires au Gabon <i>Gilles Saphou-Bivigat et Edgard Maillard Ella</i>	270
Lexikovaria / Lexicovaria	
HAT veertig jaar — 'n persoonlike oorsig <i>F.F. Odendal</i>	280
Leksikonuus / Lexiconews	
H.E. Wiegand Receives an Honorary Doctorate from the University of Stellenbosch	290
Resensieartikel / Review Article	
<i>Lexique Pove–Français / Français–Pove</i> , Mickala Manfoumbi: Seconde Note de Lecture <i>Hugues Steve Ndinga-Koumba-Binza</i>	293
Resensies / Reviews	
Julie Coleman and Anne McDermott (Editors): <i>Historical Dictionaries and Historical Dictionary Research: Papers from the International Conference on Historical Lexicography and Lexicology at the University of Leicester, 2002</i> <i>J.C.M.D. du Plessis</i>	309
Thomas Herbst and Kerstin Popp (Editors): <i>The Perfect Learners' Dictionary (?)</i> <i>Michaël Abecassis</i>	316

Thomas Herbst, Gunter Lorenz, Brigitta Mittmann and Martin Schnell (Editors): *Lexikografie, ihre Basis- und Nachbarwissenschaften: (Englische) Wörterbücher zwischen ‚common sense‘ und angewandter Theorie* 318
Maria Smit

Werner Hüllen: *English Dictionaries 800–1700: The Topical Tradition* 327
Werner Hüllen: *A History of Roget’s Thesaurus: Origins, Development, and Design*
Julie M. Coleman

Gabriele Stein: *Better Words: Evaluating EFL Dictionaries* 337
Michaël Abecassis

Ursula Wingate: *The Effectiveness of Different Learner Dictionaries: An Investigation into the Use of Dictionaries for Reading Comprehension by Intermediate Learners of German* 339
Michaël Abecassis

Voorpublikasieaankondigings / Prepublication Announcements

Rufus H. Gouws, Ulrich Heid, Wolfgang Schweickard and Herbert Ernst Wiegand (Editors): *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Special Focus on Computational Lexicography.* 341

Stefan J. Schierholz and Herbert Ernst Wiegand (Editors): *Wörterbücher zur Sprach- und Kommunikationswissenschaft (WSK) [Dictionaries of Linguistics and Communication Science]* 351

Publikasieaankondigings / Publication Announcements 354

Voorskrifte aan skrywers 356

Instructions to Authors 357

Richtlinien für Autoren 358

Instructions aux auteurs 359

Voorwoord

Vroeër vanjaar het een van die bekendste en gereeldste medewerkers van *Lexikos*, prof. H.E. Wiegand, 'n eredoktorsgraad van die Universiteit van Stellenbosch ontvang. Namens al die lesers van *Lexikos*, wil die redakteur hom gelukwens met hierdie erkenning van sy bydrae op die gebied van die metaleksikografie.

Die redakteur wil graag al die skrywers bedank vir hul bydraes tot *Lexikos*. Hy waardeer die geduld van diegene wat 'n aansienlike tyd moes wag vir hul artikels om gepubliseer te word. Aan dié lesers wat nog nie 'n bydrae tot *Lexikos* gemaak het nie, wil die redakteur 'n uitnodiging rig om hul pogings voor te lê. *Lexikos* lewer deurlopende terugvoer aan minder ervare skrywers met die oog op die verbetering van hul artikels totdat hulle geskik geag word vir beoordeling deur die keurders.

'n Woord van dank moet ook gaan aan al die keurders wat so onbaatsugtig artikels vir hierdie uitgawe van *Lexikos*, dikwels op kort kennisgewing, beoordeel het. Hul bydrae is onontbeerlik vir die handhawing van die gehalte van *Lexikos*. Dit is vir die redakteur aangenaam om twee nuwe lede op die Redaksiekomitee te verwelkom, prof. H.M. Batibo en dr. P.A. Louw, wat albei alreeds by verskeie geleenthede hulleself as skerpsinnige beoordelaars getoon het.

Soos met vorige uitgawes het me. Riette Ruthven haarself weer eens as onmisbaar bewys met die bedrewe en oordeelkundige set van al die artikels, waarvan die meeste deur verskeie stadiums gegaan het. Me. Tanja Harteveld moet weer bedank word vir die oplossing van talle rekenaarprobleme en -haakplekke.

Die redakteur wil graag die hoop uitspreek dat almal die lees van vanjaar se uitgawe van *Lexikos* lonend sal vind.

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

Foreword

Earlier this year, one of the most noted and regular contributors to *Lexikos*, Prof. H.E. Wiegand, received an honorary doctorate from the University of Stellenbosch. On behalf of all the readers of *Lexikos*, the editor wants to congratulate him on this recognition of his contribution to the field of metalexigraphy.

The editor wishes to thank all writers for their contributions to *Lexikos*. He appreciates the patience of those who had to wait a considerable time for their articles to be published. To those readers who have not yet made a contribution to *Lexikos*, the editor wants to extend an invitation to submit their efforts. *Lexikos* offers continuous feedback to less experienced writers with a view to improving their articles until they are deemed suitable to be judged by the referees.

A word of thanks must also be offered to all the referees who so unstintingly, often at short notice, evaluated articles for this issue of *Lexikos*. Their contribution is vital to upholding the quality of *Lexikos*. The editor has the pleasure of welcoming two new members to the Editorial Board, Prof. H.M. Batibo and Dr P.A. Louw, both of whom have already on several occasions showed themselves discerning referees.

As with past issues, Ms Riette Ruthven once more proved herself invaluable in expertly and discriminatingly typesetting all the articles, most of them going through several versions. Ms Tanja Harteveld must again be thanked for solving numerous computational problems and hitches.

The editor wishes to express the hope that this year's issue of *Lexikos* will be found a rewarding read by all.

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

'n Woord van AFRILEX

Wanneer 'n mens oor die aktiwiteite en suksesse van die African Association for Lexicography (AFRILEX) nadink, besef jy dat die vereniging dit regkry om op 'n baie hoë vlak te funksioneer, dit ten spyte van erge beperkinge betreffende finansies, tyd en menslike hulpbronne.

AFRILEX hou sy jaarlikse internasionale konferensies op uitnodiging by tersiêre of leksikografiese instansies. Dit bied ook seminare en slypskole aan. 'n Mens sou met veiligheid kon beweer dat AFRILEX dit kon regkry om die afgelope elf jaar hoë standaarde te handhaaf ten opsigte van leksikografiese en terminologiese ontwikkeling in Suid-Afrika en op die Afrikakontinent. Geen vakvereniging of instansie kan egter sonder lede voortbestaan nie. AFRILEX kort hierdie onmisbare komponent. Die ledetal is besig om vinnig af te neem, 'n toestand wat dringend aandag verdien.

Die AFRILEX-raad het besluit daar moet 'n georganiseerde poging wees om meer lede te werf. Daar is dus ooreengekom dat elke raadslid sal probeer om die vereniging by konferensies, seminare en soortgelyke geleenthede te bemark deur pamflette aan belangstellendes te gee. AFRILEX-lede kan ook aan hierdie ledewerwing deelneem. Pamflette is by raadslede verkrygbaar.

Elke AFRILEX-lid ontvang 'n kopie van *Lexikos* as deel van sy/haar ledegeld. *Lexikos* is die amptelike mondstuk van AFRILEX. Dit is nie net belangrik vir die vereniging om 'n geakkrediteerde tydskrif te hê waarin lede die resultate van hul navorsing kan publiseer nie, maar dit is ook belangrik dat elke lid die tydskrif ontvang om op hoogte te bly van leksikografiese ontwikkelinge. Dit kos egter baie harde werk, toewyding en fondse om *Lexikos* uit te gee. Sonder genoegsame befondsing is dit nie moontlik om die tydskrif aan AFRILEX-lede beskikbaar te stel nie. Die ledegeld dek nie die koste nie. Meer lede is nodig om *Lexikos* te help borg.

Die AFRILEX-raad wil graag alle lede aanspoor om belanghebbendes (bv. taalpraktisyns soos leksikografe, terminoloë, vertalers, redigeerders, tolke, joernaliste en studente) uit te nooi om lede van AFRILEX te word. Ons behoort te begin om inligting oor ons organisasie, ons vereniging, ons tydskrif en ons beroep te deel en daardeur mense te motiveer om lede van AFRILEX te word.

Vriendelike AFRILEX-groete

Dr. Mariëtta Alberts

President: AFRILEX

PanSAT, Privaat sak X08, ARCADIA 0007

Tel: +27 (0)12 341 9638 Faks: +27 (0)12 341 5938

E-pos: marietta@pansalb.org.za

AFRILEX-tuisblad: <<http://afrilex.africanlanguages.com>>

A Few Words from AFRILEX

Upon reflecting on the activities and achievements of the African Association for Lexicography (AFRILEX), one realises that the Association manages to operate on a very high level, this despite severe constraints regarding finances, time and human resources.

AFRILEX holds its annual international conferences by invitation at tertiary or lexicographic institutions. It also conducts seminars and workshops. One could safely say that AFRILEX managed to set good standards in the past eleven years regarding lexicography and terminology development in South Africa and on the African continent. However, no association or institution can exist without members. AFRILEX lacks this vital component. Its membership is declining rapidly, a situation which is to be addressed urgently.

The AFRILEX Board decided there should be an organized effort to enlist more members. It was therefore agreed that each Board member should try to promote the Association at conferences, seminars and similar gatherings by handing AFRILEX flyers to those interested. AFRILEX members could also contribute to the membership drive. Flyers are available from Board members.

Every AFRILEX member receives a copy of *Lexikos* as part of his/her membership fees. *Lexikos* is the official mouthpiece of AFRILEX. It is not only important for the Association to have an accredited journal in which its members can publish the results of their research, but it is also important for each member to receive the journal to stay ahead with lexicographical developments. However, it takes much hard work, dedication and funding to produce *Lexikos*. Without adequate funding it is not possible to make the journal available to AFRILEX members. The membership fees do not cover the expenses. More members are necessary to assist sponsoring *Lexikos*.

The AFRILEX Board would like to urge all members to invite stakeholders (e.g. language practitioners such as lexicographers, terminologists, translators, editors, interpreters, journalists and students) to become members of AFRILEX. We should start sharing information about our organization, our association, our journal and our profession and by doing so motivate people to become members of AFRILEX.

Kind AFRILEX regards

Dr Mariëtta Alberts
President: AFRILEX
PanSALB, Private Bag X08, ARCADIA 0007
Tel: +27 (0)12 341 9638 Fax: +27 (0)12 341 5938
Email: marietta@pansalb.org.za
AFRILEX home page: <<http://afrilex.africanlanguages.com>>

Redaksionele doelstellings

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

Die strewe van die AFRILEX-reeks is:

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

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

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

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

Editorial Objectives

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

The objectives of the AFRILEX Series are:

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

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

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

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

Redaktionelle Ziele

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

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

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

Gemäß den Zielsetzungen der AFRILEX-Serie werden:

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

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

Politique éditoriale

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

Les objectifs de la collection AFRILEX sont de :

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

Pour atteindre ces objectifs, la collection AFRILEX

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

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

The Treatment of Terms Relating to Islam and the Arab World in English Dictionaries*

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

Abdullah Al-Otaibi, *Department of English, Public Authority for Applied Education and Training, Kuwait (dr_alotaibi@yahoo.com)*

Abstract: This article investigates the treatment of vocabulary items relating to Islam and the Arab world in three major English dictionaries: *Longman Dictionary of Contemporary English*, *New Oxford Dictionary of English*, and *Merriam-Webster's Collegiate Dictionary*. The three dictionaries are compared with regard to their inclusion of 73 Arabic items (including loanwords) used in texts on Islam and the Middle East at the Google news file. With religion as theme, eight of these terms have been examined and compared across their entries in these three dictionaries to determine the closeness of the lexicographic description of these terms to their original meanings among native speakers of Arabic and to detect, from these speakers' viewpoint, any deficiencies in the defining language of their respective entries. The examination of the definitions of terms relating to Islam and the Arab world in the three dictionaries has revealed a number of pitfalls that may require editorial reconsideration. These include circularity, insufficient information, negative contexts, very general definitions, ambiguity, incomplete information, narrow definitions and encyclopedic redundancy.

Keywords: ARABIC LOANWORDS, CULTURAL TERMS, DEFINITION, ISLAM, REPRESENTATION, LEXICOGRAPHY

Opsomming: Die behandeling van terme betreffende Islam en die Arabiese wêreld in Engelse woordeboeke. Hierdie artikel ondersoek die behandeling van woordeskatitems betreffende Islam en die Arabiese wêreld in drie vooraanstaande Engelse woordeboeke: *Longman Dictionary of Contemporary English*, *New Oxford Dictionary of English*, en *Merriam-Webster's Collegiate Dictionary*. Die drie woordeboeke word vergelyk met betrekking tot hul opname van 73 Arabiese items (insluitende leenwoorde) wat in tekste oor Islam en die Midde-Ooste in die Google-nuusleërs gebruik word. Met godsdienst as tema, is agt van hierdie terme ondersoek en vergelyk met hul inskrywing in dié drie woordeboeke om die nabyheid vas te stel van die leksikografiese omskrywing van hierdie terme aan hul oorspronklike betekenis onder moedertaalsprekers van Arabiese, en om vanuit hierdie sprekers se gesigspunt enige gebreke op te spoor in die definisietaal van hul onderskeie inskrywings. Die ondersoek van die definisies van terme betref-

* This article was presented as a paper under the title 'The Presentation (or Misrepresentation) of Arabic-Islamic Culture in English Dictionaries' at the 51st Annual International Linguistic Conference, York University, Toronto, Canada, 31 March – 2 April 2006.

fende Islam en die Arabiese wêreld in die drie woordeboeke het 'n aantal slaggate blootgelê wat moontlik redaksionele heroorweging nodig maak. Dit sluit in sirkulariteit, onvoldoende inligting, negatiewe kontekste, baie algemene definisies, dubbelsinnigheid, onvolledige inligting, beperkte definisies en ensiklopediese oortolligheid.

Sleutelwoorde: ARABIESE LEENWOORDE, KULTURELE TERME, DEFINISIE, ISLAM, VERTEENWOORDIGING, LEKSIKOGRAFIE

1. Introduction

A dictionary may be defined as a reference work that documents the culture of the speakers of a given language and also reflects their knowledge and view of other cultures. Dictionaries, however, may not always provide objective, clear and accurate information about cultural terms borrowed from other languages. A related case here is the treatment of terms relating to Islam and the Arab world in English dictionaries.

Since the oil crisis of 1973, there has been a noticeable trend in the West to learn more about Islam and the Arabs. The political events of the times have shown that there is greater need for a better understanding of Islam and the Arab world where religion plays a pivotal role. Such a need has been emphasized by the UNESCO Universal Declaration on Cultural Diversity (2 November 2001) which promotes the 'fruitful diversity of ... cultures' for a more open and creative world in the new 21st century context.

Still, there exists a great deal of misunderstanding in the West of Islam and the Arab world and recent incidents are indicative of a deeply-rooted misunderstanding between Christian Westerners and Muslims. There can be referred to the use, although atypical, of the word *crusade* by Western political leaders when referring to the Afghan war and the anger that term caused among many Muslims.

It is widely believed nowadays that Western media emphasize further negative images of Muslims largely for political reasons. Poole (2002: 47) provides, 'as an example of how prevalent themes are adapted to suit the political purposes of the time, the image of the oil-rich Sheikh with his harem, which dominated the 1970s following the oil crisis'. One of Poole's conclusions about the portrayal of Muslims in the media is that 'it is a representation that appears to tell us more about the representers than the represented'.

2. Metalexigraphic background

Dictionaries, which are records of not only language and culture but also of a nation's view of other cultures, may not reflect the reality of the other. The lexicographic description of foreign cultures is supposed to be done carefully and objectively since a dictionary, as an unquestionable and supposedly impartial authority, is consulted by thousands or millions of users seeking both linguistic

and encyclopedic information about foreign cultural terms. However, as Moon (1989: 75) observes, 'dictionaries provide the world views of their lexicographers'. In turn, lexicographers are influenced by the social groups to which they belong. In the words of Landau (1989: 303), dictionaries express 'the views and prejudices of the established well-educated, upper classes'. This notion of social influence on a lexicographer's work is stressed by Cowie (1995: 283): 'both the content of a dictionary and its organization are strongly indicative of the values, judgements and priorities of the society in which the dictionary is compiled'.

A few studies have found that foreign cultures are underrepresented or misrepresented in English dictionaries. Benson (1997: 1) shows in his analysis that China is represented as peripheral at the levels of the wordlist, definitions, labels and illustrative text. He noticed that there may have been biases in the composition of the corpus and the selection of items from it and suggested that the selection of items expresses a Western perspective on China, not a Chinese perspective. Another study, Brown (1996), examined the treatment of religious terminology in four EFL dictionaries and found that there exists a great deal of inaccuracy in the definitions of Islamic terms and that these terms have been given a Western interpretation. The study asked the question: 'Whose definition of *jihad* or *Mecca* should appear in English dictionaries: Muslims or non-Muslim English speakers?' Brown (1996: 313) also refers to the role of corpora containing English speech and writing: 'any definition based on the corpus must represent the incorrect and potentially offensive meaning given to the term by those non-Muslim English speakers'.

In the metalexigraphic literature on cultural information in dictionaries, repeated mention of the negative representation of Muslims are found. Burchfield (1989: 20) refers to an incident where the word *Palestinian* was defined in COD (*Concise Oxford Dictionary*) as someone 'seeking to displace Israelis from Palestine', causing an angry protest from Palestinian groups demanding a correction of the entry. More recently, an entry, which appeared on the online version of *Roget's Thesaurus*, listing the word *Arab* as a noun meaning "beggar," and "homeless person" along with 14 other pejorative synonyms, was received with protests in the Arab world and letters of complaint were sent to the publishers who responded by removing the entry from the site.

3. Aims of the study

The present study aims to examine the treatment of vocabulary items relating to Islam and the Arab world in three major English dictionaries: *Longman Dictionary of Contemporary English* (LDCE), *New Oxford Dictionary of English* (NODE), and *Merriam-Webster's Collegiate Dictionary* (Webster's). Firstly, Arabic terms relating to religion have been gathered in a search-list and looked up in the three dictionaries to compare the dictionaries on the basis of their inclusion. Secondly, a group of eight terms have been examined and compared across

their entries in the three dictionaries to determine how close or distant the lexicographic description of these terms is to or from their original meanings among native speakers of Arabic and to detect, from these speakers' viewpoint, any deficiencies in the defining language in their respective entries.

4. The dictionaries

The three dictionaries have been chosen because they serve as good representatives of three different dictionary types used by both native speakers and foreign learners of English: EFL (LDCE), encyclopedic (NODE), and native speakers' (Webster's) dictionary. This selection is aimed at revealing differences at the levels of vocabulary inclusion, objectivity of description and adequacy of details across the three types in their representation of Islam and the Arab world.

LDCE is the representative of the EFL type of dictionary. It is probably the most popular dictionary among foreign learners of English around the world. Therefore, its simplified definitional style should be examined vis-à-vis its treatment of terms relating to Islam and the Arab world.

NODE has been chosen because this encyclopedic dictionary enjoys wide popularity among different categories of users worldwide. It is expected that such a dictionary will contain more of these terms than the other two and its more detailed etymological treatment of vocabulary makes it a reliable reference work for many users. It should be noted that the 2003 edition with the title *Oxford Dictionary of English* was also examined and no major changes were noticed with regard to the treatment of Islamic terms.

Native speakers' dictionaries are represented here by Webster's, which also represents American dictionaries. It is regarded as one of the most widely used titles in the United States by both college students and the general public.

5. Inclusion of Islamic vocabulary in the wordlist

A list of 73 arbitrarily selected common Islamic terms has been made, using as corpus the Google news file, a huge record of English news sources where the appearance of Islamic items would reflect the reality of usage and frequency of use. The selected items are of high frequency in mostly news reports and articles on Islam and Middle Eastern cultures. The three dictionaries were assessed in terms of their inclusion of the 73 items. As expected, the encyclopedic NODE has turned out to be the dictionary listing more of these items than the other two. Yet, it only contained 51 terms (69.9%) followed by Webster's with 33 (45.2%) and finally LDCE with only 25 (34.2%) (see Appendix).

This list of common cultural terms includes words that are frequently used by Muslims regardless of their native languages. Some of these terms have found their way into everyday English in recent years, e.g. *halal* the con-

cept of which has even become acceptable in most European countries (Rosen 2002: 150). The low percentages recorded by the three dictionaries indicate the need to revise the amount of this vocabulary in English dictionaries, given the fact that recent years have witnessed an increasing occurrence of these terms in the printed media and other types of literature in addition to the growing interest in the West to read more about Islam and the Middle East.

Another important observation concerns the spelling forms of many terms in the list. It was found that there are clear disagreements in the three dictionaries as to the approved spelling of terms like *haj/hadj/haji*, *Muhammed/Muhammad*, *halal/hallal*, *hegira/hijra* etc. The question here is to what extent have computerized corpora been utilized for determining the most common spellings of these terms. Standardization of these spellings would certainly be welcomed by users engaged in the production of texts on Islam or the Middle East.

6. Definitions of Islamic terms

A careful examination of the definitions of terms relating to Islam and the Arab world in the three dictionaries has revealed a number of pitfalls that require editorial reconsideration. These shortcomings include circularity, insufficient information, negative contexts, very general definitions, ambiguity, incomplete information, narrow definitions and encyclopedic redundancy.

(a) Circularity

In the definition of *Islam* in NODE, the user is told that it is the religion of Muslims. A similar definition is given in LDCE where the term is defined as the Muslim religion. NODE fails to provide a self-sufficient definition as the user is left with other Islamic terms: *Muslim* and *Allah*. Also, the three definitions differ in the inclusion of an essential feature of Islam, i.e. monotheism. LDCE does not refer to the religion as monotheistic.

LDCE *Islam* ... the Muslim religion, which was started by Muhammed and whose holy book is the Koran ...

NODE *Islam* ... the religion of the Muslims, a monotheistic faith regarded as revealed through Muhammed as the Prophet of Allah

Webster's *Islam* ... the religious faith of Muslims including belief in Allah as the sole deity and in Muhammad as his prophet ...

(b) Insufficient information

As an EFL dictionary, LDCE provides short and simplified definitions for these terms, but its definition of a common term *halal*, offers only one of two senses

of the term. Unlike the other two dictionaries, *halal* is defined as 'meat from an animal that has been killed in a way that is approved by Muslim law'. Here the dictionary reflects English speakers' knowledge of the word, and it is restricted to the gastronomic sense of the word. It should also be noted that NODE gives the original Islamic sense of the word as a secondary meaning in the compound section of the entry. Webster's provides both senses but focuses on the meaning related to food as shown in the illustrative examples *halal food* and *halal restaurant*. Users from other cultures will accordingly understand such terms from an English cultural viewpoint rather than from the perspective of the Islamic culture, which may give rise to inadequate understanding at a global level.

LDCE *halal* ... halal meat is meat from an animal that has been killed in a way that is approved by Muslim law

NODE *halal* ... denoting or relating to meat prepared or prescribed by Muslim law: halal butchers ... religiously acceptable according to Muslim law: halal banking

Webster's *halal* ... 1: sanctioned by Islamic law; esp: ritually fit for use (<~foods>) 2: selling or serving food ritually fit according to Islamic law (a ~ restaurant)

(c) Negative contexts

The negative context of war is used in the definition of the term *jihad* in all three dictionaries. They all equate the term with 'holy war', which is to Muslims the lesser form of jihad (Brown 1996). The modern informal English sense of *jihad* as a 'single-minded or obsessive campaign' is represented in NODE, but the foreign user may not know that this sense does not exist in the original usage of the term. Webster's also gives another sense, as 'crusade' which may have more than one interpretation and the original meaning of the term, 'struggle', is not specified. In its apparent attempt to avoid the mention of who Muslims would fight against, LDCE presents the shortened definition 'a holy war fought by Muslims'. NODE, on the other hand, gives the impression that *jihad* is fighting those who do not believe in the Muslim religion. This latter meaning may cause offense to Muslims as it considers all non-Muslims to be enemies and reveals the extent to which the concept of jihad, in an ill-represented reality, has been misunderstood in the West.

LDCE *jihad* ... a holy war fought by Muslims

NODE *jihad* ... a holy war undertaken by Muslims against unbelievers ... infml: a single-minded or obsessive campaign ...

Webster's *jihad* ... 1. a Muslim holy war against infidels ... 2. a crusade ...

(d) Very general definitions

Some of the terms examined in this study are of a very general nature, thus leading to possible confusion among users. The word *Arab*, being in itself the subject of successive definitions in history books, is one example that illustrates the point. In Webster's definition of this word historical specificity is lacking as it defines *Arab* as 'a member of a Semitic people in the Arabian peninsula' while actually Arabs are originally from the southern part of the peninsula. The second sense of the word in Webster's refers to Arabs as members of 'an Arabic-speaking people'. Here the user may wonder whether ethnic groups that also speak Arabic in the Middle East would be described as Arabs. NODE, on the other hand, focuses on the origin of Arabs and neglects the linguistic side. According to NODE's definition the user may think that descendants of pharaohs and Assyrians may not be regarded as Arabs because of their different origins. LDCE combines descent and language in its definition of the word, but the use of the geographical term *Middle East* is a generalization since this term may be taken as referring to non-Arab regions such as Iran, Turkey and Pakistan.

LDCE *Arab* ... someone whose language is Arabic or whose family have their origin in Arabia or the Middle East ...

NODE *Arab* ... a member of a Semitic people, originally from the Arabian Peninsula and neighboring territories, inhabiting much of the Middle East and North Africa ...

Webster's *Arab* ... 1.a. a member of the Semitic people of the Arabian peninsula b. a member of an Arabic-speaking people ...

(e) Ambiguous definitions

The old-fashioned term *Muhammedan* is defined in Webster's in a vague style in an obvious attempt to avoid the offensive meaning that relates the adjective to the followers of Muhammed or Islam in older titles by Merriam-Webster's publishers, e.g. *Webster's II New Riverside University Dictionary*. But the present definition can still be interpreted as referring to the people, not just to the prophet or the religion. In the other two dictionaries, the compilers have been careful to indicate explicitly that this term is not favoured by Muslims who do not refer to Muhammed as a founder of their religion but rather as God's messenger.

LDCE *Mohammedan* ... a word meaning Muslim, now considered offensive by most Muslims

NODE *Muhammadan* ... archaic term for Muslim (not favored by Muslims)

Webster's *Mohammadan* ... of or relating to Muhammad or Islam

(f) Incomplete information

Some definitions of Islamic terms in English dictionaries reflect their compilers' inadequate knowledge of foreign cultures and the little effort they make to find the correct information in the proper sources. A good example is the entry for *haj* in an earlier edition of Webster's where the term was defined as 'pilgrimage to Mecca during Ramadan'. It is known that *haj* takes place during another Islamic month, Dhu'l-Hijja. But it seems that the publishers have become aware of this error and the correction has been made in later editions. A clearer definition of the same term is needed in NODE where *haj* is defined as 'the pilgrimage to Mecca which takes place in the last month of the year'. The dictionary does not specify which year it is, the Gregorian or the Islamic year.

LDCE *haj* ... a journey to Mecca for religious reasons, that all Muslims try to make at least once in their life

NODE *hajj* ... the Muslim pilgrimage to Mecca which takes place in the last month of the year, and which all Muslims are expected to make at least once during their lifetime

Webster's *haj* ... the pilgrimage to Mecca prescribed as a religious duty for Muslims

(g) Narrow definitions

The Arabic meaning of *mosque* is not restricted to a building for worship but it can refer to any place where a Muslim can perform the daily prayers. Only NODE's definition is consistent with this meaning while the other two dictionaries define this common word using the words 'house' and 'building' in their definitions.

LDCE *mosque* ... a building in which Muslims worship

NODE *mosque* ... a Muslim place of worship

Webster's *mosque* ... a Muslim house of worship

(h) Encyclopedic redundancy

There does not seem to be any specific degree of encyclopedity in the definitions of Islamic terms in English dictionaries. In some entries, a dearth of essential encyclopedic details is found, while in other entries there is an abundance of encyclopedic information. For example, in the entry for *Shia* in Webster's, this smaller branch of Islam has been defined in an entry the size of

which is almost twice the size of the entry for *Sunni*, the mainstream Islam. This definition contains other Islamic terms that require further searching on the part of the user, e.g. *Ali*, *Imams*, *Muhammad* and *the last recognized Imam*.

LDCE *Shia* ... the Siite branch of the Muslim religion ...

NODE *Shia* ... one of the two main branches of Islam, followed especially in Iran, that rejects the first three caliphs and regards Ali, the fourth caliph, as Muhammad's first true successor ...

Webster's *Shia* ... the Muslims of the branch of Islam comprising sects believing in Ali and the Imams as the only rightful successors of Muhammad and in the concealment and messianic return of the last recognized Imam

6. Conclusion

Dictionaries can play a vital role in promoting cultural understanding and dispelling preconceived notions and stereotypes about other ethnicities and religions. We have found that, despite the critical remarks raised above regarding dictionary definitional inconsistencies, the three major English dictionaries examined in this study provide a comparatively more objective image of Islam and the Arab world than what people usually see and read in the mass media and many works of literature, despite the fact that the compilers of these dictionaries are influenced by their social environments and views of other cultures with obviously inadequate access to the original Islamic sources. It is not suggested in this paper that English dictionaries of different sizes should reflect Arabic native speakers' knowledge of Islamic terms. Rather, we have tried to show that there remains some room for better treatment of terms relating to Islam in English lexicography. With the help of corpus evidence which will probably include material produced by English-speaking Muslims, the list of Islamic terms can be revised in order to accommodate common Islamic terms that have recently become of wide currency in different types of printed materials. In addition, definitions of these terms can be reconsidered to provide native speakers and learners of English worldwide with clearer, more accurate, objective and sufficient information that is essential for them to understand Islam and the Arab world.

References

Dictionaries

- Kipfer, B. (Ed.). *Roget's Thesaurus* [online]. Available at <<http://www.thesaurus.com>>.
Mish, F.C. (Ed.). 2003¹¹. *Merriam-Webster's Collegiate Dictionary*. Springfield, MA: Merriam-Webster.

- Pearsall, J. and P. Hanks (Eds.).** 1998. *New Oxford Dictionary of English*. Oxford: Oxford University Press.
- Soanes, C. and A. Stevenson (Eds.).** 2003. *Oxford Dictionary of English*. Oxford: Oxford University Press.
- Soukhanov, A.H. and K. Ellis (Eds.).** 1984. *Webster's II New Riverside University Dictionary*. Boston, MA: Houghton Mifflin Company.
- Summers, D. et al. (Eds.).** 2001³. *Longman Dictionary of Contemporary English*. Harlow: Longman.

Other sources

- Benson, Philip.** 1997. *The Lexicography of English in the World: The Treatment of China in Four British Dictionaries*. Ph.D. Thesis. Exeter: University of Exeter.
- Brown, A.** 1996. The Treatment of Religious Terminology in English Dictionaries. Khan, J.U. and A.E. Hare (Eds.). *English and Islam, Creative Encounters 96: Proceedings of the International Conference Organized by the Department of English Language and Literature, International Islamic University Malaysia, Kuala Lumpur, 20–22 December 1996*: 307-314. Kuala Lumpur: Research Centre, International Islamic University Malaysia.
- Burchfield, R.W.** 1989. *Unlocking the English Language*. London: Faber and Faber.
- Cowie, A.P.** 1995. The Learner's Dictionary in a Changing Cultural Perspective. Kachru, B.B. and H. Kahane (Eds.). *Cultures, Ideologies and the Dictionary*: 283-295. *Lexicographica*. Series Maior 64. Tübingen: Max Niemeyer.
- Landau, S.I.** 1989. *Dictionaries: The Art and Craft of Lexicography*. Cambridge: Cambridge University Press.
- Moon, R.** 1989. Objective or Objectionable? Ideological Aspects of Dictionaries. *English Language Research* 3: 59-94.
- Poole, Elizabeth.** 2002. *Reporting Islam: Media Representations of British Muslims*. London: I.B. Tauris Publishers.
- Rosen, Lawrence.** 2002. *The Culture of Islam: Changing Aspects of Contemporary Muslim Life*. Chicago: Chicago University Press.
- UNESCO.** 2001. *Universal Declaration on Cultural Diversity*. Adopted by the 31st Session of the General Conference of UNESCO, Paris, 2 November 2001.

Appendix: Arab-Islamic vocabulary in the wordlist

<i>Term</i>	<i>LDCE</i>	<i>NODE</i>	<i>Webster's</i>
Ashura	–	–	–
Abbasid	–	✓	–
adhan	–	–	–
Ahlul Bayt	–	✓	–
Allah	✓	✓	✓
ayah	–	–	–
Ayatollah	✓	✓	✓
caliph	✓	✓	✓
da'wah	–	–	–
dervish	✓	✓	✓
Druze	–	✓	✓
Eid	–	✓	–
fard	–	–	–
fatwa	✓	✓	✓
fiqh	–	–	–
fitnah	–	–	–
Hadith	–	✓	✓
halal	✓	✓	✓
haj	✓	✓	✓
Hamas	–	✓	–
haram	–	✓	–
hijab	–	–	–
Hijrah	–	✓	✓
Hizb Allah	–	✓	–
Iblis	–	–	–
iftar	–	✓	–
Imam	✓	✓	✓
Injil	–	–	–
Islam	✓	✓	✓
Jannah	–	–	–
jihād	✓	✓	✓
jinn	✓	✓	✓
Jumuah	–	–	–
Ka'ba	–	✓	✓
kafir	✓	✓	–
Koran	✓	✓	✓
madrassa	–	✓	✓
Mahdi	–	✓	✓
masjid	–	✓	–
Mecca	✓	✓	✓
Medina	–	✓	–

<i>Term</i>	<i>LDCE</i>	<i>NODE</i>	<i>Webster's</i>
minaret	✓	✓	✓
Miraj	–	–	–
muezzin	✓	✓	✓
mufti	✓	✓	✓
mujahid	–	–	✓
mullah	✓	✓	✓
Muslim	✓	✓	✓
niqab	–	–	–
P.B.U.H.	–	–	–
kadi	–	✓	–
qibla	–	✓	–
Ramadhan	✓	✓	✓
salat	–	✓	–
salafism	–	–	–
salaam	–	–	–
sawm	–	–	–
shahadah	–	✓	–
shaheed	–	✓	–
sheik	✓	✓	✓
Shia	✓	✓	✓
Shi'ite	✓	✓	✓
shirk	–	–	–
shura	–	✓	–
sufi	–	✓	–
Sunnah	✓	✓	✓
Sunni	✓	✓	✓
surah	–	✓	✓
ummah	–	✓	–
umrah	–	–	–
Whhabism	–	✓	✓
zakat	–	✓	–
Zaboor	–	–	–
Total = 73	25 (34.2%)	51 (69.9%)	33 (45.2%)

How to Do Language Policy with Dictionaries

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

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

Abstract: The lexicographical presentation of terms from the field of language planning often lacks clear and unambiguous distinction and proper explanation. Too often dictionaries even fail to include these terms in the lemma list, and some central terms have not been treated in any general or special purpose dictionary. This article utilises results from research in the field of language policy to make suggestions for the lexicographical presentation and treatment of a number of relevant terms. The emphasis on a distinction between language policy as applied to the intralingual and the interlingual level and the motivation for the introduction of the notion of a communication policy should help lexicographers to give a more comprehensive account of terms from this field and it will also benefit scholars in the field of language policy. A second aspect of this article is the discussion of ways in which dictionaries participate in the implementation of language policy. It is indicated that lexicographers make ever so many decisions of a language political nature. In this regard the lexicographical influence of issues like linguistic hegemony and language purism are discussed. Suggestions are also made for future lexicographical procedures.

Keywords: COMMUNICATION POLICY, INTERLINGUAL, INTRALINGUAL, LANGUAGE PLANNING, LANGUAGE POLICY, LANGUAGE PROMOTION, LEXICOGRAPHY, LINGUISTIC HEGEMONY, PRESCRIPTIVE, PURISM

Opsomming: Hoe om taalbeleid met woordeboeke te bepaal. Die leksikografiese aanbieding van terme uit die veld van taalbeplanning kort dikwels duidelike en ondubbelsinnige onderskeiding en behoorlike verklaring. Woordeboeke laat te dikwels na om hierdie terme in die lemmalys in te sluit, en sommige kernterme word in geen algemene of vakwoordeboek behandel nie. Hierdie artikel benut resultate van navorsing op die terrein van taalbeplanning om voorstelle te maak vir die leksikografiese aanbieding en behandeling van 'n aantal relevante terme. Die klem op die onderskeid tussen taalbeleid soos dit toegepas word op die intratallige en intertallige vlak en die motivering van die invoering van die idee van 'n kommunikasiebeleid behoort leksikograwe te help om 'n meer omvattende verslag te bied van terme vanuit hierdie veld en dit sal ook navorsers op die terrein van taalpolitiek bevoordeel. 'n Tweede aspek van hierdie artikel is die bespreking van maniere waarop woordeboeke deelneem aan die implementering van taalpolitiek. Daar word aangedui dat leksikograwe talle besluite van 'n taalpolitieke aard neem. In hierdie verband word die leksikografiese invloed op kwessies soos linguistiese hegemonie en taalpurisme bespreek. Voorstelle word ook gemaak vir toekomstige leksikografiese werkswyses.

Sleutelwoorde: INTERTALIG, INTRATALIG, KOMMUNIKASIEBELEID, LEKSIKOGRAFIE, LINGUISTIESE HEGEMONIE, PRESKRIPTIEF, PURISME, TAALBELEID, TAALBEPLANNING, TAALBEVORDERING

1. How to do things with words

The title of this article is meant as a reference to the title of Austin's well-known book *How to Do Things with Words*. The title of this book on pragmalinguistics is not only famous; it is also misleading because one necessarily fulfils a speech act when one uses words. Somehow, the title of this article is also partly misleading. In practical lexicography, one can explain the different political motivations behind every single decision, but this does not mean that as a lexicographer one can act without political choices, intended or not. The basic thesis is therefore

- Every single lexicographical decision has a language policy relevance and therefore, in the end, a political dimension.

This thesis is not new because the influence of politics on lexicography has often been discussed. However, it usually applies to politically charged words like *democracy*, *terrorist*, *revolution*, *black*, *Negro*, *racism* etc. It is new to regard every lexicographical decision as the result of a political decision, more precisely a language policy decision. Although this is new in lexicography it is not new in the field of language policy (cf. Spolsky 2004: 35). We do not see the lexicographer as "a harmless drudge" who busies himself with tracing the origin and detailing the significance of words, as Dr Johnson and many lexicographers of the last 250 years have underrated themselves. Most lexicographers may be aware of their great influence on language behaviour but fail to realise their influence on cultural and political behaviour. It is also historically unfair to take the present understanding of *policy* as the only benchmark, although new interpretations in the light of our recent knowledge are required. There is a development that *policy* is frequently not only used as a term restricted to official or governmental considerations and decisions. The second half of the 20th century saw the formation of a number of new compounds with the lexical item *policy* as stem, e.g. *educational policy*, *children's policy*, *women's policy*, *environmental policy* and *immigration policy*. These new forms are related to key areas in the political decision-making processes of international, national and local authorities. From the late 1980s, the scope of such compounding was broadened to include aspects such as planning and decision-making in companies, organisations, schools and other non-political groups. Examples include *senior policy*, *smoking policy*, *alcohol policy* and *stress policy*. Along with this development, the scope of older, purely political terms has also been broadened to include planning and decision-making in companies, etc. This devel-

opment is social as well as linguistic, and it does not just apply to Denmark, Germany, the U.S.A. and South Africa, but as far as we can see to most of the world. The group of new compounds also comprises the terms *language policy* and *communication policy*.

One does not find the term *communication policy* in any contribution to lexicography. *Language policy* (or the near synonym term *language planning*) is not a term in the lemma list of most lexicographical dictionaries (e.g. the *Nordic Dictionary of Lexicography* 1997). The term is not very often mentioned or even regarded as a lexicographical relevant term in the metalexicographical literature (e.g. not at all in Zgusta 1971, Hausmann 1977, Landau 1981, Gouws 1989, Bergenholtz and Tarp 1995, Svensén 1993, Wiegand 1998, Gouws and Prinsloo 2005), and is only mentioned twice in the comprehensive three-volume *International Encyclopaedia of Lexicography* (1989–1991), and in both cases not as a real subject of discussion. Of course we have papers and monographs on politically problematic words and quotations in dictionaries and even dictionaries with such words, e.g. Strauß, Haß and Harras (1989) and *Zeitgeschichtliches Wörterbuch* (2003). We are dealing with the most obvious example indicating that language is narrowly connected with power, including political power. One can practise politics by means of words. With words one can describe facts in such a way that the factual contents become blurred. Words can also be used or misused in such a way in a given historical context that they fall into disrepute for a long time or even permanently. The non-use of specific words that would actually have been more apt could be part of a conscious political strategy. This is an old strategy, which will continue to be used. Typical examples are: Who is a *terrorist* and who is a *freedom fighter*? Is one an *anti-Semite* when as a non-Jew one criticises some decisions of the Israeli government? Should one refrain from using the word *nigger* in favour of *coloured*? Is it only politically correct language use to refer to a woman as *sex worker* and not as *prostitute*? This strategy is especially employed in political classifications, directed at race, sexuality, sex, ecology and the mental state of a person.

In all these cases, we are dealing with loaded political words that could also fall within the scope of a language policy. They regard controversial political or socio-political expressions, discussed in many metalexicographical contributions. This is typically done by examining a selection of dictionaries to determine how good or bad, politically correct or incorrect they are or whether certain words have at all been included in their lemma list. In current public discussion such words are the most obvious examples for language policy decisions in lexicography. However, this is only a small part of the problem; only one out of three in the proposed classification of intralingual language policy.

In general we want to achieve two aims in this contribution. Firstly, we want to introduce some terminological proposals from Bergenholtz (2006) to lexicography and, secondly, we want to show where and to what extent language policy plays a role in lexicography.

2. What is language policy?

In at least one sense the above-expounded view in lexicographical textbooks is incomplete: In contributions to all kinds of language regulations *language policy* is not the only term used. In English-speaking countries especially, the term *language planning*, introduced by Haugen (1959), is more often employed. In addition, other terms are sometimes used synonymously, or for special kinds of 'language regulations', e.g. in English *language engineering*, *glottopolitics*, *language development*, *language regulation* and *language management*; in German *Sprachlenkung*, *Sprachpflege*, *Sprachreinigung*; and in Afrikaans *taalbeplanning*, *taalbeleid* and *taalpolitiek*. We will not try to give a terminological overview (for this see Bergenholtz 2006). It is confusing enough when one only considers the use of the three main terms *language policy*, *language planning* and *communication policy*.

2.1 Explanations in dictionaries

Most general language dictionaries do not include the above-mentioned terms in their lemma list. *Communication policy* is not found at all, and *language policy* (or the respective equivalents) and *language planning* only in a few German, Dutch, Danish and Afrikaans dictionaries. It is interesting to note the absence of these terms in the majority of English dictionaries.

Regarding German it is actually only two dictionaries that include the term *Sprachpolitik*, i.e. the three editions of *Duden-GWB* and the *Brockhaus-Wahrig* 1980–1984. It is surprising that more comprehensive dictionaries like *Wahrig* 2000, *Duden-Universalwörterbuch* 2003 or Bünting and Karatas 1996 do not include this word as a lemma. Taking the proper and precise lemma selection of these dictionaries into account, one can conclude that *Sprachpolitik* does not belong to the 100 000 most important German words. In *Duden-GWB*¹ and almost identically in *Brockhaus-Wahrig* 1984 the following item is found:

- (1) **Sprachpolitik** staatliche Maßnahmen in Hinblick auf eine Sprache (*Duden-GWB*¹ 1981)
[**Language policy** governmental regulations regarding a language]

Firstly, it refers to governmental regulations and therefore national and not international or non-governmental language regulations. Secondly it refers to one, and not more than one language or else the lexicographers would have written "regarding one or more languages". The next two editions give completely different information. The preparation of a new edition of a dictionary often, but not always, leads to changes that are improvements. With identical items, the changes in the case of *Duden-GWB*² and *Duden-GWB*³, are not improvements. Syntactically the explanation has been constructed in such a way that one can only try to guess the contents:

- (2) **Sprachpolitik** die in einem Land gesprochene[n] Sprache[n], die in einem Land sich stellende Sprachenfrage o.ä. betreffende Politik (*Duden-GWB*³ 1999)
[**Language policy** the language(s) spoken in a country, the prevailing language or similar issues in a country regarding politics.]

When one assumes that both parts of the explanation need to be amplified, as elsewhere in this dictionary, one realizes that "the language(s) spoken in a country regarding politics" refer(s) to a national political matter with regard to either one or more languages. This corresponds to the item in *Duden-GWB*¹ with this change, however, that it now regards not only one, but more than one language. It probably still deals with a national and not an international matter and not with private but with public decisions, else it would have read: "in a country, in an international or national organisation as well as in private or public companies or institutions". It is not clear exactly what the language policy wants to influence regarding one or more than one language. It could refer to linguistic purism or linguistic norms, or it could refer to the prohibition of other languages or prescriptions for the official use or promotion of one or more languages. To be able to understand the second part of the explanation "the prevailing language or similar issues in a country regarding politics" one should know what *language issues* are. This term is explained as follows in *Duden-GWB*² and *Duden-GWB*³:

- (3) **Sprachenfrage** aus dem Zusammenleben mehrerer ethnischer Gruppen mit verschiedenen Sprachen innerhalb eines Staates herrührende Problematik (*Duden-GWB*³ 1999)
[**Language issues** problems arising from the living together of various ethnic groups with different languages within one country]

The not so industrious user who does not look for *Sprachenfrage* (language issues) will be unable to guess that *Sprachpolitik* (language policy) specifically deals with the relation between languages spoken by different ethnic groups in one country. It can be seen that the explanation given in *Duden-GWB*¹ is also complemented in the sense that language policy does not only refer to more than one language, but also to more than one language of members of different ethnic groups. This does not really improve our understanding of the term. One can assume that language policy is not a central theme in Germany and that the unclear explanation reflects the little public interest and consequently the unclear debate.

In the domain of language policy one in reality finds, with the exception of France, no comprehensive debate in the bigger European countries like Germany, England, Spain and Italy. The situation is different in smaller countries like Switzerland, Austria and especially in the Scandinavian countries. One can only suspect the reasons. We believe there are at least two reasons: The one reason is that the big countries, and that also means the big languages, in the

European Union are increasingly gaining influence to the detriment of smaller countries and smaller languages. Consequently these countries see the necessity for language political initiatives to maintain the current status of the linguistic diversity, as prescribed by the EU (Article 22 of the Charter of Constitutional Rights of the European Union), and not to let it be undermined any further. On the one hand Germany supports the small and medium countries but would like to see German, as the biggest language in the EU, also reaching the status which English and French have in practice. To follow both directions at the same time is not likely to succeed. The other reason is connected with cultural tradition regarding official language regulations. A few countries like Spain and France have in principle a long tradition of a positive attitude towards prescriptive regulations. In practice, however, the interest lies much more in general regulations and not as much in details and the implementation of sanctions for not complying with regulations. However, since 1990, the development in France follows a more descriptive direction by allowing new orthographical and inflexional morphological variants. The Nordic countries, especially Iceland, the Faeroe Islands and Norway, have another tradition. Contrary to southern Europe, governmental language interventions for changes in existing language usage are implemented here by official decisions and supported by public control.

In Holland and South Africa, as in English-speaking countries, the term *language planning* and its equivalents are frequently used. In Afrikaans, the term *taalbeleid* is also quite often employed.

- (4) **taalbeplanning** Beplanning in staatkundige verband van geleentede om 'n taal binne groepsverband of in die land as geheel te gebruik (*Verklarende Afrikaanse Woordeboek* 1993)

[**language planning** Planning in political context for opportunities to use a language in communal context or in the country as a whole]

We are dealing here with an intralingual relation and with planned opportunities within a language. It should be noted that it is considered a public matter. The term *language planning* is used in accordance with its original use (Haugen 1959), but not exactly as in current scientific contributions to language planning and language politics. Language policy is appropriately described as an interlingual matter:

- (5) **taalpolitiek** Politieke beleid t.o.v. die tale wat in 'n meertalige staat gebruik word (*Verklarende Afrikaanse Woordeboek* 1993)

[**language politics** Governmental policy with regard to languages used in a multilingual state]

Whether this explanation has been taken from the Dutch Van Dale dictionary can only be presumed:

- (6) **taalpolitiek** politiek die gevoerd wordt ter bescherming of onderdrukking van een taal, vooral in een staat met meer dan één taal (*Van Dale Groot Woordenboek der Nederlandse Taal* 1999)
[language politics Policy followed for the protection or oppression of a language, especially in a state with more than one language]

These two entries can hardly be regarded as correct. Had this been the case, monolingual countries would typically not have had a language policy or the need for such a policy. Regarding all the differences in the definitions in scientific contributions and technical dictionaries, this division and classification of interlingual language politics and intralingual language planning is quite special. This distinction is elsewhere found only in Hartmann and James (1998), one of the four existing lexicographical dictionaries. The division into inter- and intralingual regulations is important and correct but the terminological classification is not in accordance with the theoretical and practical standard. It would be more accurate to have a division of *language policy* or *language planning* in inter- or intralingual language regulations as in the Danish Internet Dictionary where *language policy* and *language planning* are regarded as synonyms, but with *language policy* being the preferred term:

- (7) **sprogpolitik**
 1. valg af et eller flere sprog og hermed også fravalg af et eller flere andre sprog
 2. valg og hermed også fravalg af bestemte stilmidler, regler, ord eller ordformer i et sprog (*Den Danske Netordbog* 2004)
[language policy
 1. selection of one or of more languages, that also means the voting out of one or more languages
 2. selection and so also the voting out of certain styles, grammatical rules, words or inflections or spellings within a language]

However, this explanation is not in full accordance with the standard, where *language planning* is often regarded as the superordinate and *language policy* as the subordinate term. This often prevails as a result of the opinion that language planning describes and develops generic theories and methods whereas language policy is the implementation of language planning, as seen in the following articles from a didactic dictionary:

- (8) **language planning** planning, usually by a government or government agency, concerning choice of national or official language(s), ways of spreading the use of a language, spelling reform, the addition of new words to language and other language problems. Through language planning, an official language policy is established and/or implemented. (Richards, Platt and Platt 1992: 203)
- (9) **language policy** see language planning. (Richards, Platt and Platt 1992: 203)

But in the end theoretical contributions and specialized dictionaries also give a confusing picture. This field has a need for applying its own aims and methods to its own terminology, e.g. a linguistic dictionary distinguishing between language planning as an intralingual regulation by someone and language policy as an inter- and intralingual language regulation with official government participation:

(10) **language planning**

(also language engineering) Making deliberate decisions about the form of a language, such as choosing among competing forms and inventing new vocabulary. This is most often carried out on some kind of official basis. (Trask 1997: 127)

(11) **language policy**

An official government policy regulating the form, teaching or use of one or more languages within the area controlled by that government. (Trask 1997: 127)

2.2 Cultural and scientific habits

Terms like *language planning*, *language policy* and *language politics* have often been used in a haphazard way and very seldom has a distinction been made between e.g. *language policy* and *communication policy*. The latter term should be more often introduced to cover an extensive domain in the language-planning field. Linguists should develop the habit of using these terms in a well-defined way.

Different terminology in different languages

One of the real problems in the discussion of language policy is the lack of a standardised terminology. This does not only apply to the intralingual level but especially to the interlingual level. As a result, confusion easily prevails, causing participants in the scientific discussion often to be unsure whether their use of a given term will be interpreted unambiguously. Confused users will not really find clarity in dictionaries. The English terms *language policy* and *language politics* may in some environments perhaps be regarded as synonyms (cf. their presentation as translation equivalents of the Dutch word *taalpolitiek* in the *Van Dale Groot Woordenboek Nederlands-Engels*). However, this does not imply that for instance their Afrikaans equivalents *taalpolitiek* and *taalbeleid* are used as synonyms. Linguists working in the field of language policy should make a serious attempt to ensure the standardisation of terminology, which should be reflected in dictionaries. Terms from the semantic field of language policy and language planning are often used in communication situations between subject-field experts and laypersons (cf. Bergenholtz and Tarp 1995). These terms should therefore be reflected not only in dictionaries dealing with languages

for special purposes but also in dictionaries dealing with language for general purposes.

Who plans what for whom and how and why?

In a language-sensitive environment there may well be a variety of language policies, ranging from the official language policy for a given country or speech community to language policies for specific institutions like schools, universities, companies, church congregations, etc. The official policy may be a political decision taken by a national government, whereas other policies may be formulated by the respective institutions. A publishing house may have its own policy regarding the language usage in its publications. Dictionaries often reflect an existing language policy, e.g. the orthography laid down by an official body, but dictionaries can also impose a biased approach on the language use of a publishing house or a lexicographer.

Language policy is directed at different communication types (e.g. lingual or non-lingual), communication channels (TV, radio, newspapers, e-mails) and communication venues (meetings, assemblies) (cf. Bergeholtz 2006).

A language policy, usually introduced by means of prescriptive rules or a language plan, should always be directed at a specific target group which may range from all the citizens of a country to a small group of people being members of a specific institution. The language policy will concern the outgoing communication of the target group or the institute-internal communication occurring between members of the group. A typical example could be the language policy of the University of Stellenbosch. This policy has led to a language plan, which gives directives to staff and students regarding the use of Afrikaans and English, especially in the lecture-hall environment. The formulation of a language plan has to be seen as one of the results of a language policy.

2.3 Different types of communication and language policies

The preceding discussion and the following definitions are meant for the metalevel, i.e. the scientific discussion of the fields of communication policy and language policy (or language planning). Under the recommended terms we have listed expressions, which seem to be synonymous with our preferred term. The distinction between the theory and the real and concrete communication policies is reflected in some of these proposals as *theory of language planning* versus *language planning* (Tauli 1974: 56) or *language planning* versus *language policy*, in the sense of Fettes (1997: 14). The most convincing terminological use seems to be the distinction between *language policy* and *language plan* (*Language Policy and Plan for South Africa* 2000). In international terminology, however, one does not find such a clear distinction. *Language policy* is used for the theory and methodology and the results are also called *language policy*. We

want to propose that the terminological use of *language plan* should be adopted for the results, and *language policy* or *language planning* for the theory. The sum of all kinds of communication and language policies should be called *communication and language plan*.

Terminological entries can serve different functions. We have to distinguish between different types of users, different assumptions and different user situations. For a successful reception of texts on language policy, we need to take into account as many frequently used terms as possible, in the following indicated as synonyms by means of an equal-sign. For experts in the field of language policy, the theoretically-systematically formed terms are recommended. For laypeople and for semi-experts, i.e. for experts in fields other than language policy (lexicographers, politicians, sociologists, etc.), we recommend other terms: those formed not quite so systematically but rather tending to be transparent and comprehensible without fully having to understand the theoretical background. Therefore the term *specific intralingual language policy* is recommended for theoretical contributions to language policy, while the synonym term *linguistic units selecting language policy* is recommended for laypeople and experts in other fields than language policy. In some cases only one term is recommended for both user groups, e.g. *communication policy*:

general communication policy (this term is generally recommended)

- = general language policy
- = general language planning

A general communication policy involves the deliberate control of an organisation's internal and external communication in order to optimise the functionality of the organisation, including its product development and sales. Usually, a communication policy takes into account the values and visions of the organisation concerned. A general communication policy is language-independent and is usually issued by, and applies to, companies, organisations or sub-national authorities.

Language-independent general communication policies may also be issued by large, international organisations, such as the UN, the EU and NATO. In some cases, however, such a policy may also apply to an entire state, as can be seen in the case of certain African countries, such as Senegal and Tanzania (Bathily 2005; Tanzania 1997). International organisations and large companies often include a general description of their basic values in their language policies. Normally, however, the term *communication policy* does not include consideration of the value system entrenched in the company's or organisation's language policy.

The result of a general communication policy should be called **general communication plan**.

specific communication policy (this term is generally recommended)

- = specific language policy
- = specific language planning

A specific communication policy involves general or thematic requirements or prohibitions concerning particular forms or channels of communication or a

prohibition against acquiring particular information or documents. Such prohibitions include requirements concerning general or limited access to, or use of, specific types of information.

Parts of a specific communication policy may not only apply to employees of the organisation or authority involved, but also, or solely, to enquiries made by customers or citizens. Thus, communication policy differs from language policy, which is always directed at the internal or external communication of employees, members or students. Whereas a general communication policy rarely applies to states, several examples of specific communication policies on a state level can be found, particularly in dictatorships. Such communication policies may include the prohibition of flyers, phone calls to foreign countries or use of the Internet. Examples of such communication policies can also be found in some existing language policies.

The result of a specific communication policy should be called **specific communication plan**.

interlingual communication policy

→ □ interlingual language policy

intralingual communication policy

→ □ intralingual language policy

Along with this definition of *communication policy*, we arrive at a definition of the term *language policy* (and *language planning*) that should be free from the flaws and inadequacies inherent in the afore-mentioned definition. Simultaneously a number of synonyms will be listed in accordance with the terminology used in various theories and practices. It should be noted, however, that many language policies contain elements, which we have classified as being part of a communication policy. In such language policies, these elements form an intrinsic part of the language policy, taking one of the following two forms:

basic values of a language policy and language planning

→ □ general communication policy

selection of communication types, communication channels and the use of non-verbal communication

→ □ specific communication policy

Despite this being the case, the terms *language policy* (or *language planning*) and *communication policy* may concern the promotion of different verbal and non-verbal languages. Whereas *communication policy* may also include guidelines on the selection and use of non-verbal languages, the term *language policy* (or *language planning*) refers solely to the selection and use of verbal communication forms in different contexts and may therefore be defined as follows:

language policy (this term is generally recommended)

= language planning

Language policy, which refers to the deliberate control of issues pertaining to language, concerns relationships between languages, consisting of interlingual relations, on the one hand, and issues specific to one language, consisting of intralingual concerns, on the other. Though a language policy may form part of a communication policy, the existence of a communication policy is not a prerequisite for the existence of a language policy.

The result of a language policy should be called a **language plan**.

interlingual language policy (this term is recommended for theoretical contributions to language policy)

= **language selection and promotion** (this term is recommended for contributions for laymen and for experts in other fields than language policy)

= status planning

= interlingual communication policy

An interlingual language policy is the clear and deliberate choice, recommendation or promotion of one or more language(s).

The result of an interlingual language policy should be called an **interlingual language plan**.

general interlingual language policy (this term is recommended for theoretical contributions to language policy)

= **language selection** (this term is recommended for contributions for laymen and for experts in other fields than language policy)

= general interlingual language planning

= status planning

= general language policy

A general interlingual language policy involves language selection, entailing the selection of certain languages at the expense of others. This type of selection concerns not only a choice of which languages are to be used in communication, but also which languages are chosen as subjects in schools and universities.

The result of a general interlingual language policy should be called a **general interlingual language plan**. A language plan such as this could result in the production of bilingual or polylingual dictionaries in the selected or supported languages.

specific interlingual language policy (this term is recommended for theoretical contributions to language policy)

= **language promotion** (this term is recommended for contributions for laymen and for experts in other fields than language policy)

= specific interlingual language planning

= status planning

A specific interlingual language policy serves as a protectionist language policy that prescribes the promotion of one or more language(s), either by means of language-directed legislation or by means of financial and political support for selected organisations that promote a particular language or culture. Language-directed legislation concerning both lingual and non-lingual languages, primarily apply within the country that promulgates the legislation in ques-

tion. Organisations promoting a particular language or culture are supported in order to promote or increase the knowledge and use of the language in question, primarily in foreign countries.

The result of a specific interlingual language policy should be called a **specific interlingual language plan**. The result of such a language plan can be the production of bilingual or polylingual dictionaries in the selected or supported languages.

special form of specific interlingual language policy (this term is recommended for theoretical contributions to language policy)

= **linguistic hegemony** (this term is recommended for contributions for laymen and for experts in other fields than language policy)

= special form of interlingual language planning

A special form of specific interlingual language policy is a hegemonic language policy, involving an opposition to one or more language(s) in favour of one or more other language(s).

The result of a specific form of interlingual language policy should be called a **special form of interlingual language plan**.

intralingual language policy (this term is recommended for theoretical contributions to language policy)

= **style and linguistic units selecting language policy** (this term is recommended for contributions for laymen and for experts in other fields than language policy)

= intralingual language planning

An intralingual language policy is the choice or recommendation of, the warning against, or the banning of certain linguistic constructions, collocations, phrases or words in a particular language.

The result of an intralingual language policy should be called an **intralingual language plan**.

general intralingual language policy (this term is recommended for theoretical contributions to language policy)

= **style selecting language policy** (this term is recommended for contributions for laymen and for experts in other fields than language policy)

= general intralingual language planning

= communication-optimising language policy

= language guide

= manual of style

= style book

= language hygiene

A general intralingual language policy involves stylistic selection, meaning a clear and deliberate choice or recommendation of specific stylistic aspects or rules pertaining to a particular language.

The result of a general intralingual language policy should be called a **general intralingual language plan**. A general intralingual language plan may result in the production of style guides or handbooks for language use.

specific intralingual language policy (this term is recommended for theoretical contributions to language policy)

= **linguistic units selecting language policy** (this term is recommended for contributions for laymen and for experts in other fields than language policy)

= specific intralingual language planning

= prescriptive linguistics

= language hygiene

= corpus planning

A specific intralingual language policy involves the selection of linguistic units, meaning a clear and deliberate choice or recommendation regarding specific grammatical constructions, words or word forms in a particular language.

The result of a specific intralingual language policy should be called a **specific intralingual language plan**. Such a language policy may result in the production of dictionaries or grammars.

special form of specific intralingual language policy (this term is recommended for theoretical contributions to language policy)

= **puristic language policy** (this term is recommended for contributions for laymen and for experts in other fields than language policy)

= special form of intralingual language planning

= purism

= ideological language policy

= language hygiene

A special form of specific intralingual language policy is a puristic language policy based on a particular historical, moral or political perception of proper language, which results in the prohibition of, or opposition to, certain grammatical constructions, words or word forms in a particular language.

The result of a special form of specific intralingual language policy should be called a **special form of intralingual language plan**. Such a language policy may result in the production of handbooks or specialised dictionaries.

3. Communication policy and lexicography

In the following sections we will give some comments on the relations between the different kinds of communication and language policy and lexicography. For every term from the above list one could write a separate contribution. We will therefore only give some general comments with examples from some concrete language policies. According to the preceding argumentation, we will use the terms for semi-experts and laypeople, i.e. in relation to the field of language policy for lexicographers.

3.1 General communication policy

What we call "general communication policy", i.e. the incorporation of values for decisions for a nation, an organisation, a company, a family or an individual, is the main topic in those policies usually called "communication policy",

but values (or visions and missions) are also part of the theoretical discussion of language policy and are part of many language policy plans:

This language policy is intended as an enabling framework for promoting South Africa's linguistic diversity and encouraging respect for language rights within the policy framework of building and consolidating a united democratic South African nation. (*Language Policy and Plan for South Africa*, 1.1)

This is a very general vision using a language policy for the consolidation of democracy or, more realistically, a step to a more democratic South Africa. In the end the choice of — only — 11 languages is not a fully democratic selection because of the omission of languages like Khoi, San, and commonly used languages like Hindi, Arabic, Tamil, etc. More concrete is the clear wish to promote and create conditions for the development and use of all official languages (The Language Plan Task Group (LANGTAG) 1996 Chapter 1, section 5, cf. Kamwangamalu 2004: 262). In the current discussion in contributions to interlingual language policy one finds a focus on or a restriction to government decisions, e.g. Webb (2002: 40). This is too narrow and also a contradiction in the practice of language policies where values are a prerequisite for every communication and language policy plan on all levels.

Democracy as the main value is not only a question of support for language diversity, but also for an easy and easily understandable language use, as it is claimed in the latest Swedish governmental proposal for a language policy:

En levande demokrati, där medborgarna deltar i det offentliga samtalet och kan göra sina röster hörda, förutsätter klara och begripliga myndighetstexter. Detta ökar också rättssäkerheten och bidrar till ökad effektivitet i förvaltningen. (Swedish Governmental Proposal 2005: 1)

[A living democracy with an active participation in the public discussion of all citizens needs clear and easily understandable official texts. Such a communication will more efficiently increase the legal security and support in the administration.]

It may be difficult to define the exact borderline between values as a part of a general communication policy and the linguistic units selecting language policy as part of the intralingual policy. We consider the view "influence two dialects minimizing the differences between them" as being a value or a political wish for equalizing differences in a language community. This has been the case in the official language policy in Norway regarding Nynorsk and Bokmaal for the last 50 years until 2005 when the Norwegian language council took the decision to promote the differences between the two official dialects (Norwegian Language Council 2005). If this decision is not reversed within the next 100 or 200 years, we can expect to have two languages in Norway and not two dialects as at present. This example is extraordinary because in almost all intralingual policies the selected value(s) is (are) the same as that in Norway before, and the opposite of that after 2005.

One kind of value selection known from linguistic discussions will — if it is chosen — be a value leading to no language policy: Leave the language alone! This post-behaviouristic, almost religious credo claims that every language change is "natural" and non-natural changes are manipulative:

For change that comes spontaneously from below, or within, our policy should be, Let your language alone, and leave its speakers alone! But other forms of language manipulation have other origins, other motives, other effects, and are far more dangerous. (Lakoff 1990: 298)

To this anticomunicative and language-political statement Cameron (1995: 20) begins to give the right answer, formulated as a question: Why is spontaneity the only value? How about truth, beauty, logic, utility? We are not sure that Cameron's possible values are the most important ones for policies. The choice of and arguments for every value as pre-condition for language policies has to be discussed separately. For every kind of communication and language policy one has quite different values: for communication policies see Kotler and Keller (2006: 534-562), for interlingual language policies see Webb (2002: 39f), and for intralingual language policies see Bergenholtz et al. (2003). Both interlingual and intralingual policy values (democracy, the creation of conditions for the development of language(s), more effective communication, etc.) invites among other decisions a discussion of the needed dictionaries for a specific nation according to its specific social conditions.

3.2 Specific communication policy

Traditionally we only discuss the need for printed language dictionaries. We have to add to it the need for sign-language dictionaries and access to such dictionaries, as well as language dictionaries on the Internet. The choice between paper and electronic dictionaries concerns not only governmental, but also personal and organisational decisions to determine whether the access to Internet lexicographical tools gets priority, partial priority or no priority. An estimate of the number of Internet dictionaries shows more than 100 000 (cf. <www.yourdictionary.com>). Many of these dictionaries are developed by individuals or produced by companies or universities. Only a minority has government support. On the given homepage we find very big differences between the large languages, which means between the 100 largest languages in the world. The largest languages (Chinese, English, Russian, German, and Spanish) have many dictionaries. In some cases one cannot explain the differences between languages in terms of the respective number of native speakers or their being well documented and standardised or even their status as official languages. Sometimes differences exist between languages spoken in the Third World, but this is not the full explanation, e.g. Somali has four dictionaries compared to the six of Dutch in the yourdictionary list. It is also not only a question of languages with many or few native speakers. Faeroese, with 45 000 native speakers, has

three dictionaries, almost the same as Afrikaans with four Internet dictionaries and more than six million native speakers. Also with regard to three well-documented languages, all in the top-200 list of the largest languages in the world, the medium of the Internet dictionary is used in varying degrees:

- Catalan has 6,7 million native speakers and five million second-language speakers (1996). With Spanish the national language in Spain, Catalan is only one of the official languages in parts of Spain. The collection of Catalan Internet dictionaries on <www.yourdictionary.com> nevertheless comprises 29 titles, among which 22 are specialized language dictionaries.
- Danish, the only official language in Denmark, has six million native speakers (2006). For Danish we find 17 dictionaries in the above-mentioned dictionary bank.
- Afrikaans has six million native speakers and 10,3 million second-language speakers (1996). It is one of the 11 official languages in South Africa. For Afrikaans we find only four titles, three English–Afrikaans dictionaries and one specialized language dictionary, but surprisingly no monolingual dictionary.

The difference between Danish and Catalan (the latter not being an official national language) on the one side and Afrikaans as an official language on the other is significant. One explanation could be that the fight for Catalan in Catalonia (a province in Spain) is so intensive, that the local government, local universities, local organisations and single individuals spend time and money on developing dictionaries as part of the fight for their language. Such considerations could and should be made for many more languages. Internet dictionaries will become still more important in future.

4. Language policy and lexicography

The main distinction made in this article is between inter- and intralingual language policy. It does not mean that one has to discuss bilingual lexicography only as part of interlingual language policy. Bilingual dictionaries are also a topic for intralingual decisions concerning the choice of linguistic units for the language in bi- or polylingual dictionaries (cf. par. 4.5–4.6).

4.1 Language selection

It would be unrealistic to plan the development of different dictionaries for every lexicographical function for every language. Even attempting to compile

only a polyfunctional dictionary (cf. Bergenholtz 1997 and Gouws and Prinsloo 2005: 54f) for every language would be difficult or even impossible to achieve. It is not only a question of money (cf. the example from Greenland in the next section). If we assume that we have about 6 800 languages in the world, we can make the following calculation: For every language one needs at least one general language dictionary, at least one general encyclopaedia, at least 150 different specialized dictionaries, e.g. a spelling, synonymy, collocation, idiom, proverb dictionary, etc. etc., an accounting, gene technology, computer dictionary, etc. etc. Furthermore we need at least two bilingual dictionaries for every language pair (L1 → L2 and L2 → L1 (6 800 × 2 = 13 600) and the same for the single specialized fields (accounting, computer science, etc.). A quite conservative assumption gives us at least 100 different fields or 100 × 6 800 × 2 = 1 360 000). For each of the estimated 6 800 languages in the world one needs at least

- 1 monolingual general language dictionary,
- 1 monolingual general encyclopaedia,
- 150 monolingual specialized dictionaries,
- 13 600 bilingual general language dictionaries, and
- 1 360 000 bilingual specialized dictionaries.

For the 11 official languages in South Africa the number of "needed" dictionaries almost equals the total population of the country. We could proceed with the same speculative calculation for all the languages in the world and will come to the same result: Dictionaries are important tools helping people with some kind of communication and knowledge problems, but we will never reach the ideal number of needed dictionaries. Some potential dictionaries will have so few users that it would not be worth the effort trying to produce them. We therefore have to decide which dictionary types for which languages and language pairs should be supported, if the market mechanism, i.e. the number of sold paper dictionaries or subscribers for Internet dictionaries, cannot pay for the necessary theoretical and practical lexicographical work.

We also have to decide which dialect is to be chosen as the standard for a certain language in a certain dictionary. This topic cannot be treated here, but it must be stated that the concrete decision is to be made as part of the specific intralingual policy (see par. 4.5).

Another language selection decision to be made is the choice of the explanation language, especially in bilingual dictionaries, and the choice of language for the outer texts (preface, user guide, dictionary grammar, etc.). This decision should be motivated by selected dictionary functions for the specific dictionary (cf. Bergenholtz and Tarp 2003). But the lexicographer (or the government or the organisation paying for the lexicographical work) could also have language policy considerations of overriding importance. During the planning of the *Malagasy–Aleman Dictionary* (1991) it was proposed from governmental side that the outer texts should be written in Malagasy, German and also French.

The reason for this proposal was that the language of instruction in many high schools is French — also for teaching German. As a matter of fact, most teachers and pupils are not used to Malagasy grammar and other Malagasy linguistic terms. They of course understand and use their mother language, but not as language of instruction. Although Henning Bergenholtz, as director of the lexicography project, had to agree to the correctness of this description of the language situation and the arguments for French as one of three dictionary explanation languages, he did not follow the proposition because he and other members of the research group pleaded for another language policy with a broader use of Malagasy as language of instruction. Such decisions for the selection of the dictionary explanation language or language pair are not unique to Madagascar but are necessary in many countries where bilingual dictionaries are to be compiled in polylingual communities.

4.2 Language promotion

In the 500-year-old history of printed dictionaries and also in the very short history of Internet dictionaries, many dictionaries could not have been produced if there had not been governmental or private promotion of the dictionary or lexicography centre. Money and trained lexicographers are needed. These needs are often referred to in theoretical discussions, e.g. by Lopes (2004: 185) and Kamvangamalu (2004: 62), but they are also part of national language policies, e.g. in the Danish Cultural Ministry's *Sprog på spil* 2003, the Swedish Governmental proposal (2005: 7ff) or the Pan South African Language Board (*About South Africa* 2005). The real need for money and the financial planning of lexicographical projects are not often the subject of scientific publications. It is a desideratum when one considers that big national monolingual dictionaries using large amounts of money for many years in long-term projects, as for the *Woordeboek van die Afrikaanse Taal* in South Africa (since 1926 and not finished yet), the *Norsk Ordbok* in Norway (since 1930 and not finished yet, cf. Bergenholtz 1996) or the *Svensk Akademiens Ordbok* (cf. Bergenholtz and Malmgren 2000). The latter dictionary project started in 1892 and 115 years later the editorial team is now working on the letter T, having had a budget with a current value of approximately 1 000 000 000 SEK or 100 000 000 Euro. One of the few publications on the topic of planning the costs for small and larger dictionary projects is the publication *Lexicography as a Financial Asset* (1996) in which many of the contributions surprisingly speaks more about the education of and need for lexicographers and not much about the financing and the calculation of expenses, e.g. Alexander 1996, Mini 1996, Prinsloo 1996a, 1996b and Gouws 1996. However, in some ways it might be the right point to start. As a matter of fact: Money is not enough. The education of lexicographers is a primary prerequisite for dictionary-making. To prove this we may use the example of the development of a Greenlandic monolingual dictionary. In 1978 Greenland's Provincial Council proposed the compilation of such a dictionary, for which

the government granted about 7 000 000 DEK (in today's value approximately 2 000 000 Euro). The project "Ordbogersuaq" started with meetings and discussions, spending about half a million crowns on refreshments (cf. Langgård 1997). After some years the rest of the money was returned to the government because the dictionary planners, being unacquainted with lexicography, did not succeed in drawing up any real schemes for or making any real progress with the dictionary.

4.3 Linguistic hegemony

Linguistic hegemony, as a special form of interlingual language policy, can influence lexicography in a bi- or multilingual community where all languages do not have the same official status or recognition as preferred language. Dictionaries can be influenced by having to reflect this situation or they can reflect a view independent of the hegemonic approach.

In Gabon, a multilingual country with 62 indigenous languages, French is the only official language. Since 1994, a much more lenient approach exists towards the indigenous languages seeing that the revised constitution states that the national languages should be promoted and protected. Until then French had not only been the sole official language but it had also been imposed as the only allowed language of communication, e.g. in the educational system. This ruling refers to the language of instruction as well as to the general communication system in schools. Strict rules were enforced to forbid learners using any language but French on the school grounds and where learners trespassed they were punished (cf. Mavoungou 2002). This language policy has been detrimental to the development of lexicography in Gabon. There was no realisation of the need for the compilation of French dictionaries because users could rely on French dictionaries compiled in France. Even though a new variety of French, i.e. Gabonese French, emerged in Gabon, the hegemony of the language of colonisation eschewed the need for a lexicographical acknowledgement of this new variety. The French dictionaries being used in schools, in public life and in households presented a variety of French not fully reflecting the day-to-day language use of the Gabonese speech communities. The position of the indigenous languages was much worse. Due to the lack of any official recognition, and the dominance and hegemony of French there was no support for lexicographical projects in any of the indigenous languages. Language policy impeded lexicographical activities in the indigenous languages and led to a situation where the official language was lexicographically represented in a super-standard not fully reflecting the real language usage.

In spite of the language policy some bilingual dictionaries with French and an indigenous language were compiled. If the remarks of Gallardo (1980: 61) that "in non-standardized language situations, dictionaries do not exist or, at best, are bilingual dictionaries, that is, compiled in function of a different

language" are considered, this compilation of bilingual dictionaries combining French with an indigenous language may seem the proper way to follow. However, the genuine purpose of these dictionaries was not to promote the relevant indigenous language or to give its speech community a source of reference. Nyangone Assam and Mavoungou (2000: 269) clearly state that most of the Gabonese bilingual dictionaries dealing with French and an indigenous language, are biased towards French. They emphasise the major role of missionaries and colonial administrators in the compilation of Gabonese dictionaries. Their overview of Gabonese lexicography leads them to remark: "As far as dictionary purposes are concerned, most earlier dictionaries were compiled to help European traders and explorers to communicate in Gabonese languages. They were also useful aids for Christian evangelisation and French colonial administration." (2000: 269).

The situation in Gabon was not unfamiliar, for similar situations prevailed in many other African countries. Kidida Awak (1990: 10) states that many dictionaries in Africa were not compiled for use by Africans but to assist the European explorers as expeditionary guides or to help missionaries in learning the African languages for evangelisation purposes.

Gouws (2005) makes a distinction between externally-motivated and internally-motivated lexicographical endeavours in Africa. Externally-motivated dictionaries are usually bilingual products, co-ordinating a local African language with a European language, typically the first language of the missionaries or the colonisation officials. The compilers of these products are not speakers of the local language but rather of the language of colonisation. Internally-motivated lexicographical endeavours are the dictionaries compiled from within the given speech community. Externally-motivated lexicographical endeavours often are the results of a prevailing language policy that protects the hegemony of the language of colonisation.

Lexicography in Africa gives evidence of varying forms of the influence of linguistic hegemony or attempted linguistic hegemony. In this regard externally-motivated dictionaries played a significant role. For many years during the nineteenth century English was the only official language of the Cape Province in South Africa. Since the late seventeenth century, Afrikaans had already been a divergent dialect of Dutch (cf. Gouws and Ponelis 1992). The early development of Afrikaans occurred in a pre-lexicographical environment and colonial Dutch received scant lexicographical attention (cf. Gouws and Ponelis 1992: 81). The emerging Afrikaans showed increasing differences from the then standard Dutch as used in the Netherlands. Although Dutch was no longer an official language in South Africa, there was a definite Dutch interest in the linguistic situation in South Africa, an interest that illustrates a form of distant linguistic hegemony. In 1831 the Dutch scholar A.N.E. Changuion came to South Africa as professor in classic and modern languages, especially Dutch literature (cf. Van der Merwe 1971). In 1844 he published his well-known book *De Nederduitsche taal in Zuid-Afrika hersteld, zijnde eene handleiding tot de kennis*

dier taal, naar de plaatselijke behoefte van het land gewijzigd. His intention with this book was to give Dutch such a firm base in the Cape that Afrikaans could make no progress. This book contained a back-matter text "Proeve van Kaapsch taal-eigen", one of the first dictionaries treating Afrikaans. In the preface Changuion said that the practical aim of this dictionary was to rid Dutch, spoken in South Africa, from the "corrupt" words and expressions that he encountered in South Africa, signalling the beginning of the new language, Afrikaans. This external motivation was a typical Euro-centred approach to African lexicography and an attempt to maintain linguistic hegemony, although from a distance.

Some externally-motivated dictionaries had also been compiled to benefit the local language of African speech communities and to go against attempted hegemony. In 1884 the Dutch academic Mansvelt published an Afrikaans dictionary *Proeve van een Kaapsch-Hollandsch Idioticon met Toelichtingen en Opmerkingen betreffende Land, Volk en Taal*. He compiled this dictionary to prove that Afrikaans differs sufficiently from the then standard Dutch to be regarded as an emerging language and not merely a dialect of Dutch that has to conform to the rules of standard Dutch. As in the case of Changuion's word list, Mansvelt's dictionary was published at a time when English was the only official language in the Cape Province. The monolingual language policy and the consequent linguistic hegemony did not deter the continuing presence of an ever-changing emerging language or the attempts to impose an older form of linguistic hegemony amongst a certain group of speakers at the Cape.

The history of Afrikaans lexicography indicates that lexicographical attempts against linguistic hegemony were complemented by attempts to substitute linguistic hegemony with linguistic harmony. Notwithstanding an official language policy that only made provision for one official language, lexicographers attempted to ensure harmony between speakers of different languages. In 1902/1904 the *Patriot Woordeboek/Patriot Dictionary* was published. This dictionary, appearing shortly after the Anglo-Boer War (1899–1902), played a major role to introduce Afrikaans as an independent language coexisting alongside English. From the perspective of the development of Afrikaans lexicography, the *Patriot Woordeboek/Patriot Dictionary* played an extremely important role, which will not be discussed in this article. Of significance here is its attempt to ensure linguistic harmony in spite of the official language policy of linguistic hegemony. In the preface the editor of the *Patriot Woordeboek/Patriot Dictionary* concludes, after stating various aims and objectives: "This is our principal aim: to promote co-operation between the leading races, English and Dutch, in South Africa, for which purpose it is indispensable that they should mutually know each other's language ..." This dictionary does not adhere to an approach to expand English to the detriment of Afrikaans but it endeavours to create a form of harmony between speakers of English and Afrikaans.

In the first decades following its formal recognition as second official language besides English, much was done to promote Afrikaans. The beginning of the work on the *Woordeboek van die Afrikaanse Taal* (the WAT) in 1926 intro-

duced a lexicographical era characterised by state support for this Afrikaans project. This support was increased substantially and the dictionary became one of the symbols of the success of Afrikaans. At a much later stage a much more limited financial allocation was made to the *Dictionary of South African English*. This support for the two official languages stood in sharp contrast to the lack of support for lexicographical projects in the other South African languages. Since the political transformation of 1994, there has been an official change in the favouring of only two languages to the detriment of the rest. The establishment of eleven national lexicography units, one for each of the official South African languages, has created the opportunity, at least in theory, for all the official languages to be promoted and to embark on lexicographical projects. In practice, however, the application of the official policy of eleven national languages leaves much to desire and it still has to be seen how this language policy will influence South African lexicography. The strong position of English as language of documentation has already resulted in the choice by a number of national lexicography units to compile bilingual dictionaries with English as partner language. If not managed in a proper way, this may lead to a new linguistic and lexicographical hegemony.

4.4 Style selecting language policy

This kind of language policy is part of some national language policies (cf. the discussion above and the citation from the Swedish Governmental proposal 2005). We are not aware of any such concrete national language policy plans in practice but in private companies and organisations below governmental level it is the most frequently used type of language policy plan. Their concern is better communicative habits to ensure a homogeneous language use (because a non-homogeneous language use sends a signal of non-professionalism) and a clear and easily understood communication. We only know of small corporate dictionaries (word lists) as part of corporate language policies, demanding e.g. the use of short words instead of long word formations or the abolishment of foreign words. Such demands are collected in small word lists, e.g. that of the Danish Pump Company (cf. *Grundfos* 2002 and Bergenholtz et al. 2003). For future corporate, but also national dictionaries we will suggest the development of an integrated outer text with style selecting proposals to which references from all relevant entries should be made.

4.5 Linguistic units selecting language policy

Dictionaries are too often regarded by their users as authoritative sources of knowledge that should never be questioned. However, as Samuel Johnson indicated, the authority of a dictionary is only that of its compiler. Yet, dictionaries are seen as sources in which the "proper" language usage is described and

if a given form has been included in a dictionary the average dictionary user regards it as an indication that it had been accepted as part of the standard language. Once again it is important to be reminded of the remark by the editor of *Webster's Third New International Dictionary of the English Language*, Philip Gove, that the task of the dictionary is to record language, not to set its style. This would imply that a dictionary reflecting the real language should not be impeded by puristic considerations resulting from a biased or prescriptive approach on the side of the lexicographer.

An intralingual language policy may be an attempt to ensure the use of the standardised spelling system, and dictionaries may be required to utilise this standardised orthography. In this regard dictionaries often adhere to the orthography rules laid down by a formal language or spelling commission (cf. par. 4.5). For Afrikaans the Taalkommissie van die Suid-Afrikaanse Akademie vir Wetenskap en Kuns is the body responsible for the official ruling of Afrikaans orthography. Its decisions are published in the successive editions of the *Afrikaanse Woordelys en Spelreëls* and dictionaries are expected to abide by the spelling forms and rules presented in this book. A language policy may determine that orthography needs to be prescribed by a given body, and for dictionaries to fulfil their obligation as authoritative instruments it is wise to adhere to the laid-down criteria. Where evidence from actual language use goes against the suggestions and rules of the ruling body, dictionaries, depending on their type and function, could and should indicate the occurrence of the other forms as variants of the officially preferred form.

Only if one has the broad understanding of language policy we have chosen, decisions about orthography and inflection will be a part of a language policy. But this understanding accords exactly to the practice of a language policy for languages with a language commission taking official decisions of this kind, as can be illustrated from a few Danish examples. By law the decisions of the Danish Language Council are to be followed by all governmental employees and pupils or students in institutions of which more than 50% of their expenses are paid by the government. All decisions made by this council are published in an official dictionary. Until November 2001 — when a new edition of this dictionary appeared — there have been two officially valid variants for English *line*: Danish *linie* (with *-ie*) or *linje* (with *-je*). But since November 2001 only one variant (*linje*) has been permitted. Another example is the translation for the English *combined power and heating plant station*: Danish *kraft-varmeværk*, whose spelling without hyphens was the only valid form until November 2001. This is now forbidden and only *kraft-varme-værk* (with two hyphens) permitted. In the Danish language law from September 1997 (Law on Orthography 1997 and Law on Language Council 1997) and in the circular to this law (Circular on Danish Language Council 1997) we find two principles the Danish Language Council should use in its decisions on Danish orthography regulations. The first principle is a rule on consistency, called "the principle

of tradition". It is a practical language-stabilizing rule, without which the changes would be so tremendous that it would be difficult to read texts written 30 or 50 years ago. The second principle is a rule of descriptivity called "the principle of language use". According to this rule the language council has to prescribe spelling according to the language use in texts written by good writers.* With such a law and with about 2 000–3 000 spelling and inflexional changes within a few years, e.g. 1986, 1995, 2001 and 2006, there is a big need for updated dictionaries. Longer periods without changes and a proscriptive method — instead of the prescriptive one used — could be a good solution. In Denmark — and equally in other Scandinavian countries — we have an example of a language policy where a certain dictionary is explicitly mentioned as part of the law in its very first paragraph:

§ 1. Dansk retskrivning fastlægges af Dansk Sprognævn og offentliggøres i Dansk Sprognævns retskrivningsordbog. (Law on Orthography 1997)

[Danish orthography is decided by the Danish Language Council and is published in the spelling dictionary *Retskrivningsordbogen*]

The argument for the choice of one variant and the prohibition of another in cases like *linie* (forbidden since 2001) and *linje* (the only form allowed since 2001) is perhaps that the latter corresponds to the pronunciation, the former does not. Such an argument accords with the intention in the process of uniformity in many languages without a long tradition of being used as written language (cf. Chebanne et al. 2003: 3). Such cannot be the case in the Danish decision because this intention is not part of the language law (cf. the above reference to the principle of tradition as a main rule for official regulations: language use in texts written by good writers). Other changes can be explained as attempts to make the language more easily readable (prescribing the use of hyphens in long word formations like *kraft-varme-værk*). Such proposed changes have indeed a better chance of acceptance, if they accord with other orthographical conventions in a certain language. But the condition for a successful proposal is not only the acceptance of but also the easy access to information about the (new) norm. Dictionaries are the normal tools for quick access to such information. In that regard internet dictionaries are much more effective because the changes can immediately be communicated.

4.6 Puristic language policy

Language policy, as a form of intralingual language policy, is based on a particular historical, moral or political conception of proper language. Such a policy resorts primarily but not exclusively on an intralingual level and it should never be interpreted as being predominantly interlingual in nature.

A puristic language policy often is an attempt to keep a language "pure" from either the influence of another language but, especially, from the influence of non-standard varieties, ideologically marked forms or derogatory words and expressions. This is the type of language policy that can lead to dictionaries eschewing the actual language use of their speech communities. A publisher may adhere to such a language policy and use its dictionaries to promote this approach. The individual lexicographer may adopt such a puristic approach in his/her dictionary, knowing quite well that the intended target user of the dictionary may very well interpret this approach as an official ruling on language.

During the first decades following the recognition of Afrikaans as an official language in South Africa there was a strong puristic approach to protect and free Afrikaans from Anglicisms. At that stage Afrikaans was attempting to maintain itself as official language alongside the world language English. Afrikaans linguists preferred forms originating from Dutch to loan-words from English. In the education system a strict puristic approach was implemented and learners often had to learn lists of Anglicisms along with their "pure" Afrikaans equivalents, in order to know which words and expressions to avoid.

The puristic urge in Afrikaans lexicography led to the inclusion of words that did not occur in typical day-to-day communication whilst forms with a high usage frequency, but stigmatised as Anglicisms, were omitted from dictionaries. In the fourteenth edition (1997) of *Groot Woordeboek/Major Dictionary* forms like *bruismelk/roomysmelk/skuimmelk* instead of *melkskommel* (for the English *milkshake*) and *posbesteller/briewebesteller* instead of *posman/posbode* (for the English *postman*) appeared. In the eighth edition (1984) of *Tweetalige Woordeboek/Bilingual Dictionary* a well-known Afrikaans word like *geboortemerk* (*birthmark*) is omitted in favour of the Dutch form *moedervlek* whilst *bookmark* gets the equivalents *boeklêer* and *leeswyser* but not the high-frequency word *boekmerk* (cf. Gouws 1993, 1995). This kind of puristic approach has impeded the lexicographical representation of actual Afrikaans usage.

True to its typological nature the comprehensive *Woordeboek van die Afrikaanse Taal* (WAT) has endeavoured to include a variety of Afrikaans forms without the same puristic approach. However, it is only in recent volumes of the WAT that informal English loan-words in Afrikaans are systematically lemmatised. In this regard many members of the Afrikaans speech community are still conservative, favouring the puristic approach in dictionaries. In the fourth edition (2000) of HAT, the *Verklarende Handwoordeboek van die Afrikaanse Taal*, a number of informal loan-words from English like *cool*, *sorry*, *oukei*, *orraait* were included and labelled to indicate their status. The inclusion of these words was the aspect of this dictionary which received the most criticism, several users insisting that the dictionary should adhere to an approach of including only pure Afrikaans words.

The inclusion of this type of loan-word was increased in the fifth edition (2005) of HAT due to a language policy on the side of the editors that aims to give a balanced account of the lexicon of Afrikaans, also allowing items coming

even from non-standardised varieties. In contrast to this approach the implementation of a puristic language policy would have resulted in a much less representative account of the Afrikaans lexicon.

Dictionaries can play an important role to maintain the established "pure" forms but still acknowledge the occurrence of loan-words and -expressions, although these loan-forms may contradict the wishes of certain sections of the speech community. Whether they are pure or impure, the lexicographer needs to include these words, for which the knowledgeable user will be looking in the dictionary. In some of the African languages traditional words and loan-words coexist. Nong et al. (2002: 3) indicate that for Sepedi the former Language Board "adopted a sensible approach in being prescriptive in the coinage and approval/disapproval of terminology on the one hand, while still placing a high premium on actual usage as criterion for acceptability on the other hand", allowing for more than one option. For the months of the year the Language Board prescribed the use of "Sothoised" terms like *Aprele* (April) instead of *Moranang* (Nong et al. 2002: 4). Investigating the preference of users when it comes to traditional words versus loan-words, Nong et al. found that Sepedi speakers prefer traditional words like *seyalemoya* (radio) and *setimela* (train) to loan-words like *radio* and *terene*. The research indicated that 29,4% of the respondents prefer the loan-word only, while 70,6% prefer the indigenous word only. However, the research also indicated that younger speakers have a stronger preference for loan-words than older speakers. Whereas the Language Board acknowledged certain loan-forms the existing Sepedi dictionaries display an even more lenient and less puristic approach towards loan-words. From the nine dictionaries investigated by Nong et al. it becomes clear that loan-words and their indigenous counterparts receive an equal treatment, with 50,6 % of the loan-words and 49,4% of the indigenous counterparts included in the dictionaries.

Where the speech community clearly has a bias towards a more puristic approach, dictionaries are progressive and guide users by presenting them with different options. As indicated in the survey, the dictionary-internal language policy may go against the preferences of the majority of the speech community. Dictionaries reflect the usage patterns of the younger generation and still make provision for the preferences of the older generation.

The role of dictionaries in the purist drive is not only to approve or condemn loan-words as counterparts to indigenous forms. Dictionaries should also reflect on derogatory forms, for instance. In this regard the function of the dictionary and the target users will determine the possible inclusion of these forms and their treatment. It would be unacceptable for a comprehensive dictionary to omit sensitive words like *Boer*, *rock-spider*, *kaffer* and *queer*. Such an omission would imply a puristic approach to the detriment of the dictionary being representative of the lexicon of the language. The derogatory nature of these words should be indicated by means of labels, comments or even in the

lexicographical definitions. Even in a school dictionary these words should be included but with an explicit warning against their use.

When it comes to a puristic language policy one cannot formulate a single approach for all dictionaries. As with the contents and structures of dictionaries the users and the lexicographical functions of each dictionary should determine the nature and extent of its puristic approach.

5. Outlook for future dictionaries

This article has indicated that further discussions of language policy issues should be preceded by a clear definition and understanding of terms from this field. As has been shown, these terms had not been efficiently treated in existing dictionaries and this applies both to dictionaries dealing with language for general purposes and those dealing with languages for special purposes. The current treatment is not really helpful but tends to be rather confusing.

In this article, we defined the terms in such a way that they can be employed in an LSP (language for special purposes) dictionary that focuses on language planning. The discussion included suggestions for terminological use by politicians, journalists and other language users participating in the discussion.

The main thesis of the article is that lexicographers of concrete dictionaries constantly make decisions relevant to language policy. This does not only regard the domain that clearly belongs to party or national politics but also the domain of general decisions regarding language and languages. This implies a wide scope of the lexical item *policy* as can be seen in its occurrence in a combination like *smoking policy*.

Language policy has three basic dimensions: the choice of communication channels, the choice between languages and the choice between an interlingual and an intralingual approach. These dimensions also apply to lexicography and we want to emphasise that interlingual and intralingual decisions should be directed at each other and that both these dimensions are indispensable. Up to now intralingual decisions have been the obvious ones for lexicographers. This article acknowledges this, but also stresses the additional necessity and relevance of decisions on the level of communication policy and interlingual language policy.

Endnote

- * In practice the Danish Language Council does not follow these clear rules (Bergenholtz 2003a, 2003b); the concrete decisions are therefore criticized in public discussion. But the debate is not as serious as a similar discussion in Germany (after their language decisions in 1996). A reason for this could be, that new general changes in Germany only take place once in a hundred years, whereas in Denmark changes occurred more than ten times in the same period.

Literature

- About South Africa*. 2005. Pan South African Language Board. <http://www.southafrica.info/ess_info/sa_glance/constitution/pansalb.htm> [January 2006].
- Alexander, N.** 1996. Introduction. *Lexicography as a Financial Asset* 1996: 3-6.
- Bathily, A.** 2005. *Strategic Thinking. Community Radio in Senegal Today. Issues and Outlooks*. <<http://www.comminit.com/africa/strategicthinking/st2005/thinking-1053.html>> [November 2005].
- Bergenholtz, H.** 1996. Norsk international ordbog [review of *Norsk ordbok. Ordbok over det norske folkemålet og det nynorske skriftmålet. Band III. Flusker-Gigla*. Edited by Reidar Bø, Arnbjørg Hageberg, Laurits Killingberth, Sigurd Norlie and Gunnar Pedersen. Oslo: Det Norske Samlaget 1994.]. *Norsk Lingvistik Tidsskrift* 14: 75-84.
- Bergenholtz, H.** 1997. Polyfunktionale ordbøger. *LexicoNordica* 4: 15-29.
- Bergenholtz, H.** 2003a. User-oriented Understanding of Descriptive, Proscriptive and Prescriptive Lexicography. *Lexikos* 13: 65-80.
- Bergenholtz, H.** 2003b. Bryder Dansk Sprognævn den danske sproglov? Sprogpolitik i teori og praksis. Lönnroth, Harry (Ed.). *Från Närpesdialekt till EU-svenska. Festskrift till Kristina Nikula*: 17-31. Tammerfors: Tampere University Press.
- Bergenholtz, H.** 2006. Toward a Definition of Communication Policy, Language Policy and Language Planning. *SPIL PLUS: Stellenbosch Papers in Linguistics Plus* 34.
- Bergenholtz, H. and S.-G. Malmgren.** 1998. Förord. *LexicoNordica* 7: 1-4.
- Bergenholtz, H. and S. Tarp.** 2003. Two Opposing Theories: On H.E. Wiegand's Recent Discovery of Lexicographic Functions. *Hermes, Journal of Linguistics* 31: 171-196.
- Bergenholtz, H. and S. Tarp (Eds.).** 1995. *Manual of Specialised Lexicography. The Preparation of Specialised Dictionaries*. Amsterdam: Benjamins.
- Bergenholtz, H., J. Bisgaard, M.B. Lauridsen and K.K. Wichmann.** 2003. Sprogpolitik: So ein Ding müssen wir auch haben. *Hermes, Journal of Linguistics* 31: 135-165.
- Brockhaus-Wahrig* = Wahrig, G., H. Krämer and H. Zimmermann (Eds.). 1980-1984. *Brockhaus-Wahrig. Deutsches Wörterbuch in sechs Bänden*. Wiesbaden: Brockhaus/Stuttgart: Deutsche Verlagsanstalt.
- Bünting and Karatas = Bünting, Karl-Dieter and Ramona Karatas. 1996. *Deutsches Wörterbuch. Mit der neuen Rechtschreibung*. Chur: Isis.
- Cameron, D.** 1995. *Verbal Hygiene*. London/New York: Routledge.
- Changuion, A.N.E.** 1848². Proeve van Kaapsch taaleigen. Changuion, A.N.E. 1848². *De Nederduitse taal in Zuid-Afrika hersteld, zijnde eene handleiding tot de kennis dier taal, naar de plaatselijke behoefte van het land gewijzigd*: i-xxvi. Rotterdam: J. van der Vliet.
- Chebanne, Andy, Mbulelo Jokweni, Makali Isabella Mokitimi and Sihawukele Ngubane.** 2003. Introduction. The Symposium and the Book. *Unifying Southern African Languages. Harmonization and Standardization*: 1-11. Cape Town: The Centre for Advanced Studies of African Society (CASAS).
- Circular on Danish Language Council = Bekendtgørelse om Dansk Sprognævns virksomhed og sammensætning. Bekendtgørelse nr. 707 af 4.9.1997. *Lovtidende, A.* 1997: 3745-3746. <<http://www.dsn.dk/>> [February 2006].
- Danish Culture Ministry.** 2003. *Sprog på spil. Et udkast til en dansk sprogpolitik*. Copenhagen: Kulturministeriet. <http://www.kum.dk/graphics/kum/downloads/Publikationer/Sprog_paa_spil.pdf> [February 2006].

- Den Danske Netordbog* = Bergenholtz, Henning and Vibeke Vrang with contributions by Lena Lund, Helle Grønberg, Maria Bruun Jensen, Signe Rixen Larsen, Rikke Refslund and Jette Pedersen. 2004. *Den Danske Netordbog*. Database and layout: Richard Almind. <<http://netordbog.asb.dk>>.
- Dictionary of South African English* = Silva, Penny (Mng. Ed.). 1996. *A Dictionary of South African English on Historical Principles*. Oxford: Oxford University Press.
- Duden-GWD¹* = *Duden. Das große Wörterbuch der deutschen Sprache in sechs Bänden*. 1976–1981. Edited and compiled by the Wissenschaftlicher Rat and the collaborators of the Dudenredaktion under the direction of Günther Drosdowski. Mannheim/Vienna/Zurich: Bibliographisches Institut.
- Duden-GWD²* = *Duden. Das große Wörterbuch der deutschen Sprache in acht Bänden*. 1993–1995. Second, completely newly edited and strongly enlarged edition. Edited and compiled by the Wissenschaftlicher Rat of the Dudenredaktion under the direction of Günther Drosdowski. Mannheim/Vienna/Zurich: Bibliographisches Institut.
- Duden-GWD³* = *Duden. Das große Wörterbuch der deutschen Sprache in zehn Bänden*. 1999. Third, completely newly edited and enlarged edition. Edited by the Wissenschaftlicher Rat of the Dudenredaktion. Mannheim/Leipzig/Vienna/Zurich: Dudenverlag.
- Duden-Universalwörterbuch* = *Duden. Deutsches Universalwörterbuch*. 2003. Edited by the Dudenredaktion. Fifth revised edition. Mannheim/Leipzig/Vienna/Zurich: Dudenverlag.
- Fettes, M.** 1997. Language Planning and Education. Wodak, Ruth and David Corson (Eds.). 1997. *Encyclopedia of Language and Education. Volume 1. Language Policy and Political Issues in Education*: 13-22. Dordrecht/Boston/London: Kluwer.
- Gallardo, A.** 1980. Dictionaries and the Standardization Process. Zgusta, L. (Ed.). 1980. *Theory and Method in Lexicography*: 59-69. Columbia: Hornbeam Press.
- Gouws, R.H.** 1989. *Leksikografie*. Pretoria/Cape Town: Academica.
- Gouws, R.H.** 1993. Normatiewe leiding in woerdeboeke: 'n nuwe benadering. *Lexikos* 3: 49-66.
- Gouws, R.H.** 1995. Dictionaries and the Dynamics of Language Change. Kachru, B. (Ed.). 1995. *Culture, Ideologies and the Dictionary*: 297-313. Tübingen: Max Niemeyer Verlag.
- Gouws, R.H.** 1996. A Sequence for Meeting Lexicographic Needs. *Lexicography as a Financial Asset*. 1996: 97-110.
- Gouws, R.H.** 2005. Lexicography in Africa. Brown, K. (Ed.). 2005. *Encyclopedia of Language and Linguistics*: 95-101. Second edition. Oxford: Elsevier.
- Gouws, R.H. and F.A. Ponielis.** 1992. The Development of Afrikaans and the Lexicographic Tradition. Zgusta, L. (Ed.). 1992. *History, Languages and Lexicographers*: 77-104. Tübingen: Max Niemeyer Verlag.
- Gouws, R.H. and D.J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.
- Gove, P.B.** 1961. *Webster's Third New International Dictionary of the English Language*. Springfield, Mass.: Merriam Webster.
- Groot Woordeboek/Major Dictionary* = Eksteen, L.C. (Ed.). 1997¹⁴. *Groot Woordeboek Afrikaans–Engels/Engels–Afrikaans / Major Dictionary Afrikaans–English/English–Afrikaans*. Cape Town: Pharos.
- Grundfos* = *Med andre ord ... Grundfos' sprogpolitik*. 2002. Bjerringbro: Grundfos.
- Hartmann, R.R.K. and G. James.** 1998. *Dictionary of Lexicography*. London/New York: Routledge.
- Haugen, E.** 1959. Planning for a Standard Language in Modern Norway. *Anthropological Linguistics* 1(3): 8-21.

- Hausmann, F.J.** 1977. *Einführung in die Benutzung der neufranzösischen Wörterbücher*. Tübingen: Max Niemeyer Verlag.
- International Encyclopedia of Lexicography* = 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.
- Kamwangamalu, N.M.** 2004. The Language Planning Situation in South Africa. Baldauf, Richard B. Jr and Robert B. Kaplan (Eds.). 2004. *Language Planning and Policy in Africa, Vol. 1. Botswana, Malawi, Mozambique and South Africa*: 197-281. Clevedon/Buffalo/Toronto: Multilingual Matters.
- Kidda Awak, M.** 1990. Historical Background, with Special Reference to Western Africa. Hartmann, R.R.K. (Ed.). 1990. *Lexicography in Africa*: 8-18. Exeter: Exeter University Press.
- Kotler, P. and K.L. Keller.** 2006. *Marketing Management*. Twelfth edition. Upper Saddle River, N.J.: Prentice Hall.
- Lakoff, R.** 1990. *Talking Power: The Politics of Language in our Lives*. New York: Basic Books.
- Landau, S.I.** 1981. *Dictionaries: The Art and Craft of Lexicography*. New York: Schribner Press.
- Langgård, P.** 1997. Ordbogersuaq, den store grønlandske ordbog, der ikke blev til noget — nogle etnopolitiske overvejelser. *LexicoNordica* 4: 83-100.
- Language Policy and Plan for South Africa = Final Draft: Language Policy and Plan for South Africa*. 2000. By the Advisory Panel on Language Policy. To the Minister of Arts, Culture, Science and Technology, 6 November 2000. <http://www.dac.gov.za/about_us/cd_nat_language/language_policy/Language%20Policy%20and%20Plan%20for%20South%20Africa.htm> [January 2006].
- Law on Language Council = Sprognævnsloven Lov nr. 320 af 14.5.1997 om Dansk Sprognævn. *Lovtidende, A*. 1997: 1497-1498. <<http://www.dsn.dk/>> [February 2006].
- Law on Orthography = Retskrivningsloven Lov nr. 332 af 14.5.1997 om Dansk Retskrivning. *Lovtidende, A*. 1997: 1521. <<http://www.dsn.dk/>> [February 2006].
- Lexicography as a Financial Asset = Lexicography as a Financial Asset in a Multilingual South Africa. Seminar held at the Bureau of the WAT, University of Stellenbosch, 12 April 1996*. 1996. Language Planning Report No. 5.3. October 1996. Pretoria: Department of Arts, Culture, Science and Technology.
- Lopes, A.J.** 2004. The Language Planning Situation in Mozambique. *Language Planning and Policy in Africa, Vol. 1. Botswana, Malawi, Mozambique and South Africa*. Baldauf, Richard B., Jr and Robert B. Kaplan (Eds.). 2004: 150-196. Clevedon/Buffalo/Toronto: Multilingual Matters.
- Malagasy–Aleman Dictionary* = Bergenholtz, H. in co-operation with S. Rajaonarivo, R. Ramasomanana, B. Radanielina as well as J. Richter-Johanningmeier, E. Olszowski, V. Zeiss in collaboration with H. Ranaivoson, N. Rasoarimanana, R. Ravololomboahangy and M. Razafiarivony. 1991. *Rakibolana Malagasy–Alema/Madagassisch–Deutsches Wörterbuch*. Antananarivo: Leximal/Moers: Aragon.
- Mansvelt, N.** 1884. *Proeve van een Kaapsch-Hollandsch idioticon met toelichtingen en opmerkingen betreffende land, volk en taal*. Cape Town: Van de Sandt de Villiers en Co.
- Mavoungou, P.A.** 2002. *Metalexical Criteria for the Compilation of a Trilingual Dictionary: Yilumbu–English–French*. Unpublished Doctoral Dissertation. Stellenbosch: University of Stellenbosch.
- Mini, B.M.** 1996. The Lexicographic Needs of Xhosa. *Lexicography as a Financial Asset*: 67-74.

- 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.
- Nordic Dictionary of Lexicography* = Bergenholtz, H., I. Cantell, R.V. Fjeld, D. Gundersen, J.H. Jónsson and B. Svensén. 1997. *NLO. Nordisk leksikografisk ordbog*. Oslo: Universitetsforlaget.
- Norwegian Language Council = *Rettskrivningsendringer fra 1. juli 2005*. 2005. Oslo. <<http://www.sprakrad.no/upload/9630/rettskriving2005.doc>> [February 2006].
- Nyangone Assam, B. and P.A. Mavoungou.** 2000. Lexicography in Gabon: A Survey. *Lexikos* 10: 252-274.
- Patriot Woordboek/Patriot Dictionary* = Anonymous (S.J. du Toit). 1902/1904. *Patriot Woordboek Afrikaans-Engels/Patriot Dictionary Cape Dutch-English*. Paarl: D.F. du Toit.
- Prinsloo, D.J.** 1996a. The Lexicographic Needs of Pedi. *Lexicography as a Financial Asset*: 37-45.
- Prinsloo, D.J.** 1996b. The Lexicographic Needs of Sotho. *Lexicography as a Financial Asset*: 47-52.
- Richards, J., J. Platt and H. Platt.** 1992. *Dictionary of Language Teaching and Applied Linguistics*. London: Longman.
- Spolsky, Bernard.** 2004. *Language Policy*. Cambridge: Cambridge University Press.
- Strauß, G., U. Haß and G. Harras.** 1989. *Brisante Wörter von Agitation bis Zeitgeist. Ein Lexikon zum öffentlichen Sprachgebrauch*. Berlin/New York: De Gruyter.
- Svensén, B.** 1993. *Practical Lexicography. Principles and Methods of Dictionary-Making*. Translated from the Swedish by John Sykes and Kerstin Schofield. Oxford/New York: Oxford University Press.
- Swedish Governmental Proposal = *Bästa språket — en samlad svensk språkpolitik*. 2005. Stockholm. <<http://www.regeringen.se/sb/d/5969/a/50740>> [February 2006].
- Tanzania = *National Information and Communication Infrastructure (NICI) Policies and Plans (e-strategies)*. 1997. <<http://www.uneca.org/aisi/nici/Tanzania/tanzania.htm>> [November 2005].
- Tauli, V.** 1974. The Theory of Language Planning. Fishman, Joshua A. (Ed.). 1974. *Advances in Language Planning*: 49-67. The Hague: Mouton.
- The Language Plan Task Group (LANGTAG).** 1996. *Toward a National Language Plan for South Africa: Final Report of the Language Plan Task Group*. Pretoria: DACST.
- Trask, R.L.** 1997. *A Student's Dictionary of Language and Linguistics*. London/New York/Sydney/Auckland: Arnold.
- Tweetalige Woordboek/Bilingual Dictionary* = Bosman, D.B. et al. (Eds.). 1984^s. *Tweetalige Woordboek/Bilingual Dictionary*. Cape Town: Tafelberg.
- Van Dale Groot Woordenboek Nederlands-Engels* = Martin, W. (Ed.). 1999. *Van Dale Groot Woordenboek Nederlands-Engels*. Utrecht: Van Dale Lexicografie.
- Van Dale Groot Woordenboek der Nederlandse Taal* = Geerts, G. and T. den Boon in co-operation with D. Geeraerts and E. Vos. Etymology by N. v.d. Sijs. 1999. *Van Dale Groot Woordenboek der Nederlandse Taal*. Thirteenth revised edition S-Z. Utrecht/Antwerp: Van Dale Lexicographie.
- Van der Merwe, H.J.J.M.** 1971. *Vroeë Afrikaanse Woordelyste*. Pretoria: J.L. van Schaik.
- Verklarende Afrikaanse Woordboek* = Labuschagne, F.J. and L.C. Eksteen. 1993^s. *Verklarende Afrikaanse Woordboek*. Eighth revised and augmented edition. Pretoria: Van Schaik.
- Verklarende Handwoordeboek van die Afrikaanse Taal* = Odendal, F.F. and R.H. Gouws. 2000⁴/2005⁵. *Verklarende Handwoordeboek van die Afrikaanse Taal*. Cape Town: Pearson Education.
- Wahrig = Wahrig, Gerhard. 2000. *Deutsches Wörterbuch*. Newly edited by Renate Wahrig-Burfeind. With a "Lexikon der deutschen Sprachlehre". Gütersloh/Munich: Bertelsmann.

- Webb, V.** 2002. *Language in South Africa: The Role of Language in National Transformation, Reconstruction and Development*. Amsterdam/Philadelphia: Benjamins.
- Wiegand, H.E.** 1998. *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie. 1. Teilband*. Berlin/New York: De Gruyter.
- Woordeboek van die Afrikaanse Taal*. Stellenbosch: Buro van die Woordeboek van die Afrikaanse Taal. <<http://www.yourdictionary.com>>.
- Zeitgeschichtliches Wörterbuch* = Stötzel, G. and T. Eitz in collaboration with A. Jährling-Marienfeld, L. Plate a.o. 2003. *Zeitgeschichtliches Wörterbuch der deutschen Gegenwartssprache*. Second enlarged and updated edition. Hildesheim/Zurich/New York: Georg Olms Verlag.
- Zgusta, L.** 1971. *Manual of Lexicography*. Prague: Academia / The Hague/Paris: Mouton.

'n Voorlopige model vir die sistematiese beskrywing van gebruikersvriendelikheid in woordeboeke*

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

Opsomming: Die gebruikersperspektief staan tans sentraal in die leksikografiese gesprek. Dit spruit uit die beginsel dat enige woordeboek in die eerste plek vir gebruik deur 'n bepaalde teikengebruikersgroep bedoel moet wees, en as sodanig moet die woordeboek op daardie gebruikersgroep afgestem wees — dit moet gebruikersvriendelik wees. In hierdie artikel word van die standpunt uitgegaan dat die gebruik van 'n woordeboek as 'n kommunikatiewe handeling begryp moet word, wat die geleentheid bied om insigte uit die kommunikasiewetenskap in die metaleksikografie te benut. Na aanleiding van die algemene kommunikasie-model vir interpersoonlike kommunikasie wat in die literatuur oor kommunikasiewetenskap voorkom, word 'n leksikografiese kommunikasie-model voorgestel. Uit hierdie model kan minstens elf leksikografiese parameters afgelei word aan die hand waarvan gebruikersvriendelikheid in woordeboeke sistematies beskryf kan word. Die struktuur wat hierdie parameters aan 'n gesprek oor gebruikersvriendelikheid in woordeboeke kan gee, bemiddel die formulering van die ware doel van 'n gebruikersvriendelike woordeboek.

Slutelwoorde: GEBRUIK, GEBRUIKER, GEBRUIKERSPERSPEKTIEF, GEBRUIKERSVERWYSINGSRAAMWERK, GEBRUIKERSVRIENDELIKHEID, KANAAL, KOMMUNIKASIE-MODEL, KONTEKS, LEKSIKOGRAAF, LEKSIKOGRAFIESE KOMMUNIKASIE-MODEL, MEDIUM, META-TEKS, STYLGIDS, TERUGVOER, WERKLIKE DOEL, WOORDEBOEK, WOORDEBOEKPLAN

Abstract: A Preliminary Model for the Systematic Description of User-friendliness in Dictionaries. The user perspective is a central notion in current lexicographical discourse. This stems from the principle that a dictionary should in the first place be

* Hierdie artikel is 'n verwerking van 'n gedeelte van 'n hoofstuk uit 'n D.Litt.-proefskrif, *'n Metaleksikografiese ondersoek na konteksleiding in Afrikaanse vertalende woordeboeke*, wat in April 2006 deur die Universiteit van Stellenbosch, Stellenbosch, Republiek van Suid-Afrika, aanvaar is. 'n Verkorte Engelse weergawe is as referaat gelewer by die Elfde Internasionale Konferensie van die African Association for Lexicography, georganiseer deur die Tshivenda Nasionale Leksikografie-eenheid, Universiteit van Venda vir Wetenskap en Tegnologie, Thohoyandou, Republiek van Suid-Afrika, 5–7 Julie 2006.

meant for use by a specific target user group, and as such the dictionary should be aimed at that target user group — it should be user friendly. In this article the point of departure is that the use of a dictionary should be regarded as a communicative act, which presents the opportunity to utilise insights gained in communication science in the metalexigraphy. From the general communication model for interpersonal communication, as found in literature on communication science, a lexicographical communication model is developed. From this model at least eleven lexicographical parameters can be derived according to which user-friendliness in dictionaries can be described in a systematic way. The structure that these parameters can offer in debating user-friendliness in dictionaries informs the formulation of a definition for the genuine purpose of a user-friendly dictionary.

Keywords: CHANNEL, COMMUNICATION MODEL, CONTEXT, DICTIONARY PLAN, DICTIONARY, FEEDBACK, GENUINE PURPOSE, LEXICOGRAPHER, LEXICOGRAPHICAL COMMUNICATION MODEL, MEDIUM, METATEXT, STYLE GUIDE, USE, USER, USER'S FRAME OF REFERENCE, USER'S PERSPECTIVE, USER-FRIENDLINESS

1. Inleiding

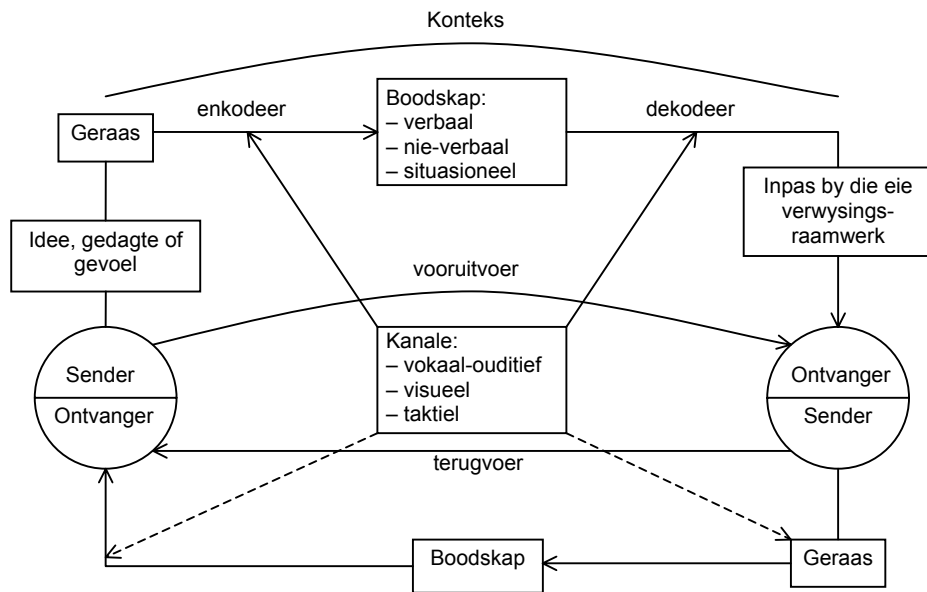
Die gebruikersperspektief staan tans sentraal in die leksikografiese gesprek. Dit spruit uit die beginsel dat enige woordeboek in die eerste plek vir gebruik deur 'n bepaalde teikengebruikersgroep bedoel is, en as sodanig moet die woordeboek op daardie gebruikersgroep afgestem wees — dit moet gebruikersvriendelik wees (vgl. Gouws en Prinsloo 2005: 39).

Volgens Kühn (1989: 112) moet die gebruik van 'n woordeboek as 'n kommunikatiewe handeling begryp word. Wiegand (1996: 135) betoog dat die woordeboek 'n draer van tekssoorte is, en dit veronderstel 'n kommunikasiesituasie, want tekste kan as kommunikasie beskou word (vgl. Van Cuilenburg et al. 1992: 21), of, meer presies, as media vir kommunikasie (vgl. Steinberg 1994: 15; en Smit 1996: 171). Hierdie siening bied die geleentheid om insigte uit die kommunikasiewetenskap in die metaleksikografie te benut om 'n raamwerk vir die sistematiese beskrywing van gebruikersvriendelikheid in woordeboeke te ontwerp. Dit is wat met hierdie artikel beoog word.

2. 'n Interpersoonlike kommunikasiemodel

Variante van die algemene kommunikasiemodel vir interpersoonlike kommunikasie word byna sonder uitsondering in die literatuur oor kommunikasiewetenskap voorgehou (vgl. o.m. Van Cuilenburg et al. 1992: 9-11; Frost et al. 1993: 5-49; Janssen 1996: 23-30; Steehouder et al. 1999: 22-25; en Guffey 2001: 4-5). Die vertaalde voorstelling van Van der Molen et al. (1995: 13-19) sal vir die doel van hierdie bespreking gebruik word. Hoewel die model op interpersoonlike kommunikasie gerig is, word die nut daarvan vir die metaleksikografie nie uitgesluit nie; die naslaanproses kan immers as 'n soort "gesprek" tussen leksi-

kograaf en woordeboekgebruiker beskou word, hoewel oënskynlik slegs een-richtingkommunikasie ter sprake is (vgl. Gouws 1989: 14). Vergelyk die skematiese voorstelling van die model in figuur 1.



Figuur 1: Die volledige kommunikasiemodel van Van der Molen et al. (1995)

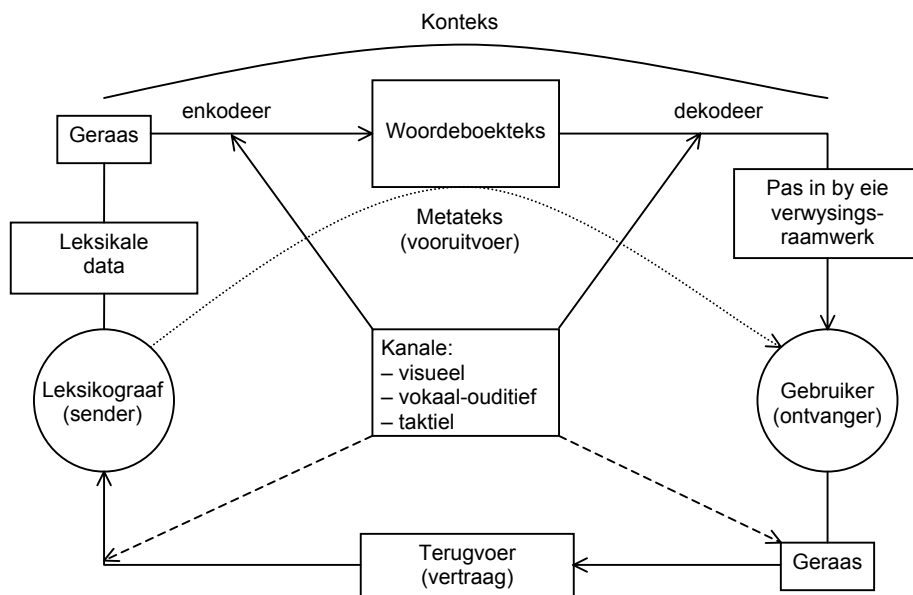
Die skema hierbo kan soos volg toegelig word:

Die sender het 'n bepaalde idee, gedagte of gevoel wat hy/sy aan die ontvanger wil oordra. Hierdie idee, gedagte of gevoel benodig 'n verpakking of voertuig, en meestal word van taal gebruik gemaak. Die sender skakel dus die idee, gedagte of gevoel (in taal) om — hy/sy enkodeer die inligting en die boodskap (in verbale, nie-verbale of situasionele vorm) kom tot stand. Ten einde die boodskap te verstaan, moet die ontvanger dit dekodeer. 'n Voorvereiste vir suksesvolle dekodering is dat beide die ontvanger en die sender die kode wat laasgenoemde gebruik, moet beheers. Die interpretasie van die gedekodeerde boodskap vind plaas binne die verwysingsraamwerk van die ontvanger — dit is die geheel van norme, waardes, opvattinge en vanselfsprekendhede op grond waarvan die ontvanger handel en sy/haar omgewing beoordeel. Die kommunikasieproses vind deur die vokaal-ouditiewe, die visuele en die taktiele kanale plaas. Geraas is alles wat die ontvangs en suksesvolle interpretasie van die boodskap versteur. Drie soorte geraas kan onderskei word: fisiese geraas (bv. as iemand met 'n lugdrukboor buite die oop venster werk wanneer twee persone in die kamer 'n gesprek probeer voer), psigologiese geraas (vooroordele en stereotiepe opvattinge wat die deurgang van die sein belemmer) en semantiese geraas (wanneer die twee partye betrokke by die kommunikasie nie

dieselfde kode beheers nie, maar ook byvoorbeeld die gebruik van jargon teenoor 'n leek). 'n Boodskap het 'n vooruitvoer-funksie indien dit iets aankondig oor die boodskap wat gaan volg, bv. "Ek hoop jy sal my nou nie kwalik neem nie, maar ..." Terugvoer behels die ontvanger se reaksie op die boodskap, en gee die sender gewoonlik die geleentheid om vas te stel of sy/haar boodskap korrek geïnterpreteer is. Dan keer die rolle om, en die oorspronklike ontvanger word nou die sender, terwyl die oorspronklike sender nou die ontvanger is. Die konteks verwys na die omgewing waarbinne die kommunikasie plaasvind, hetsy fisies, hetsy psigologies-sosiaal.

3. 'n Leksikografiese kommunikasiemodel

Die algemene model vir interpersoonlike kommunikasie kan aangepas word om die kommunikatiewe handeling van woordeboekgebruik sistematies voor te stel. Die produk van dié aanpassing is 'n *leksikografiese* kommunikasiemodel, wat in figuur 2 aangebied word:



Figuur 2: 'n Leksikografiese kommunikasiemodel

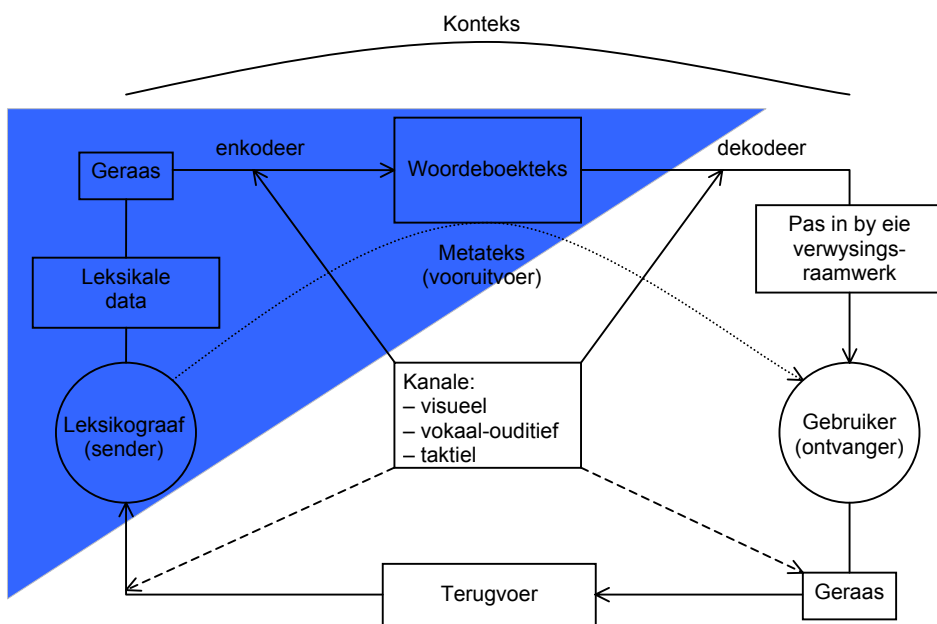
Die skema in figuur 2 kan soos volg toegelig word:

Die leksikograaf het die taak om leksikale data tot 'n woordeboekteks te verwerk wat vir die teikengebruiker nuttig sal wees. Geraas kan die leksikograaf verhinder om 'n optimaal benutbare woordeboekteks saam te stel: gewoon onkundigheid; gebrek aan kennis oor die behoeftes van die woordeboek-

gebruikers, hulle naslaanvaardighede en die doel waarvoor die woordeboek gebruik word; ens. Die leksikograaf encodeer die leksikale data volgens die stylgids tot 'n woordeboekteks met 'n bepaalde struktuur. Die woordeboekteks word deur 'n bepaalde medium aangebied, waarvan die gebruiklikste die gedrukte en elektroniese media is. Die woordeboekgebruiker lees die woordeboekteks en dekodeer dit deur die teksstruktuur te agterhaal en die relevante datakategorieë in hulle soeksones te identifiseer en die gesoekte inligting te ekstraheer aan die hand van leiding wat in die metateks aangebied word. Die metateks (gebruikersgids) geld as vooruitvoer: Die gebruiker wat die metateks raadpleeg, weet wat die aard van die woordeboek is en watter datakategorieë hy/sy in woordeboekartikels kan vind, ook hoe dit aangebied word. Hierdie vooruitvoer geld egter in weinig gevalle, aangesien min gebruikers die metateks van 'n woordeboek raadpleeg. In te veel gevalle bestaan 'n bruikbare metateks ook glad nie; derhalwe word die verband in figuur 2 deur 'n stippellyn aangedui. Funktionele vooruitvoer kan geraas minimaliseer. Geraas kan die woordeboekgebruiker verhinder om die woordeboekteks optimaal te dekodeer as gevolg van gebrekkige kennis van woordeboektipies wat aanleiding gee tot die gebruik van die verkeerde woordeboeke, gebrekkige kennis van die teksstruktuur, die nie-raadpleging van die metateks, ens. Die woordeboekgebruiker pas die interpretasie van die data in die woordeboekteks by sy/haar verwysingsraamwerk in deur byvoorbeeld te glo dat indien 'n leksikale item nie in 'n woordeboek gelemmatiseer is nie (d.w.s. indien die betrokke woordeboekartikel afwesig is), daardie leksikale item nie bestaan nie (verwysingsraamwerk: 'n woordeboek bevat al die leksikale items van die behandelde taal se leksikon). Wanneer woordeboekgebruikers terugvoer gee, is hulle dalk nie in staat om dit optimaal te doen nie, aangesien hulle nie met die terminologie en werkswyses van die leksikografie vertrou is nie. Terugvoer word gewoonlik in die vorm van resensies in die pers en korrespondensie tussen gebruikers en die uitgewer en/of die leksikograaf gegee. Sodanige terugvoer is gewoonlik beperk en vertraag (anders as in die geval van die interpersoonlike gespreksituasie). Leksikograwe kan ook persoonlike telefoonoproepe en besoeke ontvang, in welke geval die vokaal-ouditiewe kanaal gebruik word. Geraas kan veroorsaak dat leksikograwe kritiek kan waninterpreteer of as onbelangrik kan beskou, of dit gewoon nie verstaan nie. Die konteks waarbinne die kommunikasie tussen leksikograaf en woordeboekgebruiker afspeel, is die omgewing waarbinne sowel die saamstel as die gebruik van die woordeboekteks plaasvind, bv. in 'n multitalige, demokratiese Suid-Afrika. Dit kan ook verwys na die onmiddellike gebruiksituasie waarin die gebruiker hom/haar bevind, bv. in 'n onderrigsituasie in 'n klaskamer. Woordeboektekste word by uitstek deur die visuele kanaal aangebied, aangesien gebruikers die artikels visueel waarneem wanneer hulle dit lees. 'n Uitsondering is braille-woordeboeke, waar die taktiele kanaal gebruik word. Die vokaal-ouditiewe kanaal kan ook by die gebruik van elektroniese woordeboeke benut word.

Die leksikografiese kommunikasie-model akkommodeer die woordeboekplan soos deur Prinsloo en Gouws (2005: 13-19) beskryf. Binne die model

enkapsuleer die woordeboekplan die leksikograaf as sender, die leksikale data wat verwerk moet word, die enkoderingsproses en die ontwerp en saamstel van die metateks te midde van geraas, om 'n woordeboekteks tot stand te bring. Sowel die organisasieplan as die woordeboekkonseptualiseringsplan as komponente van die woordeboekplan kom binne die model tot hul reg. In figuur 3 hieronder word die terrein van die woordeboekplan in die leksikografiese kommunikasiemodel deur die donker geskakeerde area aangedui:



Figuur 3: Die terrein van die woordeboekplan in die leksikografiese kommunikasiemodel

Hoewel kommunikasiemodelkomponente soos die gebruiker nie direk binne die sfeer van die woordeboekplan as deel van die leksikografiese proses val nie, beteken dit nie dat die waardes van hierdie komponente in die uitvoering van die woordeboekplan genegeer word nie; intendeel, die leksikografiese kommunikasiemodel moet juis ten doel hê om die leksikografiese beplanner in sy/haar taak te rig deurdat spesifieke waardes aan komponente soos "dekodeer", "gebruiker", "geraas" en "terugvoer" in die beplanningsproses toegeken kan word. Die waardes wat so aan hierdie komponente toegeken word, bepaal juis die waardes wat tydens die beplanningsproses toegeken sal word aan die komponente wat wel deel uitmaak van die woordeboekplan. Hierdie tipe holistiese leksikografies-kommunikatiewe benadering lei daartoe dat spesifieke beïndrukkende leksikografiese parameters geïdentifiseer kan word wat op hulle

beurt die komponente van 'n leksikografiese model vir gebruikersvriendelike woordeboeke kan vorm.

4. 'n Leksikografiese model vir gebruikersvriendelike woordeboeke

Van der Molen et al. (1995: 19) erken dat hulle algemene kommunikasiemodel "een vereenvoudigde weergawe van die werklikheid en bovendien een erg algemene en statiese weergawe" is, en dieselfde geld waarskynlik in 'n mindere of meerdere mate die leksikografiese kommunikasiemodel wat daarvan afgelei is, maar dié model skep 'n funksionele raamwerk waarbinne die leksikografiese gebruikersperspektief op 'n wetenskaplike basis gedefinieer en aangepak kan word. Na aanleiding daarvan kan die volgende parameters geïdentifiseer word wat bepalend is in 'n gebruikersgedrewe leksikografie:

1. Die woordeboekparameter
2. Die leksikograafparameter
3. Die stylgidparameter
4. Die gebruikersparameter
5. Die gebruikparameter
6. Die gebruikersverwysingsraamwerkparameter
7. Die mediumparameter
8. Die kanaalparameter
9. Die metateksparameter
10. Die kontekstparameter
11. Die terugvoerparameter

Elkeen van hierdie parameters word vervolgens kortliks beskryf.

4.1 Die woordeboekparameter

Die woordeboekparameter kan as die sentrale parameter beskou word. Al die ander parameters bepaal direk en in kombinasie met mekaar die waardes van die veranderlikes binne hierdie parameter. Die verhouding is egter nie eensydig nie: Die woordeboekparameter kan ook die waardes van die veranderlikes binne elk van die ander parameters bepaal, soos wat uit die volgende bespreking sal blyk.

Die woordeboekparameter sluit daardie veranderlikes in wat te doen het met die tipologiese plasing, werklike doel en funksies van die woordeboek, die dataverspreidingsprogram en die mikrostrukturele program (vgl. Gouws 2001:

68). Die invloed van die ander parameters bepaal hoe effektief, hoe *gebruikersvriendelik* die waardes van hierdie veranderlikes in die woordeboek eindprodukt neerslag sal vind, terwyl die waardes van die veranderlikes binne hierdie parameter die raamwerk skep wat bepaal watter veranderlikes binne die ander parameters geld.

4.2 Die leksikograafparameter

Die leksikograafparameter is waarskynlik dié belangrikste parameter; dit blyk uit die feit dat die metaleksikografiese literatuur oor die gebruikersaspek oorwegend op hierdie parameter fokus, en dat die leksikograaf die rolspeler is wat uiteindelik vir die saamstel van die woordeboekteks verantwoordelik is. Kromann et al. (1991: 2713) som hierdie parameter goed op deur te sê dat dit gaan oor "*consideration on the part of the lexicographer for the dictionary's target group, its needs and the types of user situations*" [my kursivering — HLB].

Die leksikograafparameter kan dus beskryf word in terme van alles wat die leksikograaf kan en moet doen om die teikengebruiker ter wille te wees in die saamstel van die woordeboekteks, maar ook alles wat die leksikograaf nie (kan) doen nie as gevolg van 'n gebrek aan kennis oor die teikengebruiker. Hierby moet gevoeg word alles wat die metaleksikograaf van die leksikograaf weet, en behoort te weet, bv. hoe hy/sy te werk gaan in die ontwerp van 'n woordeboek(teks), ten einde hom/haar van hulp te kan wees (vgl. Swanepoel 2001: 170). Die mees fundamentele vereiste hier is dat die leksikograaf opgelei moet wees (vgl. in hierdie verband o.a. Martin 1994; Botha en Botha 1998; en Gouws 2001: 60-65). Trouens, Bergenholtz en Gouws (2006: 31) is van mening dat die leksikografiese proses by die opleiding van leksikograwe kan begin, selfs voordat te veel aandag aan die finansiële implikasies van die beplande woordeboekprojek bestee word. Verder maak die leksikograaf staat op die resultate van navorsing oor die gebruiker, sy/haar behoeftes, naslaanvaardighede, taalvaardighede in die behandelde taal of tale, ens. 'n Gebrek aan kennis oor en insig in hierdie faktore veroorsaak dat die leksikograaf woordeboeke en woordeboektekste ontwerp wat nie aan die behoeftes van die teikengebruiker voldoen nie. Ongelukkig bestaan daar juis weinig studies oor die woordeboekgebruiker en sy/haar behoeftes en vaardighede; derhalwe behoort die aanpak van sulke studies as prioriteit beskou te word indien die doel is om die leksikograafparameter 'n positiewe invloed op gebruikersgedrewe leksikografie te laat uitoefen (vgl. 4.4). Navorsing van dié aard behoort ook op wetenskaplik-empiriese basis gedoen te word, wat eksperimentele studies insluit; indien dit nie die geval is nie, "much of these research results can only be used as basis for preliminary hypotheses about the design of a functional dictionary of a specific kind" (Swanepoel 2001: 173).

Na aanleiding van die voorafgaande bespreking sou die belangrikste veranderlikes wat binne die leksikograafparameter geld, kwessies insluit soos die leksikograaf se opleiding, sy/haar konsep van die teikengebruiker, sy/haar

kennis van die leksikografieteorie, kennis van die behandelde taal of tale (en die relevante leksikale domein in die geval van 'n vakwoordeboek) en kennis van die taalwetenskap.

4.3 Die stylgidsparameter

Die stylgidsparameter kan van die leksikograafparameter onderskei word, aangesien dit nie altyd die leksikograaf is wat die stylgids opstel nie. Veral in die geval van omvattende woordeboeke, waarvan die produksie oor dekades strek, word nuwe leksikograwe wat by sodanige projekte aansluit, bekend gemaak met die toepassing van die stylgids — hulle was self nie by die opstel daarvan betrokke nie. Hierdie stand van sake stel minstens twee vereistes aan (die opsteller van) die stylgids: (a) die stylgids moet die dataverspreidingsprogram van die woordeboek duidelik uiteensit, en (b) dit moet vir die leksikograaf verstaanbaar en toepasbaar (d.w.s. gebruikersvriendelik) wees. Hierby kan die kritiese vereiste gevoeg word dat die stylgids in die metateks verantwoord moet kan word. Die stylgidsparameter is by uitstek dié parameter wat die konsekwente aanbod van leksikografiese data in die woordeboek moet reguleer en sodoende die woordeboek gebruikersvriendelik maak (vgl. Gouws 2001: 67).

4.4 Die gebruikersparameter

Teenoor die leksikograaf as sender staan die gebruiker as ontvanger in die woordeboekkommunikasiesituasie.

Literatuur wat handel oor die behoefte aan en vereiste van navorsing oor die teikengebruiker en die vereiste dat 'n woordeboek op die teikengebruiker afgestem moet wees, sowel as die feit dat onvoldoende navorsing tot op hede hieroor gedoen is, sluit in Wiegand (o.a. 1984; 1985), Gouws (1989: 49; 1990: 55; 1996; 1996a; 1999; 2001: 66, 72), Hartmann (1989), Kromann et al. (1991), Alberts (1992), Svensén (1993: 4, 16, 17), Lombard (1994), Louw (1997: 2 e.v.), Fouché-Van der Merwe (1999), Prinsloo en Gouws (2000), en Swanepoel (2001). Uit hierdie literatuur kan afgelei word dat die leksikograaf inligting oor minstens die volgende aspekte van die teikengebruiker moet inwin en integreer in die beplanning van sy/haar woordeboek:

- (a) die teikengebruiker se taalkennis, beide van die bron- en die doeltaal van die woordeboek,
- (b) die gebruiker se naslaanvaardighede, en
- (c) die taal- en inligtingsbehoefte van die gebruiker.

Dat dié inligting die leksikograaf tans nog ontbreek, word duidelik uit waarnemings soos die volgende:

- (a) Gebruikers beskik nie oor genoegsame linguistiese en pragmatiese insig om die inligting in die woordeboek suksesvol te ontsluit nie (vgl. Gouws 1996: 21).
- (b) Gebruikers toon 'n weersin in woordeboekgebruik (vgl. Carstens 1995: 107).
- (c) Gebruikers ken nie die woordeboektipies wat bestaan nie (vgl. Carstens 1995: 108).
- (d) Gebruikers beskik oor onvoldoende naslaanvaardighede (vgl. Carstens 1995: 113).
- (e) Gebruikers beskik nie oor voldoende ondervinding in die gebruik van woordeboeke nie.

Die onderskeid tussen die leksikograafparameter en gebruikersparameter ondersteun die insig dat 'n gebruikersvriendelike woordeboek die produk van 'n die leksikograaf én die kundige teikengebruiker is. Die vestiging van 'n woordeboekcultuur sal van die teikengebruiker 'n kundige teikengebruiker maak (vgl. Gouws en Prinsloo 2005: 42).

4.5 Die gebruiksparemeter

Die gebruiksparemeter verwys na die onmiddellike situasie wat woordeboekgebruik stimuleer, bv. in die vreemdetaalklaslokaal of in 'n vertaalsituasie om 'n gepaste moedertaalekwivalent vir 'n vreemdetaalitem te vind ten einde 'n vreemdetaaltekst te verstaan (en te vertaal). Tydens die beplanning van 'n woordeboek moet die tipiese gebruikssituasies waarin die beplande woordeboek gebruik sal word, so ver moontlik voorsien (en tydens die bemerking van die woordeboek aangedui) word, sodat die strukture in die woordeboek op daardie situasies afgestem kan word.

Uit die literatuur wil dit soms lyk asof die gebruikssituasie as deel van die gebruikerskenmerke beskou word. Hier is dit egter belangrik om dié twee parameters van mekaar te onderskei, aangesien die veranderlikes wat die gebruiksparemeter bepaal, nie noodwendig 'n invloed op die veranderlikes wat die gebruikersparemeter bepaal, uitoefen nie. 'n Geografiewoordeboek kan vir die spesifieke gebruikssituasie van die geografieklas bedoel wees, hoewel die gebruikersparemeter se veranderlikes grootliks identies kan wees aan dié wat vir die standaard verklarende woordeboek vir spesifieke gebruik in die taalklas geïdentifiseer is. Hoewel die gebruikersparameters vir die twee woordeboeke grootliks identies is, verskil die gebruiksparemeteers beduidend. Die feit dat die twee parameters van mekaar onderskei word, beteken egter nie dat hulle sonder mekaar kan funksioneer nie; soos die ander parameters, figureer dié twee parameters as onmisbare elemente in die bepaling van die werklike doel van 'n woordeboek.

Vanuit 'n kognitief-funksionele benadering identifiseer Swanepoel (2001: 176) agt vrae in terme waarvan woordeboekgebruiksituasies beskryf kan word:

- (a) Vir watter presiese take (lees, skryf, vertaling, woordeskatverwerwing, ens.) word verskillende soorte woordeboeke gebruik?
- (b) Watter kognitiewe prosesse, vaardighede, strategieë en mentale voorstelling is betrokke by die suksesvolle uitvoering van so 'n taak?
- (c) Watter rol speel leksikale kennis in die uitvoering van die taak?
- (d) Wat is leksikale kennis?
- (e) Hoe word leksikale kennis in die mentale leksikon geberg en hoe word daartoe kennis verkry?
- (f) Watter soorte gapings kan in die gebruiker se leksikale kennis ontstaan en watter implikasies het dit vir die suksesvolle uitvoering van die taak?
- (g) Watter strategieë, behalwe om 'n woordeboek te raadpleeg, kan die gebruiker inspan om vir hierdie leksikale leemte te kompenseer?
- (h) Onder watter omstandighede sal gebruikers hulle tot 'n woordeboek wend om 'n gaping in hulle leksikale kennis te hanteer?

Navorsing oor hierdie belangrike aspek ontbreek in die metaleksikografie (vgl. Swanepoel 2001: 186).

4.6 Die gebruikersverwysingsraamwerkparameter

Hierdie parameter onderskei hom van die gebruikersparameter daarin dat die gebruikersparameter hoofsaaklik verwys na vaardighede en kennis wat redelikerwys van die woordeboekgebruiker verwag kan word om voordeel uit 'n gebruikersvriendelike woordeboek te trek, d.w.s. om 'n kundige gebruiker te wees. Die gebruikersverwysingsraamwerkparameter het 'n breër bestek. Dit gaan hier naamlik oor veronderstellings, gewoontes, persepsies en houdings. Bepaalde veronderstellings is byvoorbeeld agterhaalbaar uit die volgende aanhalings van taalgebruikers:

"As daardie woord nie in die woordeboek staan nie, bestaan dit nie."

"As daardie woord in die woordeboek staan, kan ek dit gebruik waar en wanneer ek wil."

Gouws (1989: 13) wys op die algemene persepsie dat 'n woordeboek dié bron van gesag oor taalsake is:

Vir die gewone taalgebruiker is 'n woordeboek die belangrikste boek wat oor taal geskryf is. Hierdie onbevangte gebruiker dink aan sy taalkundige hulpbron

as **die** woordeboek en nie as 'n woordeboek nie. 'n Publikasie met **woordeboek** in sy titel verkondig die absolute waarheid.

Deur verantwoordelike en gebruikersvriendelike leksikografiebeoefening kan die leksikograaf ook sekere wanpersepsies teenwerk, soos dié wat deur die aanhalings hierbo verteenwoordig word.

Die gebruikersverwysingsraamwerkparameter word egter ook deur ander veranderlikes bepaal. Woordeboekgebruikers het naamlik bepaalde gewoontes en verwagtinge van woordeboeke, bv. dat die lemmata daarin alfabeties gerangskik is. Read (1973: 178) maak die invloed van hierdie veranderlike duidelik:

There is a constant pulling and hauling between what a lexicographer would like to do on scientific grounds, and what he is compelled to do by the habitual demands of the public.

Vir 'n woordeboek om optimaal gebruikersvriendelik te wees, moet dit binne die (woordeboek)verwysingsraamwerk van die teikengebruiker figureer. Gouws (1999: 7) verwys hierna as die aanpassing van die leksikografie by die maatskappy. Indien die leksikograaf daarin slaag om aspekte van die verwysingsraamwerk van die gebruikers ten gunste van gebruikersvriendelike woordeboeke aan te pas, sal daar sprake wees van die aanpassing van die maatskappy by die leksikografie (vgl. nogmaals Gouws 1999: 7). Dit sal byvoorbeeld gebeur indien onderwysers en onderwyskundiges oortuig kan word van die waarde wat 'n komponent oor woordeboekgebruik in enige taalleerplan het (vgl. Gouws 1996: 182). Die gebruikersverwysingsraamwerkparameter kan dus inspeel op die gebruikersparameter. Die omgekeerde is ook waar: Indien woordeboekgebruikers byvoorbeeld konstant in hulle naslaanpogings deur gebruikersonvriendelike woordeboeke gefrustreer word, kan hulle 'n weersin in woordeboeke en woordeboekgebruik ontwikkel — 'n verskynsel wat volgens Carstens (1995: 107) ongelukkig reeds voorkom.

Die mate van aanpassing van die leksikografie by die maatskappy word deur die leksikograafparameter bepaal, bv. die saamstel van 'n nuwe tipe woordeboek nadat die behoefte daartoe deur empiriese navorsing vasgestel en onder die leksikograaf se aandag gebring is. Die mate van aanpassing van die maatskappy by die leksikografie word deur die gebruikersverwysingsraamwerkparameter bepaal.

4.7 Die mediumparameter

Woordeboeke word vandag hoofsaaklik deur twee media versprei, naamlik die gedrukte medium en die elektroniese medium. 'n Gebruikersvriendelike woordeboek sal in die gepaste medium beskikbaar wees: Die reisende joernalis kan 'n elektroniese woordeboek wat op sy/haar skootrekenaar geïnstalleer is meer gebruikersvriendelik vind as 'n papierwoordeboek; daarenteen kan skole woor-

deboeke in die gedrukte medium verkies, aangesien rekenaarfasiliteite relatief beperk is en dus toegang tot 'n elektroniese woordeboek moeilik (gebruikersonvriendelik) maak.

Veranderlikes wat die waarde van die mediumparameter bepaal, is byvoorbeeld papierkwaliteit en bindkwaliteit in die geval van gedrukte woordeboeke, en die programmatuur wat gebruik is in die saamstel van die woordeboek en waarmee die woordeboek versoenbaar is in die geval van elektroniese woordeboeke. In die geval van internetwoordeboeke is die beskikbaarheid van internettoegang vir die teikengebruiker 'n kernveranderlike.

Die gekose medium stel egter ook vereistes aan die leksikograaf. Indien die gedrukte medium gekies word, word die leksikograaf 'n beperking op ruimte opgelê, aangesien 'n gedrukte woordeboek wat gebruikersvriendelik wil wees, ook fisies hanteerbaar en bekostigbaar moet wees. Die beperking op ruimte lei die leksikograaf om bepaalde strategieë aan te wend om ruimte te bespaar, waarvan die algemeenste prosedure dié van teksverdigtiging is. Teksverdigtiging word nie primêr deur die strewe na gebruikersvriendelikheid gemotiveer nie, maar deur die beperkte ruimte wat deur die gekose medium opgelê word. Teksverdigtiging as sodanig werk teoreties juis gebruikersvriendelikheid teen. Die ruimtebesparende strategie van teksverdigtiging het 'n diep gewortelde leksikografiese tradisie geword en kom selfs voor in media wat nie die mate van ruimtebeperking op die leksikograaf plaas wat deur die gedrukte medium op hom/haar geplaas word nie. Dit word duidelik as elektroniese woordeboeke beskou word: In die meeste gevalle word tradisionele teksverdigtigingstrategieë toegepas, hoewel die medium dit nie dikteer nie. In sommige gevalle is elektroniese woordeboeke (soos die *Elektroniese Verklarende Handwoordeboek van die Afrikaanse Taal 2.00* (2000) en *Pharos Woordeboeke/Dictionaries: 5 in 1* (2000)) bloot rekenaarmatige aanpassings van reeds bestaande gedrukte woordeboeke (Louw 2001: 324).

Hoewel 'n sekere mate van teksverdigtiging sal voorkom selfs waar die medium nie ernstige beperkinge op ruimte plaas nie, juis omdat die totale afwesigheid van teksverdigtiging ook gebruikersonvriendelik sal wees omdat dit die naslaanproses sal kompliseer, behoort die samestellers van elektroniese woordeboeke nie teksverdigtigingstrategieë wat in die gedrukte medium voorkom, slaafs te dupliseer nie, maar behoort hulle die virtueel onbeperkte ruimte tot hulle beskikking in dié verband innoverend te benut. Teksverdigtigingstrategieë in gedrukte woordeboeke wat ten doel het om gebruikersvriendelikheid in die hand werk, sal nie noodwendig dieselfde effek in elektroniese woordeboeke hê nie. Dit blyk dus dat navorsing oor elektroniese woordeboekgebruik noodsaaklik is. Louw (2000: 292) wys ook op die dringende behoefte aan 'n metaleksikografiese gesprek oor elektroniese woordeboeke in die algemeen. In hierdie verband is dit verblydend dat daar tans by SIGNA, dit is die Stellenbosse Instituut vir Gevorderde Navorsing, deur Suid-Afrikaanse en buitelandse navorsers gewerk word aan 'n uitvoerige navorsingsprojek om 'n teoretiese model vir elektroniese woordeboeke te ontwikkel.

4.8 Die kanaalparameter

Tradisioneel is die waarde van die kanaalparameter die visuele kanaal (woordeboekgebruikers neem die woordeboekartikels visueel waar wanneer hulle dit lees). Die uitsondering op hierdie reël is braille-woordeboeke, waar die waarde van die kanaalparameter die taktiele kanaal is. Uiteraard sal 'n woordeboek wat blinde gebruikers teiken, hoogs gebruikersonvriendelik wees indien dit deur die visuele kanaal aangebied word. Met die ontwikkeling van spesifiek die rekenaartegnologie is dit ook moontlik om die vokaal-ouditiewe kanaal naas die visuele kanaal in 'n elektroniese woordeboek te gebruik: Deur byvoorbeeld op die fonetiese transkripsie van 'n lemma of op 'n spesiale ikoon op die rekenaarskerm te klik, kan 'n klanklêer geaktiveer word en die gebruiker kan die uitspraak van die lemma of 'n ander mikrostrukturele inskrywing ouditief waarneem. Vergelyk in hierdie verband byvoorbeeld die *Oxford Talking Dictionary* (1998).

4.9 Die metateksparemeter

Die metateks (gebruikersgids) verduidelik die werkswyse van die woordeboek aan die gebruiker. Indien die waarde van die metateksparemeter "afwesig" is, word die gebruikersvriendelikheid van die woordeboek sterk negatief beïnvloed. Die omgekeerde is nie noodwendig waar indien die metateksparemeter se waarde "aanwesig" is nie, aangesien die kwaliteit van die metateks so minderwaardig kan wees dat dit vir die woordeboekgebruiker geen waarde inhoud nie. Die metateksparemeter moet dus as 'n skalêre veranderlike geld. Vergelyk Kirkpatrick (1989) vir 'n bespreking van metatekste in woordeboeke.

Die metateks is in wese 'n instruktiewe teks, aangesien die teikengebruiker daarin instruksies vind oor hoe om die woordeboek effektief te gebruik. Insigte uit die tekswetenskap kan benut word om gebruikersvriendelike metatekste te skep, wat op hulle beurt die woordeboeke wat hulle toelig gebruikersvriendeliker sal maak. Volgens Steehouder et al. (1999: 37) is die doel van instruktiewe kommunikasie om die ontvanger in staat te stel om bepaalde take uit te voer — in die geval van 'n woordeboek: om die aangebode inligting te ontsluit. Verder word die volgende eise aan instruktiewe tekste gestel (Steehouder et al. 1999: 37-38):

- (a) Die teks moet help om die taak *effektief* uit te voer, d.w.s. met so min as moontlik moeite.
- (b) Die teks moet insig gee in die taak. Die leser moet nie slegs in staat wees om die taak uit te voer nie, maar ook om te verstaan waarmee hy/sy besig is.
- (c) Soms is dit belangrik dat die taak aangeleer word. Dan moet die inligting *leerbaar* wees. Die woordeboekgebruiker wil byvoorbeeld nie elke

keer wanneer hy/sy 'n bepaalde woordeboek raadpleeg, eers weer in die metateks nalees hoe om die woordeboek te gebruik nie.

- (d) Die leser van 'n instruktiewe teks moet ook *gemotiveer* word om die taak uit te voer.

Om aan hierdie eise te voldoen, behoort 'n instruktiewe teks (soos die metateks in 'n woordeboek) *volledig* te wees: dit moet die gepaste hoeveelheid *detail* bevat, die stappe moet in die korrekte *volgorde* aangebied word en die taalgebruik moet so duidelik wees dat die leser verstaan wat hom/haar te doen staan (Steehouder et al. 1999: 38).

Dit is duidelik dat bogenoemde vereistes ook geld vir die stylgids wat as instruktiewe teks aan die leksikograaf gerig is.

4.10 Die konteksparameter

Die konteksparameter omvat daardie versameling veranderlikes wat normaalweg buite die beheer van die leksikograaf en die gebruiker in hulle onderskeie kapasiteite as sodanig val, maar waarbinne én die saamstel én die gebruik van die woordeboek plaasvind. Hoewel die leksikograaf en die gebruiker gewoonlik min aan die waardes van hierdie veranderlikes kan doen, moet dié veranderlikes so ver moontlik in die beplanning van 'n gebruikersvriendelike woordeboek verreken word. Hieronder volg enkele veranderlikes se waardes soos dit in die literatuur voorkom:

- (a) Die vertalende woordeboek is in baie huise in 'n multitalige samelewing (soos in Suid-Afrika) die enigste bron van semantiese inligting in beide die moedertaal en die vreemde taal (Gouws 1996: 16).
- (b) Daar is 'n gebrek aan woordeboekonderrig in skole in Suid-Afrika (Fouché-Van der Merwe 1999: 144).
- (c) Die ouderdomme waarop kinders skool toe gaan, sowel as die standaard van onderrig, wissel van instelling tot instelling (Fouché-Van der Merwe 1999: 145).
- (d) Suid-Afrika is 'n multitalige samelewing (Gouws 1996: 18).
- (e) Woordeboeke is nie gereedlik op skool beskikbaar nie (Carstens 1995: 107).
- (f) Onderwysers se benadering tot, kennis van en houding jeens woordeboeke is nie altyd na wense nie (Carstens 1995: 109).
- (g) Woordeboeke word binne die konteks van 'n (nasionale) taalbeleid saamgestel (Bergenholtz en Gouws 2006: 14).

Die beskrywing van die konteksparameter suggereer dat die verhouding tussen dié parameter en veral die woordeboekparameter 'n eensydige een is, m.a.w. die waardes van die veranderlikes binne die konteksparameter bepaal die waardes van die veranderlikes binne die woordeboekparameter, en nie andersom nie. Hoewel dit in baie gevalle inderdaad die geldende verhouding is, bestaan daar wel instansies waar die veranderlikes binne die woordeboekparameter 'n beduidende invloed op die waardes van die veranderlikes binne die konteksparameter uitgeoefen het. 'n Sprekende voorbeeld van so 'n instansie is die verskyning van *An American Dictionary of the English Language* (Webster 1828), wat 'n beduidende en blywende invloed op die spelling van Amerikaanse Engels gehad het (vgl. Gouws 1989: 19; Landau 2001: 69-74). So ver as wat woordeboeke as normeers van taal beskou en gebruik word, kan die waardes van die veranderlikes binne die konteksparameters wel deur die leksikograaf en woordeboekgebruiker in hulle onderskeie kapasiteite as sodanig beïnvloed word. Vergelyk in hierdie verband byvoorbeeld Bergenholtz en Gouws (2006) vir 'n bespreking oor die invloed van taalbeleid op die leksikografie. 'n Verdere voorbeeld wat die invloed van die konteksparameter op die gebruikersvriendelikheid van 'n bestaande woordeboek het, kan in terme van die ortografie van Engels in Namibië voorgehou word: In Namibië word die Britse ortografie van Engels ingevolge 'n kabinetsbesluit toegepas. Daar is egter ook Engelse woordeboeke wat deur Amerikaanse uitgewers versprei word, in die Namibiese boekhandel beskikbaar. Hierdie woordeboeke gee uiteraard leiding oor die Amerikaanse ortografie van Engels, wat in bepaalde opsigte beduidend van die Britse ortografie afwyk. Binne die Namibiese konteks kan 'n woordeboek wat Amerikaanse Engels behandel, derhalwe nie in terme van hierdie parameter as gebruikersvriendelik beskou word nie.

4.11 Die terugvoerparameter

Tradisioneel figureer terugvoer nie in besprekings oor woordeboeke nie, hoofsaaklik omdat terugvoer deur die gebruiker nie onmiddellik kan plaasvind nie en omdat geen terugvoer enige onmiddellike effek op die woordeboek sal uitoefen nie, aangesien kritiek eers by 'n volgende uitgawe verreken kan word. Terugvoer is wel waardevol en kan bydra tot die verhoging in gebruikersvriendelikheid van 'n woordeboek. Terugvoer geskied gewoonlik in die vorm van resensies wat in die pers en in vaktydskrifte verskyn, en in die vorm van studies, referate en wetenskaplike artikels wat tydens leksikografiekongresse en in vaktydskrifte die resultate van woordeboeknavorsing en -kritiek weergee.

Die terugvoerparameter het nie slegs 'n invloed op die leksikograaf(parameter) nie, maar kan ook invloed op die gebruikersparameter, die gebruikersverwysingsraamwerkparameter én die konteksparameter uitoefen. Woordeboekresensies wat in die openbare pers verskyn, word immers nie slegs deur leksikograwe gelees nie (trouens, leksikograwe maak slegs 'n mikroskopiese

fraksie van byvoorbeeld koerantlesers uit), maar kan ook op sekondêre vlak 'n leksikografies-didaktiese funksie en selfs 'n voorligtingsfunksie in die breër gemeenskap vervul. 'n Leksikografies-didaktiese funksie word vervul indien die (koerant)leser iets oor die samestelling en nut van die woordeboek wat geresenseer word (en by implikasie van ander woordeboeke ook), leer; 'n voorligtingsfunksie word vervul indien die breër gemeenskap (weer) deur die media bewus gemaak word van 'n bepaalde woordeboek en van woordeboeke oor die algemeen. In die vestiging van 'n woordeboekkultuur sal die verskyning van woordeboekresensies in die openbare pers 'n belangrike rol vervul.

In 'n relatief onlangse ontwikkeling in die metaleksikografie het De Schrijver en Prinsloo (2000) 'n model voorgestel waarvolgens die leksikograaf terugvoer van die teikengebruikers kan ontvang terwyl hy/sy nog met die saamstel van die woordeboek besig is. Hierdie tegniek noem De Schrijver en Prinsloo "gelyktydige terugvoering", en hoewel hier nie verder daarop ingegaan word nie, hou dié tegniek groot potensiaal vir die saamstel van gebruikersvriendelike woordeboeke in.

Die belang van die terugvoerparameter word daardeur onderstreep dat woordeboekkritiek een van die navorsingsgebiede in die metaleksikografie verteenwoordig.

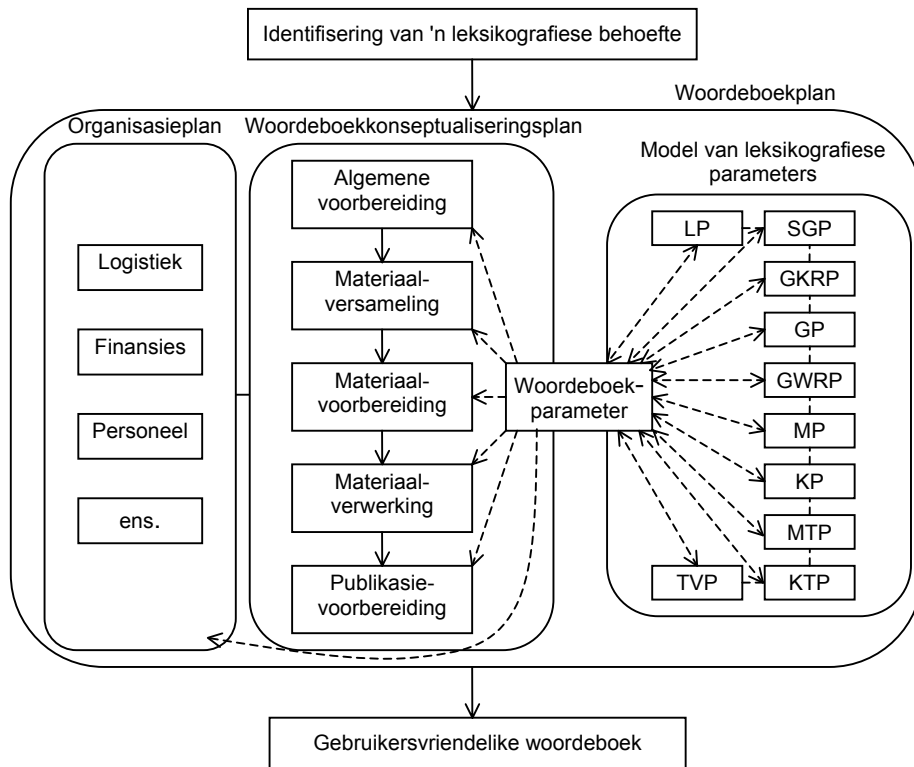
4.12 Samevattend

Die toepassing van die aangepaste kommunikasiemodel op leksikografiese kommunikasie stel 'n nuttige teoretiese raamwerk daar waarbinne die gebruikersgedrewe leksikografie die veranderlikes wat dit bepaal, volledig kan verreken.

Soos reeds geblyk het, is dit belangrik om daarop te let dat die parameters binne die model nie in isolasie of onafhanklik van mekaar funksioneer nie, maar dat elke parameter in 'n interafhanklike verhouding tot die ander parameters staan.

Die model wat steeds ontwikkel word, sal voorlopig *die model van leksikografiese parameters vir gebruikersvriendelike woordeboeke* heet. As model wat wil bydra tot die produksie van gebruikersvriendelike woordeboeke, moet dit binne die leksikografiese proses optree; in hierdie opsig vertoon dit 'n sterk verband met die woordeboekkonseptualiseringsplan binne die woordeboekplan, hoewel dit ook op die organisasieplan betrekking kan hê. Die optrede van die model van leksikografiese parameters vir gebruikersvriendelike woordeboeke kan gevolglik, soos in figuur 4 hieronder uitgebeeld, binne die leksikografiese proses geplaas word.

Soos figuur 4 toon, vind die skakeling tussen die model van leksikografiese parameters vir gebruikersvriendelike woordeboeke en veral die woordeboekkonseptualiseringsplan primêr deur die woordeboekparameter van die model plaas.



Figuur 4: Die funksionering van die model van leksikografiese parameters vir gebruikersvriendelike woordeboeke in die leksikografiese proses

Notasiekonvensie: " \downarrow " rigting van die leksikografiese proses; " $x \leftarrow y$ " y beïnvloed x ; " $x \dashrightarrow y$ " x funksioneer in samehang met y . Afkortings: LP = leksikograafparameter; SGP = stylgidparameter; GKRP = gebruikersparameter; GP = gebruikparameter; GWRP = gebruikersverwysingsraamwerkparameter; MP = mediumparameter; KP = kanaalparameter; MTP = metateksparameter; KTP = konteksparameter; TVP = terugvoerparameter.

5. Perspektief: 'n Definisie vir die werklike doel van 'n gebruikersvriendelike woordeboek

Na aanleiding van die model van leksikografiese parameters vir gebruikersvriendelike woordeboeke kan die formulering van 'n definisie vir die werklike doel van 'n gebruikersvriendelike woordeboek aangepak word.

Tarp (2000: 193) haal Wiegand se algemene formulering van die ware doel van 'n woordeboek soos volg aan:

The genuine purpose of a lexicographic reference work is, according to Wiegand, "that it can be used to obtain information from its lexicographic data about the respective subject of the reference work (daß sie benutzt werden können, um aus ihren lexikographischen Daten Informationen über den jeweiligen Gegenstand der Nachschlagenwerk zu gewinnen)".

Tarp (2000: 193) kritiseer hierdie definisie omdat dit na sy mening te algemeen is (hoewel hy toegee dat sodanige definisie in sy aard algemeen sal en behoort te wees). Na aanleiding van sy bespreking oor leksikografie en terminologie, die gebruikers en hulle eienskappe, gebruiksituasies, gebruikersbehoefes en leksikografiese funksies, stel Tarp (2000: 198) 'n alternatiewe definisie voor:

The dictionary covers this or that area and is conceived to assist users with these or those characteristics in this or that situation in order to solve problems of this or that sort.

Hoewel Tarp se definisie nie veel aan eksaktheid wen nie, word die gebruikersparameter en die gebruikparameter tereg ingereken.

Gouws (2001: 65) verskaf 'n vergelykbare definisie, maar voeg die rol van die teikengebruiker by:

[A dictionary's] genuine purpose is to transfer, by means of lexical data, information regarding the set of lexical items included as treatment units in order to ensure the linguistic empowerment of the intended target user.

Met die voordeel van die insig wat die model van leksikografiese parameters vir gebruikersvriendelike woordeboeke bied, kan vervolgens 'n generiese definisie vir die werklike doel van 'n gebruikersvriendelike woordeboek voorgestel word, wat ook elemente van die voorgaande definisies bevat:

Die werklike doel van 'n bepaalde tipe woordeboek is om binne 'n bepaalde konteks deur middel van 'n bepaalde medium via 'n bepaalde kanaal vanuit die woordeboekbasis, na aanleiding van die stylgids, wat in die metateks verantwoord is, in 'n bepaalde woordeboekteks (gewoonlik 'n woordeboekartikel) 'n bepaalde gebruikersvraag van die teikengebruiker, wat in 'n bepaalde gebruiksituasie verkeer en wat die woordeboek vanuit 'n bepaalde gebruikersverwysingsraamwerk benader, te beantwoord ten einde die talige bemagtiging van die teikengebruiker te verseker.

Bronnelys

Woordeboeke

Elektroniese Verklarende Handwoordeboek van die Afrikaanse Taal 2.00. 2000. Midrand: Perskor. (CD-ROM)

Pharos Woordeboeke / Dictionaries: 5 in 1. 2000. Kaapstad: Pharos. (CD-ROM)

The Oxford Talking Dictionary. 1998. Oxford: Oxford University Press. (CD-ROM)
Webster, N. 1828. *An American Dictionary of the English Language*. New York: S. Converse.

Ander bronne

- Alberts, M.** 1992. Behoeftebepaling in die leksikografie. *Lexikos* 2: 1-27.
- Bergenholtz, H en R.H. Gouws.** 2006. How to Do Language Policy with Dictionaries. *Lexikos* 16: 13-45.
- Bergenholtz, H. en J. Mugdan (Reds.).** 1985. *Lexikographie und Grammatik. Akten des Essener Kolloquiums zur Grammatik im Wörterbuch 28.-30.6.1984*. Lexicographica. Series Maior 3. Tübingen: Max Niemeyer.
- Botha, W.F. en E. Botha.** 1998. Lexicographic Training at the Bureau of the Woordeboek van die Afrikaanse Taal. *Lexikos* 8: 276-281.
- Carstens, A.** 1995. Language Teaching and Dictionary Use: An Overview. *Lexikos* 5: 105-116.
- De Schryver, G.-M. en D.J. Prinsloo.** 2000. The Concept of "Simultaneous Feedback": Towards a New Methodology for Compiling Dictionaries. *Lexikos* 10: 1-31.
- Emejulu, J.D. (Red.).** 2001. *Éléments de lexicographie gabonaise. Tome I*. New York: Jimacs-Hillman.
- Fouché-Van der Merwe, M.** 1999. *Gebruikersperspektief in die Afrikaanse leksikografie*. Ongepubliseerde D.Litt.-proefskrif. Stellenbosch: Universiteit van Stellenbosch.
- Frost, R., H. Vos en M. Dreyer.** 1993. *Kommunikasie dinamiek*. Johannesburg: Lexicon.
- Gouws, R.H.** 1989. *Leksikografie*. Pretoria/Kaapstad: Academica.
- Gouws, R.H.** 1990. Information Categories in Dictionaries, with Special Reference to Southern Africa. Hartmann, R.R.K. (Red.). 1990: 52-65.
- Gouws, R.H.** 1996. Leksikografiese behoeftevervulling. *Lexikos* 6: 171-183.
- Gouws, R.H.** 1996a. 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: Universiteit van Stellenbosch Drukkery.
- Gouws, R.H.** 2001. Lexicographic Training: Approaches and Topics. Emejulu, J.D. (Red.). 2001: 58-94.
- Gouws, R.H. en D.J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.
- Guffey, M.E.** 2001. *Essentials of Business Communication*. Cincinnati: South-Western College Publishing.
- Hartmann, R.R.K. (Red.).** 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.** 1989. Sociology of the Dictionary User: Hypotheses and Empirical Studies. Hausmann, F.J. et al. (Reds.). 1989-1991: 102-111.
- Hartmann, R.R.K. (Red.).** 1990. *Lexicography in Africa. Progress Reports from the Dictionary Research Centre Workshop at Exeter, 24-25 March 1989*. Exeter Linguistic Studies 15. Exeter: University of Exeter Press.
- 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. Handbücher zur Sprach- und Kommunikationswissenschaft* 5.1-5.3. Berlin/New York: Walter de Gruyter.

- Janssen, D. (Red.).** 1996³. *Zakelijke communicatie 1*. Groningen: Wolters-Noordhoff.
- Kirkpatrick, B.** 1989. User's Guides in Dictionaries. Hausmann, F.J. et al. (Reds.). 1989-1991: 754-761.
- Kromann, H.-P., T. Riiber en P. Rosbach.** 1991. Principles of Bilingual Lexicography. Hausmann, F.J. et al. (Reds.). 1989-1991: 2711-2728.
- Kühn, P.** 1989. Typologie der Wörterbücher nach Benutzungsmöglichkeiten. Hausmann, F.J. et al. (Reds.). 1989-1991: 111-127.
- Landau, S.I.** 2001². *Dictionaries. The Art and Craft of Lexicography*. Cambridge: Cambridge University Press.
- Lombard, F.J.** 1994. Lexicographer, Linguist and Dictionary User: An Uneasy Triangle? *Lexikos* 4: 204-214.
- Louw, P.** 2000. Guide Structures in CD-ROM Dictionaries, with Specific Reference to the EWED and the ELHAT. *Lexikos* 10: 291-305.
- Louw, P.** 2001. Resensie van *Pharos Woordeboeke/Dictionaries 5 in 1: 5 ten volle geïntegreerde woerdeboeke/Fully Integrated Dictionaries* op CD-ROM. *Lexikos* 11: 324-327.
- Louw, P.A.** 1997. *Kriteria vir 'n standaard vertalende woerdeboek*. Ongepubliseerde M.A.-skripsie. Stellenbosch: Universiteit van Stellenbosch.
- Martin, W.** 1994. Lexicologie als universitaire vakopleiding. *Lexikos* 4: 215-229.
- Prinsloo, D.J. en R.H. Gouws.** 2000. The Use of Examples in Polyfunctional Dictionaries. *Lexikos* 10: 138-156.
- Read, A.W.** 1973. Approaches to Lexicography and Semantics. Sebeok, T.A. (Red.). 1973: 145-205.
- Sebeok, T.A. (Red.).** 1973. *Current Trends in Linguistics 10*. Den Haag/Parys: Mouton.
- Smit, M.** 1996. *Wiegand's Metalexigraphy as a Framework for a Multilingual, Multicultural, Explanatory Music Education Dictionary for South Africa*. Ongepubliseerde D.Litt.-proefskrif. Stellenbosch: Universiteit van Stellenbosch.
- Steehouder, M., C. Jansen, K. Maat, J. van der Staak, D. de Vet, M. Witteveen, E. Woudstra.** 1999⁴. *Leren kommuniseren*. Groningen: Wolters-Noordhoff.
- Steinberg, S.** 1994. *Introduction to Communication*. Kaapstad: Juta.
- Svensén, B.** 1993. *Practical Lexicography. Principles and Methods of Dictionary-Making*. Oxford: Oxford University Press.
- Swanepoel, P.** 2001. Dictionary Quality and Dictionary Design: A Methodology for Improving the Functional Quality of Dictionaries. *Lexikos* 11: 160-190.
- Tarp, S.** 2000. Theoretical Challenges to Practical Specialised Lexicography. *Lexikos* 10: 189-208.
- Van Cuilenburg, J.J., O. Scholten en G.W. Noomen.** 1992. *Communicatiewetenskap*. Bussum: Coutinho.
- Van der Molen, H.T., F. Kluytmans en M. Kramer.** 1995. *Gespreksvoering: Vaardigheden en modellen*. Groningen: Wolters-Noordhoff.
- Wiegand, H.E.** 1984. On the Structure and Contents of a General Theory of Lexicography. Hartmann, R.R.K. (Red.). 1984: 13-30.
- Wiegand, H.E.** 1985. Fragen zur Grammatik in Wörterbuchbenutzungsprotokollen. Ein Beitrag zur empirischen Erforschung der Benutzung einsprachiger Wörterbücher. Bergenholtz, H. en J. Mugdan (Reds.). 1985: 20-98.
- Wiegand, H.E.** 1996. A Theory of Lexicographic Texts: An Overview. *South African Journal of Linguistics* 14(4): 134-149.

Do Dictionary Users Really Look Up Frequent Words? — On the Overestimation of the Value of Corpus-based Lexicography*

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),*

David Joffe, *TshwaneDJe HLT, Pretoria, Republic of South Africa (david.joffe@tshwanedje.com),*

Pitta Joffe, *Pretoria, Republic of South Africa (leonjoffe@gmail.com), and*

Sarah Hillewaert, *Department of Anthropology, University of Michigan, Michigan, United States of America (sarahhil@umich.edu)*

Abstract: An innovative online Swahili–English dictionary project is presented. A careful study of some of the log files attached to this reference work reveals some hitherto unknown aspects of true dictionary look-up behaviour, which results in the depreciation of the importance of corpora for dictionary making. Three lexicography software modules are advanced to further enhance the success of the online dictionary.

Keywords: LEXICOGRAPHY, SOFTWARE, ONLINE, DICTIONARY, LOG FILE, CORPUS, FREQUENCY, RANK, CORRELATION, SWAHILI, ENGLISH, TSHWANELEX

Samenvatting: Zoeken woordenboekgebruikers werkelijk frequente woorden op? — Over de overschatting van de waarde van corpusgebaseerde lexicografie. Een vernieuwend online Swahili–Engels woordenboekproject wordt voorgesteld. Een minutieuze studie van enkele van de log bestanden gekoppeld aan dit referentiewerk onthult tot dusver onbekende aspecten van het echte opzoekgedrag van woordenboekgebruikers, wat leidt tot een devaluatie van het belang van corpora voor het maken van woordenboeken. Drie lexicografische softwaremodules worden naar voor geschoven om het succes van het online woordenboek verder te vergroten.

Sleutelwoorden: LEXICOGRAFIE, SOFTWARE, ONLINE, WOORDENBOEK, LOG BESTAND, CORPUS, FREQUENTIE, RANG, CORRELATIE, SWAHILI, ENGELS, TSHWANELEX

* Sections of this article are based on a presentation by G.-M. de Schryver at the Tenth International Conference of the African Association for Lexicography, organised by the Sesu sa Sesotho Lexicography Unit, University of the Free State, Bloemfontein, Republic of South Africa, 13–15 July 2005.

1. Corpus-based Lexicography

Since the beginning of modern corpus-based lexicography with the COBUILD project in the early 1980s (Sinclair 1987), hardly anyone has doubted the value of using electronic corpora on both the macrostructural and microstructural levels during the entire compilation process of reference works. Literally hundreds of research papers exalt the benefits of using corpora in lexicography, and upon studying the arguments one also immediately and intuitively tends to agree.

In oversimplified terms the main advantages can be summarised as follows. When one draws a lemma-sign list from the top section of a corpus-derived (lemmatised) frequency list, then the resulting macrostructure (i.e. 'the list of headwords' plus their parts of speech and some morphological guidance) will also provide the dictionary user with what he/she is most likely to want to look up. Focussing on the microstructure of reference works, it further seems sound to accept that when senses are arranged according to their occurrences in corpora, that when examples are chosen from the living language, and that when also all other aspects such as collocationality issues are based on corpus statistics, that the user will again be most satisfied.

When, back in 2003, two of the current authors applied for the *Kernerman Dictionary Research Grants* to support 'The Creation of an Innovative Kiswahili-English Online Dictionary', those same assumptions could also be described. Given all the data in the current article will be based on this online dictionary project, a summary of that proposal seems in order.

2. Project Proposal Summary

Swahili (or Kiswahili in the language itself) is one of Africa's major languages, is the official language of Tanzania, and is also spoken throughout East Africa as a lingua franca by several dozen million people. Since well over a century ago, numerous mono- and bilingual dictionaries have been compiled for Swahili. Given its official status, the substantial number of speakers, and a relatively long lexicographic tradition, one would assume an advanced state of lexicographical research as well as the availability of modern and up-to-date dictionaries for Swahili. This is unfortunately not really the case. Western compilation principles were largely transferred, and up to today the most commonly used Swahili dictionaries remain rooted in lexica originally compiled by missionaries. Overall, one also notices a restricted to non-existent dictionary culture.

Swahili is an agglutinating language, which means that morphemes are juxtaposed to form linguistic words. In all current Swahili dictionaries, 'orthographic words' have been decomposed into their formatives, with only the latter being lemmatised. Not all primary speakers of Swahili can look up 'words' in their own language (as this implies being able to cut off pre- and suffixes),

and even trained learners and scholars often need more than one look-up round before they find what they are looking for (as sound changes between formatives are not always predictable).

In this research project the idea is to deal with all these problems simultaneously. The aim is to create the first corpus-based dictionary that is also intuitive in nature, and to research the feasibility of this approach in real time. Instead of lemmatising stems as in traditional Swahili dictionaries, the suggestion is to lemmatise full orthographic words in addition to stems, and to provide full translations for these strings. In order to sensibly limit the number of items one can physically treat, the intention is to select the items from a frequency list derived from a large corpus. Concordance lines will be culled from the corpus for each frequent orthographic word, and the various translations will be recorded in order of frequency.

A user will thus be able to directly look up words as they are spoken or written, and the translations found will be from most likely to least likely. An English search index will additionally enable searches in the other direction. Such an approach will obviously require much more 'space' than in a traditional stem-based dictionary, which is why the dictionary will be developed and made available in an electronic environment right from the start. This environment will primarily be on the Internet, where it is possible to keep a log of all searches. The analysis of such log files will enable the team to research whether or not this hybrid approach is feasible and to amend the approach if need be.

Given the intuitive lemmatisation approach, especially primary speakers and learners at the elementary and intermediate levels will for the first time be able to effectively look up words, and find meanings of 'real' words, which should come some way in combating the lack of a dictionary culture. Furthermore, log files in an electronic environment are a notoriously underused tool, a tool that will be utilised to its full potential in this project. Each visitor to the dictionary will automatically receive a user ID with which dictionary-using behaviour, including vocabulary retention, will be tracked. For the first time, truly unobtrusive data will be collected and true look-up behaviour in an electronic environment will be recorded. Finally, this project will ensure that Swahili, an increasingly popular language on the Internet, is also kept alive in a modern online reference work based on sound lexicographical principles.

3. The Result: An Innovative Online Swahili–English Dictionary

Three years later, this online Swahili–English dictionary has indeed been produced, and is available at <<http://africanlanguages.com/swahili/>>. As is clear from the project proposal summary above, it has always been (and it still remains) the intention to view the project as a 'work in progress' — or even a 'research environment' — where the project members can freely try out

different approaches to making better (online) dictionaries. Adaptations and changes are based on the searches and look-up behaviour seen in the log files, and involve the development of new lexicography software modules.

As a random example, Addendum 1 shows a screenshot of what a primary speaker of Swahili will see when looking up the English phrase 'that one over there'. One may firstly notice that the entire dictionary interface text and the dictionary's metalanguage have both been 'adapted' (customised) to Swahili, the language of the user. If one browses the same dictionary in English, as is done in Addendum 2, all of these aspects will of course be in English. This is an example of 'dynamic metalanguage customisation', as explained in De Schryver and Joffe (2005).

In a stem-based dictionary for Swahili, only the root *-le* would have been lemmatised, with, in the better dictionaries, also an indication that this root is used to form 'demonstratives of position 3'. If one does not know the concordial agreement system of Swahili, however, this information is not of much help, as during actual usage, this root takes different forms according to the class (cl.) of the noun: *yule* (cl. 1), *wale* (cl. 2), *ule* (cl. 3 and 11), *ile* (cl. 4 and 9), *lile* (cl. 5), *yale* (cl. 6), *kile* (cl. 7), *vile* (cl. 8) and *zile* (cl. 10).

The data shown in Addendum 1 were directly output from *TshwaneDJe HLT's* dictionary compilation software *TshwaneLex*, and as one can see, each of the possible (frequent) forms has been lemmatised, *in addition* to the inclusion of the root. Cross-references (or 'hyperlinks' online) moreover link each of the various forms with the root. Related material is shown on the same output screen, which means that no matter whether users start at any of the full forms (as in Addendum 2) or at the root, they will receive guidance, see the meaning, and be provided with (corpus-based) usage examples.

This hybrid approach to lemmatisation has the advantage that words are 'restored' to their actual appearance in written text, and thus that meaningful and pronounceable text strings are shown rather than mere linguistic concepts (i.e. roots) only.

Showing full forms has another advantage, as seen in Addendum 2, from which one can derive that *ule* as 'that one over there' (for class 3 and class 11 nouns) is homonymous with another *ule* with the meaning 'that you eat' (derived from, and linked to, the root *-la* 'eat'). The latter is exactly one of the innovative aspects introduced into this project from the early stages. Given there are literally thousands of possible forms for each verb, the idea was to only treat the frequent full forms.

Likewise for the other parts of speech: In each case the team would 'focus on' and physically 'enter' full forms only when the corpus frequencies would warrant doing so. The corpus would thus be used as the ultimate arbiter, and with a balanced and representative Swahili corpus of around fifteen million running words, the assumption was that this approach would indeed also answer most users' queries. The remainder of this article now studies how successful this assumption was.

4. **Bergenholtz and Johnsen (2005): Furthering the Discussion**

In response to an article by De Schryver and Joffe (2004), the metalexigraphers Bergenholtz and Johnsen (2005: 122) wrote:

There are only a few published scholarly descriptions of internet dictionary log files. The most interesting contribution from de Schryver/Joffe (2004) describes the log file for a South-African bilingual dictionary, a Sesotho sa Leboa–English dictionary. The number of visitors and the number of lookups is not very high: 21,337 lookups made by 2,530 different visitors. [...] De Schryver/Joffe (2004) fail to mention one very interesting point: With 28,000 English lemmas and 25,000 Sesotho sa Leboa lemmas, the users cannot have looked up all lemmas (with only 21,337 lookups). It would be most interesting to know which types of words are not looked up: Is about 90% or 80% of the dictionary never used at all? The very limited number of lookups indicates that no more than 40–50% of the dictionary is actually being used. Will all lemmas in the dictionary be looked up in time when the dictionary has had many more users? Or are there some lemmas that will never be looked up? If future dictionary makers knew the answers to those questions, they would not have to waste time describing words of no interest to the users.

These are indeed intriguing questions, and questions one can answer once one has enough data. These questions also link in well with what the current project team wanted to find out through a thorough study of the log files of the online Swahili–English dictionary.

A preliminary remark is in order, however, and it concerns word-status. As is well known, users of electronic dictionaries, whether on CD-ROM, an intranet, or the Internet, search for much more than just 'words'. Log files attached to such dictionaries clearly show that users increasingly assume that electronic dictionaries behave like Web search engines such as *Google*, and type in concatenations of keywords, combinations and phrases surrounded by quotes, entire sentences, and even dump full paragraphs (lifted from other sources) into the search field. In addition to that, an increasing number of people do not care about spelling, even type in SMS-like words and smileys, and search for a variety of languages other than the one(s) the dictionary is treating.

For languages such as Swahili and Northern Sotho (Sesotho sa Leboa), there is the additional problem of word-status on word-level itself, with a difference between 'linguistic word' and 'orthographic word'. With for instance thousands of verbal forms for a single verb root, is one dealing with thousands of (orthographic) words, or with just one (linguistic) word? When is which 'word' described? With this in mind, the (Indo-European-biased) statement by Bergenholtz and Johnsen above is at least slightly naive.

In other words, in order to 'solve' the problem posed, one must actually make sure that users of a dictionary 'understand' the lemmatisation approach used in a certain dictionary. Imagine a user wants to find the translation of the

Swahili *mchezaji* 'player'. In a stem-based dictionary, the user will have to know that this word is a deverbative, and thus look up the verbal root *-cheza* 'to play', but in the hybrid approach of the online Swahili–English dictionary either a search for the verbal root or the full noun will reveal the answer. This thus means that one should, in order to see which percentage of this dictionary is being returned over time, also take the cross-referenced material into account, as this material is shown concurrently with the searched-for-item.

The Swahili dictionary was placed online in May 2004, and a little over two years later, over half a million searches were made in both 'directions'. Observe that the latter is actually another novelty introduced in this online dictionary, as only one side is being prepared: Swahili to English. To allow for English searches to be made, a smart English index is generated. This is described online as follows:

Note: This is not your 'standard' English to Swahili dictionary. Instead, the results you see here have been generated following a search in the Swahili to English side. This is a novel type of electronic dictionary we are experimenting with, one that shows how senses in one language are spread all over the lexicon in another, and how these then again spread out, etc.

At the time of writing, there are around 6 500 articles in the Swahili dictionary, and around 11 500 items in the English search index. Each of the Swahili lemma signs treated was chosen for its (high) frequency in the corpus. Assuming that the current (in-progress) data would have been online since 'day one', one can simulate — using the *real* search queries, as the logs have been kept since the very beginning — which percentage of the dictionary is being returned as the number of searches grows. Taking a snapshot every one thousand searches, which at the current look-up rate roughly means 'one snapshot a day', the graph shown in Figure 1 is obtained.

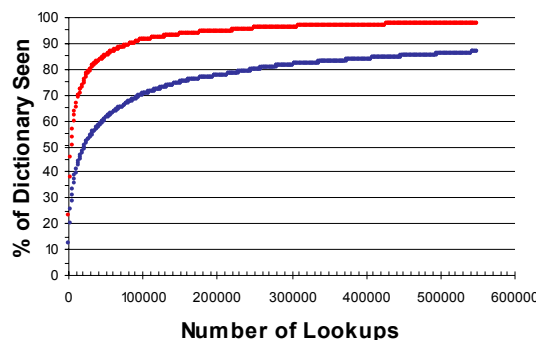


Figure 1: Percentage of dictionary returned ('seen') in function of the total number of searches (bottom: directly, top: with cross-references)

The bottom line indicates that over 86% (86.55%) of the material has been searched for directly, while the top line indicates that close to 98% (97.81%) of the dictionary data have been returned when one also includes the cross-referenced material. Looking at the trend of these lines, it should be clear that *all* dictionary data will indeed be seen over time.

Users search for far more than what is returned, of course, so one would like an indication of the 'hit rate'. Here one can again use the actual half a million searches done so far. The outcome is that 'only' 53.1% of the Swahili searches return one or more hits, while this value climbs to 68.5% for searches in the English index. These values are in line with what was observed for the online Northern Sotho–English dictionary. According to De Schryver and Joffe (oral communication at EURALEX 2004, 7 July 2004), only about 16% of all the misses are 'real misses'. This value was arrived at following a detailed study of each and every miss in their Northern Sotho dictionary. Real misses are those items that should/could have been in the dictionary, and these are basically easy to handle, as one must only make time to compile the necessary articles for them.

Based on all the above and taken at face value, therefore, it seems as if treating just the top-frequent orthographic words in a dictionary will indeed satisfy most users, and this in turn seems to indicate that a corpus-based approach to the macrostructural treatment of the 'words' of a language is an excellent strategy. This conclusion, however, is *not* correct, as will be shown in the next section.

5. The Relation between Corpus Ranks and Actual Dictionary Lookup Ranks

With over half a million real dictionary searches at one's disposal on the one hand, and with corpus-derived frequencies on the other, it becomes possible to calculate various correlation coefficients between the two sets of data. Reformulated, one can effectively take a corpus list of words, and compare that list word for word with actual dictionary searches, and/or one can take searched-for items in a dictionary, and compare those with the corpus. In a way, De Schryver and Joffe (2004: 190) already tried to look into this type of correlation when they sought to answer the following research question:

'Are the top 100 searches also the top 100 in a corpus?' If it would turn out that there is indeed a large overlap, this finding would provide substantial support for the practice of including or omitting lemma signs in a dictionary based on frequency considerations (and by extension for corpus-based lexicography in general).

Based on the fact that 30 of the top 100 Northern Sotho searches could also be found in the corpus top 100, while as many as 63 could be found in the corpus

top 1 000, they came to the conclusion that users indeed look up the frequent words of the language. While this observation is also true for the Swahili-English data at hand, it is only part of the story. It is and remains true that the top few thousand words of a language are also those that users most frequently look up, but the real question one wishes to answer is what happens beyond that point. In the bold words of Bergenholtz and Johnsen one would like to know whether there are indeed words that lexicographers should 'not have to waste time describing' as they are 'of no interest to the users'.

There are different ways to approach this question, but one of the most straightforward ones is as follows. In a two-dimensional plane, one could have the corpus data (as frequencies or ranks) on one axis, and the corresponding actual dictionary lookups (expressed as a count or also as a rank) on the other axis. If corpus-based lexicography indeed reflects (or rather 'pre-empts') what users look up (or 'will look up') in a real dictionary, then the most frequent word in the corpus should also correspond with the word most frequently searched for, the tenth most frequent corpus item should correspond with the tenth most frequent lookup, the one hundredth with the one hundredth, etc. In this ideal situation, the result would be a straight line out of the intersection of the axes in the two-dimensional plane. Allowing for (small) deviations, the straight line would turn into a 'scatter plot', with a cloud of dots 'around' the imaginary straight line. Mathematically, the straight line corresponds with a Pearson correlation coefficient of 1.0, while deviations result in lower values.

Before turning to the actual scatter plot, Table 1 lists the top 10 Swahili words, according to their frequency in the fifteen-million-word TshwaneDJe Swahili Corpus (TSC), and the ranks of these are contrasted with the lookup ranks derived from the actual searches made in the online Swahili dictionary.

Table 1: Comparing corpus ranks with dictionary lookup ranks for Swahili

Item	Corpus frequency	Lookup frequency	Corpus rank	Lookup rank
<i>na</i>	399 663	1 236	1	2
<i>ya</i>	384 813	781	2	6
<i>wa</i>	282 625	683	3	9
<i>kwa</i>	190 645	980	4	4
<i>katika</i>	104 859	472	5	17
<i>za</i>	88 488	244	6	57
<i>ni</i>	87 585	1 173	7	3
<i>kuwa</i>	70 267	469	8	18
<i>la</i>	68 857	239	9	59
<i>hiyo</i>	55 888	117	10	173

A brief look at this top 10 seems to indicate that the full scatter plot might indeed revolve around a straight line. However, and again using the ranks, the outcome of this exercise on a much larger scale as displayed in Figure 2, is at least highly surprising. Note that, given full orthographic words are entered/treated in the online Swahili dictionary, these are compared with *unlemmatised* corpus data. In Figure 2, each dot thus represents the dictionary lookup rank versus the unlemmatised corpus rank of a particular word.

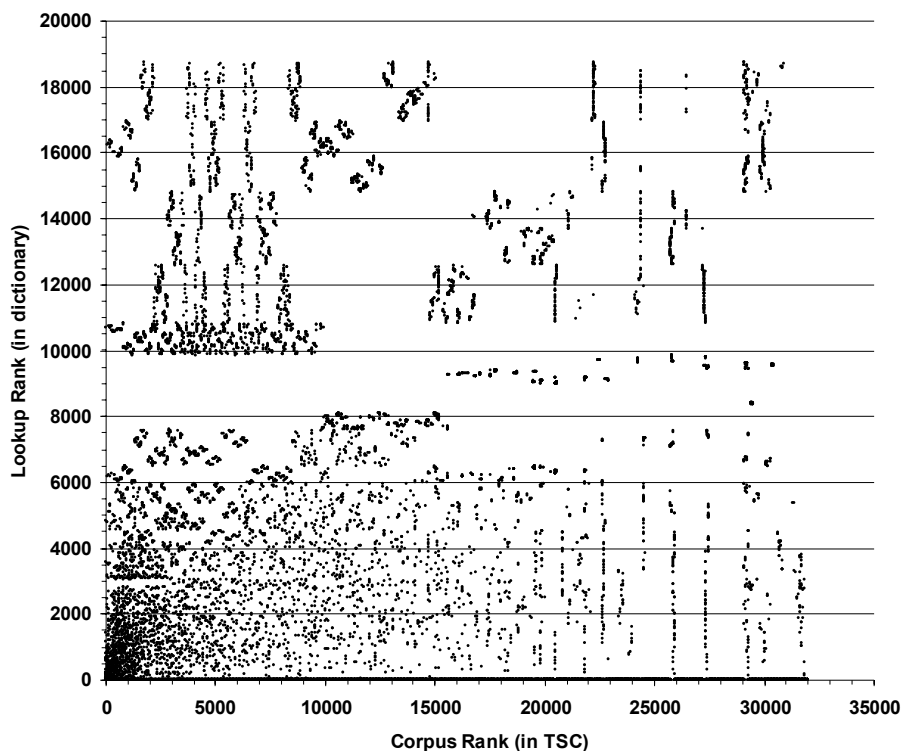


Figure 2: Ranks of the Swahili 'dictionary lookups' versus their corresponding 'corpus frequency' ranks in the TshwaneDJe Swahili Corpus (TSC)

These data for Swahili can immediately be contrasted with those for English. In Table 2, the top 10 English corpus words as seen in the one-hundred-million-word British National Corpus (BNC) are contrasted with their lookup occurrences in the English index, and Figure 3 displays the scatter plot for English. Note, again, that given the nature of the online Swahili–English dictionary, whereby users are 'allowed' to search for non-canonical dictionary forms in both directions, the *unlemmatised* corpus statistics were used in both Table 2 and Figure 3.

Table 2: Comparing corpus ranks with dictionary lookup ranks for English

Item	Corpus frequency	Lookup frequency	Corpus rank	Lookup rank
the	6 187 925	1 100	1	12
of	2 941 786	652	2	41
and	2 682 874	681	3	36
to	2 560 344	1 050	4	14
a	2 150 872	553	5	56
in	1 883 290	642	6	43
that	1 115 377	343	7	127
it	1 089 558	395	8	103
is	998 857	1 037	9	15
was	923 972	200	10	258

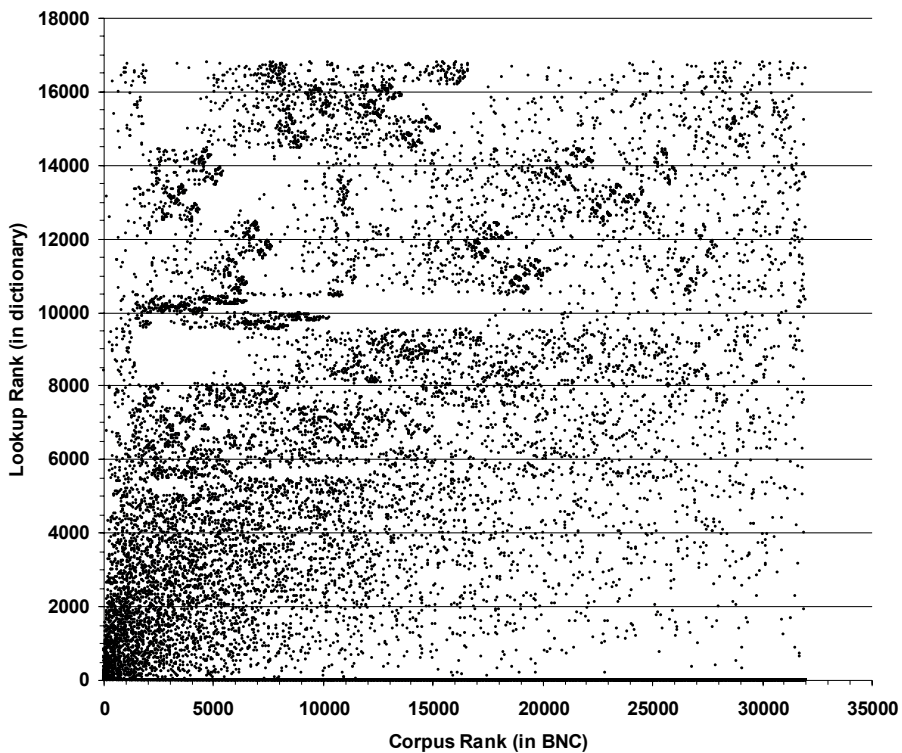


Figure 3: Ranks of the English 'dictionary lookups' versus their corresponding 'corpus frequency' ranks in the British National Corpus (BNC)

If one zooms in on the area around the intersection of the axes in Figures 2 and 3, or thus the top ranks, then one *does* see some kind of vague correlation, but as one moves along the axes, this correlation vanishes entirely. Actually, this too can conveniently be displayed. Figure 4 shows a graph plotting the Pearson correlation coefficient for Swahili word rankings, with each point being the correlation recalculated from 1 to the N*100th point (so each one 'includes' the previous one). Figure 5 shows the equivalent for English.

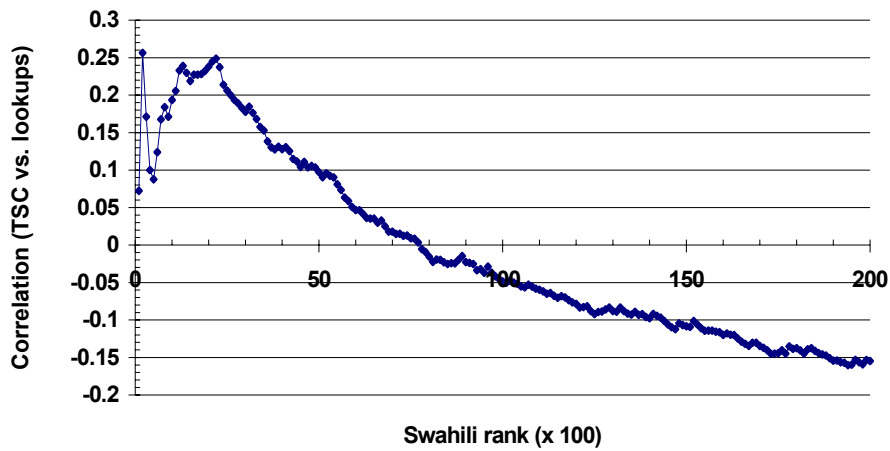


Figure 4: Correlation between corpus ranks and actual dictionary lookup ranks for Swahili (recalculated after every increase of the rank with one hundred)

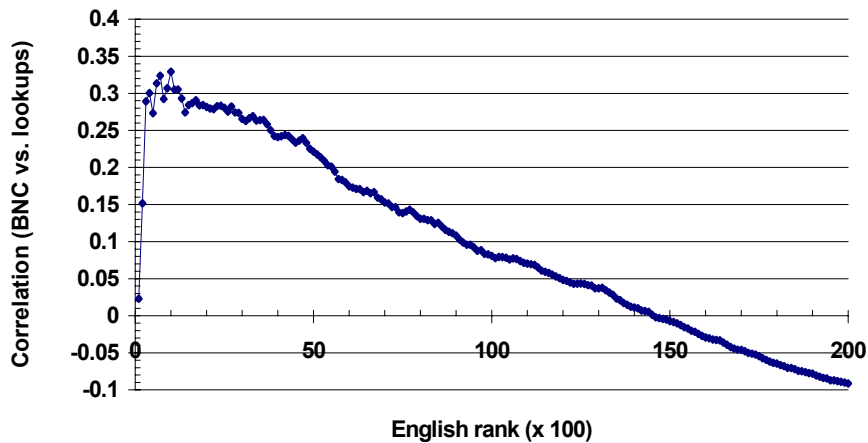


Figure 5: Correlation between corpus ranks and actual dictionary lookup ranks for English (recalculated after every increase of the rank with one hundred)

Figures 4 and 5 clearly reveal that there is indeed some minor correlation between corpus ranks and actual dictionary lookup ranks for the first few thousand words (up to around 3 000 for Swahili, and up to around 5 000 for English), but beyond that point there simply is no correlation whatsoever.

This is a hugely important — albeit shocking — revelation, as it means that it is simply impossible to 'predict' which words will be of interest to the dictionary user. Given the nature of the Internet, it is safe to assume that this dictionary user is a 'general dictionary user' with 'general needs'. To make this conclusion more tangible, take for example Figure 3 at the BNC rank 15 000, which could be the cut-off point for a dictionary with an upper limit of roughly fifteen thousand entries. Looking upwards from that point in Figure 3, it should be clear that it is unfortunately so that virtually any word may be looked up with any frequency at this cut-off point.

6. Additional Lexicography Software Modules

If one were to summarise the outcomes of the research so far — and against the background of other existing studies into (paper) dictionary use, an overview of which may be found in De Schryver and Joffe (2004: 187-188) — then one can make two statements:

- (a) If one needs to prepare a small dictionary, for example a pocket school dictionary, with only a few thousand entries, then it is indeed good practice to base the selection of the lemmas on corpus data.
- (b) If one needs to prepare a large dictionary, for example a large desktop dictionary, with several tens of thousands of entries, then the use of a corpus as an arbiter on what to include in and what to exclude from the dictionary makes little sense for all low-frequency lemmas.

These conclusions, then, pose great difficulties to lexicographers, as the corpus does not provide the 'magic answer' every dictionary maker was hoping for. When one compiles a small (school) dictionary one tends to 'throw out'/'skip' function words and so-called easy and basic words, but those are precisely the ones needed in such a dictionary. When one compiles a large (desktop) dictionary one has become accustomed to using corpus frequencies for selecting material, but it turns out that this is by no means a guarantee for look-up success. For want of any better/other approach at this stage, however, the corpus 'may' continue to be used, but as a guidance only.

Bringing these outcomes back to the online Swahili-English dictionary, and to conclude, it is obvious that 'progressing down the (unlemmatised) corpus frequency list' when selecting lemmas to be treated during compilation is not the way to go (anymore). Of course, in order to increase the hit rate (cf. the

16% 'real misses', mentioned in section 4) one must carry on with the addition of more lemmas. Instead of continuing the manual treatment of 'full orthographic forms', however, it seems more advantageous to call in an extra lexicography software module that could do some level of morphological analysis of those lookups that are not treated in the dictionary. Given a detailed and linguistically exhaustive decomposition is not needed for lexicographic purposes, the project team is currently experimenting with what could be termed 'clumped morphotactic decomposition'. Whenever a particular search is 'not found' in the dictionary, the clumped morphotactics module kicks in, and tries to decompose the search item. This is best illustrated with an example. An actual word that was searched for but not found in the past is the Swahili *yakai-sha* '(and then) they stopped'. From Addendum 3 one can see how guidance is currently given in this regard. At the time of writing, around 100 'Swahili rules' have been stored in the 'clumped morphotactic decomposition' module, with which the Swahili hit rate has increased with another 4%.

From Addendum 1 it can be derived that multi-words have also been stored in the English index. Whenever such items have been stored, the respective Swahili article(s) is (are) simply offered to the user. All Swahili and English combinations can also directly be looked up. However, and as pointed out earlier (cf. section 4 above), users also increasingly search for phrases and entire sentences, and these are of course more often than not missing from the dictionary/index. In order to meet the dictionary user halfway in this regard, another lexicography software module was written that takes the input text, and when no matches are found, presents 'answers' (i.e. displays articles) for up to the first 10 words of an input string. This is illustrated in Addendum 4 for 'I love chicken' (which is an actual search string that was flagged as 'not found' earlier).

A third new lexicography software module that has been developed is an additional custom search index, which aims to re-route frequent misspellings to the most likely form. With this module the hit rate continues to climb. A further extension of this module is illustrated in Addendum 5. The dictionary team is rather sure that the gross majority of the dictionary users who look up *jambo* 'matter, affair, thing' are actually searching for the meaning of the pair *hujambo* 'how are you?' / *sijambo* 'I'm fine!'. Without further ado, by presenting the pair *hujambo/sijambo* together with *jambo*, the user will be in a position to distinguish between these forms, and will hopefully start using the non-corrupted forms.

Looking back, therefore, it is clear that it is simply impossible to know in advance which words users will want to look up in a large dictionary. Corpus frequencies do not predict look-up behaviour beyond the top few thousand words of a language. There is thus no such thing as words a lexicographer better not treat. Instead, and in an electronic environment, it will be more advantageous to add lexicography software modules that help increase the hit rate. As these modules may reuse the already compiled material, properly treating

and covering the top few thousand words of a language, however, remains an important core component of any reference work. Beyond that, lexicographers will have to be inventive.

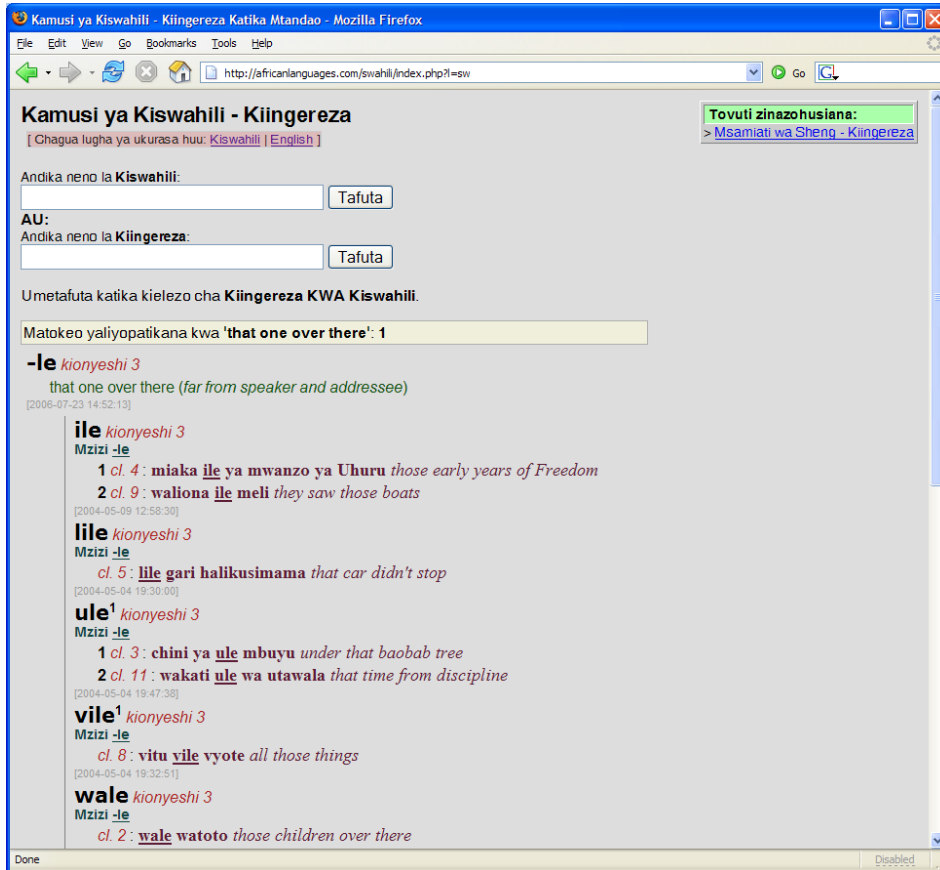
Acknowledgements

Gilles-Maurice de Schryver and Sarah Hillewaert would like to extend their gratitude to Ari Kernerman, who made the 2003 *Kernerman Dictionary Research Grants* available to them for their three-year project 'The Creation of an Innovative Kiswahili-English Online Dictionary'. The work on this dictionary has since been transferred to Pitta Joffe and David Joffe, and the continued development is currently sponsored by *TshwaneDJe HLT*.

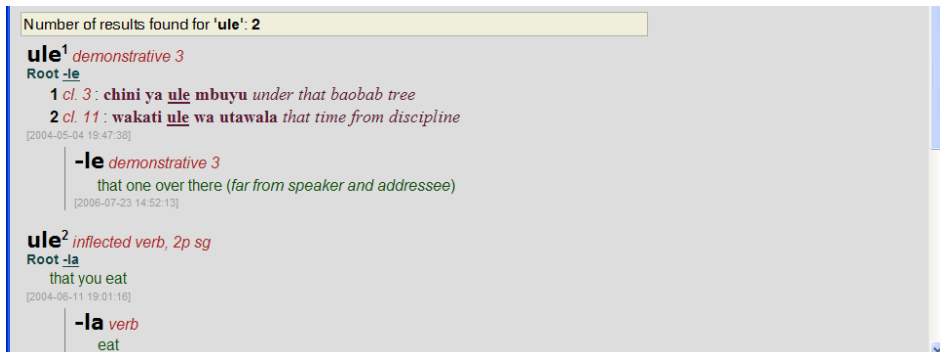
References

- Bergenholtz, Henning and Mia Johnsen.** 2005. Log Files as a Tool for Improving Internet Dictionaries. *Hermes, Journal of Linguistics* 34: 117-141.
- BNC.** 1995–2006. British National Corpus [online]. Available: <<http://www.natcorp.ox.ac.uk>>.
- De Schryver, Gilles-Maurice and David Joffe.** 2004. On How Electronic Dictionaries are Really Used. Williams, G. and S. Vessier (Eds.). 2004. *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004, Lorient, France, July 6–10, 2004*: 187-196. Lorient: Faculté des Lettres et des Sciences Humaines, Université de Bretagne Sud.
- De Schryver, Gilles-Maurice and David Joffe.** 2005. Dynamic Metalanguage Customisation with the Dictionary Application TshwaneLex. Kiefer, F., G. Kiss and J. Pajzs (Eds.). 2005. *Papers in Computational Lexicography, COMPLEX 2005*: 190-199. Budapest: Linguistics Institute, Hungarian Academy of Sciences.
- Google.** 1998–2006. Google Search Engine [online]. Available: <<http://www.google.com/>>.
- Hillewaert, Sarah, Pitta Joffe and Gilles-Maurice de Schryver.** 2004–2006. *Kamusi ya Kiswahili-Kiingereza Katika Mtandao/Online Swahili-English Dictionary* [online]. Available: <<http://africanlanguages.com/swahili/>>.
- Sinclair, John McH.** (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.
- TshwaneDJe HLT.** 2002–2006. TshwaneDJe Human Language Technology [online]. Available: <<http://tshwanedje.com/>>.

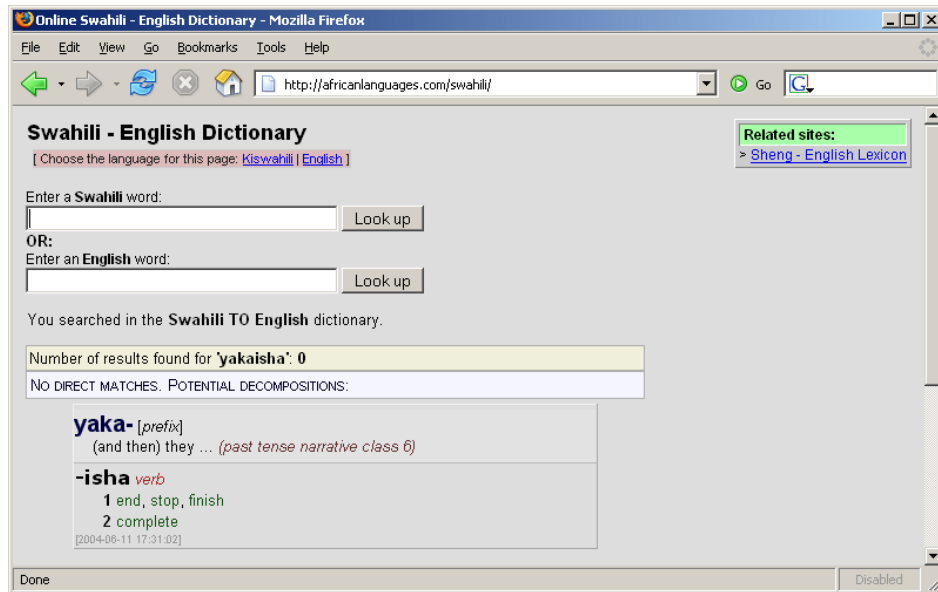
Addendum 1: Looking up 'that one over there' in the online English to Swahili index (with interface in Swahili)



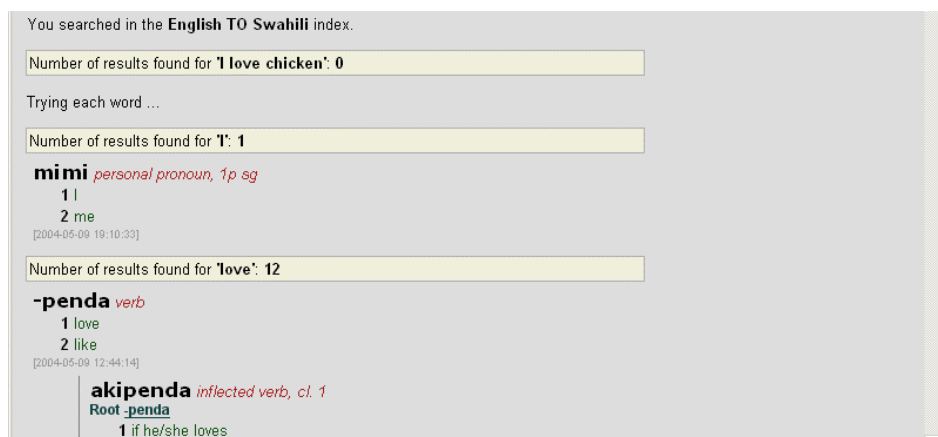
Addendum 2: Looking up *ule* in the online Swahili to English dictionary (with interface in English)



Addendum 3: Looking up *yakaisha* '(and then) they stopped' in the online Swahili to English dictionary (Note the morphological decomposition)



Addendum 4: Looking up 'I love chicken' in the online English to Swahili index (Note that each word is being handled separately)



Addendum 5: Looking up *jambo* 'matter, affair, thing' in the online Swahili to English dictionary (Note that also the pair *hujambo* 'how are you?' / *sijambo* 'I'm fine!' is shown, thanks to the re-router)

Number of results found for 'jambo': 2

jambo *noun 5/6*
matter, affair, thing
[2004-03-02 12:37:58]

mambo *pl noun 5/6*
[See singular jambo](#)
[2004-05-09 19:27:39]

hujambo
[Answered with sijambo](#)
1 how are you?
2 hi, hello
3 greetings
[2006-07-29 16:18:00]

sijambo
[In reply to hujambo](#)
1 I'm fine!
2 hi, hello

Discipline, Dilemmas, Decisions and Data Distribution in the Planning and Compilation of Monolingual Dictionaries*

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

Abstract: Bilingual dictionaries play an important role in the standardisation of a language and are often the first dictionary type to be compiled for a given speech community. However, this may never lead to an underestimation of the role and importance of monolingual descriptive dictionaries in the early lexicographic development of a language. In the planning of first descriptive dictionaries the choice of the proper subtype and a consistent application of theoretical principles should be regarded as of extreme importance. Even the compilation of a restricted descriptive dictionary should be done according to similar theoretical principles as those applying to comprehensive dictionaries. This contribution indicates a number of dilemmas confronting the lexicographer during the compilation of restricted monolingual descriptive dictionaries. Attention is given to the role of lexicographic functions and the choice and presentation of lexicographic data, with special reference to the presentation of certain types of polysemous senses which are subjected to frequency of use restrictions. Emphasis is placed on the value of a heterogeneous article structure and a micro-architecture in the articles of restricted dictionaries.

Keywords: ACCESS STRUCTURE, DATA DISTRIBUTION, FRAME STRUCTURE, FREQUENCY OF USE, HETEROGENEOUS ARTICLE STRUCTURE, LEXICOGRAPHIC FUNCTIONS, LEXICOGRAPHIC PROCESS, MICRO-ARCHITECTURE, MONOLINGUAL DICTIONARY, POLYSEMY, SEMANTIC DATA, TEXT BLOCK, USER-FRIENDLINESS, USER-PERSPECTIVE, VERTICAL ARCHITECTONIC EXTENSION

Opsomming: *Dissipline, dilemmas, besluite en dataverspreiding in die beplanning en samestelling van eentalige woordeboeke.* Tweetalige woordeboeke speel 'n belangrike rol in die standaardisering van taal en is dikwels die eerste woordeboektype wat vir 'n bepaalde taalgemeenskap saamgestel word. Dit mag egter nie tot 'n geringskatting lei van die rol en waarde van eentalige verklarende woordeboeke in die vroeë leksikografiese ontwikkeling van 'n taal nie. Van belang in die samestelling van eerste eentalige verklarende woordeboeke is die keuse van 'n korrekte sub tipe en 'n konsekwente toepassing van gesonde teoretiese beginsels in die beplanning van die woordeboek. Selfs in die samestelling van 'n beperkte verklarende woordeboek moet die beplanning volgens soortgelyke teoretiese beginsels gedoen word as wat die geval is in

* This article was presented as a paper at the Tenth International Conference of the African Association for Lexicography, organised by the Sesiu sa Sesotho Lexicography Unit, University of the Free State, Bloemfontein, Republic of South Africa, 13–15 July 2005.

die samestelling van 'n omvattende woordeboek. Hierdie bydrae wys op 'n aantal dilemmas waarmee die leksikograaf gekonfronteer word in die samestelling van 'n beperkte eentalige verklarende woordeboek. Aandag word gegee aan die rol van leksikografiese funksies, die keuse en aanbod van leksikografiese data met spesifieke verwysing na die aanbieding van sekere tipes betekenisonderskeidinge wat aan gebruiksfrekwensiebeperkings onderhewig is. Die waarde van 'n heterogene artikelstruktuur en 'n mikro-argitektuur in die artikels van beperkte woordeboeke word beklemtoon.

Slutelwoorde: DATAVERSPREIDING, EENTALIGE WOORDEBOEK, GEBRUIKERSERSPEKTIEF, GEBRUIKERSVRIENDELIKHEID, GEBRUIKSFREKWENSIE, HETEROGENE ARTIKELSTRUKTUUR, LEKSIKOGRAFIESE FUNKSIES, LEKSIKOGRAFIESE PROSES, MIKRO-ARGITEKTUUR, POLISEMIE, RAAMSTRUKTUUR, SEMANTIESE DATA, TEKSBLOK, TOEGANGSTRUKTUUR, VERTIKALE ARGITEKTONIESE UITBREIDING

1. Introduction

The planning and compilation of a dictionary may never be done randomly. At the centre of all the decisions regarding the compilation of dictionaries stands the person who has to open the dictionary and use it. This dictionary user is no longer the well-known stranger it used to be because lexicographers are fully aware of the compelling need to identify their potential target user before they even start with the planning of a dictionary. Modern-day lexicography demands a clear indication of the genuine purpose of each dictionary and the fulfilment of that genuine purpose is only possible if the intended target user can achieve an optimal retrieval of information from the dictionary.

Within a multilingual environment it is extremely difficult to decide whether a monolingual or a multilingual dictionary would be the better choice for a given speech community. This question should not be isolated from the bigger lexicographic picture and the real lexicographic needs of the relevant speech community. Looking at these needs, one should also distinguish between the needs of the users in terms of their day-to-day communication and their needs in terms of having a standardised language. Gallardo (1980: 61) maintains that, when a non-standardised language is involved, bilingual dictionaries would be the typical type available, whereas the existence of monolingual dictionaries implies that the cultural properties of the standard language are well established. According to Gallardo, the very existence of monolingual dictionaries testifies to a highly sophisticated native linguistic scholarship and a sizeable literate speech community. However, one should never underestimate the important role of bilingual dictionaries, especially in a multilingual society. Moreover, one should never underestimate the importance of moving on and working towards monolingual dictionaries. "Working towards" implies that it is not an either/or choice when it comes to bilingual versus monolingual dictionaries. The emergence of the bilingualised dictionary or the monolingual dictionary with a bilingual dimension, i.e. a hybrid form which contains features of both monolingual and bilingual dictionaries, has given

enough proof that such a dictionary type can often be regarded as the best lexicographic response to the diverse needs of a specific user group.

The realisation of the need, also in the South African national lexicography units, for bilingual dictionaries will direct the focus of the remainder of this article to various aspects regarding decisions in the planning and compilation of monolingual dictionaries.

2. Dilemmas and decisions in the planning and compilation of a monolingual dictionary

2.1 Lexicographic functions

The compilation of any dictionary needs to be preceded by a thorough planning phase which gives attention to the users, the type, the functions, the structures and the contents of the envisaged dictionary. This planning should be done in a structured way. Research in the field of metalexicography has provided enough guidelines not only to assist lexicographers in their endeavours to formulate the necessary dictionary conceptualisation plan, but also to ensure that the lexicographic process follows the necessary direction for the eventual success of the dictionary. For the lexicographic process of any dictionary there are a number of crucial route markers which cannot be ignored without lexicographers running into difficulties. The lexicographer should not only know at whom the dictionary is directed but also how the target user will use it and in what user situation it will be used. This implies that once the target user, the needs of the target user and the situation of use have been ascertained, the lexicographer can determine the dictionary type and the functions of the dictionary. Regarding the functions it has to be decided whether the dictionary should help to fulfil a knowledge-directed function or a communication-directed function, and, if the latter is the case, most importantly whether it should aim at helping with text production, text reception or both these functions. These questions regarding lexicographic functions (cf. Tarp 2000, Bergenholtz and Tarp 2002, Tarp and Gouws 2004) have to be dealt with at a very early stage of the lexicographic process. Once the functions have been determined the lexicographer can decide on the relevant lexicographic structures to provide a venue for the contents so that the functions can be fulfilled (cf. Gouws to appear).

2.2 Subtypological dilemmas

One of the decisions to be made at an early stage regards the typological nature of the dictionary. The typological category of monolingual dictionaries includes a number of subtypes. Even a broad subtypological category like "general monolingual dictionary" has a number of members. These dictionary subtypes do not represent distinctly isolated categories but rather a continuum of dic-

tionaries ranging from the comprehensive multivolume product to the smallest of dictionaries for tourists, or an initial source of reference like *My First Three Hundred Words* for a young child acquiring his/her mother-tongue.

Although the comprehensive multivolume product, especially if it is based on historical principles, represents the ultimate lexicographic aim, it should not, on account of its size or typological classification, be regarded as a better dictionary than any other monolingual dictionary of a much more modest extent. The compilation of a monolingual dictionary should be the result of a well-informed decision based on the needs of a specific user group. These needs should also determine the subtypological category of the dictionary to be compiled. In this regard it is important to recognise the fact that the lexicographic process of any dictionary has to be regarded as a sophisticated endeavour that should reflect the relevant theoretical approach. Although the compilation phase of the lexicographic process of a comprehensive dictionary takes much longer than that of a standard or a school dictionary, the planning of all dictionary types needs to be done along the same lines. To ensure that the compilation of a small monolingual dictionary results in the publication of a good product a planning exercise similar to that of a multivolume dictionary is demanded.

2.3 Selection and presentation of data and the lexicographic treatment

The nature and extent of the treatment on offer in different types of monolingual dictionaries differ vastly and this creates various dilemmas for the lexicographer. One of the first dilemmas regards the macrostructural selection. When confronted with all the interesting words in one's corpus it is difficult to limit the list of lemma candidates to a small number. One should, however, resist the temptation to incorporate more lemmas than originally planned. Discipline to stick to one's assignment is extremely necessary. A further dilemma is posed by the microstructural entries. Here it is of paramount importance that a microstructural programme should be devised in the early phases of the lexicographic process and that lexicographers should adhere to the constraints of such a programme. Yet another dilemma confronts the lexicographer when it comes to decisions regarding the data-distribution structure. Modern-day lexicography (cf. Kammerer and Wiegand 1998, Gouws 2004) has emphasised the need for a frame structure with front- and back-matter texts complementing the central list and adding venues that can accommodate lexicographic data. The use of front- and back-matter texts should not be seen as applying exclusively to bigger dictionaries. Users in an environment where a dictionary culture thrives, should be made aware that the dictionary as a carrier of text types has much more on offer than the data included in the central list. What better way of helping to establish a dictionary culture than by introducing the use of a frame structure in the first dictionary products with which a member of a speech community is confronted?

For the average dictionary user the central list remains the primary target of the dictionary consultation procedures. Consequently the presentation and treatment in the central list should be of such a nature that the intended target user can find answers to the questions that prompted the consultation. This has implications for both the macrostructural and microstructural selection and presentation. As far as the macrostructure is concerned, a choice has to be made between a straight alphabetical ordering and a sinuous lemma file with nested and niched lemmata (cf. Gouws 2005, to appear). On a microstructural level, a variety of relevant topics demands a thorough planning during the early phases of the lexicographic process (cf. Gouws 2003). These comprise the type of microstructure, the data categories to be included and the organisation and presentation of the different entries in the dictionary article. When it comes to the data on offer, it is important that lexicographers should select data in accordance with the criteria applicable to the subtypological classification of the dictionary. Besides the data categories, the extent of the treatment allocated to each class and the presentation thereof are crucial factors. When the organisation and presentation of entries are planned, the notion of the micro-architecture of the article structure comes to the fore.

The main focus of the remainder of this article will be on the selection of some types of semantic data and the development of a micro-architecture for smaller monolingual dictionaries.

2.4 Aspects of the selection of semantic data

The most typical consultation procedure in monolingual dictionaries is directed at the retrieval of information regarding the meaning of the word represented by the lemma sign. Consequently, in spite of the importance of all the other microstructural entries, lexicographers may never neglect the treatment allocated to the meaning of a word. Attention should not only be given to an explanation of the meaning of the word but it should also be shown that the lexicon consists of a network of semantic relations. For the lexical items being treated relations like semantic opposition, e.g. antonymy and complementarity, and semantic inclusion, e.g. synonymy and hyponymy, should therefore be indicated. The paraphrase of meaning, often referred to as the lexicographic definition, remains the most prominent microstructural entry to convey semantic data.

The dynamics of language leads to constant changes on all levels of language and language use. This includes the lexicon, with new items being added and others being omitted, as well as the semantic load of single lexical items, with changes in the polysemous paradigm by the addition or omission of polysemous senses. Decisions regarding the number of polysemous senses to be presented in the treatment of a lexical item pose a real challenge to lexicographers. In this regard the lexicographers of a comprehensive multivolume dictionary have fewer problems. Their historical approach compels them to

give account of both old and new senses and even where the dictionary does not follow an historical approach they can rely on a balanced and representative corpus to identify all the senses that pass the minimum frequency of use test, and they have to include these senses. A standard dictionary has a lower density of data and usually fewer co-text entries but it still contains a reasonable number of senses based on frequency of use. In using this criterion one often finds that figurative and metaphoric senses are listed before the literal senses of a word, and that the original senses display a lesser frequency of use than new senses. Bigger monolingual dictionaries have the luxury to include the lesser used literal and older senses along with the figurative or newer senses, allowing users to make a link between figurative and literal and between old and new.

In learners', desk and smaller dictionaries the strict adherence to usage frequency can become a problem. Fox (2005) indicates that a sole reliance on usage frequency often leaves the user stranded for lack of a link between a metaphoric sense of a word and its literal sense. This also applies to new literal senses. In the *Oxford Basic English Dictionary* the word *tablet* is treated as a monosemous item with only one paraphrase of meaning, i.e. "a small hard piece of medicine that you swallow". The user is left stranded regarding the well-known older sense "flat piece of clay, wood or stone with words inscribed on it". The *South African Oxford School Dictionary* assists its users by including three senses, i.e. "a pill", "a solid piece of soap" and "a flat piece of stone or wood etc. with words carved or written on it". In the compilation of a small dictionary the lexicographer needs clear guidelines to decide how the choice of polysemous senses should be made and whether it is sufficient only to rely on frequency of use as a criterion for inclusion. Consistency remains a cornerstone of the lexicographic practice and lexicographers should not follow two different approaches in their selection and ordering of senses. However, it may become necessary to complement the criterion of frequency of use with an additional procedure in order to give users a fuller picture of the meaning of the word. This is especially true in the treatment of words where the metaphoric sense is much more frequent than the literal sense.

Where the lexicographer opts for a selection and ordering of senses according to frequency of usage this criterion should be applied in all articles. The number of senses allowed in the article of a lemma representing a polysemous lexical item should not only be determined purely mathematically but also on usage frequency. Instead of allowing only two senses per article provision should be made that the treatment of a given word with three frequently used senses should have three subcomments on semantics to include these senses. However, depending on the size of the dictionary, the lexicographer could decide beforehand not to go further than, say, three or four senses in any given article. The inclusion of a given sense should satisfy a minimum frequency of usage. This may mean that the metaphoric sense qualifies for inclusion but not the literal sense. If the lexicographer feels it is important that users

should get information on a literal or an old sense of the word, the article structure should make provision for the inclusion of an additional text block in which the complementing data is presented. This should not be an additional subcomment on semantics, because all senses allocated to subcomments on semantics qualify on the grounds of their frequency of use. In modern-day Afrikaans the word *pitkos* is primarily used in a figurative sense, i.e. "substantial food for thought". The literal sense of this word is "concentrated food like grain". Very few people actually use, or even know, the word *pitkos* in this sense. In a comprehensive dictionary each one of these senses should be allocated a separate subcomment on semantics but in a dictionary of limited extent only the figurative sense will qualify if frequency of use is the criterion. For the user of such a dictionary it might be necessary to link the figurative sense to a literal one. However, the lexicographer should adhere to the principles governing the sense selection and should not include the literal sense if it does not satisfy these principles, e.g. the frequency of usage criterion.

One of the valuable insights of modern-day metalexigraphy is that dictionaries do not necessarily have to display a homogeneous article structure. A specific article structure needs to be devised to which the default article should adhere but provision should also be made for motivated deviations from this structure. The introduction of a heterogeneous article structure allows the lexicographer to include some additional or alternative article-specific entries, irrespective of whether they have a lemmatic or a non-lemmatic addressing.

Where the criteria for the selection of polysemous senses for inclusion in the subcomments on semantics exclude a specific sense which the lexicographer deems necessary for the user, the lexicographer may opt for a heterogeneous article structure to include an additional entry complementing the default article structure. This could, for instance, be done by means of an inserted inner text, or a comment block. In the case of *pitkos* the treatment of the figurative sense could be followed by a postcomment in which an indication is given of the original meaning of the word.

pitkos s.nw.

Kragtige voedsel vir die gees ...

Oorspr.: gesonde kos wat bestaan uit mieliepitte.

Dictionary article 1

If the lexicographer regards this as important data, it should be presented in a position of salience. This demands that the planning of the article structure should make provision for positioning data in such a way that the user can be made aware of it and have unimpeded access to it. In this regard the lexicographer of any dictionary, even a restricted monolingual dictionary, should embrace the notion of a micro-architecture.

2.5 The micro-architecture of dictionary articles

Bergenholtz, Tarp and Wiegand (1999) make a distinction between dictionary articles with and those without a micro-architecture. In an article with an architecture the article-internal items display a left-right as well as a top-bottom relation. The following article from *The South African Pocket Oxford Dictionary of Current English* does not display a micro-architecture:

ventilate ... *v.* **(-ting)** **1** cause air to circulate freely in (a room etc.). **2** air (a question, grievance, etc.). **3 Med.** **a** oxygenate (the blood). **b** admit or force air into (the lungs).

Dictionary article 2

The introduction of a micro-architecture could lead to the following presentation:

ventilate ... *v.* **(-ting)**
1 cause air to circulate freely in (a room etc.).
2 air (a question, grievance, etc.).
3 Med. **a** oxygenate (the blood). **b** admit or force air into (the lungs).

Dictionary article 3

From a user perspective dictionary article 3 is more user-friendly than dictionary article 2. This is primarily the result of article 3 displaying four text blocks compared to the single text block of article 2. Because of these text blocks article 3 has a micro-architecture (cf. Bergenholtz, Tarp and Wiegand 1999: 1770). These text blocks play an important role in the recognition of the entries accommodated in them. Bergenholtz, Tarp and Wiegand (1999: 1770) argue that they increase the access confidence of the user and diminish the access time.

In dictionary article 3 the first text block only contains the comment on form. It is separated from the comment on semantics, presented in the following three text blocks where a subcomment on semantics, containing a polysemous sense of the lexical item represented by the lemma sign, appears in each one of the text blocks, preceded by a polysemy marker. The comment on form precedes the comment on semantics in the top-bottom relation and within the comment on semantics the different subcomments on semantics are also in a top-bottom relation.

Following the arguments of Bergenholtz, Tarp and Wiegand (1999: 1771) the higher degree of user-friendliness of dictionary article 3 compared to dictionary article 2 is also evident from the rapid inner access structure flowing vertically in article 3 compared to the sinuous flowing in article 2. The rapid inner access structure of article 3 shows a vertical-architectonic extension, and vertical-architectonic access structures are regarded as straight access structures.

Compare dictionary articles 4 and 5 below. Dictionary article 4 from the electronic version of the Afrikaans monolingual dictionary HAT has no micro-architecture:

bad ☒ s.nw. (-dens; -jie) **1** Groot bak waarin 'n mens jou liggaam kan was. **2** Die was van die liggaam in so 'n bak: *'n Bad neem.* **3** Hoeveelheid water, vloeistof waarin die liggaam gewas word: *Jou bad is glad te warm.* **4** Bak waarin bv. olie gehou word en waarin 'n deel van 'n masjien werk: *Oliebak* **5** (mv. baaie) Blootstelling van die liggaam aan water, sonstrale, stoom, modder, elektrisiteit, ens.: *'n Sonbad, modderbad.* **6** (mv. baaie) Badplek met geneeskragtige water: *Die baaie gebruik.* Vgl. WARMBAD. UITDR.: *Iemand 'n koue bad gee, hom ontgoël.* ☒ ww. (gebad) Die liggaam in 'n bad was: *'n Mens behoort elke oggend te bad.*

Dictionary article 4

In dictionary article 5, the above HAT article has been adapted to display a micro-architecture:

bad ☒ s.nw. (-dens; -jie)
1 Groot bak waarin 'n mens jou liggaam kan was.
2 Die was van die liggaam in so 'n bak: *'n Bad neem.*
3 Hoeveelheid water, vloeistof waarin die liggaam gewas word: *Jou bad is glad te warm.*
4 Bak waarin bv. olie gehou word en waarin 'n deel van 'n masjien werk: *Oliebak*
5 (mv. baaie) Blootstelling van die liggaam aan water, sonstrale, stoom, modder, elektrisiteit, ens.: *'n Sonbad, modderbad.*
6 (mv. baaie) Badplek met geneeskragtige water: *Die baaie gebruik.*
Vgl. WARMBAD.
UITDR.: *Iemand 'n koue bad gee, hom ontgoël.*
☒ ww. (gebad) Die liggaam in 'n bad was: *'n Mens behoort elke oggend te bad.*

Dictionary article 5

In dictionary article 5 the comment on form is yet again ordered in a top-bottom relation to the comment on semantics. Within the comment on semantics a vertical-architectonic relation exists between the different subcomments on semantics. Separate text blocks for fixed expressions and the occurrence of the lemma in a different part of speech function follow the subcomments on semantics.

The user-friendly nature of a dictionary article with a micro-architecture is also seen in the following excerpt from the article of the lemma sign **because** in the *Collins COBUILD on CD Rom*:

because
1 because
You use because when stating the reason for something.
He is called Mitch, because his name is Mitchell.
Because it is an area of outstanding natural beauty, the number of boats available

for hire on the river is limited.

Women are doing the job well. This is partly because women are increasingly moving into a man's world.

'Why didn't you tell me, Archie?' 'Because you might have casually mentioned it to somebody else.'

CONJ-SUBORD

2 because

You use because when stating the explanation for a statement you have just made.

Maybe they just didn't want to ask too many questions, because they rented us a room without even asking to see our papers.

The President has played a shrewd diplomatic game because from the outset he called for direct talks with the United States.

*I had a sense of *deja vu* because I could recognise everything in London.*

CONJ-SUBORD

3 because

If an event or situation occurs because of something, that thing is the reason or cause.

Many families break up because of a lack of money.

Because of the law in Ireland, we had to work out a way of getting her over to Britain.

PHR-PREP

4 because

You use just because when you want to say that a particular situation should not necessarily make you come to a particular conclusion. (SPOKEN, INFORMAL)

Just because it has a good tune does not mean it is great music.

Just because something has always been done a certain way does not make it right.

PHR-CONJ-SUBORD

Dictionary article 6

Using a micro-architecture does have space and therefore also cost implications. But it also has implications for the success of the dictionary consultation procedures of the target users and it can enhance the rapid and unambiguous retrieval of information and play an important role in fulfilling the lexicographic functions and the genuine purpose of the dictionary.

3. In conclusion

The compilation of monolingual dictionaries confronts lexicographers with decisions and dilemmas, and demands a well-disciplined approach to the task. It is important to regard all dictionaries as sophisticated lexicographic products and not to argue that a dictionary of limited extent is inferior to its more comprehensive family members. Lexicographers need to conduct the planning and compilation of smaller dictionaries in a similar theoretically-based fashion as they would use in the planning and compilation of comprehensive dictionaries.

Care should be taken that the data distribution and presentation reflect the needs of the users and the typological criteria of the specific dictionary. A consistent application of these criteria is of extreme importance. Provision should be made for a deviation from the default article structure in order to present lexical item-specific data needed by the user. To ensure the best possible presentation of all data types the use of a micro-architecture will lead to macro benefits for the user.

References

- 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., S. Tarp and H.E. Wiegand.** 1999. Datendistributionsstrukturen, Makro- und Mikrostrukturen in neueren Fachwörterbüchern. Hoffmann, L. et al. (Eds.). 1999. *Fachsprachen. Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft/Languages for Special Purposes. An International Handbook of Special-Language and Terminology Research: 1762-1832.* Berlin: De Gruyter.
- Collins COBUILD on CD Rom.* London: HarperCollins.
- Crawley, E. (Ed.).** 2004¹¹. *Oxford Basic English Dictionary.* Oxford: Oxford University Press.
- Fox, G.** 2005. Metaphor in Learners' Dictionaries. Unpublished paper read at the Fourth Asialex Conference, organised by the National University of Singapore, Singapore, 1-3 June 2005.
- Gallardo, A.** 1980. Dictionaries and the Standardization Process. Zgusta, L. (Ed.). 1980. *Theory and Method in Lexicography: 59-69.* Columbia: Hornbeam Press.
- Gouws, R.H.** 2003. Aspekte van mikrostrukturele verskeidenheid en inkonsekwentheid in woordeboeke. *Lexikos* 13: 92-110.
- Gouws, R.H.** 2004. Outer Texts in Bilingual Dictionaries. *Lexikos* 14: 67-88.
- Gouws, R.H.** 2005. Oor die verhouding tussen woordeboekstrukture, woordeboekinhoud en leksikografiese funksies. *Lexikos* 15: 52-69.
- Gouws, R.H.** To appear. Sublemmata or Main Lemmata — A Critical Look at the Presentation of Some Macrostructural Elements. Mogensen, E., H. Gottlieb and A. Zettersten (Eds.). To appear. *Symposium on Lexicography XII: Proceedings of the Twelfth International Symposium on Lexicography, April 29–May 2, 2004 at the University of Copenhagen.* Amsterdam: Benjamins.
- Hawkins, J.M.** 1996. *The South African Oxford School Dictionary.* Cape Town: Oxford University Press.
- Kammerer, M. and H.E. Wiegand.** 1998. Über die textuelle Rahmenstruktur von Printwörterbüchern: Präzisierungen und weiterführende Überlegungen. *Lexicographica* 14: 224-238.
- Odendal, F.F. and R.H. Gouws.** 1998. *Die Elektroniese Verklarende Handwoordeboek van die Afrikaanse Taal.* Midrand: Perskor.
- Tarp, S.** 2000. Theoretical Challenges to Practical Specialised Lexicography. *Lexikos* 10: 189-208.
- Tarp, S. and R.H. Gouws.** 2004. Wie leer wat uit Afrikaanse (aan)leerderwoordeboeke. *Tydskrif vir Geesteswetenskappe* 44(4): 276-298.
- Thompson, D.** 1987. *The South African Pocket Oxford Dictionary of Current English.* Cape Town: Oxford University Press.

Inclusion Strategies for Multi-word Units in Monolingual Dictionaries*

Phillip Louw, *Department of Afrikaans and Dutch, University of Stellenbosch, Stellenbosch, Republic of South Africa (pal@sun.ac.za)*

Abstract: This article focuses on inclusion strategies for different types of multi-word units, be it as part of the macrostructure or embedded as treatment units in the microstructure of a specific dictionary. The types of multi-word units discussed range from multi-word lexical items to collocations and multi-word compound lexical items. The general principles set out in this article are applied specifically to monolingual school dictionaries that target learners of English in the junior secondary phase.

In order to discuss inclusion strategies adequately it is, however, necessary to make a cursory distinction between idioms and collocations, on the one hand, and between collocations and multi-word compound lexical items, on the other. It is shown that current monolingual dictionaries often fail to distinguish between these types and therefore apply potentially confusing inclusion strategies.

In the discussion of inclusion strategies for multi-word lexical items that follows, it is shown that, whereas loan groups and group prepositions require lemmatisation as full multilexical lemmas, the strategy for idioms is not as simple. The problems with a full lemmatisation of idioms are pointed out and an alternative system, whereby idioms are consistently included as sublemmas with full microstructural treatment, is proposed.

Next it is shown that collocations do not have lexical item status and can therefore not be treated in the same way as multi-word lexical items. However, provision must be made that some collocations may need additional microstructural treatment addressed to them.

Lastly, inclusion strategies for multi-word compound lexical items, which frequently occur in English, are discussed. The practice of sublemmatising so-called "transparent" compound lexical items and giving them no or little microstructural treatment, is shown to be inappropriate for school dictionaries.

Hopefully the guidelines provided in this article can be of some help in clearing up the muddled approaches currently followed in some South African monolingual school dictionaries.

Keywords: COLLOCATIONS, COMPOUND NOUNS, GROUP PREPOSITIONS, IDIOMS, INCLUSION STRATEGIES, LEMMA, LOAN GROUPS, MACROSTRUCTURE, MICROSTRUCTURE, MULTI-WORD COMPOUND LEXICAL ITEMS, MULTI-WORD UNITS, SUBLEMMA, TRANSPARENCY

Opsomming: **Opnamestrategieë vir veelwoordige eenhede in eentalige woordeboeke.** Hierdie artikel fokus op opnamestrategieë vir verskillende veelwoordige een-

* This article is an adaptation of a part of a chapter from a D.Litt. Dissertation *Criteria for a Multifunctional, Monolingual Dictionary in Junior Secondary Education*, which was accepted by the University of Stellenbosch, Stellenbosch, Republic of South Africa in April 2004.

hede, hetsy as deel van die makrostruktuur of ingebed as behandelingseenhede in die mikrostruktuur van 'n bepaalde woordeboek. Die tipes veelwoordige eenhede wat bespreek word, wissel van veelwoordige leksikale items tot kollokasies en veelwoordige samestellings. Die algemene beginsels wat in hierdie artikel uiteengesit word, word spesifiek op eentalige skoolwoordeboeke wat leerders van Engels in die junior sekondêre fase teiken, toegepas.

Om opnamestrategieë toereikend te bespreek, is dit egter nodig om eers 'n onderskeid te tref tussen idioeme en kollokasies aan die een kant, en tussen kollokasies en veelwoordige samestellings aan die ander. Daar word aangetoon dat die huidige eentalige woordeboeke dikwels nalaat om tussen hierdie tipes te onderskei en derhalwe potensieel verwarrende opnamestrategieë toepas.

In die bespreking van opnamestrategieë wat daarop volg, word aangetoon dat, terwyl leenwoordgroepe en groepvoorsetsels lemmatisering as volle multileksikale lemmas benodig, die strategie vir idioeme nie so eenvoudig is nie. Die probleme met 'n volle lemmatisering van idioeme word uitgewys en 'n alternatiewe stelsel waardeur idioeme konsekwent as sublemmas met volle mikrostrukturele behandeling opgeneem word, word voorgestel.

Vervolgens word aangetoon dat kollokasies nie leksikale-itemstatus het nie en dus nie op 'n soortgelyke wyse as meerwoordige leksikale items behandel kan word nie. Daar moet egter daarvoor voorsiening gemaak word dat sommige kollokasies wel addisionele mikrostrukturele behandeling mag benodig.

Laastens word opnamestrategieë vir meerwoordige samestellings, wat dikwels in Engels voorkom, bespreek. Die gewoonte om sogenaamde "deursigtige" samestellings te sublemmatiseer en dan van min of geen mikrostrukturele behandeling te voorsien nie, word as onvanpas vir skoolwoordeboeke getoon.

Hopelik kan die riglyne wat in hierdie artikel verskaf word, van hulp wees om die verwarde benaderings wat tans in sommige Suid-Afrikaanse eentalige woordeboeke gevolg word, op te klaar.

Slutelwoorde: DEURSIGTIGHEID, GROEPVOORSETSELS, IDIOME, KOLLOKASIES, LEENWOORDGROEPE, LEMMA, MAKROSTRUKTUUR, MEERWOORDIGE EENHEDE, MEERWOORDIGE SAMESTELLINGS, MIKROSTRUKTUUR, NAAMWOORDSAMESTELLINGS, OPNAMESTRATEGIEË, SUBLEMMA

Introduction

Multi-word units present many problems to practical lexicographers, ranging from criteria for their selection through to the actual microstructural treatment afforded to each type. In this article the focus is, however, on inclusion strategies for different types of multi-word units, be it as part of the macrostructure or embedded as treatment units in the microstructure of a specific dictionary. The general principles set out in this article are applied specifically to monolingual school dictionaries that target both mother tongue and non-mother tongue learners of English in the junior secondary or senior phase. The types of multi-word units discussed range from multi-word lexical items (with specific reference to idioms), to collocations and multi-word compound lexical items (with specific reference to compound nouns).

Multi-word lexical items

Multi-word lexical items should be considered for inclusion in a monolingual school dictionary and can be lemmatised as multilexical lemmas. Gouws (1989: 97) states that "as a lexical item a multilexical lemma represents a single semantic unit, and the meaning of this unit cannot be deduced from the sum of the meaning of the constituent parts" [my translation]. Zgusta (1971: 154) adds that "for the lexicographer, the detection and correct presentation of multi-word lexical units is one of his most important tasks". Yet, what this correct presentation should be is a polemic issue. It may depend on the type of dictionary and may even differ for different types of multilexical lemmas, as illustrated in the following discussion. On a macrostructural level the lexicographer needs to decide whether these items should be lemmatised as main lemmas, or whether they can be listed under the first prominent constituent of the multi-word lexical item. Should the latter option be preferred, methods should be found not to perpetuate the confusing practice of grouping multi-word lexical items with collocations and examples.

Loan groups

Loan groups are perhaps the multi-word lexical items most consistently lemmatised as main lemmas. They are lemmatised in full and the space between constituent parts simply ignored when determining their place in the dictionary's sort order. This is also the practice in the *Chambers-Macmillan South African Dictionary Junior Secondary* (henceforth SADJS) and *The South African Oxford School Dictionary* (henceforth SAOSD), where loan groups such as **et cetera** and **post mortem** are lemmatised in full.

Group prepositions

The lexicographic treatment of group prepositions is a more disputed matter. Generally, the status of group prepositions as lexical items is not fully recognised in current monolingual standard and school dictionaries. They are often included as collocations and not as lemmas. Furthermore, even in dictionaries where their value and search priority are recognised and they are given sublemmatic status, group prepositions are often not sublemmatised under their first constituent parts.

If the group preposition **in aid of** is taken as an example, it soon becomes obvious that lexicographers identify **aid** as its main element and accordingly use that lemma as point of inclusion. This is the case in both SAOSD and SADJS where **in aid of** is given as a sublemma under **aid**. This practice unfortunately leads to the disruption of the initial alphabetical ordering principle adhered to elsewhere in these dictionaries. Such a disruption may be justifiable

in the case of idioms, where the dictionary culture leads users to look up idioms under the first prominent constituent. In the case of group prepositions, though, it is uncertain whether the dictionary culture dictates this practice to the same measure. It could therefore be contended that the users of a school dictionary may well be better served by listing group prepositions consistently as multilexical main lemmas. This would also be a lexicologically sound lexicographic practice.

Idioms

Should the model employed for other multi-word lexical items be perpetuated, idioms should also be lemmatised as multilexical main lemmas. However, this is not a practical solution for a school dictionary. Firstly, it will not always be possible to identify the initial component of the idiom. Articles are often interchangeable or optional at the start of an idiom and other subtle variations can occur. Secondly, lemmatisation of idioms can take up more space than alternative methods. Thirdly, the current dictionary culture (perpetuated by the available dictionaries) is one in which users of school dictionaries will probably expect to find the idiom as a sublemma under the lemma corresponding to the first word in the idiom that is considered to be semantically prominent, especially a noun, verb or adjective. This practice probably stems from the assumption that words function as independent lexical items in an idiom, rather than as constituent parts of an encompassing multi-word lexical item.

A case could be made out that a school dictionary is the ideal place to start changing the dictionary culture in subtle ways and that the lemmatisation of idioms should therefore be considered, as it is a lexicologically and lexicographically sounder method. However, the practice of including idioms as sublemmas is so strongly entrenched that such a move may be experienced as too unconventional and therefore result in users not finding the data they are looking for. Furthermore, the standard dictionaries, which these users are likely to use when school dictionaries no longer meet their needs, also predominantly give idioms as sublemmas, one of the functions of the junior secondary school dictionary being to prepare its users for a seamless transition to standard dictionaries. It may, therefore, be more advisable to work within these confines by ensuring that idioms are clearly marked and that the microstructural treatment of these idioms is as user-friendly as possible.

SAOSD and SADJS have opted to conform to the often-used practice of including idioms as sublemmas. There is, however, a problem with their approach. Should idioms be included as sublemmas, they need to be clearly distinguishable from microstructural data categories. This is unfortunately not the case in SADJS and SAOSD. Multi-word compound lexical items, idioms and collocations are treated similarly, making it difficult for the user to discern between these data types. The lemmatisation of multi-word compound lexical items will help to alleviate this situation, but the problem of possible confusion

between collocations and idioms remains. SAOSD bears witness to this confusion at the lemma **stick**, where the idiom **stick up for** is presented in exactly the same way as the collocations **stick out** and **stick to**. All three are included as sublemmas and given near-full microstructural treatment. As neither the collocations nor the idiom are deemed as fully transparent, items giving the meaning description are provided throughout.

Gouws (1996: 5) proposes the following solution:

Belonging to separate information categories the collocations and idioms ... should be accommodated in different article positions which will leave the user with different search areas allocated to each information category. By using different typefaces or structural markers the user could be lead to a clear distinction between these two information categories.

This solution can be modified in that idioms should rather maintain their lexical item status and function as sublemmas instead of entries within data categories in the microstructure. Collocations, on the other hand, will fit into that part of the comment or subcomment on semantics reserved for examples, but the possibility should still be there for less transparent collocations to be treatment units. The compiler(s) of a dictionary can also consider using an explicit structural indicator to show the start of the idiom group, as is practiced in WAT and HAT, for example, to ensure swift access. In terms of micro-architecture, it would also be advisable that each idiom, as well as the structural indicator introducing the idioms, start on a new line.

Collocations

Collocations are lexical combinations usually included in the microstructure as co-text entries in order to illustrate, what Gouws (1989: 227) refers to as "the typical microsyntactic context of the lemma" [my translation]. These combinations are typical and usually transparent. They therefore do not have lexical status as a whole, but comprehensive inclusion is still a necessity, "especially in pedagogical and translation dictionaries" [my translation] (Gouws 1989: 227). Cop (1991: 2776) states as reason for their inclusion that "even transparent collocations must be present, because they are not predictable". This sentiment is echoed by Svénson (1993: 101): "Information about collocations is important in both monolingual and active bilingual dictionaries, since the user cannot be expected to know which words customarily occur together." Data on collocations provides microsyntactic empowerment, especially to users employing their dictionaries in an encoding task.

Transparency is, however, a problematic concept as users' perceptions of what is and is not transparent can differ greatly. It would therefore be wise for the lexicographer to err on the side of caution and ensure that collocations of which the transparency is at all doubtful, be included as treatment units. The

extent of the treatment will depend on the perceived lack of transparency (the lexicographer has to exercise sound judgement, but empirical research could also be of value here). It can include a short item giving a paraphrase of meaning, constructed examples showing the macrosyntactic use of the collocation, or a combination of these two data types.

If collocations are to be truly user-friendly, they "must reflect natural language" [my translation] (Gouws 1989: 227). It is therefore important that corpus data is analysed in order to identify possible collocations. The superior sorting abilities of the new generation of corpus-querying tools make this a more or less standard task for the lexicographer. These programs have the additional advantage of indicating to the lexicographer the frequency of use of each collocation. Should there then be a need to only select certain collocations, due to there being too many to include, the lexicographer can choose the most typical ones.

There is a marked difference in the treatment of collocations in SADJS and SAOSD. SADJS presents detailed example material in the form of collocations and constructed examples, whereas SAOSD opts for a larger macrostructure at the cost of linguistic examples. SAOSD does present some collocations and a very limited number of constructed example sentences.

There is, however, a significant problem in both these dictionaries' placement of less transparent collocations acting as treatment units. As has been mentioned in the discussion of idioms, these collocations are displaced from the normal search zone for syntactic data and moved to the end of the article to be lumped together with compound nouns consisting of more than one constituent, and idioms. This move has various implications. Firstly, it is very difficult for the target user to determine which type of data is being dealt with. This treatment therefore clashes with a basic lexicographic principle, i.e. that each data type should be treated distinctly. Secondly, the displacement disrupts the coherence in the search zone for examples, in that the micro- and macrosyntactic data are now distant.

Lastly, and perhaps most importantly, SAOSD and SADJS's placement of collocations at the end of an article complicates the search path at polysemous lemmas or lemmas with more than one syntactic function. Collocations may vary in transparency, but the guiding principle in determining whether a phrase is a collocation is still that there should be a discernable correspondence between the lexical item represented by the lemma's manifestation in the collocation and the meaning of either the lexical item represented by the lemma or the sense of the lexical item represented by the lemma. At **long** in SADJS, for example, the collocation **before long** corresponds to the second sense of **long**, but interspersed between them are the third sense, a compound noun (**long jump**) and an idiom. The user has to follow a complicated remote addressing procedure to bring all the relevant data together. It would therefore be much more sensible to include the collocation at the specific sense or syntactic function it corresponds to.

Compound lexical items

Besides the fundamental distinction between idioms and collocations, there is also another necessary distinction, i.e. between collocations (as microstructural items) and multi-word compound lexical items (as macrostructural items). The inclusion of compound lexical items provides another difficult macrostructural challenge to the compiler of a monolingual English dictionary. This challenge is specifically rooted in the variation in spelling of these compound lexical items, the appropriateness of sublemmatisation to the target user group and the question of transparency. In the following paragraphs, the treatment of compound nouns is used as an example, as this category best illustrates the variety of problems faced in the macrostructural treatment of compound lexical items.

Béjoint (1999: 81) comments that using "graphic cohesion" as a criterion to distinguish compound nouns "is difficult to apply, particularly in English, because of the variations in spelling: an English compound noun like *paper clip* can have the forms XY, X Y, or X-Y". He (1999: 82) adds that "this makes the automatic extraction of compounds particularly difficult in English". The lexicographer should obviously not have trouble identifying one-word or hyphenated compounds and considering them for lemmatisation or sublemmatisation, but the so-called "open compounds" (written as two words) can be a more challenging prospect. As is hinted at by Béjoint, it is often difficult to extract these from corpora, especially without having sophisticated software with corpus-querying tools that can sort according to context on the right of the search term. This problem underlines, once again, the need for such software. Furthermore, it can be very difficult to determine whether the combination dealt with is a compound noun or a collocation. Here the lexicographer's intuition, as an advanced language user, will play an important role, but more scientific criteria can be identified to aid in the task. Béjoint (1999: 82) lists some of these criteria as "non-compositionality", "position of the stress", "frequency", and "lexical unity". (For a more detailed discussion of these criteria, see Béjoint 1999: 82.)

Once a method of distinguishing between collocations and multi-word compound nouns has been found, the treatment of this type of compound noun can be contemplated. These compound nouns should be treated in the same way as single-word or hyphenated compound nouns, because "despite the blank, these compounds will be identified as one concept and therefore one base form ..." (Schnorr 1991: 2815). The question then arises whether compound nouns should be included as lemmas or as sublemmas.

The use of sublemmas can be an important space-saving mechanism when they are given a limited microstructural treatment (e.g. just part-of-speech indication). Yet, there are serious reservations regarding their appropriateness for use in a school dictionary, which must be addressed by any prospective compiler. Firstly, either nesting or niching must be identified as the user-friendliest ordering method. Secondly, the compiler(s) must discern which types of mor-

phologically complex items can be sublemmatised. Thirdly, the level of textual condensation of the specific sublemma signs must be addressed, as, for example, omitting the part of the sublemma that corresponds to the lemma can save space, but could also alienate the target user group if it does not understand this procedure. Compound nouns present a particularly taxing task to the lexicographer when considered for sublemmatisation.

In many dictionary projects, the decision to lemmatise or sublemmatise depends on the level of transparency of the compound noun. Should a compound noun be deemed transparent, i.e. that the sum of the meanings of its constituent parts is equal to the meaning of the whole, it is often sublemmatised and given a limited microstructural treatment. Transparency is, however, a highly subjective criterion that requires the lexicographer to make assumptions as to which compound nouns the target users of the dictionary may experience as transparent. Béjoint (1999: 84) correctly surmises that "the actual transparency of a compound noun varies according to the ability of each language user to understand its elements."

The assumption of transparency is particularly difficult in the compilation of a junior secondary school dictionary, as there is a considerable difference between the linguistic skills and intuition of the lexicographer and the target user group. Furthermore, the target user group is a diverse group with great variation anticipated in the linguistic skills and intuition of its individual members. To this dilemma could be added that, even if the constituent parts of a compound noun are recognised by the user, confusion could still arise as to which senses of the constituents are activated by their functions in the whole. A good case can therefore be made out that the compiler(s) of a dictionary should not readily assume transparency, but rather give a full microstructural treatment to each compound noun that meets the frequency requirements for inclusion. The lemmatisation of compound nouns, as guiding elements of default single articles, would be one way of achieving this goal.

Both SADJS and SAOSD have a somewhat unusual approach to the treatment of compound nouns. In both dictionaries single-word or hyphenated compound nouns are usually lemmatised, whereas those of the multi-word variety are sublemmatised, but, strangely, given detailed microstructural treatment. Examples of this type of treatment are provided by **grandfather clock** (sublemmatised under **grandfather** in both SADJS and SAOSD), **further education** (sublemmatised under **further** in SADJS) and **rat race** (sublemmatised under **rat** in SAOSD). This treatment is not always applied consistently. SAOSD, for example, gives **sulphuric acid** full lemma status, whereas it could (following similar examples in SAOSD) have been sublemmatised under **sulphur**.

The sublemmatisation procedures for compound nouns followed in SADJS and SAOSD can be motivated from the point of view that it places these items where they morphologically belong. More research needs, however, to be done to determine whether the target users of the dictionary would expect to find

these items as sublemmas. It may be contended that giving these items full lemma status and ignoring the space between constituent parts when ordering them in with single-word lemmas would better meet the user expectations and be less confusing. It would also solve the problems presented by decisions based on perceived transparency.

Conclusion

Hopefully the guidelines provided in this article can be of some help in clearing up the muddled approaches currently followed in some South African monolingual dictionaries. Such improvements will, in turn, be to the benefit of the users of these dictionaries, especially those whose linguistic ability is still at a formative stage.

Bibliography

Dictionaries

- Chambers-Macmillan.** 1996. *Chambers-Macmillan South African Dictionary Junior Secondary*. Manzini: Macmillan Boleswa Publishers.
- Hawkins, J.M.** 1996. *The South African Oxford School Dictionary*. Cape Town: Oxford University Press.

Other sources

- Béjoint, H.** 1999. Compound Nouns in Learners' Dictionaries. Herbst, T. and K. Popp (Eds.). 1999: 81-99.
- Cop, M.** 1991. Collocations in the Bilingual Dictionary. Hausmann, F.J. et al. (Eds.). 1989-1991: 2775-2778.
- Gouws, R.H.** 1989. *Leksikografie*. Cape Town: Academica.
- Gouws, R.H.** 1996. Idioms and Collocations in Bilingual Dictionaries and their Afrikaans Translation Equivalents. *Lexicographica* 12: 54-88.
- Hausmann, F.J. et al. (Eds.).** 1989-1991. *Wörterbücher. Ein internationales Handbuch zur Lexikographie/Dictionaries. An International Encyclopedia of Lexicography/Dictionnaires. Encyclopédie internationale de lexicographie*. Berlin/New York: Walter de Gruyter.
- Herbst, T. and K. Popp (Eds.).** 1999. *The Perfect Learners' Dictionary (?)*. Lexicographica. Series Maior 95. Tübingen: Max Niemeyer Verlag.
- Schnorr, V.** 1991. Problems of Lemmatization in the Bilingual Dictionary. Hausmann, F.J. et al. (Eds.). 1989-1991: 2813-2817.
- Svensén, Bo.** 1993. *Practical Lexicography: Principles and Methods of Dictionary-Making*. Oxford: Oxford University Press.
- Zgusta, L.** 1971. *Manual of Lexicography*. Prague: Academia / The Hague/Paris: Mouton.

The *Dictionnaire français–yipounou / yipounou–français* of the Église Évangélique du Sud-Gabon (CMA) 1966*

Ludwine Mabika Mbokou, *Department of Afrikaans and Dutch, University of Stellenbosch, Stellenbosch, Republic of South Africa*
(ludy2475@yahoo.com)

Abstract: The promotion and development of a language involves a wide usage of dictionaries and reference works. The native languages of Gabon are no exception. Although lexicography is still a young discipline in the Gabonese domain, remarkable works have been produced by non-lexicographers. They have therefore left a considerable amount of data available for the development of modern lexicography in Gabon. The different lexicons and dictionaries compiled during this time offer an important treatment of the Gabonese languages. Or, seen from another angle, everything should be done, considering the state of these works, to revise them. For some of the Gabonese languages like Yipunu, Fan and Myeng, lexicons, dictionaries, grammars and even encyclopedias already exist. However, several questions remain pending: Why are these dictionaries not used? Is it possible for the non-native speaker to learn the language by means of a dictionary? Is it possible for the native speaker to have quick and easy access to the wanted information? This article therefore has as main aim to give tentative answers to these questions by examining the *Dictionnaire français–yipounou / yipounou–français* of the CMA.

Keywords: ACCESS STRUCTURE, BACK MATTER, BILINGUAL DICTIONARY, CENTRAL LIST, CORPUS, DICTIONARY, FRONT MATTER, HOMONYMS, HOMOGRAPHS, INCLUSION POLICY, LEXICOGRAPHIC TREATMENT, LEXICOGRAPHY, MACROSTRUCTURE, MICROSTRUCTURE, POLYSEMOUS SENSES, TARGET USERS, REVERSIBILITY, REVISION, SOURCES, TRANSLATION DICTIONARY, TRANSLATION EQUIVALENT

Résumé: Le *Dictionnaire français–yipounou / yipounou–français de l'Église Évangélique du Sud-Gabon (CMA) 1966*. La promotion et le développement d'une langue va de pair avec une utilisation importante de dictionnaires et d'ouvrages de référence. Les langues locales du Gabon ne font pas exception. Bien que la lexicographie soit une discipline encore jeune dans la sphère gabonaise, de remarquables travaux ont été produits par des non-lexicographes. Ces derniers ont ainsi légués un nombre considérable de données disponibles pour

* This article is a revised version of a paper presented at the Ninth International Conference of the African Association for Lexicography hosted by the Groupe de Recherche en Langues et Cultures Orales (GRELACO) and held at the Omar Bongo University in Libreville, Gabon, 21–23 July 2004.

le développement de la lexicographie moderne au Gabon. Les différents lexiques et dictionnaires compilés durant cette période présentent un important traitement des langues du Gabon. Ou, pris sous un autre angle, tout est à refaire vu l'état de ces ouvrages. Dans certaines langues du Gabon telles que le yipunu, le fan et le myeng, des lexiques, des dictionnaires, des grammaires et même des encyclopédies existent déjà. Pourtant plusieurs questions restent en suspens: Pourquoi ces dictionnaires ne sont-ils pas utilisés? Est-il possible pour le locuteur natif d'apprendre la langue au moyen d'un dictionnaire? Est-il possible pour un locuteur non-natif d'avoir un rapide et facile accès aux informations désirées? Cet article a donc pour objectif principal d'apporter des tentatives de réponses à ces questions, tout en faisant un examen du *Dictionnaire français-yipounou / yipounou-français* de la CMA.

Mots-clés: CORPUS, DICTIONNAIRE, DICTIONNAIRE BILINGUE, DICTIONNAIRE DE TRADUCTION, DONNÉES, EQUIVALENCES, HOMONYMES, HOMOGRAPHE, INCLUSION, LEXICOGRAPHIE, MACROSTRUCTURE, MICROSTRUCTURE, NOMENCLATURE, PRÉ-TEXTES, POSTTEXTES, RÉVISION, REVERSIBILITÉ, SENS POLYSÉMIQUE, SOURCES, STRUCTURE D'ACCÈS, TRAITEMENT LEXICOGRAPHIQUE, USAGERS CIBLES

1. Introduction

In his survey of the lexicographic work in the Gabonese languages, Mihindou (2001: 7-37) critically discusses some of the existing dictionaries and encyclopedias. He points out the lack of lexicographic principles, the practice of dictionary making in Gabon dating since long before the establishment of metalexigraphy.¹ The existing lexicographic work would therefore benefit from having a good theoretical basis.

The *Dictionnaire français-yipounou / yipounou-français* (henceforth DFY/YF) is a translation dictionary, strictly presenting translation equivalents with no further lexicographic treatment. With this dictionary, the compilers wanted to assist the Bapunu people in studying the Bible. Not having any linguistic or lexicographic background, they collaborated with the Bapunu people among whom they were living at that time. The users targeted with the dictionary are all those who believe in the Christian faith. These include a variety of users with different backgrounds. Old, young, and every Mupunu person whose desire it is to understand the Bible is meant to use the dictionary. This specification made by the compilers narrowed the data distribution process to the vocabulary concerning the Christian faith and everything dealing with the church. Knowing that their dictionary would not be of the desired standard, the compilers made provision for suggestions to improve its quality. With the help of these suggestions, they expected to publish a new revised edition of the dictionary. This version was supposed to be more "complete" and more "correct" as they mention in the front matter. However, after almost forty years, no such revised version has been published.

The aim of this article is to characterize the DFY/YF and to consider how, in the light of metalexigraphical principles, it can be revised.

2. The *Dictionnaire français–yipounou / yipounou–français*

In translation dictionaries, the emphasis is often on providing only the translation equivalents of a given language. No further treatment is required although it is possible to find such lexicographic treatment in some translation dictionaries. The DFY/YF is a bidirectional bilingual dictionary arranged alphabetically. As such, it is composed of two main sections that represent the two central lists of the dictionary: the French–Yipunu (F–Y) section and the Yipunu–French (Y–F) section. The dictionary has no back matter and the front matter consists only of some acknowledgements made by the compilers. As the focus of this article will be on macro- and microstructural issues, no further attention will therefore be paid to the content of the front matter.

The fact that the main central list is divided into two sections makes it possible for the dictionary to accommodate two different macrostructures. The dictionary as a whole contains 8 829 entries divided between the central lists. Central list A (F–Y) has 6 541 entries and central list B (Y–F) 2 288 entries. The unequal division between the number of entries in each list is immediately noticeable. The F–Y section contains almost 74% of the entries while the Y–F section contains only 26%.

2.1 The target users

The structuring of the macrostructural elements of any dictionary highly depends on the needs of the intended target users of that specific dictionary. The amount of data to be included and the ordering and arrangement are all determined by the needs of a certain user group. It is therefore important for any compiler of a specific dictionary to define the target users. Principles of user-friendliness should determine any effort or attempt of collecting data, compiling a corpus and designing or planning a dictionary. The selection of macrostructural elements and their inclusion in the structure of the dictionary should also be determined by the design plan of the dictionary which should be based on a well-defined target user group (Mabika Mbokou 2002).

According to the missionary compilers, the intended users of the DFY/YF would be persons belonging to the Christian church and interested in acquiring Biblical knowledge. These persons would include two categories of users: young and adult users. If a dictionary is compiled to meet the needs of a wide range of target users, the objective to address any group satisfactorily would be difficult. However, when dealing with a wide range of users, lexicographers tend to turn to polyfunctional dictionaries. This would only have been a partial solution for the compilers of the DFY/YF, because polyfunctional dictionaries cannot meet the needs of very different types of users. And it is the case with this particular dictionary. A child attending a Sunday school class has a completely different need from an adult listening to a sermon. These are two different types of users who each needs different data information. The main object of the compilers of

the DFY/YF was to produce a tool to help in performing their evangelical task. Although the DFY/YF forms a good basis for the compilation of further corpora in Yipunu, it was available to any Punu and/or French speaker who wanted to use it. Whether the user was able to access and retrieve the needed information easily was not the main concern.

Translation dictionaries are often the result of a bilingual situation. It is the perceived needs of the users that effects the compilation of a particular dictionary. Thus, the way the languages are treated in any bilingual dictionary should be tested among the users. A quasi-bilingual range of users has used the DFY/YF, but it is not intended to meet the needs of monolingual French speakers. The lexicographic treatment in this dictionary could have been directed at Yipunu as well as French speakers. In doing so, the compilers could have narrowed the profile of their intended target users, yet, these users still form a very wide range with different needs.

Along with the language issue, another restriction could have been made concerning the age of the users as well as their social status. During 1966 when the dictionary was compiled, not all Bapunu people in the rural areas were able to read or write French, not to mention Yipunu. The dictionary was only accessible to people who could read and write in these languages. This was in contrast to the type of members attending church services. It was the time of the early independence of Gabon when the church was most often attended by old people. When the target users are considered from this perspective, it seems difficult to really say for whom the dictionary was intended.

2.2 The access structure

In metalexigraphy, the access structure of a dictionary is the search route a user follows in an attempt to retrieve the needed information. It is one of the most important structural components of the dictionary. The only access structure that the DFY/YF presents consists of the lemma candidate list. The dictionary was made available in a typed version. Although no bold or italic type-script was used to distinguish between the lemma sign and its translation equivalents, a short dash delimits the one language from the other. The compilers relied on the knowledge of the users to make this distinction. For French mother-tongue users with sound dictionary skills it would have been easy to differentiate between Yipunu and French, and thus to identify the needed lemma. But it would have been more difficult for Yipunu mother-tongue users with no experience of consulting a dictionary.

The success of any dictionary, and a bilingual one in particular, highly depends on the ease of the consultation process. If the target users find it easy to retrieve the looked-for information, the dictionary has achieved its main goal. Gouws (2001: 102) explains this as follows: "The rapid and unimpeded access of the user to the relevant data presented in the dictionary has to be regarded as a prerequisite for a successful lexicographic product in a user-

driven approach." No integrated texts are found in the front matter of the DFY/YF. Due to this, the access structure of the dictionary is limited. However, the pattern followed throughout the dictionary is applied in a consistent way. It is during the consultation process that the user acquires the skill of retrieving the needed information. He/she is led by the alphabetical ordering and the short dash, and his/her feeling for and knowledge of the language(s).

2.3 The ordering and inclusion policy

As has been mentioned above, the DFY/YF is a bilingual dictionary without outer texts as such. The lemmas are arranged alphabetically in both central lists. Nevertheless, some comments concerning the order and the inclusion policy of macro- and microstructural elements should be made.

2.3.1 The macrostructure

To be able to fulfil the needs of its intended target users the macrostructure of any synchronic dictionary should reflect the treated language(s) as they are spoken at the time of its compilation. It should include lexical creations newly introduced into and adopted by the language(s). However, even words concerning general matters of the church and everything associated with the church environment and activities were not fully treated in the DFY/YF. For example, items such as *Sunday school*, "catechism", or *choir* "chorale" are not part of the macrostructural element of the central lists. The reason for this is probably that the Bible was the main source for the words included in the dictionary. The following remarks concern the inclusion of lemmas in the macrostructure of the DFY/YF.

Firstly, the alphabetical order of the lexical items included as main lemmas is interrupted now and then. In the F-Y section, transitive verbs are often included in their infinitive form. Prepositional phrases are also listed alphabetically along with the preposition that precedes them.

Example:

habile - na kuijulu, uguku	travail - yisalu	troisième - ajimurieru
habilement - na uguku	travailler - usala	trompe - mumbanda, yilanga
habilité - yiuku	travailleur - musalitsi	tromper - udunsa
habillement - yiduaru	à travers - puangala	se tromper - utsimbu
s'habiller - uduara	traverser - uvioga, usabuga	tromperie - ndunsulu
habit - biduaru	trébucher - uduku dibaku	trompette - mumbanda
habitable - ayivu na mandagu	treillis - diguyi na maluta	trompeur - mudunsitsi
habitant - muisi	treize - yigumi na birieru	tronc - muiri
habitation - yitsanunu	treizième - ajimuyigumi na	trône - yitsiga yibufumu
habiter - utsana	birieru	
habitude - yifu	tremblant - uregama, usisana	
s'habituer - uguku	tremblement - disisana	
hache - diumbi	trembler - uregama, usisana	
hacher - uriaba, usaka		

The above examples show a break in the alphabetical ordering of the lemmas listed under the letters *h* and *t*. The prepositional phrase **à travers** follows the item **travailler** and precedes the item **traverser**. The transitive verbs **s'habiller**, **s'habituer** and **se tromper** are treated in the same way. They are listed in between other items starting with the letters *h* and *t* in combination with the reflexive forms *s'/se* that categorize them. Being presented in this way, they break the flow of the alphabetical ordering of the macrostructural elements. The following extracts from the above examples show the breaks in the alphabetical order:

habillement	travail	trompe
s'habiller	travailler	tromper
habit	travailleur	se tromper
[...]	à travers	tromperie
habitude	traverser	
s'habituer		
hache		

Secondly, not all the letters of the alphabet are treated in the dictionary. For the F–Y section, there are no entries under the letter *w*, while in the Y–F section, several letters are left untreated. The compilers do not give any explanation in the front matter or anywhere else in the dictionary for this omission. However, this omission can be explained by the fact that the dictionary was directed at the learning of the Bible. The Bible was the first source on which a corpus for this purpose was based.

Thirdly, all the Yipunu verbs are presented in their stem form. The verbal prefix *u* attached to them is not part of the macrostructural presentation. The advantage of this is that a balance in the number of lemmas under each letter is maintained. If the compilers had grouped the verbs together with their verbal prefixes, half of the dictionary would have been under the letter *u*. However, this approach could cause a certain confusion for users, because the compilers do not explain the treatment of verbs in the front matter. Within the present lexicographic treatment, they appear like homonyms while they are in fact different words not written or pronounced the same. By following the consistent pattern, a lexicographer or an experienced user would know that the verbs are presented according to the stem form. But the common user has to rely on his/her knowledge of the language to distinguish between the pairs.

The lexical items **belusa** and **belusa** are presented with no lexicographic indicators. The user can consider them as homographs because they are not pronounced the same, the one meaning "healed", the other "to heal". The main tone for the word meaning "healed" is high, while the one meaning "to heal" has a low tone. The same pattern is followed throughout the entire dictionary.

Example:

belusa/belusa	should in fact be	belusa [bélúsà] / ubelusa [úbélúsà]
doba/doba		doba [dòbà] / udoba [údòbà]

kanda/kanda	kanda [kāndà]/ukanda [úkāndà]
kuasa/kuasa	kuasa [kwàsà]/ukuasa [úkwàsà]
paka/paka	paka [pākà]/upaka [úpākà]

Fourthly, homonyms and/or homographs are given in the central list of the Yipunu section without any lexicographic conventions. Due to the high frequency of tones, there is a great occurrence of homographs in Yipunu. These homographs are treated under the letter to which they belong with no lexicographic indicators, subscripts or numbering. When it comes to homonyms, some of them are presented under a single lemma. The following are examples of homographs where the pronunciation clarifies the difference between them.

Example:

digondi [dìghòndì]	digondi [dìghòndì]
bukulu [búkùlù]	bukulu [búkùlù]
funda [fùndà]	funda [fùndà]
kala [kālà]	kala [kālà]

Fifthly, the choice of compound words listed as macrostructural elements seems to have been determined by the meaning of the headword.

Example:

laba - voir
laba goma - avoir peur
laba kari - être triste
laba múru - avoir mal à la tête
laba ubueji - être heureux
laba yisonyi - être embarrassé
labana - apparaître, arriver
labasana - rencontre quelqu'un
la díbandu - c'est pourquoi

The word **laba** means not only "to see" but also "to look". It has been chosen as headword for the following phrases:

laba goma	avoir peur (to be afraid; literally: to look afraid)
laba kari	être triste (to be sad; literally: to look sad)
laba ubuedji	être heureux (to be happy; literally: to look happy)

The same pattern can be seen in the following lexical items:

tabula	couper (to cut)
tabula bunsonsi	juger (to judge; literally: to divide or to solve the case)
tabula diambu	decider (to decide; literally: to divide or to solve the case)
tabula kiga	fausser un serment (to give false testimony)
musamu	nouvelles, message (news)
musamu uboti	Évangile (Gospel; literally: good news)

mfula	pluie (rain)
mfula mamanyi	grêle (ice rain; literally: rock rain)
mfula jivema	neige (snow; literally: white rain)
mfula unoga	pleuvoir (to rain)

2.3.2 The microstructure

The only microstructural elements found in the DFY/YF are translation equivalents. They are listed in a row without illustrative examples. This pattern is consistently used throughout the dictionary. The different translation equivalents are separated by a comma. Here again, the lexicographic treatment of the translation equivalents is not as effective as it should be. By giving no indications about synonyms and/or polysemous senses, the user is left to guess with which type of information he/she is confronted. It is a hindrance to the access structure, making the consultation process very difficult. Some polysemous senses and meaning extensions are even listed as macrostructural elements. This results in complicating the consultation process.

2.4 The reversibility of the dictionary

In bilingual lexicography, the reversibility of a bidirectional dictionary is one of the vital components. It is important that the dictionary should be produced in a well-balanced way. The two sections cannot be equal because of the linguistic differences between the two languages, but the discrepancy should not exceed 60%. But, as has been noted, in the DFY/YF, the F–Y section contains 74% of the lemmas while the other 26% is found in the Y–F section. It can be seen that reversibility was not taken into account during the compilation process or the planning phase. Being a translation dictionary, the DFY/YF should have been expected to give a record in the Y–F section of all the lemmas treated in the F–Y section. In other words, all the lemmas treated in the F–Y section should also have been found in the Y–F section. But the fact is that of the 8 829 lemmas treated in the dictionary only 2 288 are found in the Y–F section, while the F–Y section has 6 541 lemmas. This means that the F–Y section contains more than twice the number of entries given in the Y–F section.

To illustrate this the letter *u* can be taken as an example. There are only seven lemmas under the letter *u* in the Y–F section. A look at the F–Y section provides far more than seven translation equivalents starting with *u*. Only on one page of the F–Y section there is a record of 19 translation equivalents starting with *u*. If the verbs (*u-* being the verbal prefix) are calculated, the user is left with five other lexical items starting with the letter *u*, namely **usagama**, **ureyimisa**, **uregma**, **udueji** and **unanguga**. These are almost half the number of lemmas found in the Y–F section under the letter *u*. More examples of lemmas starting with *u* can be found throughout the F–Y section. Why these words were omitted is unclear, for no reason is given in the front matter.

As far as the reversibility of the DFY/YF is concerned, another inconsistency can be noticed. If the letter *u* is again used as an example, it is noticed that the first lemmas in the Y–F section are presented as follows:

ubuedji - joie, heureux
uganyi - battement, palpitation
ulelama - bienséant, decent, proper
unangama - haut, élevé, grand

Looking at the F–Y section, the user should expect to find the above Yipunu lexical items as part of the translation equivalents. But this is not the case:

- there is no lemma **battement** in the F–Y section;
- the lemma **bienséant** has a completely different translation equivalent **uwelimina** which is the same equivalent found at the lemma **décent**;
- **élevé** is not found in the F–Y section while **élever** is included as main lemma;²
- **grand** has **uneni** as translation equivalent, a lexical item not listed in the Y–F section;
- **haut** has **julu** as translation equivalent; and
- **palpitation** and **propre** are not listed as part of the lemmas included in the F–Y section.

From all the examples mentioned, only **joie** and **heureux** meet the requirements of reversibility.

heureux - ubueji, uboti
joie - ubuedji, nsangu

However, **uboti** is not listed as part of the lemmas under the letter *u* in the Y–F section. Why did the compilers choose not to include **uneni**, **uwelimina** and **uboti** in the central list of the Y–F section? Are the translation equivalents presented as synonyms? Again the user is left to deduce which form to choose during the consultation process.

Mashamaite (2001: 114) mentions that "bi-directional refers to a condition whereby a bilingual dictionary is structured in such a manner that the speakers of both languages may use it for either encoding, decoding or any other purpose, including translating". And from the way the lexicographic treatment of the items is effected, it appears that the DFY/YF was intended to serve both French and Yipunu speakers in translation matters. The encoding part of the consultation process was not meant to be fulfilled. With the wide range of users the compilers have left the dictionary open. It is difficult to provide good and relevant translation information. Regardless of the different sections, the lexicographic treatment of each lemma presents a series of equivalents without any

further information. The user is left with the choice of deciding which translation would satisfy his/her needs. Neither the French user nor the Yipunu user can make such a choice when it comes to synonyms if there are no illustrative examples to direct the decision.

In bilingual lexicography, it is difficult to find equal synonymy because of the differences between the treated languages, and the cultural environment in which they are spoken. All the synonyms for a particular word in a specific language cannot be used equally. Some synonyms will better suit a particular context while they will be less appropriate in another. If the lemma **joie** from the F–Y section is taken as an example, the translation equivalents are **ubuedji** and **nsangu**. In the Y–F section, **joie** along with **heureux** is listed as translation equivalents of **ubuedji**. How can the user know in which context to employ **joie** and in which **heureux** when he/she comes across the word **ubuedji**? It is impossible. The compilers rely too much on the user's knowledge. In this way the DFY/YF fails in satisfying the principle of user-friendliness because it does not properly assist the user in finding the needed information. When the information is found, the dictionary does not provide assistance in interpreting this information, nor in applying this information in a specific usage context.

3. A new edition of the DFY/YF

3.1 The sources

The 1966 edition of the DFY/YF was compiled relying on informants. The compilers were English-speaking missionaries who learnt both treated languages as foreign languages. This is stated in the single page constituting the front matter.³ To build a corpus, the compilers of the dictionary sought help from the Bapunu people among whom they lived. It can be deduced that those who helped in the compilation process were Bapunu members of the congregation. Thus for a new edition to be compiled, the use of a well-devised corpus would be the proper policy to adopt. It would help the compilers in the choice of the data to include as macro- and microstructural elements. The Bible being the first written source of data, the compilers would have to build a significant and relevant corpus that will have to be tested among a wide range of Bapunu speakers. As a start, the compilers could use the knowledge of the informants who helped in the compilation of the first edition. In the meantime, Bapunu mother-tongue speakers will check the data collected in order to obtain a good frequency list. This list will be used as a tool for the inclusion of macrostructural elements. Furthermore, the inclusion policy based on frequency occurrences in the corpus has the advantage of allowing an easy way to refine the selected amount of data. Elements that meet the frequency requirement will thus be included as macro- and microstructural elements. It would be a challenge for the compilers to extract the necessary data from the corpus, but advice from linguists and lexicographers could be used.

3.2 The corpus

The first step for the establishment of a reliable corpus is data collection. A corpus should be representative of the oral repertoire of the lexicon of a language, as well as of the written texts found in that language. The mother-tongue speaker of the treated language(s) must master these repertoires. It should also give a record of the characteristics of the lexical items of that lexicon. As a result, it should not be selective, but objective, for it could serve as basis for the compilation of bigger corpora. The corpus for the compilation of a new edition of the DFY/YF should provide more data than the initial purpose for which it will be built.

In the case of Yipunu, using data from existing linguistic and lexicographic works is necessary. French has a long and well-established lexicographic tradition. No corpus needs to be compiled because corpora already exist, and using existing general French dictionaries would save a considerable amount of time. Thus the compilers will not have to compile two different corpora for the two treated languages.

Although small and specific, with a concentration on Biblical and religious terms, the corpus built for the purpose of the DFY/YF should be relevant and accurate with a scope to expand the range of the contents of the dictionary. It is important that a reliable corpus for the DFY/YF should reflect a true synchronic picture of the Yipunu lexicon. It should also give a fair survey of the lexicological status of the lexical items included as lemmas in the dictionary. On the other hand, the compilation of any corpus does not mean that all the data contained in the corpus has to be included in the dictionary. The typological nature of the specific dictionary, its volume and its content will mostly determine the amount of data to be included. The establishment of a corpus for a new edition of the DFY/YF is no exception. This type of dictionary belongs to the category of LSP dictionaries. As such, it has a limited scope. It will contain all Biblical terms as well as terms related to church activities. Newly entered words should be taken into consideration.

The compilers of the DFY/YF have the benefit of having access to advanced technology in dictionary making. Some of these new kinds of technology are corpus queries. These computer programmes are increasingly used in modern lexicography. They are less time-consuming and some of them are useful in providing frequency counts. They can run several frequency counts of words and phrases. With some programming input, detailed data can be extracted from the corpus. Thus the compilers are able to make correct decisions on inclusion procedures. The use of a corpus query like WordSmith, for example, could be of great help in the compilation process of a version of a dictionary such as the DFY/YF.

Another kind of computational technology that could be helpful in the compilation of a new edition of the DFY/YF is the use of the "ruler", a new lexicographic concept developed by Prinsloo and De Schryver (2002, 2004) consisting of the creation of an alphabetic scale that limits the amount of work

done during the data collection and the compilation phase. With this programme, lexicographers working on any dictionary project are able to devise a ruler showing roughly the amount of data each letter of the alphabet of the described language should receive. It is thus possible to know if a particular letter would require more data, and another one less. Whatever the language may be, and however the treated languages may differ, the use of a ruler can set a more realistic balance of letters regarding the right amount of incorporated data. By using such a tool, the compilers of a new version of the DFY/YF could avoid one of the mistakes made concerning the amount of data allocated to the letter *u* for example.

An additional way to enhance the corpus for the compilation of a new edition of the DFY/YF is by using existing Yipunu dictionaries. Although it should be a restricted corpus, considering existing lexicographic works will help with the inclusion policy in terms of the treatment of microstructural elements. The *Grammaire Pounoue et Lexique Pounou* (henceforth GPLP) is a good example of an existing reference work that can provide additional data. It is a monodirectional bilingual dictionary of Yipunu spoken in the Nyanga province, mainly in Tchibanga. The data that can be drawn from this dictionary concern the clear grammatical explanation given in the front matter, presenting a section on Yipunu morphology and phonology. The compiler even offers an alphabet of Yipunu, being the basis for the written form of the Yipunu lemmas contained in the central list of the dictionary.

The GPLP was compiled ten years before the old version of the DFY/YF. A look at this dictionary which provides good and relevant data, would have been of great help to the compilers of the DFY/YF. Thus, the lexicographic treatment would not only have been based on the knowledge of informants. Furthermore, to avoid the lack of Yipunu entries under the missing letters as in the first edition, the compilers can refer to research conducted on the language. Although based on the Latin alphabet, some attempts to provide a well-established alphabet for the Gabonese languages were made. The compilers of the new version of the DFY/YF can use one of the existing alphabets. In 1999 the Raponda-Walker Foundation revised Raponda-Walker's alphabet composed in 1932, consisting of 38 letters. It was reduced to 33 letters. This is used by the Foundation to produce schoolbooks (in the Rapidolangue series) in five of the most spoken Gabonese languages. The alphabet⁴ used in the Rapidolangue series consists of the following letters and letter combinations: the vowels a, e, e [ɛ], ë [ə], i, o, o [ɔ], u, u [y], and the consonants b, d, dy, dj, f, g, h, j, k, l, m, mb, mp, n, nd, ng, ny or ng [ŋ], p, r, s, t, v, w, z.

In the same year, the Orthography for Gabonese Languages (Orthographie des Langues Gabonaises, OLG) was compiled. The OLG has 35 letters and letter combinations written as follows: the vowels a, e, e [ɛ], ə, i, o, o [ɔ], u, u [y], and the consonants b, c [tʃ], d, d [ð], f, g, gh [ʏ], h, j, jh [ʒ], k, l, ml, n, ny [ɲ], n [ŋ], p, r, s, sh [ʃ], t, v, vh [β], w, y, z. The alphabet makes provision for the length of vowels by doubling any of them, if this is necessary in a particular language.

Chuwa (1996: 327) states that, when "a lexicographer depends only on informants and on his/her intuition, it leads to a poor choice of entry words". As mentioned in the front matter of the first edition, the DFY/YF is supposed to reflect the Yipunu language ("langue des Bapounou"). The variety of Yipunu described in this version does not represent the entire Yipunu lexicon because the Yipunu Bible the compilers used was written in the specific variety of Yipunu spoken in Mouila. For a new edition, the compilers should test the lemma candidate list in other areas where Yipunu is spoken. In doing so, they should check that, where, in the first version, the informants produced a word starting with *y*, it should, in most cases, be an *i*.

Example:

yidwaru	idwaru
yikola	ikola
yilinga	ilinga
yibamba	ibamba
yibedu	ibedu
yibiku	ibuku
yibadangu	ibadangu
yikoru	ikoru
yikutu	ikutu

A consideration of earlier lexicographic work could be helpful and all varieties could be taken into account where there are major differences. To confirm that the letter *i* should have been part of the letters included in the dictionary, modern linguistic research on Yipunu can be referred to: Bonneau 1940, 1947, 1956, Mboumba 1985, and Kwenzi-Mikala 1990, 1998, to name but a few.

3.4 The frame structure of the dictionary

As has been said in the previous pages, the frame structure of the actual DFY/YF does not have a proper outer texts section. Its front matter only consists of one page of non-integrated text containing the preface of the dictionary. The frame structure of the dictionary thus only refers to the central lists where the different languages are treated. However, in lexicography, the use of outer texts, that is the front and back matter, is a vital tool for the lexicographer to subscribe to the requirements of user-friendliness. The compilers of the DFY/YF in many ways fail in fulfilling these principles. In the following paragraphs, a different pattern is proposed that will allow a new edition to enhance the content of the dictionary as well as to facilitate the consultation process.

The DFY/YF being an LSP limited-sized dictionary, there would be no need for the compilers to design an inner structure such as the one found in some bidirectional bilingual dictionaries. Only the front and back matter could in this case fulfil the role of such a section in a new edition of the dictionary.

3.5.1 The front matter

An effective way to divide the data in the dictionary would be to incorporate some of the data into outer texts. This approach is effective where it allows the compiler(s) to include a larger amount of data into the dictionary; for example, the use of an expanded mini-grammar in the front matter, where rule formulations and examples of word-formation procedures could help the compiler(s) to incorporate morphological information on the treated language(s). In the case of the DFY/YF, morphological information on Yipunu is needed for a better understanding of the grammar of the language. Such a section could be used to connect information in the outer texts with information in the central list(s) of the dictionary. Thus by means of cross-referencing between the outer texts and the central list, the dictionary actually provides an unambiguous selection of information. This can be illustrated by the treatment of the nominal classes. Because they are important in the system of Yipunu, they should be presented directly after the lemma sign in the Y-F section. Information on nominal classes could be further emphasised in the mini-grammar. In this regard Louw (2004: 53) says: "The consistent implementation of such text-external cross-references will be imperative to the success of the dictionary as an interactive communicative ... aid."

Furthermore, the access structure of the dictionary could be improved by means of its front matter. Explanations on how to use the dictionary as well as on how to retrieve the needed information should be given within this section. The treatment concerning the verbs could be a good example of the complementary relation existing between the front matter and the central list. The effectiveness of the approach used in the first edition could be enhanced if the compilers could also explain the way simple compound words are used.

The front matter of the dictionary could contain a table of contents helping the user in finding and retrieving the needed information. According to Gouws (2001: 105), such a table of contents can be of great help in the access structure of any dictionary.

The dictionary should also contain a Yipunu pronunciation and grammar outline. The major role of this section should be to explain the method used for the lexicographic treatment in each entry. This should not only concern Yipunu but the compilers should also leave room for French. It may be helpful for Yipunu mother-tongue speakers with less knowledge of French. As the compilers of the first edition of the DFY/YF were non mother-tongue speakers of both treated languages, these issues should be dealt with in the front matter of the dictionary. The front matter should give a record of Yipunu ensuring a better transfer of linguistic information. Thus it should show the difference between the two treated languages. Except for the mini-grammar, the compilers of the DFY/YF should present a detailed alphabetical list of the speech sounds of Yipunu along with their pronunciation. The same should be done for French to show the difference between the two languages. A clear explanation and sample of the way entries are presented within the central list(s) should also be provided in the front matter.

3.5.2 The back matter

The compilers of the DFY/YF should add back matter within the frame structure of their dictionary. Should a new edition be compiled, the back matter could be used to incorporate data on Hebrew culture in Biblical times. Here the names of Biblical characters and Jewish customs and ceremonies could also be included and translated into Yipunu. The information contained in the glossary presented at the back of the 1992 version of the Yipunu Bible could furthermore be given in the back matter.

3.6.1 The macrostructure

Although some adjustments should be made, all the lemmas contained in the first edition of the DFY/YF should be included as macrostructural elements in a new edition. The change will appear in the orthography. The compilers should make use of the Orthography for Gabonese Languages (Orthographie des Langues Gabonaises, OLG) for the written transcription of the lemmas. With the establishment of a well-designed corpus, the new edition should give a fair record of the lexicon of the treated languages, and more particularly of the Yipunu lexicon. Moreover, the missing letters could receive a new lexicographic treatment and the discrepancy between the amounts of data could be reduced.

In the Y-F section, homonyms and homographs should be properly treated allowing the user to make the distinctions between the two categories. One way could be to treat homonyms and homographs similarly with numeral indicators. The fact that homographs have a greater occurrence in Yipunu will not affect the lexicographic treatment. The compilers should only ensure that the user grasps the difference between the two words.

Example:

Old version	New version
digondi: régime de banane	dighondi ¹
digondi: charnière, gond	dighondi ²
bukulu: oseille	bukulu ¹
bukulu: généalogie, chronique	bukulu ²
kala: crabe	kalə ¹
kala: canere	kalə ²
kala: passé	kalə ³

It is also advisable for the compilers to use indicators for the treatment of verbs. The verb form should be preceded by a hyphen as indication for the user that he/she is dealing with a verb and not a noun. This approach will have the advantage of helping the user in such a way that he/she will not confuse homonyms and what can appear to be homonyms because of the presentation.

Example:

Old version	New version
kanda (na): exprès, à dessein	kanda (na)
kanda: bloquer, enfermer	-kanda
paka: Pâques	pakə
paka: tailler, graver, scruter	-pakə

3.6.2 The microstructure

As far as the microstructural elements are concerned, the translation equivalents should be given together with illustrative examples. In this way, it could guide the user in the choice of the right word for the right usage context. Therefore, all the synonyms should be treated according to their senses. The use of lexicographic conventions such as bold, italics and numbering should be applied in order to help the user access the needed information. These types of markers should also help the compilers in avoiding the use of any confusing pattern.

To treat a fair amount of the Yipunu lexicon, the compilers should include other varieties as well as the pronunciation and the nominal class of the nouns. The microstructure of the dictionary should also name the parts of speech in order to provide the right lexicological status to the data given in the central lists. Because of its typology, there is no need for definitions in the DFY/YF.

4. Conclusion

The benefit of already having an existing lexicographic work is that the lexicographer will not have to start from the very beginning. There are great possibilities in extending and improving the existing work by using a proper lexicographic methodology. A dictionary such as the DFY/YF provides a good basis for a theoretically sound work. However, for a dictionary to be used, the lexicographer should combine theory and practice in order to respond to the need of the target users. Although practical work provides good data information, the relevancy of that information is made possible by a theory adjusted to a particular context. By mixing theory and practice, the compilers of a new version of the DFY/YF will give the Bapunu a dictionary that can meet their needs. In the meantime, it can help in promoting, developing and standardizing Yipunu.

Notes

1. The first dictionary *Dictionnaire français-pongoué/pongoué-français* was compiled in 1847 by Msr J.-R. Bésieux.
2. **Élevé** is the past participle of the verb **élever** which is included in its infinitive form. The motivation behind the choice of one form to the other is unclear.

3. Although the Bapunu people helped in the compilation of the dictionary, it is possible that it contains errors because the missionaries who compiled it were foreigners.
4. The Rapidolangue alphabet presented here is part of the history of the development of the Gabonese languages examined by Idiata (2002).

Bibliography

- Bonneau, J.** 1940. Grammaire Pounoue. *Journal de la Société des Africanistes* 10: 131-161.
- Bonneau, J.** 1947. Grammaire Pounoue. Sequel. *Journal de la Société des Africanistes* 17: 23-50.
- Bonneau, J.** 1956. *Grammaire pounoue et lexique pounoue-français*. Brazzaville: Institut des Études Centrafricaines (A.E.F.).
- Chuwa A.R.** 1996. Problems in Swahili Lexicography. *Lexikos* 6: 323-329.
- Église Évangélique du Sud-Gabon (CMA).** 1966. *Dictionnaire français-yipounou/yipounou-français*. Mouila.
- Gouws, R.H.** 2001. Lexicographic Training: Approaches and Topics. Emejulu, J.D. (Ed.). 2001. *Éléments de lexicographie gabonaise. Tome I*: 58-94. New York: Jimacs-Hillman.
- Idiata, D.F.** 2002. *Il était une fois les langues gabonaises*. Libreville: Éditions Raponda-Walker.
- Kwenzi-Mikala, J.T.** 1990. Quel avenir pour les langues gabonaises? *Revue Gabonaise des Sciences de l'Homme* 2: 121-124.
- Kwenzi-Mikala, J.T.** 1998. Parlers du Gabon: classification du 11.12.97. Raponda-Walker, A. (Ed.). 1998. *Les langues du Gabon*: 217. Libreville: Éditions Raponda-Walker.
- Louw P.A.** 2004. Criteria for a Multifunctional, Monolingual Dictionary in Junior Secondary Education. Unpublished Doctoral Thesis. Stellenbosch: University of Stellenbosch.
- Mabika Mbokou, L.** 2002. Lexicographie, dictionnaire et usagers. Emejulu, J.D. (Ed.). 2001. *Éléments de lexicographie gabonaise. Tome I*: 36-53. New York: Jimacs-Hillman.
- Mashamaite, K.J.** 2001. The Compilation of Bilingual Dictionaries between African Languages in South Africa: The Case of Northern Sotho and Tshivenda. *Lexikos* 11: 112-121.
- Mboumba, F.** 1985. *Dictionnaire Français-Pounou*. Unpublished Manuscript.
- 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.
- Prinsloo, D.J. and G.-M. de Schryver.** 2002. Designing a Measurement Instrument for the Relative Length of Alphabetical Stretches in Dictionaries, with Special Reference to Afrikaans and English. Braasch, A. and A. and C. Povlsen (Eds.). 2002. *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002, Copenhagen, Denmark, August 13-17, 2002*: 483-494. Copenhagen: Center for Sprogteknologi, University of Copenhagen.
- Prinsloo, D.J. and G.-M. de Schryver.** 2004. Crafting a Multidimensional Ruler for the Compilation of Sesotho sa Leboa Dictionaries. Mojalefa, J. (Ed.). 2004. *Rabadia Ratšhatšha: In-depth Literature, Linguistics, Translation and Lexicography Studies in African Languages. Festschrift in Honour of P.S. Groenewald*. Pretoria: J.L. van Schaik.

A Trilingual Dictionary Yilumbu– French–English: An Ongoing Project

Paul Achille Mavoungou, *Department of Afrikaans and Dutch,
University of Stellenbosch, Stellenbosch, Republic of South Africa and
Department of Linguistics, Omar Bongo University, Libreville, Gabon
(moudika2@yahoo.fr)*

Abstract: In this article, an account is given of the planning of a trilingual dictionary Yilumbu–French–English. The focus is on the target user, the purpose, nature and typology of the planned dictionary. Attention is also paid to some macro- and microstructural issues. For example, all types of lexical items, including multiword lexical items, are given lemma status. Moreover all items are included according to the word tradition and on account of their usage frequency in the corpus. Apart from these aspects, types of dialectal forms as well as the type of special-field lexical items are also discussed. From a microstructural point of view, this article investigates different kinds of data types to be considered for inclusion in complex articles in particular. User-friendliness parameters and innovative access structure procedures also come into play.

Keywords: DICTIONARY, LEXICOGRAPHY, DICTIONARY PLAN, METALEXICOGRAPHY, TARGET GROUP, GABON, SOURCE LANGUAGE, TARGET LANGUAGE, ENGLISH, FRENCH, YILUMBU

Résumé: **Un dictionnaire trilingue yilumbu–français–anglais: projet en cours.** Cet article rend compte de la planification d'un dictionnaire trilingue yilumbu–français–anglais. Le centre d'intérêt réside au niveau du public cible, l'objectif, la nature et la typologie du dictionnaire proposé. Une attention est aussi accordée à quelques problèmes macro- et microstructuraux. Par exemple, tous les types d'items lexicaux, y compris les items lexicaux formés de plusieurs mots, reçoivent le statut de lemme. En outre, tous les termes sont inclus selon la tradition du mot et sur la base de leur fréquence d'emploi dans le corpus. Hormis ces aspects, les types de formes dialectales ainsi que le type d'items lexicaux de spécialité sont également discutés. D'un point de vue microstructurel, cet article explore différents types de données à considérer pour inclusion dans les articles complexes en particulier. Les paramètres de clarté et de lisibilité ainsi que des procédés propre à la structure d'accès sont également pris en compte.

Mots-clés: DICTIONNAIRE, LEXICOGRAPHIE, PLAN DE DICTIONNAIRE, MÉTALEXICOGRAPHIE, PUBLIC CIBLE, GABON, LANGUE SOURCE, LANGUE CIBLE, ANGLAIS, FRANÇAIS, YILUMBU

Introduction

Yilumbu¹ is a developing Bantu language² spoken by about 12 000 to 20 000 people (Grimes 1996) in Gabon, in the Republic of the Congo, and in the Democratic Republic of the Congo (DRC). Up to the present, French remains the only official language³ in Gabon. Therefore, most Balumbu speakers also use French as their lingua franca. The literacy rate ranges from 70% to 77%. However, French is not the only language that is studied in Gabonese schools, colleges and universities. For example, Italian, German, Spanish, Portuguese, Arabic and especially English are school subjects and they are usually taught as optional second or third languages. Moreover, some Balumbu also speak neighbouring languages such as Yipunu and Civili, to name but two. This article focuses on the Yilumbu variety spoken in Gabon, particularly in three towns, namely Mayumba, Gamba and Setté-Cama.

Linguistically, Yilumbu (B44 Guthrie 1953) is part of the Sira-Punu group (B40) of the Bantu linguistic branch, which includes Ghisira⁴ (B41), Yisangu (B42) and Yipunu (B43) that are all mutually relatively intelligible. Many Yilumbu speakers also speak these languages. There are two main varieties of Yilumbu, Yilumbu yí menaáne (Yilumbu of 'I say that'), which shares some features with Yipunu and Yilumbu yi ghângu (Yilumbu of 'fry'),⁵ which shares features with Civili (see Blanchon 1984 and Mavoungou 2002b). The language has been the subject of academic enquiry since the late nineteenth century, including Garnier (1897, 1900 and 1904) (based on the dialect spoken in the Nyanga province and the Banio lagoon) and Murard (1903, 1903a) (based on the dialect spoken in the Ogooue-Maritime province and in the cantons of Ndugu and Basse Nyanga). A brief outline of the Yilumbu language is given by Emejulu and Pambo-Loueya (1990).

Yilumbu has been written since 1900, although the orthography has changed since then. Yilumbu orthography is in accordance with the *Rapport Final de la Session de Concertation sur l'Orthographe des Langues Gabonaises* (1999). However, the latter is not specific with regard to a number of important orthographical issues such as word division, stem tradition versus word tradition, and writing of variants. These aspects will receive some attention in the present article. The Raponda Walker Foundation⁶ is currently conducting research with a view to promoting the introduction of local Gabonese languages into the educational system. At present some Gabonese languages (including Yilumbu, Yipunu and Civili) are being taught in a number of high schools in Libreville and elsewhere thanks to the support of a relatively well-elaborated textbook called *Rapidolangue* whose writing system combines the alphabet of Raponda-Walker (1932) and the set of symbols from the *Rapport Final de la Session de Concertation sur l'Orthographe des Langues Gabonaises* (1999). According to the research currently being conducted (and started in 1995) by the Raponda Walker Foundation, very little is known about how many people can, for example, actually read and write Yilumbu, Yipunu or Civili. A significant number of Yilumbu speakers are literate in Civili and Yipunu, whose spelling systems resemble that of Yilumbu. At present, there are no books, children's books and

newspapers written in Yilumbu. The Republic of Gabon attaches great importance to the development of local Gabonese languages as well as their introduction into the education system⁷. However, there is no dictionary of Yilumbu. This hinders the introduction of the language into schools. Contrariwise, some lexicographic reference works do exist in Yipunu (Bonneau 1956, CMA⁸ 1966 and Rittaud-Hutinet 1980, to list but a few) and in Civili (Derouet n.d. and Marichelle 1902).

1. Project background

According to Wiegand (1998), lexicographic processes refer to all the activities leading to the compilation and publication of a specific dictionary. Any undertaking must therefore begin with a realistic assessment of the pragmatic aspects inherent to that particular project. The current dictionary project can be traced back to a seminar organized by the Ministry of National Education, namely Les États généraux de l'Éducation et de la Formation that took place from 12 to 23 December 1983 in Libreville. One of the recommendations of the workshop was the introduction of Gabonese languages into the national educational system. Another important recommendation was the necessity to provide the Gabonese educational system with manuals, textbooks and good quality dictionaries in all the local languages of the country. These two recommendations were stressed again during the Table ronde sur les recherches linguistiques et l'Enseignement des langues au Gabon that was held from 9 to 11 December 1997 as well as during the Séminaire sur la standardisation de l'orthographe des langues gabonaises that took place in April 1999, also organized by the Ministry of National Education.

In accordance with the terms of these recommendations, three field trips to the Yilumbu-speaking areas in Gabon were successively undertaken during June–September 1995, 1996 and 1997 to obtain data.

Research at grass-roots level has shown that within the Balumbu speech community people urgently need various manuals and textbooks. But most importantly, the lack of a single dictionary could lead to a situation of frustration. It is generally accepted that the existence of dictionaries within a given speech community is a matter of great pride because dictionaries empower. Not surprisingly, the project has received good support from the Balumbu speech community including local administrative authorities (a.o. the mayor of Mayumba and members of the Mayumba Town Council) and leading personalities originating from Mayumba, Gamba and Setté-Cama (ministers, their advisors and collaborators, Members of Parliament as well as businessmen).

2. On the choice of French and English as co-target languages

The choice of French as one of the treated languages of the planned dictionary can easily be taken for granted because of the fact that up to the present French

remains the only official language of Gabon. Given the relatively poor representation of English in the linguistic situation in and the language policy of Gabon, the value of English as the co-target language of the planned dictionary together with French may raise some questions. As a matter of fact, English has a low recognition and is mainly used for communication with foreigners. As far as education is concerned, English has been integrated into high school curricula and is consequently taught as first foreign language. Taking the international standing of English into consideration, Busane (1990: 32) emphasizes that, "with the increasing spread of English in Central Africa, it appears that French will have to compete not only with the African languages but also with English". This is a point of great importance, and a lexicographer would do well to heed this observation of Busane.

Busane's statement does not only have implications for the L2 speakers but also for the L1 speakers of the planned dictionary. Although target language data is primarily directed at L2 speakers who wish to have access to Yilumbu it may be a bonus for L1 speakers who can empower themselves in French, the official language, and English whose international status is important, especially nowadays that multilingual education is the key to globalization.

3. The target users of the trilingual dictionary Yilumbu–French–English and dictionary use

3.1 Target users

In modern-day lexicography, more and more attention is given to the concept of "dictionary awareness" (Hartmann 1999). Dictionaries are produced so that they can be used by the speech communities at which they are directed. A number of metalexicographers such as Hartmann (1989) and Kromann et al. (1991) have correctly emphasized that the assessment of users' needs should always precede dictionary design. The target users of the planned dictionary range, on the one hand, from senior high school pupils to academics who have Yilumbu as first language and a relatively good command of French or English, and, on the other hand, pupils and scholars who want to improve or learn Yilumbu as a second language (Mavoungou 2002a: 185).

3.2 Dictionary use

Reference skills are "the abilities required on the part of the dictionary user to find the information being sought" (Hartmann and James 1998: 117). As far as the target user groups of the planned dictionary are concerned, it would not be expected that senior high school pupils and academics have much in common with respect to dictionary use. As a matter of fact, up to the present the teaching of dictionary skills does not form part of the curricula at school level.

Therefore, for most high school pupils of the current dictionary project, it can be assumed that they are not equipped with the necessary dictionary using skills. However, a very few high school pupils may have some knowledge of dictionary use. These pupils may have received informal instruction in dictionary use either from friends or relatives at home, but not in school from a teacher.

As far as academics are concerned, they do have some kind of dictionary using skills, being rudimentary reference skills acquired through informal learning and individual experience. Although it is impossible to satisfy the needs of everyone, it becomes clear from the foregoing that it is the lexicographer's responsibility to find ways of accommodating users. The users' often rudimentary reference skills (on the side of academics) or their absence (on the part of pupils) will have a definite influence on the dictionary conceptualization plan. In other words, dictionary data should be presented as explicitly as possible in order to take into account the needs and reference skills of both experienced and inexperienced users.

3.3 The users' socio-cultural background

The potential users of the proposed dictionary represent a culturally and linguistically heterogeneous community. In fact, from a sociolinguistic point of view Balumbu, Bavili and Bapunu, to name a few, often intermarry. As a result, some prospective Balumbu users of the planned dictionary will approach it with different mixes of languages, cultural backgrounds, world views, linguistic competence, etc. In the Mayumba area in particular, polyglossia is the norm rather than the exception. Quite a significant number of Balumbu speakers are literate in Civili and Yipunu, among others, and vice versa. Lexicographers of a Yilumbu dictionary will not have major difficulties incorporating the sociolinguistic aspects of people who have Yilumbu as first language or even those who are literate in more than two or three Gabonese languages (e.g. Civili, Yipunu or even Omyene) since they all share the same historical and linguistic background. Besides native speakers of Yilumbu who might need to use the dictionary, it might also be used by speakers of French or English (living in Gabon and abroad). These two potential user groups will obviously have different socio-cultural backgrounds. In fact, the prospective user living in Gabon and the prime target group (senior high school pupils and academics) will have the same or much the same socio-cultural background. With the users living abroad, the socio-cultural background (linguistic system, world view, etc.) may be different. As far as this particular area is concerned, this will require some attention on the part of the lexicographer. In other words, to deal with this range of backgrounds, a lexicographer must adapt his/her methods and presentation to the sociolinguistic requirements of the potential users. Whatever the decision may be, prospective users must be determined realistically. In fact, a lexicographer cannot satisfy all users' sociolinguistic differences on an equal

basis. The middle way for the lexicographer is to plan his/her dictionary in accordance with the sociolinguistic profile of the most typical target users.

4. The purpose, functions, nature and typology of the planned dictionary

Any dictionary project should clearly identify its genuine purpose (cf. Wiegand 1999: 299) prior to the compilation phase. At this stage the genuine purpose of the planned dictionary can be described as follows: On account of the encoding and decoding functions, the planned dictionary should assist native and foreign-language speakers on an equal basis in retrieving relevant information as quickly as possible. More specifically, the planned dictionary should fulfil two main functions, namely: (a) a dictionary for the Balumbu helping them to understand Yilumbu texts, and (b) a dictionary for the Balumbu helping them to produce Yilumbu texts. Apart from these primary functions, the planned dictionary may also fulfil three other functions, namely: (c) a dictionary for the Balumbu translating from English and French into Yilumbu, (d) a dictionary for French translating from Yilumbu into French, and (e) a dictionary for English translating from Yilumbu into English. The dictionary should fulfil these last two functions (from which it will earn its poly-accessible character) because it will have two separate alphabetical registers, one focusing on the French language and one focusing on the English language. Moreover, the planned dictionary will be a typological hybrid (cf. Gouws 1999: 39) because it will have features of both translation and monolingual dictionaries. As far as monolingual features are concerned, the planned dictionary will provide users with a brief paraphrase of the meaning of the lemma sign in the source language, Yilumbu. With regard to translation features, the planned dictionary will present translation equivalents for lemmata where available (Mavoungou 2002a).

5. Scope, direction and function of the planned dictionary

In metalexigraphic discourse, the concept of mono-directional is commonly used in the sense of monoscopal. In other words, a mono-directional dictionary is a publication including only one section dealing with the source language and the target languages, i.e. from A to B. In the same way, bi-directional dictionaries are publications encompassing both directions, i.e. from A to B and from B to A. This should, however, not be confused with what Hausmann and Werner (1991) regard as mono-directional/bi-directional and monoscopal/biscopeal. According to Hausmann and Werner (1991: 274), a mono-directional dictionary is a publication directed at the speakers of one of the treated languages, whereas a bi-directional dictionary is meant to serve the speakers of both the treated languages. As far as the planned dictionary is concerned, Hausmann and Werner's terminology is adhered to. In other words, the dictionary under discussion will be monoscopal in that it is only conceived for the

following language direction A>B, namely Yilumbu>French–English. The choice of this language direction has many implications in terms of users' needs. As already stated, the current dictionary project is meant in the first place for people who want to have access to the Yilumbu word, its spelling, its meaning, etc. The westernization of communities and the change in lifestyle have already been responsible for a serious decrease in the proficiency of mother-tongue speakers of Gabonese languages. This is why Yilumbu macrostructural elements of the planned dictionary will act as a reference for younger Yilumbu speakers (or learners of Yilumbu) unsure of the meaning of more difficult words.

In addition to the foregoing, the major problem with monoscopal reference works is that only source language forms are lemmatized in the central list. This shortcoming can be remedied by making provision for the inclusion of back matter texts, e.g. an alphabetical register, giving the dictionary a poly-accessible character. Moreover, it will be poly-directional in that it will be directed at the speakers of both the source language and the target languages. As already mentioned, it will earn its poly-accessible character through the inclusion of an additional outer access structure: the alphabetical equivalent register in both French and English as back matter text. Practically every translation equivalent given in the central list will be lemmatized in the register section. This can be done by means of various computer programmes, e.g. the Omkeerbare Bilinguale Lexicale Databanken (OMBI), the Reversible Bilingual Lexical Databases, which is a language editor developed by the Dutch software house Software Engineering Research Centre (SERC) in Utrecht in the Netherlands (Martin 1996). By applying the reversibility principle, OMBI is able to create a secondary macrostructure by merely giving each translation equivalent in the central list a lemma status in the register section. Practically, while building the Yilumbu corpus, a computer programme can reverse all the English and French translation equivalents and the reversed counterparts that are automatically created and stored constitute a second database. The alphabetical registers mentioned above are linked to the concept of the dictionary as text type carrier. In fact, a dictionary may contain different text types, e.g. front matter texts, central list and back matter texts. In accordance with the data distribution structure and in order to add to the poly-accessibility of the dictionary it is part of its editorial policy to include an alphabetical register as one of the back matter texts in the dictionary. As already mentioned, it is difficult to treat both the source language and target languages on an equal basis but provision for the inclusion of back matter registers in both French and English gives the lexicographer the opportunity to treat the three languages of the dictionary in more or less the same way. During the dictionary conceptualization phase it is also important to determine the extent of the utilization of the register section. For example, the back matter could merely contain a text with a list of all the items, which are translation equivalents in the articles of the central list, plus a cross-reference indicating their respective lemmata. Apart from the lexical items

that are translation equivalents in the articles of the central list, the lexicographer could also propose a set of words and phrases used in international discourse, such as *globalization*, *International Monetary Fund*, *information technology*, etc., together with their meanings. This is relevant especially nowadays that the world is more and more referred to as a "global village".

6. Word categorization and metalanguage used in the planned dictionary

Given the fact that Yilumbu has been selected as the language of primary lemmatization in the planned dictionary, it might be correctly expected that the metalanguage for word categorisation would be given in Yilumbu (the source language). However, the specification of the part of speech to which a lemma belongs is a relatively simple exercise in English and in French but not in Yilumbu. The reason for this is that French and English have long and strong established lexicographic terminologies and a relatively good foundation of metalexigraphy. Unfortunately, word-class categorization is still a challenge in most African languages. Word categorization related issues have been tackled afresh by linguists and lexicographers working in Zulu and Sepedi in particular (cf. Taljard and Gauton 2001). As far as Yilumbu is concerned, it is part of the editorial policy to use the standardized sets of abbreviations available in English and French in combination with Yilumbu on the metalinguistic level. In other words, the sets of abbreviations *n.* (*noun/nom*), *v.* (*verb/verbe*), *adv.* (*adverb/adverbe*), etc. will be used to indicate the part of speech to which the lemma belongs. However, if French or English is chosen as the metalanguage (e.g. the language used for the texts of definitions as well as metalanguage notes in the form of labels) of the planned dictionary, this will be less informative for Yilumbu speakers than for their French and English counterparts whose language will be used predominantly to reflect on the meanings of lexical items as well as the culture underlying them. That is the reason why Yilumbu has been chosen as the metalanguage for the paraphrases of meaning, the so-called diatopic or regional labels such as *Ghâng.* (short for *Yilumbu yi ghângu*) and *Men.* (short for *Yilumbu yí menaáne*) and sphere of usage labels such as (*mangumba*, dance held at a funeral), and encyclopaedic information in complex articles (cf. Gouws 2002) addressed at mother-tongue speakers. With regard to the indication of the part of speech to which the lemma sign belongs, priority has been given to the standardized set of labels in both French and English simply because of the trilingual nature of the work which obviously entails space limitations. The major problem with abbreviations is that abbreviations like *n.*, *v.*, etc. have much less value for the average dictionary user than their unabbreviated equivalents *noun/nom*, *verb/verbe*, etc. In other words, when given implicitly in the central list, the use of a comprehensive system of grammatical coding where all the codes are explained in the users' guidelines has its disadvantages because the user has to turn to the front matter section to retrieve the information he/she is expecting to find in the alphabetical section of the dic-

tionary (Gouws 1993: 36). Ideally, the word class should be given as an unabbreviated entry and this will obviously result in greater explicitness. However, this chiefly depends on the number of pages the lexicographer can allow for his/her dictionary.

With regard to this point, it is part of the editorial policy not to convey word categorization in a too implicit way. In other words, priority will be given to part of speech markers that may be regarded as less problematic for the average person because it will be easier for him/her to reconstruct their full forms. Given the fact that part of speech labels (italicized words in parenthesis or written with small capitals) will encompass a lesser degree of textual condensation, some will be longer than others in order to achieve greater explicitness.

A further aspect related to the problem of the metalanguage used in the planned dictionary is the explanation of the structure of the dictionary. As dictionaries should ideally contribute towards the development of international metalexigraphic terminology, the outer texts explaining the structure of the dictionary should be given in the three languages of the dictionary.

7. The layout and format of the planned dictionary

7.1 The problem

The layout of the data on the dictionary page is a very important component of the dictionary conceptualization phase. Some dictionaries lack a clear indication of the boundaries of the different search zones in a dictionary article. For example, the definition is not clearly perceived by the user, and illustrative examples and idiomatic expressions are hardly ever discriminated from one another.

7.2 Decisions taken

7.2.1 With regard to the format

Instead of focusing on the way font sizes, styles and layouts of the different text examples will look like in the planned dictionary, this article will merely focus on the way data are presented, on the different principles underlying this presentation, and on the relative ease with which information can be found. Moreover, the data in question can best be presented in tables or frames. For a trilingual dictionary of the kind currently being planned, the ideal will be to have a final product combining dictionary and encyclopedia. In this regard, the format of various paper dictionaries, e.g. the *Greater Dictionary of Xhosa*, can be used. This dictionary, which is being compiled at the University of Fort Hare, is an explanatory trilingual dictionary with Xhosa, English and Afrikaans as treated

languages, of which Volumes 2 and 3 (K–P and Q–Z), published in 2004 and 1989 and edited by B.M. Mini et al. and H.W. Pahl et al. respectively, have so far been published. Within the central lists of these volumes the user will find Xhosa lemmata as well as their paraphrases of meaning and illustrative examples in the first column of each page. These data categories are then translated into English and Afrikaans in two further columns.

7.2.2 With regard to the use of structural indicators

In keeping with Wiegand's (1996) concept of micro-architecture, the black square symbol ■ will be used to mark the occurrence of groupings of sublemmata. On a similar basis another information type, the paraphrases of meaning, will be marked with a small diamond ♦. Co-text entries in the source language starting on a new line will always be introduced by the symbol Σ||. The symbols < > will be used to mark a specific search zone within a partial article stretch whereas the black upside down triangle ▼ will appear in complex articles, marking an article zone in which the user will be provided with an additional text or lexicographic description of an encyclopedic or extra-linguistic nature.

Moreover, these diamonds, upside down triangles and other structural indicators are determined by the data distribution structure of the dictionary. When dealing with the treatment of lemma signs representing polysemous lexical items, boldface numbers will introduce the different senses and these distinctions in meaning will be numbered in separate sequences after each part of speech. These polysemic senses should not be arranged on a random basis, but according to fixed criteria accounted for in the compulsory text containing the users' guidelines. Consider the lexicographic treatment of the article of the lemma **mudika** below:

mudika, mi (+ dika) [müdík... /midik...] (aussi/also mudiku) <i>n.</i> (cl. 3/4) < *- di♦ka			
1 ♦ Miluunda mi mwiba ♦ ⇒ mwiba; ndogha	<F>	1 Les fruits du manguier sauvage (Irvingia gabonensis). <T> Mangues sauvages.	<E> 1 The fruits of the wild mango tree (Irvingia gabonensis). <T> Wild mangoes.
Σ Mib'aba mitsibura na ilime! Baana batsyeboola mudika.	<F>	Ces manguiers sauvages là-bas ont vraiment produit cette année! Les enfants sont allés ramasser les mangues sauvages.	<E> These wild mango trees over there have really produced this year! The children went to pick up some wild mangoes.
2 Dipe (di mudika) vho mweembu bavalaamba na miluunda mi mwiba batsiletuula vho nyangu na ughaanga.	<F>	2 Pâte et sauce obtenue à partir des amandes de mangues sauvages pilées préalablement séchées et frites. <T> (frGab) Pain d' <i>odika</i> ou chocolat indigène; sauce à l' <i>odika</i> .	<E> 2 Paste and sauce obtained from the pounded kernels of wild mangoes previously dried and fried. <T> (frGab) <i>Dika</i> bread; <i>dika</i> sauce.

Σ	Maama atsilaamba mbitsi na mudika.	⟨F⟩	Maman a préparé de la viande à la sauce <i>odika</i> .	⟨E⟩	Mother has cooked some meat together with <i>dika</i> sauce.
<p>▼ Bavatebaanga muluunda mudika yo bavatela <i>ibilu mudika</i>. Mwiri uvayidisi midika agho bavatela <i>mwiba</i>. Mudika vaandi, <i>mweembu mudika</i>. Nana tununama na iwuma i suupu mudika, ndogha. Byootsu bina a tsiuupu tso tseyetu tsi bakaata balaamba. Byootsu bina a myeembu myo bakaata bo beetu balaamba vho teemu ughulu. Avha teemu ayi yetu avha ika tsitomaata. Bakaata basanayaaba tsitomaata. ⟨F⟩ On casse la noix de la mangue sauvage (pour en extraire l'amande). L'arbre qui produit les mangues sauvages s'appelle <i>mwiba</i> (le manguier sauvage). Le terme <i>mudika</i> renvoie également à la <i>sauce odika</i>. De nos jours, on utilise beaucoup plus le terme d'emprunt "soupe" d'<i>odika</i>, le <i>chocolat indigène</i>. Tous ces termes renvoient aux sauces que préparaient nos ancêtres. Quoique la sauce tomate n'ait pas fait partie des habitudes culinaires de nos ancêtres, aujourd'hui c'est beaucoup plus elle qui tend à être utilisée de façon prédominante (par la population locale). ⟨E⟩ The nut of the wild mango is cracked (to remove the kernel). The tree which produces the wild mangoes is called <i>mwiba</i> (the wild mango tree). The lexical item <i>mudika</i> also refers to the <i>dika</i> sauce. Nowadays the borrowed lexical item <i>dika</i> "soup", the <i>indigenous chocolate</i>, is more often used. All these items refer to sauces, which our ancestors used to prepare. Although tomato sauce did not form part of the culinary art of our ancestors, today it is predominantly used (by the native population).</p>					
<p>Quot/cit ⟨F⟩ D'après Raponda-Walker et Sillans (1961: 207-208), le bois (du manguier sauvage), difficile à travailler, mais très résistant, est utilisable pour la grosse charpente, à l'abri de la pluie. Les indigènes l'emploient pour la confection des pilons à mortiers ou auges. Les ménagères recueillent les amandes des fruits pour préparer le <i>pain d'odika</i> (mpongwè, galoa, nkomi, orungu), <i>ndòc</i> (fang), <i>mudika</i> (éshira, bavili, bapunu, masango, banzabi), d'un usage courant dans la cuisine gabonaise. La pulpe amère se mange aussi. Le chocolat indigène, réduit en poudre, peut s'appliquer sur les brûlures. Il intervient encore dans la composition de certains remèdes astringents. En outre, on consomme les râpures de l'écorce dans une grosse banane verte cuite sous la braise, pour enrayer la diarrhée ou la dysenterie; ou bien on la fait macérer pour s'en servir en lavement. À cause de sa fécondité, cette essence est considérée comme un porte-bonheur pour les plantations vivrières. Avec son écorce on fait des talismans pour donner ou rendre la fertilité au sol.</p> <p>⟨E⟩ According to Raponda-Walker and Sillans (1961: 207-208), the wood (of the wild mango tree), difficult to work but very strong, is used to construct frames for huts. The indigenous population uses it to make pestles and mortars or plates. Housewives collect the kernels of the fruits in order to prepare <i>dika bread</i> (mpongwè, galoa, nkomi, orungu), <i>ndòc</i> (fang), <i>mudika</i> (éshira, bavili, bapunu, masango, banzabi), commonly used in Gabonese cuisine. The bitter pulp (of the fruit) is also edible. The powdered kernels can be applied to burns. It is also used in the preparation of astringent medicines. Moreover, raspings of the bark together with a big green banana (plantain) baked in the embers are used as a remedy for diarrhoea or dysentery; or a decoction of the bark can also be used for enemas. Because of its prolificity, this tree species is believed to bring luck to alimentary plantations. Talismans for giving or restoring fertility to the soil are made from its bark.</p>					

Text example 1: Article mudika

In addition to the foregoing, lexical items with identical spelling but different etymologies (see Table 1) will be entered separately and will have superscript or raised numbers. There should be some guidelines or criteria motivating in which order these homonyms will be presented. The frequency of use of the homonyms in question may be one of the criteria of incorporation and this basis should ideally be reflected in the *guide*. Merely from the way they create divisions in the lexicographic text, all these typographical, structural and non-typographical elements form part of the rapid access structure of the dictionary because the user is immediately guided to the data he/she is seeking. How-

ever, the lexicographer should be warned against what may be called *over-ordering*. A lexicographer may find him-/herself in a situation of over-ordering when using too many structural markers or ordering devices. The application of over-ordering results in lengthier articles with an overwhelming number of structural indicators and search zones to the displeasure of the confused user. Therefore, lexicographers will do well to keep the use of ordering devices to a strict minimum because if they utilize too many it could easily become confusing for the user.

Yilumbu lexical items	Meaning	Etymology ⁹
neela ¹	window	borrowed from the Portuguese janela "window"
neela ²	ring	borrowed from the Portuguese anel "ring"
ureesi ¹	to divine	
ureesi ²	rice	borrowed from the French riz "rice"
pali ¹	type of bush rat	
pali ²	malaria	borrowed from the French palu (abbreviation of paludisme "malaria")
kala ¹	crab	*N-kádá
kala ²	related to the past	*-kádà, *kà-dàì◊◊
kafi ¹	paddle	*-kápì◊
kafi ²	coffee	borrowed from the French café "coffee"

Table 1: Some lexical item homonyms in Yilumbu

8. The macrostructure of the planned dictionary

8.1 Introductory remarks

While planning and structuring the macrostructure of a particular dictionary, lexicographers are usually confronted with the question of which lexical items to include and which ones to exclude. Obviously, there is a limit to what can be included in a dictionary. Due to the limited space in a dictionary, the answer to this question should be determined by the target user profile. That at this stage no dictionary exists in Yilumbu should lead the lexicographer towards compiling a standard dictionary, which could play a valuable role in the standardization of the Yilumbu language. In this regard, standard dictionaries must adhere to a certain set of criteria as Gouws (2001: 76) clearly states:

Standard dictionaries can be regarded as products resulting from a well-established lexicographic environment. These dictionaries are the most commonly used monolingual lexicographic instruments and display a wide range of lemmata and microstructural categories. Standard dictionaries usually are single volume products in which a synchronic and normative approach prevails. The macrostructure represents the standard variety of the treated language although a number of high usage frequency items from non-standard varieties will also be included. These items will be marked by lexicographic labels indicating stylistic, chronolectic, regional or other deviations from the standard variety. Standard dictionaries include a representative selection of macrostructural items and an

extensive treatment of these items. These dictionaries consequently have a high data density.

The choice of the typological category of standard dictionaries is relevant for a number of reasons. First of all, Yilumbu has yet to be standardized and no dictionaries are currently available to the speakers of the different speech communities.

With regard to the choice of a standard variety, the macrostructure of a dictionary should not necessarily represent one variety of the language dealt with in the dictionary. When a dialect is chosen to be the standard form it becomes the form for the written language as well as the language of administration and the media, whereas other dialects are confined to spoken and informal communication. The choice of the standard dialect between Yilumbu yi ghāngu and Yilumbu yi menaáne will obviously favour those who already use the chosen variety and impair those who do not. In order to avoid social unrest, a compromise solution is to be preferred to the choice of one single variety. Drame (2001: 235-236) has pointed out that the development of a writing system for isiXhosa by John Bennie in 1824 was based on two dialects, namely Ngqika and Gcaleka. The same principle may be applied to Yilumbu since its two major dialects are mutually intelligible despite variations that might occur in their lexicons. In addition to this, language or dialect standardization is most often initiated from governmental level (language planners) rather than from grass-roots level (speech communities). Put differently, it is the political authorities and language policy makers who should provide metalexigraphers with answers or solutions regarding issues relating to dialect choices. The dictionary conceptualization plan also forces the lexicographer into making early decisions as to the types of lexical items to be lemmatized. Given the fact that the planned dictionary does not seek to stick to one single variety for its macrostructural elements, it can be argued that the dictionary in question should function as basis for a series of other dictionaries. This means that from a macrostructural perspective, it is important for the lemma selection of the planned dictionary not to be too restricted. On the contrary, it should include a fairly representative selection of macrostructural items as well as their extensive treatment. In other words, all types of lexical items (nouns, verbs, prefixes, adjectives, possessives, etc. including multiword lexical items) should be included as lemmata. This has a lot of implications as far as lemmatization is concerned (cf. Hausmann and Wiegand 1989: 337 and Gouws 1999: 72). Nouns and verbs are usually regarded as universal categories. Thus their inclusion in a dictionary is often taken for granted. Hausmann and Wiegand (1989: 329) correctly state that in the European tradition, verbs are lemmatized under their infinitive form or under the first person singular of the present indicative. With regard to African languages, two lexicographic traditions exist, namely the word tradition and the stem tradition. According to the word tradition, lexical items are entered in their complete forms, i.e. prefix plus stem, while in the stem tradition lexical items are lemmatized under the stem without their prefixes. It is

part of the editorial policy of the planned dictionary to lemmatize all the lexical items according to the word tradition. However, it is often claimed that the word tradition displays a bias towards prefixation. In other words, the adoption of the word tradition causes some sectors of the dictionary to be overcrowded and others nearly empty. This is especially true of certain data categories such as verbs, adjectives, possessives and so on. One of the ways for the lexicographer to avoid a bias towards prefixation is to combine features of both the word and the stem traditions. In the proposed dictionary verbs will be entered under their imperatives whereas adverbs and adjectives will be lemmatized according to the stem tradition. Consider the following examples in this regard.

Wenda! (+wenda) [ùwě̀nd...] v. (cl. 15) ...
 (bi)-**ngi** [bí̀ŋgi] adv. (cl. 8)...
 (u)-**diidi** [ùdí̀di] adj. (cl. 1)...

Text example 2: Articles Wenda!, -ngi, and -diidi

Despite the fact that a decision has been taken to include all types of lexical items, they should not be included on a random basis but according to fixed criteria accounted for in the compulsory target users' text. Among these criteria, usage frequency should be one of the dominant factors. As far as criteria of incorporation of lemmata are concerned and as part of the planning of the macrostructure, the dictionary should deal with all possible dialectal forms. However, they will not be included on account of their occurrence as dialectal forms in the lexicon but on account of their usage frequency in the Yilumbu corpus. For example, the lexical item *ngeeyu* (you) accounts for 84 or 0,24% of the records, whereas *eyyu* (you) appears three times in the entire text of the Yilumbu corpus. This simply means that the focus for the lexicographic treatment of the lexical item meaning 'you' should be on *ngeeyu* with a cross-reference to *eyyu*.

As already said, special care should also be taken at the level of the meta-language to ensure that the dialect in which the lexical items in question are used on a daily basis by mother-tongue speakers are acknowledged by the use of the so-called diatopic or regional labels such as (*Ghâng*) and (*Men*).

With regard to the macrostructural treatment of borrowed technical terms in various fields such as Mathematics, Biology, Physics, Astronomy, Chemistry, etc., there are two arguments to be considered:

- (a) A lexicographer should consider well in advance the delimitation of the macrostructure of the special fields earlier mentioned. This can only be achieved if he/she works in cooperation with universities, institutions of higher learning and high schools by requesting departments of the special fields and teachers to comment on a preliminary term list.

- (b) Given the fact that the planned dictionary will be compiled for use by the Yilumbu speech community, they should also be involved in the project as early as possible.

8.2 Dictionary basis, computational support and selection of lemma candidate list

As part of the secondary comprehensive lexicographic process (Wiegand 1998) of the planned dictionary, the gathering of speech material at grass-roots level has resulted in the compilation of the dictionary basis (Wiegand 1998: 139) or lexicographic corpus of the dictionary. The recordings were transcribed electronically. The transcriptions of these oral interviews (or data corpus collected so far and referred to in the project background above) have been the object of a frequency study conducted at the Bureau of the WAT (Mavoungou 2000). Relying on the frequency lists obtained, a selection of core vocabulary items was made.

9. The microstructure of the planned dictionary

When devising the microstructure of a dictionary, the lexicographer has firstly to consider the issue of what sort of data should be included and what excluded. One criterion for this selection is the dichotomy between what has been called *communicative-directed function* and *knowledge-directed function*. According to Bergenholtz and Tarp (2002), communicative-directed functions are meant to assist users in solving problems that might occur during the communication process. As such, they can be referred to as both the production and reception of texts (what is described as decoding and encoding functions by some authors). Knowledge-directed functions provide the user with explanations of the meaning related to the knowledge. Central to these functions is the culture that underlies the language. One aspect of a users' needs assessment encompasses the obligation for the lexicographer to know the users' general cultural and encyclopedic knowledge.

So far, questions about encyclopedic data have been addressed in terms of its relevance in the planned dictionary as well as the extent to which it should be treated. Therefore the question to be asked is: Where should encyclopedic data be included? This question connects with the data distribution structure (Bergenholtz, Tarp and Wiegand 1999: 1778) of the intended dictionary. The easy answer to the lexicographer's dilemma with regard to linguists criticizing the dictionary for not imparting linguistic data only in the definition may be found in the notion of integrated and non-integrated outer texts, as developed by Bergenholtz, Tarp and Wiegand (1999). This concept gives him/her the opportunity to account for encyclopedic data in the front and back matter texts together with a link to the central list. Such a treatment also creates another access possibility for the user and therefore gives the dictionary a poly-accessi-

ble character. For example, the *Deutsch–Madagassisches Wörterbuch/Rakibolana Alema–Malagasy* by Bergenholtz et al. (1994) presents cultural data (i.e. dealing with routine formulas) in the outer texts of the dictionary (see also Gouws 2002 for a more comprehensive account in this regard). As far as the planned dictionary is concerned, this notion of integrated and non-integrated outer texts will be used to account for proverbs and idiomatic expressions. In other words, these lexical categories will not be treated in the central text but in back matter texts instead. On the contrary, encyclopedic data will not be accounted for in the lexicographic definition but in a specific slot of a given lemma in the central list. This decision has influenced the dictionary conceptualization plan in terms of the microstructure, access structure, addressing structure, and so on. Moreover, in order to incorporate real-world knowledge in the planned dictionary, the treatment chosen is to include in the so-called complex articles all cultural data that would not be generally known to an English or a French user in particular. Consider the following dictionary article in this regard:

malamu (+ lamu) [màlámù] n. (cl. 6) < *-dámù		
1 ♦ Dingiba di yilu (⇒ malamu ma yilu) ♦	1 ⟨F⟩ Vin récolté sur un palmier non abattu.	1 ⟨E⟩ Palm wine obtained from a palm tree that has not been felled.
∑ Nge malamu utsinu?	⟨F⟩ Es-tu soûle ?	⟨E⟩ Are you drunk?
2 Wootsu ke dingiba nana malamu mabeenga vho malamu mangolu.	2 ⟨F⟩ N'importe quel type de boisson alcoolisée tel que le vin ou le whisky.	2 ⟨E⟩ Any kind of intoxicating drink such as wine or whisky.
∑ Mi sabenu malamu.	⟨F⟩ J'ai arrêté de boire.	⟨E⟩ I have stopped drinking.
3 (<i>pěta</i>) (<i>Ghâng.</i>) Muyinu bakaata bayinaanga mo teemu dufu (<i>Men.</i>) Muyinu bivhuunda bayinaanga mo dufu.	3 (<i>ancien.</i>) ⟨F⟩ Danse organisée à l'occasion des funérailles pour pacifier l'esprit du défunt (ou de la défunte) ⇒ Yip. ikóóku.	3 (<i>archaic</i>) ⟨E⟩ Dance held at funerals to pacify the spirit of the deceased. ⇒ Yip. ikóóku.
∑ (<i>Ghâng.</i>) Vho teemu ighulu bayinaanga malamu (<i>Men.</i>) Vho ighoomba i ghaala bayinaanga malamu (⇒ ▼).	⟨F⟩ Autrefois, on dansait le <i>malamu</i> (⇒ ▼).	⟨E⟩ In the olden days, people danced the <i>malamu</i> (⇒ ▼).
(<i>cf. 2</i>) (<i>Men.</i>) malamu ma mbari (aussi/also ditutu) (⇒ (<i>Ghâng.</i>) mbula) Malamu bavaboonga mu diba/mbari ili vhotsi.	⟨F⟩ Vin de palme récolté sur un palmier abattu.	⟨E⟩ Palm wine obtained from a palm tree that has been felled.
∑ (<i>Men.</i>) Ubentsuroomba malamu ma mbari vhavha iki isyemu.	⟨F⟩ Il est en ce moment difficile de trouver du vin de palme.	⟨E⟩ It is now difficult to find palm wine.
(<i>cf. 1</i>) malamu ma yilu (⇒ tsaamba) Malamu bavaboonga mu diba/mbari ili vho yilu.	⟨F⟩ Vin récolté sur un palmier non abattu.	⟨E⟩ Palm wine obtained from a palm tree that has not been felled.
∑ Yisyeeeli avasumbisi malamu ma yilu.	⟨F⟩ Le <i>malafoutier</i> vend le vin de palme récolté sur un palmier non abattu.	⟨E⟩ The <i>malafoutier</i> is selling the palm wine obtained from a palm tree that has not been felled.

▼ (*Ghâng*) Vho teemu ighulu bayinaanga malamu. A Mutrafu. Batelaanga dina di mughisi a Mutrafu. Vho yilu malamu ika mangumba. Vho yilu mangumba ke kwagha (F) Au temps des ancêtres, on dansait le *malamu*. *Mutrafu* était le nom du génie. Après le *malamu* vint le *mangumba* et ensuite le *kwagha* (E) In the ancestors' time, the *mangumba* was danced. *Mutrafu* was the name of the spirit. After the *malamu* came the *mangumba* and thereafter the *kwagha* ⇒ **dingumba**.

■ **yisyeeli yi malamu** (⇒ **yisyeeli**); **mulaangi malamu** (⇒ **mulaangi**); **malamu Mutrafu** (aussi/also **malamu ma muvhyaanga**) (⇒ ▼); (*cf. II*) **malamu ma myeenga** (F) Vin de palme frais (E) Fresh palm wine.

Text example 3: Article malamu

From the data presented in this article given as Text example 3, the user can retrieve much information. In fact, data has been presented in four search zones which can be clearly perceived by the user and which displays an integrated microstructure. The first search zone presents the lemma sign followed by its stem, the phonetic transcription item, the part of speech indicator, and the item giving the class number as well as the proto-Bantu reconstruction. The second search area starting with the Arabic numerals presents a system of direct addressing between the paraphrases of meaning/translation equivalents and their co-text entries. In the third search zone introduced by the black upside down triangle ▼, the user is provided with lexicographic data of an extra-linguistic nature. The fourth section of the article (introduced by the black square symbol ■) presents groupings of sublemmata. This paradigm of compounds with **malamu** as one of the component parts displays all the characteristic features of a case of nesting because the internal alphabetical ordering is not maintained in the sinuous file. More specifically, this presentation is referred to as *second level nesting* (Gouws, 2001: 106). In order to anticipate any problem in the retrieval of the information sought by the user, he/she is assisted by entries clearly indicating which sense of **malamu** applies to a specific treated sublemma.

With regard to the issue of labelling, the temporal label (*ancien./archaic*) actually tells the user that although the lexical item "malamu" is used on a daily basis by the Balumbu, sense 3 of the lemma sign is now obsolete. It was in use in the past but has now fallen into disuse. The matter to which it refers (funeral dance) is now referred to by another lexical item: "**dingumba** (pl. **mangumba**)".

The next text example accounts for the treatment of the lemma **ngulubu**:

ngulubu [ngùlùbù/tsingùlùbù] <i>n.</i> (cl. 9/10) *-gùdùbè			
◆ Ibulu yi musiru ili nana ngulu ◆ ⇒ (<i>Men.</i>) ngulubi	(F)	Animal ressemblant au porc domestique ou cochon vivant en forêt (T) Potamochère (Potamochoerus porcus), (<i>frGab</i>) sanglier (voir ▼ et Fig. 1 et Fig. 2).	(E) Pig-like animal that lives in the forest (T) Bush-pig (Potamochoerus porcus), (<i>frGab</i>) sanglier (see ▼ and Fig. 1 and Fig. 2).
Σ Kumba atsiboka ngulubu.	(F)	Kumba a tué un potamochère.	(E) Kumba has killed a bush-pig.

<p>Σ Mureela na ditengu batsibengunu na muraanga ngulubu o musiru (<i>kughu yi mureela o ditengu na ghu mutu</i>).</p>	<p>⟨F⟩ Le revenant et le chasseur ont rencontré un troupeau de potamochères dans la forêt (cf. <i>conte du chasseur fantôme et du chasseur humain</i>).</p>	<p>⟨E⟩ The ghost and the hunter have come across a group of bush-pigs in the forest (cf. <i>story of the ghost hunter and the human hunter</i>).</p>
<p>▼ ⟨F⟩ Les termes potamochère et sanglier doivent être soulignés à cause notamment des différences référentielles. Le <i>potamochère</i> (<i>Potamochoerus porcus</i>) dénote l'un des mammifères de la faune gabonaise en particulier. Tandis que le <i>sanglier</i> (<i>Singularis porcus</i>) est un mammifère qu'on retrouve essentiellement en Europe. Par conséquent, <i>sanglier</i> est le terme du français local pour désigner le <i>Potamochoerus porcus</i> au Gabon en particulier.</p> <p>⟨E⟩ The lexical items potamochère and sanglier need to be clarified mainly because of the differences in denotation. The <i>potamochère</i> (<i>Potamochoerus porcus</i>) refers to one of the mammals of the Gabonese fauna in particular, whereas the <i>sanglier</i> (<i>Singularis porcus</i>) is a mammal mainly found in Europe. Therefore, the lexical item <i>sanglier</i> is the local French form used to refer to the <i>Potamochoerus porcus</i> in Gabon in particular.</p>		

Text example 4: Article ngulubu

In the planned dictionary, the focus should be on one variety of English, e.g. British or American English. As far as French is concerned, the emphasis is obviously on the variety of French spoken in France (in Paris in particular). But when including lexical items restricted to a variety exclusively used by the target users of the dictionary, the lexicographer should indicate this restricted usage by employing a system of geographical labels. These labels could have both a lemmatic and non-lemmatic address but help to shift the focus from the source language to the target language when they display non-lemmatic addressing procedures. Accordingly, the treatment of the article of the lemma **ngulubu** displays some aspects of the so-called non-lemmatic addressing procedures. In the section dealing with target language data in French in particular, the entry *potamochère* is given as a translation equivalent of the lemma sign **ngulubu**. With regard to **ngulubu**, the entry *potamochère* therefore displays a lemmatic addressing procedure. The entry *sanglier* follows *potamochère* from which it is separated by a comma used to achieve a metacommunicative function. In other words, commas are used to separate synonymous translation equivalents. Although *sanglier* is also presented as a translation equivalent of the lemma, it displays a non-lemmatic addressing procedure for it is a synonym of *potamochère*. This already creates topic switching for each non-lemmatic address is a new topic within the article (Hausmann and Wiegand 1989: 329). The topic switching is even strengthened by the occurrence of the entry *frGab* (short for *Gabonese French*, cf. Mavoungou 2002, 2002c) given in parentheses and addressed to the translation equivalent *sanglier*. The so-called diatopic or regional labels (*frGab*) actually tells the user that *sanglier* is the local form of French used to refer to *Potamochoerus porcus* in Gabon in particular. This is important because in monoscopal dictionaries (like the one that is currently being planned), the source language usually remains the only language of treatment. This often results in a lemmatic bias because the lemma receives the

primary focus. However, the target languages should also be seen as languages of treatment shifting the focus from the source language to the target languages (cf. Gouws 1996: 158). This shift in the treatment from a source language dominated approach to an approach where the target languages also come into play is taken somewhat further with the occurrence of the cross-reference entry *sanglier*. This cross-reference entry consists of two text segments: the reference marker *voir* (see) and the entry marking the reference address ▼. The upside down triangle takes the user to the distinction between *potamochère* and *sanglier* in terms of denotation.

However, if the lexicographer chooses not to make use of the specific slot in which extra-linguistic data are presented (also referred to in some literature as *usage note*), he/she can rely on the use of pictorial illustrations as a strategy to convey encyclopedic knowledge to the user. Consider Fig. 1 and Fig. 2 in this regard:



Fig. 1: Sanglier (extracted from NPL, 1971: 923)



Fig. 2: Potamochère (extracted from www.samaria.co.za/bushpig.htm)

As visual aids, Fig. 1 and Fig. 2 help the user to fully distinguish *sanglier* (*Singularis porcus*) and *potamochère* (*Potamochoerus porcus*) from one another.

If a lexicographer wants to take the user to the crux of the physical distinction between *potamochère* and *sanglier* in terms of denotation he/she may, for example, use the following encyclopedic definition of *Potamochoerus porcus* taken from <http://www.samaria.co.za/bushpig.htm>:

Pig-like animal with elongated face and a short, laterally flattened body. Colour varies from reddish to dark brown or nearly black. Covered with hairy bristles with a crest of yellowish hair along back from head to base of tail, hanging over sides of shoulders and body. Ears covered with long tassels of hair, tusks short but very sharp.

If the lexicographic text already gives an account of the distinctive physical features of *Singularis porcus* and *Potamochoerus porcus* then pictorial illustrations

should not be used. In fact, there is no need for lexicographers to use both of these devices. A lexicographer can either use a usage note in order to reflect on the physical features distinguishing *sanglier* and *potamochère* from one another or he/she may simply use pictorial illustrations. If a lexicographer uses both these devices, the pictorial illustrations will be of more cosmetic or ornamental value than functional components of the dictionary article. In this regard, Gouws (1993: 45-46) has rightfully emphasized:

Dictionary critics have the right to regard pictorial illustrations as redundant entries if they perform no distinctive function. However, the lexicographical validity of their employment is unquestionable when these entries enhance the comprehension of the target user.

Conclusion

No dictionary should be attempted without a well-devised dictionary conceptualization plan as part of the dictionary specific lexicographic process. In order to lay a sound lexicographic basis for the planning and compilation of their dictionaries, lexicographers may rely e.g. on Wiegand's theoretical framework. As far as the planned dictionary is concerned, it will be monoscopal in that it is only conceived for the following language direction A>B, namely Yilumbu>French-English. However, reverse word lists for French and English will then be supplied after the central list giving the dictionary a poly-accessible character. The discussion above has merely focused on lemma signs representing a high degree of cultural information. This was motivated on the ground of the lexicographer's responsibility for imparting or enhancing the cultural knowledge of the source language. The application of Wiegand's concept of micro-architecture will both result in a decrease in the density of information in the microstructure and an increase in explicitness and user-friendliness of the dictionary. This dictionary project will only fulfil its purpose when completed.

Acknowledgements

This article was extracted from my D.Litt. dissertation entitled *Metalexicographical Criteria for the Compilation of a Trilingual Dictionary Yilumbu-English-French* completed in the Department of Afrikaans and Dutch at the University of Stellenbosch. In this dissertation, I used Wiegand's metalexigraphy in order to establish, e.g., the dictionary basis, the macrostructure, the microstructure, the access structure, the addressing structure and the mediostructure of a trilingual dictionary Yilumbu-English-French. The dissertation was written under the supervision of Prof. R.H. Gouws, a leading authority on metalexigraphy, for whose directive comments I am grateful. Apart from Prof. Gouws, there are other theoretical lexicographers participating in the development of Wiegand's theory, e.g. Prof. Henning Bergenholtz and Prof. Sven Tarp, the influence of

numerous fruitful discussions I had with them during the work on the dissertation I herewith gladly acknowledge. I also extend my thanks to Prof. D.J. Prinsloo, known for his first-hand knowledge of African languages and lexicography, whose valuable comments from his regular lexicography seminars at the University of Stellenbosch could be incorporated into the final version of the dissertation.

Endnotes

1. The Yilumbu language is often referred to as *Ilumbu* or *Yiloumbou*. The latter is a glossonym written according to a French-based orthography. Yilumbu is spoken by the ethnolinguistic group called *Balumbu*. This group is also referred to as *Baloumbou* or *Loumbou*.
2. A developing language is a language that has yet to be standardized and in which no textbooks, newspapers, dictionaries, etc. are currently available to the speakers of the different speech communities.
3. In Gabon, French is the language of wider communication in business, trade, media, primary and higher education as well as in international communication.
4. Ghisira (also referred to as *Gisira* or *Eshira*), Yipunu (also known as *Ipunu*, *Yipounou* or *Apono*), Yisangu (also referred to as *Isangu* or *Masango*) and Civili (also known as *Vili*, *Tchivili*, *Fiote* or *Ciloango*) are sister languages of Yilumbu. Yilumbu, Yipunu, Yisangu and Ghisira are all part of the Sira-Punu group (B40), whereas Civili (H12a) is part of the Kongo group, which includes Bembe (H11), Kiyoombi (H12b), Kunyi (H13a) and Kikongo (H16c).
5. In this article, the official Gabonese spelling is used and not the semiphonetic spelling employed by Guthrie, Jacquot, Kwenzi-Mikala a.o. The underlined e, o and u represent [ɛ], [ɔ] and [ɯ] respectively. The phonetic sound [ə] is represented by the symbol ə in Gabonese orthography.
6. The Raponda Walker Foundation is a non-governmental organization whose main purpose is the promotion of the languages of Gabon and its different cultures.
7. "The Republic of Gabon has been in the process of promoting its local languages for at least the last 15 years. Through the content of his recommendations, the Gabonese States-General for Education and Training (Libreville, 17–23 September 1983) has shown that the Republic of Gabon attaches great importance to the development of local languages as well as their integration into the education system. This willpower is clearly seen through the acts of the Ministry of National Education on the one hand, and the Ministry of Higher Education, Scientific Research and Technological Innovation on the other hand. As an example, the following can be mentioned:
 - (1) Organization of seminars and conferences
 - (a) Séminaire sur l'alphabet scientifique des langues du Gabon (20–24 February 1989)
 - (b) Première table ronde sur les politiques linguistiques et l'enseignement des langues gabonaises (9–11 December 1997)
 - (c) Session de concertation sur l'orthographe des langues gabonaises (8–10 April 1999)
 - (2) Creation of the Department of Language Sciences at the University Omar Bongo (1994)
 - (3) Creation of the Department of Applied Linguistics at the Teachers' Training College (ENS, 1998)

- (4) Creation of the Department of National Languages at the National Educational Institute (IPN, 1999)."
(cf. Ndinga-Koumba-Binza 2002).
8. The Christian and Missionary Alliance (CMA) was instrumental in producing religious and lexicographic publications in Yipunu (a sister language of Yilumbu). For example, the *Dictionnaire français-yipounou/yipounou-français* (1966) is the result of the input from the Language and Literature Committee of the CMA.
9. Most proto-Bantu forms in this Table are from Meeussen (1965) and Nsuka Nkutsi (1980).

References

Dictionaries and lexicons

- Bergenholtz, Henning in co-operation with Suzy Rajaonarivo, Rolande Ramasomanana, Baovola Radanielina a.o.** 1994. *Deutsch-Madagassisches Wörterbuch/Rakibolana Alema-Malagasy*. Antananarivo: Tsipika/Moers.
- Bonneau, J.** 1956. *Grammaire pounoue et lexique pounoue-français*. Brazzaville: Institut d'Études Centrafricaines.
- CMA = Église Évangélique du Sud-Gabon. 1966. *Dictionnaire français-yipounou/yipounou-français*. Mouila: Christian Missionary Alliance.
- Derouet, R.P.** n.d. *Dictionnaire français-vili*. Loango: Imprimerie de la Mission.
- Hartmann, R.R.K. and G. James.** 1998. *Dictionary of Lexicography*. London/New York: Routledge.
- Marichelle, C.** 1902. *Dictionnaire vili-français*. Loango: Imprimerie de la Mission.
- Mini, B.M. et al. (Eds.)**. 2004. *The Greater Dictionary of Xhosa (IsiXhosa/English/Afrikaans). Volume 2. K-P*. Alice: University of Fort Hare.
- NPL = *Nouveau Petit Larousse*. 1971. Paris: Librairie Larousse.
- Pahl, H.W. et al. (Eds.)**. 1989. *The Greater Dictionary of Xhosa (IsiXhosa/English/Afrikaans). Volume 3. Q-Z*. Alice: University of Fort Hare.
- Rittaud-Hutinet, C.** 1980. Lexique. Nsuka-Nkutsi, F. (Ed.). 1980. *Éléments de description du Punu*: 193-245. Lyons: Presses Universitaires de Lyon.

Other literature

- 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., S. Tarp and H.E. Wiegand.** 1999. Datendistributionsstrukturen, Makro- und Mikrostrukturen in neueren Fachwörterbüchern. Hoffman, L., H. Kalverkämper and H.E. Wiegand (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.
- 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: Revue du Laboratoire de Phonétique et Linguistique Africaine* 1: 7-35.

- Busane, M.** 1990. Lexicography in Central Africa: The User Perspective, with Special Reference to Zaire. Hartmann, R.R.K. (Ed.). 1990. *Lexicography in Africa. Progress Reports from the Dictionary Research Centre Workshop at Exeter, 24–26 March 1989*: 19-35. Exeter Linguistic Studies 15. Exeter: University of Exeter Press.
- Drame, A.** 2001. Foreign Words as Problems in Standardisation/Lxicography: English and Afrikaans Loan-words in IsiXhosa. *Lexikos* 11: 231-241.
- Emejulu, J.D. and F. Pambo Loueya.** 1990. Yilumbu. *Revue Gabonaise des Sciences de l'Homme* 2: 197-201.
- Garnier, A.** 1897. *Katesisa i gheghe nesi ma loghi ma dzambi mu mbembu 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.** 1904. *M'ambu ma nzambi mo make mu katesisa*. Loango: Imprimerie de la Mission.
- Gouws, R.H.** 1993. Afrikaans Learner's Dictionaries for a Multilingual South Africa. *Lexikos* 3: 29-48.
- Gouws, R.H.** 1996. Bilingual Dictionaries and Communicative Equivalence for a Multilingual Society. *Lexikos* 6: 14-31.
- Gouws, R.H.** 1999. *A Theoretically Motivated Model for the Lexicographic Processes of the National Lexicography Units*. Research report submitted to the Pan South African Language Board.
- Gouws, R.H.** 2001. Lexicographic Training: Approaches and Topics. Emejulu, J.D. (Ed.). 2001. *Éléments de lexicographie gabonaise. Tome I*: 58-94. New York: Jimacs-Hillman Publishers.
- Gouws, R.H.** 2002. Using a Frame Structure to Accommodate Cultural Data. Emejulu, J.D. (Ed.). *Éléments de lexicographie gabonaise. Tome II*: 54-69. New York: Jimacs-Hillman Publishers.
- Grimes, B.F.** 1996¹³. *Ethnologue*. Dallas: Summer Institute of Linguistics, Inc.
- Guthrie, Malcolm.** 1953. *The Bantu Languages of Western Equatorial Africa*. London: Oxford University Press.
- Hartmann, R.R.K.** 1989. Sociology of the Dictionary User: Hypothesis and Empirical Studies. Hausmann, F.J. et al. (Eds.). 1989-1991: 102-111.
- Hartmann, R.R.K.** 1999. Lexical Reference Books — What Are the Issues? *International Journal of Lexicography* 12(1): 5-12.
- Hausmann, F.J. and R.O. Werner.** 1991. Spezifische Bauteile und Strukturen zweisprachiger Wörterbücher: Eine Übersicht. Hausmann, F.J. et al. (Eds.). 1989-1991: 2729-2769.
- Hausmann, F.J. and H.E. Wiegand.** 1989. Component Parts and Structures of General Monolingual Dictionaries: A Survey. Hausmann, F.J. et al. (Eds.). 1989-1991: 328-360.
- Hausmann, F.J. et al. (Eds.).** 1989-1991. *Wörterbücher: Ein internationales Handbuch zur Lexicographie/Dictionaries: An International Encyclopedia of Lexicography/Dictionnaires: Encyclopédie internationale de lexicographie*. Handbücher zur Sprach- und Kommunikationswissenschaft, vol. 5.1.-5.3. Berlin: Walter de Gruyter.
- Kromann, H.-P., T. Riiber and P. Rosbach.** 1991. Principles of Bilingual Lexicography. Hausmann, F.J. et al. (Eds.). 1989-1991: 2711-2728.
- Martin, W.** 1996. Lexicographic Resources in a Multilingual Environment: An Orientation. *Lexicography as a Financial Asset in a Multilingual South Africa*: 9-24. Pretoria: Department of Arts, Culture, Science and Technology.
- Mavoungou, P.A.** 2000. *A Frequency List of the Yilumbu Language*. Unpublished study conducted at the Bureau of the Woordeboek van die Afrikaanse Taal (WAT), Stellenbosch.

- Mavoungou, P.A.** 2002. Synopsis Articles in the Planning of a Trilingual Dictionary: Yilumbu–French–English. *Lexikos* 12: 181-200.
- Mavoungou, P.A.** 2002a. Sociolinguistic and Linguistic Aspects of Borrowing in Yilumbu. *South African Journal of African Languages* 22(1): 41-58.
- Mavoungou, P.A.** 2002b. *Metalexicographical Criteria for the Compilation of a Trilingual Dictionary: Yilumbu–English–French*. Unpublished D.Litt Dissertation. Stellenbosch: University of Stellenbosch.
- Mavoungou, P.A.** 2002c. Vers un dictionnaire du français du Gabon. Emejulu, J.D. (Ed.). 2002: *Éléments de lexicographie gabonaise. Tome II*: 230-262. New York: Jimacs-Hillman Publishers.
- Meeussen, A.E.** 1965. *Reconstructions grammaticales du bantou*. (Traduction française). Tervuren: Musée Royal de l'Afrique Centrale.
- Ministère de l'Éducation Nationale.** 1999. *Rapport Final de la Session de Concertation sur l'Orthographe des Langues Gabonaises*. Libreville: Education Nationale.
- Murard, P. (Tr.)**. 1903. *Katsisu i keki i rendilu mu mbembo bis 'Sette-Cama (petit catéchisme)*. Lyons: Imprimerie Paquet.
- Murard, P. (Tr.)**. 1903a. *Katsisu i neni i rendilu mu mbembo bis 'Sette-Cama (grand catéchisme)*. Lyons: Imprimerie Paquet.
- Ndinga-Koumba-Binza, H.S.** 2002. *The Bold and the Beautiful Languages in Gabon*. Unpublished paper presented at the Department of General Linguistics, University of Stellenbosch, Stellenbosch, 13 August 2002.
- Nsuka-Nkutsi, F. (Ed.)**. 1980. *Éléments de description du Punu*. Lyons: Presses Universitaires de Lyon.
- Raponda-Walker, A.** 1932. L'alphabet des idiomes gabonais. *Journal de la Société des Africanistes* 2(2): 139-146.
- Raponda-Walker, A. and 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.
- Taljard, E. and R. Gauton.** 2000. Supplying Syntactic Information in a Quadilingual Explanatory Dictionary of Chemistry (English, Afrikaans, isiZulu, Sepedi): A Preliminary Investigation. *Lexikos* 10: 191-208.
- Wiegand, H.E.** 1998. *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie*. Berlin/New York: Walter de Gruyter.
- Wiegand, H.E.** 1999. *Semantics and Lexicography. Selected Studies (1976–1996)*. Edited by Anntje Immken and Werner Wolski. Lexicographica. Series Maior 87. Tübingen: Max Niemeyer.

Challenges to Issues of Balance and Representativeness in African Lexicography*

Thapelo Joseph Otlogetswe, *Information Technology Research Institute, University of Brighton, Brighton, United Kingdom and Department of English, University of Botswana, Gaborone, Botswana*
(otlogetswe@mopipi.ub.bw)

Abstract: Modern dictionaries depend on corpora of different sizes and types for frequency listings, concordances and collocations, illustrative sentences and grammatical information. With the help of computer software, retrieving such information has increasingly become relatively easy. However, the quality of retrieved information for lexicographic purposes depends on the information input at the stage of corpus construction. If corpora are not representative of the different language usages of a speech community, they may prove to be unreliable sources of lexicographic information. There are, however, issues in African languages which make many African corpora questionable. These issues include a lack of texts of different genres, the unavailability of balanced and representative written texts, a complete absence of spoken texts as well as literacy problems in African societies. This article therefore explores the different challenges to the construction of reliable corpora in African languages. It argues that African languages face peculiar challenges and corpus research may require a different treatment compared to European and American corpus research. It finally concludes that issues of balance and representativeness appear theoretically impossible when looking at the results of sociolinguistic research on the different existing language varieties which are difficult to represent accurately in a corpus.

Keywords: AFRICAN LANGUAGES, BALANCE, BANK OF ENGLISH, BORROWING, BRITISH NATIONAL CORPUS, COBUILD, CODE-SWITCHING, COMPUTERS, CORPORA, DIALECT, DICTIONARIES, FREQUENCY, LANGUAGE VARIETY, REPRESENTATIVENESS, SETSWANA, SOCIOLINGUISTICS, SPEECH, TEXT

Opsomming: Uitdagings betreffende kwessies van balans en verteenwoordigendheid in Afrikaleksikografie. Moderne woordeboeke steun op korpusse van verskillende groottes en soorte vir frekwensielyste, konkordansies en kollokasies, voorbeeldsinne en taalkundige inligting. Met die hulp van rekenaarprogrammatuur het die herwinning van sulke inligting toenemend redelik maklik geword. Die gehalte van herwonne inligting vir leksikografiese doeleindes steun egter op die inligtingsinset by die korpusboufase. Indien korpusse nie verteenwoordigend is van die verskillende taalgebruike van 'n spraakgemeenskap nie, mag hulle blyk

* This article is a revised version of a paper presented at the Eighth International Conference of the African Association for Lexicography, organised by the Department of German and Romance Languages, University of Namibia, Windhoek, Namibia, 7–9 July 2003.

onbetroubare bronne van leksikografiese inligting te wees. Daar is egter kwessies in Afrikatale wat baie Afrikakorpuse problematies maak. Hierdie kwessies sluit in die tekort aan tekste van verskillende genres, die niebeskikbaarheid van gebalanseerde en verteenwoordigende geskrewe tekste, die volkome afwesigheid van gesproke tekste asook geletterdheidsprobleme in Afrikagemeenskappe. Hierdie artikel ondersoek derhalwe die verskillende uitdagings betreffende die bou van betroubare Afrikataalkorpuse. Dit voer aan dat Afrikatale teenoor besondere uitdagings staan en korpusnavorsing 'n verskillende behandeling mag vereis in vergelyking met Europese en Amerikaanse korpusnavorsing. Ten slotte kom dit tot die gevolgtrekking dat kwessies van balans en verteenwoordigheid teoreties onmoontlik lyk wanneer gekyk word na die resultate van sosiolinguïstiese navorsing oor die verskillende bestaande taalvariëteite wat moeilik is om presies in 'n korpus te verteenwoordig.

Sleutelwoorde: AFRIKATALE, BALANS, BANK OF ENGLISH, BRITISH NATIONAL CORPUS, COBUILD, DIALEK, FREKWENSIE, KODEWISSELING, KORPUSSE, ONTLENING, REKENAARS, SETSWANA, SOSIOLINGUISTIEK, SPRAAK, TAALVERSCHEIDENHEID, TEKS, VERTEENWOORDIGENDHEID, WOORDEBOEKE

Introduction

More and more lexicographers realise the inevitability of using a corpus or corpora in the compilation of dictionaries. Leech (1991: 8) defines a corpus as "a sufficiently large body of naturally occurring data of the language to be investigated". Renouf (1987: 1) refers to the use of computers in the storing and analysis of corpora in his definition: "a collection of texts, of written or spoken words, which is stored and processed on computer for the purpose of linguistic research". McEnery and Wilson (1996: 24) similarly mention a reliance on computers in their definition of a corpus as "a finite-sized body of machine-readable text, sampled in order to be maximally representative of the language variety under consideration". Leech (1991: 5), however, insists that a corpus has to be differentiated from an "archive",

the latter being a repository of available language materials, and the former being a systematic collection of material for given purposes. A corpus draws upon the resources of an archive and therefore both are important. The systematic compilation of a structured corpus however is the primary objective.

Leech points to the systematicity of the collection of material as an important characteristic of a corpus. In this regard he does not conflate the substance for study with the tools used for its analysis and storage. However, whether the insistence on systematicity is crucial to the definition of a corpus may be subject to debate. Maybe "corpus" should be seen as textual data collected for linguistic research, usually stored in computers for quick analysis. But the fact that it is machine-readable, although important for its analysis, does not make it a corpus, for long before the introduction of computers there was much robust corpus research as exemplified by Kading's 1897 German corpus of some

11 million words for collating the frequency distribution of letters and sequences of letters.

For ages, lexicographers contended with ways and means of producing authentic and reliable reflections of the lexicon. Most of these lexicographers depended on their ability to remember words existing in the languages under study, something that De Schryver and Prinsloo (2000: 219) call "the random approach" and Kilgarriff (2000: 109) "the lexicographer's intuition". Others again, in the Oxford tradition, depended on readers, who searched texts for occurrences of words and submitted these for lemmatisation in the dictionary. For many years, these readers' contribution made the *Oxford English Dictionary* (OED) the unparalleled authority on the English language. More than any other English dictionary existing at the time, it included words from different genres and stylistic and regional varieties with reliable etymological information. Later developments in lexicography proved that readers were not very reliable sources of dictionary material since not only was their processing of data too slow, but it was also impossible for them to authoritatively deliver information on matters of frequency across texts and genres (see the *Longman Dictionary of Contemporary English* (1995³), the *Collins COBUILD English Dictionary* (1995²) or Kilgarriff (1997: 1)).

Over the past 20 years, a rapid growth of corpus lexicography has been witnessed, which was championed and popularised, more than by any other group, by the COBUILD (Collins Birmingham University International Language Database) group in Birmingham, led by John Sinclair. The earlier Birmingham school of corpus lexicography adhered strictly to the corpus as a source of dictionary evidence (Sinclair 1987). It was argued that corpora were the sole source of lemmatisation, frequency information and word lists. If a word was not in a corpus, it was not recognised as legitimate dictionary material. However, as corpus lexicography develops, there is a greater focus on its composition. Issues of balance and representativeness are continuously engaging theoretical and practical lexicographers. Researchers want to know the kinds of texts forming corpora and in what percentage they exist. These questions and concerns are not trivial since they put the credence and reputation of a dependency on corpus lexicography in question. Therefore the greatest challenge lies not so much in what can be obtained from a corpus, but rather in its construction.

Against this background, this article attempts to investigate the problems associated with the construction of corpora for dictionary making, particularly in many African contexts. It argues that some of the challenges facing the construction of robust corpora to be used in language research are the poverty of data, that is, the lack of texts to construct corpora representative of the different instances of language usage in a specific speech community. High illiteracy levels in African countries too pose great challenges to researchers hoping to collect written texts read by specific populations. Added to this, is the fact that, even where levels of literacy have increased, the literate members of a society

read and write texts written in English or French and not in their native languages. Even where such texts could be found in African languages, they mostly belong to a certain genre, like novels, plays and poetry, to the exclusion of other genres, like newspapers and academic texts. Even if the use of such data is attempted, the contention would still be with "sanitised" data, purified by the editorial policies and stylistic dictates of many publishing houses and newspaper offices, calling into question its authenticity as original and credible texts. The problem of representing speech still stands as one of the great challenges not only to African lexicographic research but also to research in many Western countries. At first, balance and representativeness must be investigated.

Balance and Representativeness

Most of the latest corpus-based lexicography researches consider issues of representativeness and balance (Ooi 1998) as marking standards of authenticity and robustness in corpus construction. A language corpus must be balanced and representative of the language from which it is extracted. By representativeness is meant "the extent to which a sample [text] includes the full range of variability in a population" (Biber 1993: 243), and as Summers (1993: 186) stresses "unless the corpus is representative, it is *ipso facto* unreliable as a means of acquiring lexical knowledge". Therefore, for a corpus to be representative, it must reflect the typical cross-spectrum of language use of a defined language community or period (see Ooi 1998: 49). But Summers's (1993) claim will be returned to since it raises considerable difficulties, particularly for corpus building in many African contexts and for certain linguistic theories.

A balanced corpus is one that includes proportions of a range of different text types of a language as they are reflected in the language studied.

The problem of what constitutes balanced and representative corpora still remains controversial. The selection of language from different genres to include in the language database is largely unresolved. The compilation of text must finally capture language from a specified population from which a sample is taken, which reflects how that particular language community uses language. This is significant since, as Summers (1993: 186, 190) points out, the results of corpora analysis must be generalised to the language community from which the samples were abstracted. Kennedy (1998: 94) argues for a pedagogical purpose to corpus research by noting that "high frequency of occurrence as determined by the analysis of texts should be a major determinant of lexical content of language instruction".

In a way, it is clear that issues of balance and representativeness of corpora are related. A representative corpus must reflect a representation of different genres of language use in a language community, while a balanced corpus should attempt to capture those different percentage levels or ratios in the way they occur in the specified language community. This obviously is difficult

to achieve, mainly because it is difficult to precisely know all the text types and their proportions of use in a population with its ever-changing dimensions. The difficulties are compounded when the building of a corpus of spoken language is attempted. As Kilgarriff (1997: 137) points out, dialectal varieties stand at different ratios to one another and should be represented within a corpus that attempts to accurately capture the language characteristics as a whole. There must also be contended with whether spoken texts can be accurately sampled and represented along the same lines as written texts. How many words are being looked for and what percentage of the spoken language do such words constitute? Whether spoken texts can be sampled in a representative manner is greatly questionable. Although a sample of Sengwaketse, Sekgatla, Sekwena and Sengwato can establish an acceptable representative percentage of the spoken form of these Setswana dialects, speech is a flood that refuses to be adequately accounted for numerically, for even when an attempt is made to quantify it, more of it is produced. It is Kennedy (1998: 62) who casts doubt on whether the representativeness of a corpus can confidently be argued for:

In light of the perspectives on variation offered by several decades of research in discourse analysis and sociolinguistics, it is not easy to be confident that a sample of texts can be thoroughly representative of all possible genres or even of a particular genre or subject field or topic.

By "perspectives on variation" Kennedy refers to different speech varieties existing in a speech community. Problems are faced with sampling the standard against non-standard varieties, various sociolects covering status, gender, ethnicity, age, occupation, and others, different regional varieties, like Sengwaketse, Sekgatla, Sekwena, and Sengwato in the case of Botswana, and different registers like casual, formal, technical and others. Such variations are difficult to represent in a corpus. By noting this difficulty, Kennedy does not imply that representativeness should not be attempted, but that perhaps theoretically an attempt at representativeness may not conclusively capture the nuances of existing varieties as outlined by linguistic research.

Because of practical constraints, such as a shortage of time and money, the unavailability of machine-readable text, and copyright restrictions, it is not always possible to assemble the representative and balanced corpus ideally wanted. It is precisely these problems that stand out as some of the major stumbling blocks particularly in the African context of corpus construction.

Two English Corpora

This section will bring to the fore the composition of more influential corpora which have been considered by many lexicographers and numerous language researchers as examples of "good" corpora. What should particularly be noted is the percentage of spoken text against written text since it is central to subsequent arguments made in this article.

In 1991, COBUILD launched the Bank of English (BoE), which currently has over 450 million words and continues to grow as more material is published and deposited into it. It forms the basis for the compilation of the COBUILD dictionaries (Sinclair 1991). The BoE does not claim any balance or representativeness of usage, but it does claim to provide evidence of the way everyday English is used. The spoken word is represented by transcriptions of everyday casual conversation, radio broadcasts, meetings, interviews, discussions, etc. However, even with the seemingly impressive 450 million words, the BoE is only a small sample of human speech produced on a daily basis.

The other corpus that has extensively been used is the British National Corpus (BNC) which has "a 100 million collection of samples of written and spoken British English of the late twentieth century from a wide range of sources designed to represent a wide cross-section of current British English both written and spoken" (BNC website). Ninety per cent of its composition consists of written texts including amongst other kinds of texts, extracts from regional and national newspapers, academic books and popular fiction, essays and letters (75% from informative writing such as fields of applied science and commerce and finance; 25% from imaginative, i.e. literary and creative, works). Spoken texts, which include unscripted informal conversation, government meetings and radio shows, constitute only 10%. The corpus has 4 124 texts, of which 863 are transcribed from spoken conversation and monologues. It was developed by the Oxford University Press, the Longman Group Ltd, Chambers Harrap, the Unit for Computer Research on the English Language (Lancaster University), the Oxford University Computing Services, and the British Library Research and Development Department. It has been used for a wide variety of research in language, including lexicography, as in the making of the third edition of the *Longman Dictionary of Contemporary English*.

The Primacy of Speech

It is a widely held fact that children speak before they write and that speech is primary to human communication (Aitchison 1998). It is also generally agreed that in a speech community the spoken word exists in abundance compared to written texts. Taking these linguistic arguments as base and applying them by implication to issues of balance and representativeness, it can be concluded that if corpus construction has to reflect the different ratios between spoken and written texts, different text genres and various dialectal varieties, then the percentage of spoken language has to be much greater than that of written language in a corpus. Such a greater occurrence of spoken over written texts would approximate the ratios of written and spoken texts in the real world and would be likely to produce corpora that accurately represent language as used in a speech community. However, in none of the corpora discussed in the previous section the percentage of spoken texts exceed that of written texts. Ten per cent of the data of the BNC consists of spoken texts. Leech et al. (2001: 1)

recognise the inadequacy of speech in the BNC which contains about 90 per cent written data and 10 per cent spoken data:

Although spoken language, as the primary channel of communication, should by rights be given more prominence than this, in practice this has not been possible, since 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. In view of this problem, these proportions were chosen as realistic targets which, given the size of the BNC, are also sufficiently large to be broadly representative.

According to Leech et al., the percentage of the speech text in the BNC was reached by determining what was possible to the compilers and not by making allowance for the proportion of speech to written language in a speech community. If corpora do not reflect in their composition that the spoken word is more common in real life than the written text, it calls the power and authority of corpora as sources of evidence for linguistic research in question and opens them to possible doubt.

A Newspaper versus the Purchase of a Pair of Shoes

While Kennedy (1998: 63) acknowledges the common occurrence of speech in daily discourse, he argues against it by noting:

No one knows what proportion of the words produced in a language on any given day are spoken or written. Individually speech makes up a greater proportion than does writing of the language most of us receive or produce on a typical day. However, a written text (say in a newspaper article) may be read by 10 million people, whereas a spoken dialogue involving the purchase of a pair of shoes may never be heard by any person other than the two original interlocutors.

Kennedy introduces a dimension to corpus creation that raises great controversy. It is true that a newspaper is likely to be read by many people and that its circulation can be obtained from reliable sources. However, it is not true that newspaper buyers equally read different sections of a newspaper. Some readers pass over the business section, classifieds, cartoons, letters to the editor and many other sections. Although circulation numbers might be available to assist corpus builders sample newspaper text, they are heavily unreliable because though a newspaper might be selling 40 000 copies, those copies might be read by over 100 000 individuals while others might be bought and never be read!

A similar point may be made that although lots of corpora depend on published texts, there is indeed no guarantee that such texts are widely read (or read at all). This is particularly so in the Setswana language situation where the majority of Batswana do not read Setswana texts, except at elementary school. Kennedy (1998: 52) suggests that to fix this problem "best seller lists, library

lending, statistics and periodical circulation figures can only partially reflect receptive use and influence". For many readers of texts in African languages "best seller lists, library lending, statistics and periodical circulation figures" are foreign concepts unheard of in African literature. Kennedy's use of "partially" is an indication of the immensity of problems surrounding attempts to construct corpora on the basis of common and influential texts. If "receptive use and influence" are taken as determinants of text inclusion in a corpus, varying *degrees* of such use and influence will have to be contended with. School textbooks and creative texts read by thousands of students across the country would be in use more than a library text which is rarely read. How would such a distinction be represented in a corpus? Is it not the case that textbooks would have been read more widely and therefore their texts should somehow reflect the fact that they have been seen more than other texts? This argument can be pursued further. This would mean that a sign reading "Welcome to Gaborone" would make "welcome" "to" and "Gaborone" very high in a frequency list since they have been seen many times by many people entering the city. Words like "stop", used on traffic signs and seen again and again, would be amongst some of the most common terms. Such conclusions would certainly distort the way language is used since the word "stop" does not occur frequently in daily discourse. The problem of how its commonality is represented in a corpus therefore remains.

It would appear that Kennedy's argument against spoken texts on the basis that they are private while written texts are in the public domain, is not very convincing but rather raises new problems and challenges. Spoken texts are as important as written texts in corpus creation and attempts should be made to reflect approximate ratios between written and spoken texts, ratios which are problematic to establish.

Can Anything Good Come out of Spoken Texts?

Much would be lost if a corpus does not reflect spoken texts in their right ratios. One such loss would be instances of borrowing common in written texts but censured by editors and publishers in communities where there is much code-switching, language contact and borrowing, particularly in many African countries where both native languages and former colonial languages like English or French are used. An observation of spoken Setswana texts will show a high degree of borrowing from English and Afrikaans. Borrowing is here used in Nevejina's (1998) sense of "the element of an alien language which is carried from one language to another as a result of language contact". The documentation of this phenomenon in Setswana is not recent. Cole (1955) noted words like *beke* (*week*) "week", *baki* (*baadjie*) "jacket", *gouta* (*goud*) "gold", *heke* (*hek*) "gate", *hempe* (*hemp*) "shirt", *kofi* (*koffie*) "coffee", *pena(e)* (*pen*) "pen", *peipe* (*pyp*) "pipe", *sukiri* (*suiker*) "sugar" from Afrikaans and *baesekele* "bicycle", *buka* "book", *ofisi* "office", *šeleng* "shilling" from English. There are other more recent borrowings

which reveal a certain layering in the nature of what is considered borrowed words. For instance, many Setswana speakers are not aware that *baki* and *heke* are borrowed from Afrikaans, while *jakete* "jacket" and *geiti*(?)[†] "gate" are recognised as borrowings from English. The result is that *baki* and *heke* are considered by some as "good" established Setswana, while the more recent borrowings *jakete* and *geiti* are condemned. Spoken Setswana is interspersed with instances of code-switching and borrowing in sentences such as the following:

Go shapo! (Good-bye!)
O tsile in the afternoon. (He came in the afternoon.)
Ke bra/sistere ya gagwe. (It is his brother/sister.)
O apere jase. (He is wearing his coat.)

Greater levels of code-switching and borrowing are also evident in naming the days of the week and the months of the year, and in naming the numerals. For instance, many Setswana speakers would say *Monday* or *Mantaga* (from Afrikaans *Maandag*), *Tuesday*, *Wednesday* ... *Saturday* or *Sateretaga* (from Afrikaans *Saterdag*) and *Sunday* or *Sontaga* (from Afrikaans *Sondag*). Reference to the months by Setswana speakers is also usually in English, and most would have difficulties in saying them in Setswana. In many instances Batswana speakers use the English instead of the Setswana names for the numerals. Many speakers would find it difficult saying 1 567 in Setswana since numbers are generally expressed in English. It is common for Batswana to use one, two, three, fifteen, two thousand, or one million in their speech instead of the Setswana terms.

These are some of the problems a Setswana lexicographer would have to face if he/she depends on a corpus with greater levels of spoken data rather than a corpus with written data or with smaller levels of spoken text. The lexicographer would grapple with decisions on the kind of borrowed words that should be lemmatised and the kind of stylistic information that should be derived from borrowed words. Obviously the kind of dictionary being compiled would influence such decisions: whether it is monolingual or bilingual, for learner's or general use, of table or pocket size, etc.

Dealing with borrowings and code-switching in lexicography is not a new phenomenon. Lichtenberk (2003) considers the question of which borrowed words qualify as belonging to the borrowing language and therefore deserving inclusion in a dictionary. In his report of the dictionary of Toqabiqita, an Austronesian language spoken in the Solomon Islands, he points out that the central point in determining the wordlist of a dictionary is "the prospective audience", that is, the intended users of a dictionary, and "its expectations", that is, the purposes the dictionary will be expected to serve in the society. This view is shared by Zgusta who contends that decisions of what to include are determined by "fundamental decisions concerning the type of dictionary which is to be prepared" (Zgusta 1971: 243). For instance, if the dictionary is intended to contribute to historical and comparative studies, it may list archaic and obsolete words while the inclusion of loanwords may prove to be of interest to pho-

nologists. But the greater part of Lichtenberk's article is devoted to a discussion of the inclusion or not of loanwords in the dictionary of Toqabaqita. There are comparisons which may be drawn between Setswana and Toqabaqita. Lichtenberk is confronted with a language situation where he has to decide whether to include Pijin words in the dictionary of Toqabaqita since some of them fit the phonological and phonotactic constraints of Toqabaqita while others do not. A similar challenge faces Setswana: whether to include borrowings from English or Afrikaans. Like Setswana, Toqabaqita does not permit consonantal clusters or syllable-final consonants and has a simple syllable structure of CV and V. This characteristic of Toqabaqita guides Lichtenberk (2003: 395) in deciding what to include:

Pijin words used in Toqabaqita are listed provided they fit the phonological and phonotactic patterns of Toqabaqita, either because they fit them already in Pijin or because they have been accommodated to them. Words which do not fit the patterns are not listed.

According to this principle, certain words in common use are excluded, because they are, in Lichtenberk's view, instances of code-mixing. Not satisfying the phonotactic constraints of Toqabaqita, they are not listed in the dictionary. Similar to the Setswana situation, code-mixing in Toqabaqita is common. Lichtenberk (2003: 396) argues:

Considering such words to be part of Toqabaqita lexicon would amount to claiming that the phonological inventory and the phonotactic patterns of the language have undergone some major changes.

Therefore Lichtenberk decided to restrict the matter of code-mixing to the front matter where the common but non-accommodated words would be listed. There are also problems concerning pairs of words which, though accommodated from Pijin, have variants which do not conform to the phonotactics of Toqabaqita. In these instances, the variant that does not conform to the phonotactic constraints is not listed. But it gets more complicated when the non-accommodated variant is more common than the accommodated one. In such cases, Lichtenberk ignores the most frequently used word, since it violates the phonotactic constraints of the language, and instead chooses to enter the less common one on the principle that the non-accommodated variant, though frequent, is an instance of code-mixing.

Lichtenberk (2003: 396) develops further principles which determine what to list. These are:

1. "Words that belong in well-circumscribed and relatively small sets are not listed if some other members of the same set do not occur in an accommodated form and so are not listed."

2. "A Pijin word that has been encountered only once is not listed even if it fits the phonological and phonotactic pattern of Toqabaqita."

The question of what has to be listed in the dictionary raises an issue of what are the boundaries of the lexicon of a language. Lichtenberk therefore divides Toqabaqita words into three categories: (a) native Toqabaqita words, (b) accommodated borrowings from Pijin, and (c) Pijin words used without being accommodated. Lichtenberk (2003: 397) concludes that:

Only the first two types are to be listed in the dictionary, which amounts to saying that only those words are part of the Toqabaqita lexicon, while the non-accommodated words are not.

He gives proper criticism to his approach when he says:

The principle, while explicit and applicable in a straightforward way, is nevertheless arbitrary. It gives priority to the phonological and phonotactic patterns of Toqabaqita over usage. Pijin words that are not accommodated are, by fiat, placed outside the circumference of the Toqabaqita lexicon, although by virtue of their usage they could be inside.

Some of Lichtenberk's principles are better not followed, particularly the preference of phonology over usage. Take for instance his first principle for listing sets of words. Such sets include the names of numerals, the days of the week and the months of the year. This principle creates problems for accounting for the class days of the week in Setswana.

Days such as *Sateretaga*, *Sontaga* and *Mantaga* are colloquial and more common in spoken than in written language, while *Matlhatso*, *Tshipi* and *Mosupologo* are common in written texts and formal addresses. This stylistic information is significant, particularly in dictionaries which attempt to achieve a broader coverage and a fuller understanding of a word's meaning and usage. When both formal and informal terms are lemmatised, they may provide, except stylistic information, significant information for future research on when a word has entered the language or changed its meaning.

Additionally, cases where certain terms, although known in the native language, are rarely used in speech, but are replaced by borrowings and code-switchings, cannot be ignored. This is particularly true of numerals where sentences such as *O rekisitse dinamune di le ten* "He has sold ten oranges" and *Mmiting o ka thene kamoso* "The meeting is at ten tomorrow" are found. In these examples, the speaker has chosen the English word *ten*, instead of the Setswana term *lesome/some*. The transcription of the term *ten* as either *ten* or *thene*, as in the above examples, is based on the theoretical question of whether such a term has gained currency as an instance of borrowing or of code-switching. Are lexicographers to assume that such language usages do not exist in the language and that they do not have any relevance to dictionary compilation? Any answer to these questions would lead to disagreements among lexicographers.

It is important to note that although *lesome* and *ten* refer to the same number, they usually have different usages. *Lesome* would be more common amongst the elderly, in written texts and in very formal "tribal" meetings. *Lesome* is also used to refer to P1 (one Pula). *Ten* is much more common in colloquial exchanges, spoken language and amongst the educated.

This hopefully shows the importance of including greater occurrences of spoken text in a corpus since the spoken word occupies a greater level of language usage in human communication. Next the lack of data and the available data for lexicographic research will be considered.

The Poverty of Data

While Western lexicographers enjoy an abundance of data for the construction of huge corpora running into millions of texts of different genres covering newspapers, magazines, novels, academic texts, parliamentary pronouncements, and legal texts, African lexicographers work under great constraints because of the lack of data. Unlike their Western counterparts, they usually do not possess the luxury to be discriminative and selective of texts in electronic form since in the first place such texts are nonexistent. Many African countries do not use their indigenous languages in parliamentary debates, the publication of laws, instruction at schools and journalistic publications. This is certainly the situation in Botswana where there exist very little text in Setswana. In comparison with English, there are very few Setswana novels and plays. There is also little instructional material in Setswana for lower primary school levels and virtually none for higher education. The only newspaper which wrote exclusively in Setswana, *Mokgosi*, closed down in 2005 because of poor advertising and sales. Another, *Mmegi*, which had a three and a half page Setswana insert, called "Naledi", also no longer publishes these pages. These low levels of written text give an idea of the seriousness of the problem confronting African lexicographers if they were to adopt the Western approach to corpus creation. They face practical constraints similar to those outlined above, such as a shortage of time and money, the unavailability of machine-readable text, and copyright restrictions.

Although there are few written texts in African languages, their existence does not guarantee that they are accessible to both native speakers and corpus researchers, or that the literate native speakers of the language read them. Many literate Africans rarely read texts in their own languages, although they may communicate extensively in them. The reason is not only because there is not enough written material in the African languages, but also because there is no culture of reading African literature in many African communities. African lexicographers therefore face great hurdles in attempting to access both written and spoken texts for corpus construction. In cases where they have access to

written texts, they run the risk of basing their research on the shaky foundations of the attitudes of language purists and prescriptivists who remain wedded to a linguistic world that has never existed.

This leads to the question of whether many corpora created for lexicographic research in Africa could be considered balanced and representative to the extent that they could be taken as bases for generalisations about the general language. This is greatly doubtful since most African corpora are biased towards one language variety as African languages are generally not used to render a variety of social contexts like the writing of laws, medical texts, government or official communications, academic books and business texts. Although these languages may not be used for *writing* about these topics and areas, in many occasions they are used to *speak* about them. A corpus of an African language constructed on a dependency on spoken texts is, however, likely to cover a rather restricted scope of language usage partly because of the unavailability of machine-readable data (MRD). It is also a well-known fact in natural language processing and computational linguistics that the transcription of spoken text is time-consuming and expensive, and cannot be afforded by many researchers, both Western and African. This further narrows the amount of text that could be included in many African languages corpora.

The Sanitised Data

Still on issues of written text, consideration need to be given to the involvement of publishers and editors and the power of stylebooks on the written word, resulting in what can be called "sanitised data". Many publishers and editors have very rigid principles of which words should be used in their publications. They are heavily prescriptive, as in the newspaper *Mokgosi*, for example, where the rare Setswana words *Mosupologo* (Monday), *Tshipi* (Sunday), *dira* (work, v.), and *kgwele* (ball) were preferred to the much more common *Mantaga*, *Sontaga*, *bereka*, and *bolo* respectively. Such preferences show the biased prescriptive stance adopted by numerous publishers and editors who believe that borrowed language is not authentic and not part of the language. Their control of language does not reflect how the people use language, but rather reflects *how they wish it to be used*. A dependency on such language for the construction of corpora brings serious questions to the kind of corpora whose results have to be generalised to the entire language. This is especially so since corpora provide information about what to include and exclude, guide the lexicographer towards sharper sense distinction, and assist in selecting corpus-based examples. While "sanitised data" may be unavoidable, it is greatly unsatisfactory for dictionary research where generalisations about language use must be made. Instead, it should be considered together with spoken texts to obtain a clearer picture of the language use of a speech community.

Conclusion

In this article, an attempt has been made to show that, while corpus research remains one of the most useful approaches to language research in that it can speedily offer information for addressing language-related issues and problems, a critical look at the process of corpus construction and inclusion would help determine if generalisations drawn from its results can be trusted as a true reflection of language use. The bias against spoken texts, for whatever reason, results in the greatest weakness of many corpora. The African context is unique in that, unlike Western communities, many African countries do not use their languages for academic purposes, in the media, and for governmental and official communication, making MRD difficult to access. Slow developments in computer software automatically changing spoken text into written text means that approaches to building corpora of spoken texts may remain challenged for a long time to come.

The future of a rigorous corpus research in Africa appears to be to approach issues of representativeness and balance with great caution. Kilgarriff and Grefenstette (2003: 334, 340), echoing Kennedy (1998: 62), state that "representativeness" begs the question, 'representative of what?' since, as they point out, "a corpus comprising the complete published works of Jane Austen is not a sample, nor is it representative of anything else". Although considered a language event, it is still unclear whether it is a matter of language production or of language reception. With the uncertainty surrounding matters of representativeness and balance, and with no convincing research of what precisely constitutes corpus material, it can be concluded with Kilgarriff and Grefenstette's (2003: 343) sentiments on web language that:

The Web is not representative of anything else. But nor other corpora, in any well-understood sense. Picking away at the question exposes how primitive our understanding of the topic is and leads inexorably to larger and altogether more interesting questions about the nature of language, and how it may be modeled.

For many African lexicographic projects there is a need to build organic corpora along the lines of the Bank of English (that currently has over 450 million words and continues to grow), which, in spite of attempts to update the corpus frequently to maintain a balance between written and spoken forms, does not claim to be balanced and representative. Such an approach would be sensitive to the current situation of many African languages that require a certain systematicity in their study, but would also recognise the fact that certain demands and expectations common to Western lexicography cannot be met in the African context. What goes into the compilation of a corpus must also be accounted for as much as what is extracted from it. In addition to pursuing corpus research, there is also a need for African lexicographers to look towards old and new approaches within theories of word meaning and analysis that would assist them in the collection and classification of words. A case in point is

WordNet, a University of Princeton's systematic analysis of words, whose design and execution were inspired by psycholinguistic theories of human lexical memory. It is crucial that lexicographers should not lose direction of what they want to achieve by sacrificing it to the quest of theoretical substantiality. The aim is to achieve the knowledge base of the lexical system of a language.

Note

- † A question mark is put after "geiti", borrowed from the English "gate", since Setswana does not have the voiced, velar plosive as part of its sound system, which in this instance occupies the initial word position in "geiti". There is therefore no agreed orthographic representation of such a sound in Setswana.

References

- Aitchison, J. 1998⁴. *The Articulate Mammal: An Introduction to Psycholinguistics*. London: Routledge.
- Biber, D. 1993 Using Register-Diversified Corpora for General Language Studies. *Computational Linguistics* 19(2): 219-241.
- Cole, D.T. 1955⁹. *An Introduction to Tswana Grammar*. Cape Town: Longman
- De Schryver, G.-M. and D.J. Prinsloo. 2000. Towards a Sound Lemmatisation Strategy for the Bantu Verb through the Use of *Frequency-based Tail Slots* — with Special Reference to Cilubà, Sepedi and Kiswahili. Mdee, J.S. and H.J.M. Mwansoko (Eds.). 2001. *Makala ya Kongamano la Kimataifa Kiswahili 2000: Proceedings*: 216-242, 372. Dar es Salaam: TUKI, Chuo Kikuu cha Dar es Salaam. Also available at: <<http://tshwanedje.com/publications/kiswahili2000fbts.pdf>>.
- Kennedy, G. 1998. *An Introduction to Corpus Linguistics*. London: Longman.
- Kilgarriff, A. 1997. Putting Frequencies in the Dictionary. *International Journal of Lexicography* 10(2): 135-155.
- Kilgarriff, A. 2000. Business Models for Dictionaries and NLP. *International Journal of Lexicography* 13(2): 107-118.
- Kilgarriff, A. and G. Grefenstette. 2003. Introduction to the Special Issue on the Web as Corpus. *Computational Linguistics* 29(3): 333-347.
- Leech, G. 1991. The State of the Art in Corpus Linguistics. Aijmer, K. and B. Altenberg (Eds.). 1991. *English Corpus Linguistics: Essays in Honour of Jan Svartvik*: 8-29. London: Longman.
- Leech, G., P. Rayson and A. Wilson. 2001. *Word Frequencies in Written and Spoken English: Based on the British National Corpus*. London: Pearson Education.
- Lichtenberk, F. 2003. To List or Not to List: Writing a Dictionary of a Language Undergoing Rapid and Extensive Lexical Changes. *International Journal of Lexicography* 16(4): 387-401.
- McEnery, T. and A. Wilson. 1996. *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Moe, R. 2003. Compiling Dictionaries Using Semantic Domains. *Lexikos* 13: 215-223.
- Nevegina, S.B. 1998. Some Problems of Borrowing in the Russian Language. *Vestnik Omskogo Universiteta* 1: 72-75.
- Ooi, V.B.Y. 1998. *Computer Corpus Lexicography*. Edinburgh: Edinburgh University Press.

- Renouf, A.** 1987. Corpus Development. Sinclair, J.M. (Ed.). 1987. *Looking Up: An Account of the COBUILD Project in Lexical Computing*: 1-40. London: Collins ELT.
- Sinclair, J.M.** 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sinclair, J.M. (Ed.)**. 1995². *Collins COBUILD English Dictionary*. London: HarperCollins.
- Summers, D.** 1993. Longman/Lancaster English Language Corpus — Criteria and Design. *International Journal of Lexicography* 6(3): 181-208.
- Summers, D. (Ed.)**. 1995³. *Longman Dictionary of Contemporary English*. Harlow: Longman.

Websites

Bank of English: <<http://www.titania.bham.ac.uk/docs/svenguide.html>>.

The British National Corpus: <<http://www.natcorp.ox.ac.uk/what/index.html>>

Die hantering van teenoorgesteldheid in enkele tipes woordeboeke, toegelig met Afrikaanse voorbeelde

Anna Nel Otto, *Departement Afrikaans en Nederlands, Nelson Mandela
Metropolitaanse Universiteit, Vistakampus, Port Elizabeth, Republiek van
Suid-Afrika (annanel.otto@nmmu.ac.za)*

Opsomming: Hoewel teenoorgesteldheid tot dusver deur baie leksikograwe as 'n bonus beskou is in die meeste tipes woordeboeke, word in hierdie artikel aangevoer dat hul insluiting verpligtend behoort te wees in aanleerderwoordeboeke en dat dit ook van groot nut kan wees in verskeie ander tipes woordeboeke. Eers word 'n uiteensetting gegee van verskillende tipes teenstellingsverhoudings. Daarna word die hantering van teenoorgesteldheid toegelig met Afrikaanse voorbeelde, in drie tipes woordeboeke bespreek met betrekking tot twee kategorieë: inligting in die toeligtingsgedeelte en in die alfabetiese gedeelte.

Met verwysing na die pluspunte sowel as die tekortkominge in die bespreekte woordeboeke, word voorstelle ter verbetering van die hantering van teenoorgesteldheid in hierdie woordeboeke gemaak, bv.: (1) die term *teenoorgesteldheid* behoort duidelik omskryf te word in die toeligtingsgedeelte, (2) die leksikale items waarby teenoorgesteldes aangedui kan word, behoort as lemmas op hulle alfabetiese plekke opgeneem te word, (3) die teenoorgesteldes van hierdie lemmas behoort by dié lemmas aangedui te word, en (4) die mees frekwente teenoorgesteldes kan deur middel van korpusontleding geïdentifiseer word.

Sleutelwoorde: AANLEERDERWOORDEBOEK, ALFABETIESE GEDEELTE, ANTONIMIE, DEFINIËRING, EENTALIGE WOORDEBOEK, KRUISVERWYSINGS, LEKSIKOGRAFIE, LEMMA, SEMANTIESE OPPOSISIE, TEENOORGESTELDHEID, TOELIGTINGSGEDEELTE, TWEENTALIGE WOORDEBOEK

Abstract: The Treatment of Opposites in a Few Types of Dictionaries, Illustrated by Afrikaans Examples. Although lexical items indicating semantic opposition have, until now, been regarded by many lexicographers as a bonus in most types of dictionaries, it is argued in this article that their inclusion should be compulsory in learner's dictionaries and that it can also be of great value in many other types of dictionaries. First an exposition is given of different types of semantic opposition. Then the treatment of opposition illustrated by Afrikaans examples, is discussed in three types of dictionaries by looking at two categories: the information in the usage guide and in the alphabetical part.

With reference to the strong points as well as the shortcomings in the discussed dictionaries, suggestions are made for the improvement of the treatment of semantic opposition in these dictionaries, e.g. (1) the term *semantic opposition* should be clearly defined in the usage guide, (2) the

lexical items indicating semantic opposition should be recorded in their alphabetical places as lemmata, (3) the opposites should be indicated at these lemmata, and (4) the most frequent opposites can be identified by using a corpus analysis.

Keywords: ALPHABETICAL PART, ANTONYMY, ANTONYMOUS RELATIONS, BILINGUAL DICTIONARY, CROSS-REFERENCES, DEFINING, LEARNER'S DICTIONARY, LEMMA, LEXICOGRAPHY, MONOLINGUAL DICTIONARY, SEMANTIC OPPOSITION, USAGE GUIDE

1. Inleiding

In die breër konteks van die leksikale semantiek maak antonimie deel uit van verskeie ander leksikale betrekkinge waarvan sinonimie, polisemie, homonimie en hiponimie as die bekendste semantiese verhoudings beskou word. *Antonimie* word in leketaal dikwels as oorhoofse term en sinoniem vir "teenoor-gestelde betekenis" gebruik (Gouws 1989: 158).

Ofskoon antonimie 'n natuurlike en vaste leksikale betrekking van taalstruktuur is, is daar volgens Palmer (1976: 78) tog 'n merkbare verskil in die status wat sinonimie en antonimie relatief ten opsigte van mekaar beklee. Hy beweer dat hierdie verskil in status veral sigbaar is in soverre antonimie in semantiekhandboeke soms minder aandag geniet, en in die meeste gevalle selfs geen plek in woordeboeke vind nie.

In hierdie artikel sal daar egter geprobeer word om aan te toon dat die insluiting van teenoorgesteldes nie op 'n lukrake wyse gehanteer behoort te word nie en dat dit, indien konsekwent aangedui, van groot waarde vir die woordeboekgebruiker kan wees. Vergelyk hier ook Gouws (1990: 58), wat tereg daarop wys dat daar nie op 'n gedetailleerde wyse aangedui hoef te word watter tipe teenstelling van toepassing is nie. Hy stel voor dat die leksikograaf 'n generiese etiket soos OPP gebruik om die mikrostrukturele elemente wat semantiese teenoorgesteldes van 'n lemma verteenwoordig, aan te dui.

Een van die kenmerke wat Nation (2001: 295) as belangrik beskou ten einde die gepaste woord in 'n aktiewe aanleerderwoordeboek te vind, is: "See if the dictionary provides opposites, synonyms, superordinates and other related words as part of an entry." Met betrekking tot aanleerderwoordeboeke ondersteun Moulin (1983: 149) onder andere kruisverwysings na sinonieme, teenoorgesteldes, e.s.m. by vertaalekwivalente, wat die leerder sal aanmoedig om paradigmatiese ondersoeke te doen. Paradis en Willners (2006) argumenteer dat woordeboeke oor die algemeen, maar veral aanleerderwoordeboeke, belangrike instrumente is by die aanleer van vreemde tale. Antoniemoeffeninge kom algemeen voor in taalaanleerkurrikula en kennis van antonimie is 'n noodsaaklike deel van tekstuele kompetensie (Halliday en Hasan 1976). Jones (2002: 118-119) raam dat een uit elke vyftig sinne 'n antoniempaar bevat en het ook soortgelyke resultate vir gesproke Engels gevind (vgl. Jones 2006).

Aangesien semantiekteorie 'n goeie basis vir die leksikografie is, word daar eerstens 'n teoretiese bespreking van die verskillende teenstellingsbetrek-

kinge aangebied. Daarna word dit opgevolg met 'n toegepaste ondersoek na die hantering van hierdie betrekkinge in 'n drietal verskillende woordboek-tipes.

2. 'n Uiteensetting van verskillende teenstellingsbetrekkinge

2.1 Antonimie

Volgens De Stadler (1989: 81 e.v.) is daar verskillende tipes teenoorgesteldheid, waarvan antonimie maar een is.

Die meeste linguïste definieer antonieme in terme van gradeerbaarheid by adjektiewe (vgl. Palmer 1976, Lyons 1977, Lehrer en Lehrer 1982, Cruse 1986 en De Stadler 1989).

- (a) De Stadler (1989: 82) wys daarop dat verskeie kenmerke van antonieme voortvloei uit die feit dat hulle gradeerbaar is:
 - (i) Hulle kan trappe van vergelyking neem, bv. *mooi, mooier* en *kort, korter*, ens.
 - (ii) Hulle kan met graadwoorde verbind, bv. *baie mooi, te kort*, ens.
 - (iii) Indien die een gestel word, word die ander ontken, bv. *Sy is lank* impliseer *Sy is nie kort nie*, maar die omgekeerde is nie waar nie: die ontkenning van die een beteken nie noodwendig dat die ander daarmee gestel word nie: *Sy is nie kort nie* impliseer nie noodwendig *Sy is lank* nie.
 - (iv) Eksplisiete of implisiete gradering (vergelyking in terme van 'n norm) kom altyd voor. Daarom beteken *Ons huis is groot* eintlik "Ons huis is groter as die normale huis".
 - (v) Antoniepare het 'n positiewe en 'n negatiewe pool. Die positiewe pool verteenwoordig die ongemerkte waarde van die betrokke eienskap. By die eienskap hitte, is die positiewe pool byvoorbeeld *warm* en die negatiewe pool *koud*.
- (b) Volgens Muehleisen (1997: 8) het antonieme ook die eienskap "committedness", hier vertaal as "geïmpliseerde waarde". 'n Adjektief het 'n geïmpliseerde waarde indien dit 'n bepaalde waarde impliseer wanneer dit in 'n vraag gebruik word, en dit is "impartial/uncommitted", hier vertaal as "neutraal", indien dit nie so 'n implikasie het nie. *Lank* is byvoorbeeld neutraal in 'n vraag soos *Hoe lank is Jan?* Hierdie vraag is neutraal en kan gebruik word ongeag of die spreker Jan se benaderde lengte ken en ongeag of Jan lank, kort of van gemiddelde lengte is. Hierteenoor het die adjektief *kort* 'n geïmpliseerde waarde. 'n Spreker sal slegs vra *Hoe kort is Jan?* as daar rede is om te glo dat Jan korter is as die normale lengte.

(Vgl. ook De Stadler 1989: 83.) Muehleisen (1997: 8) wys verder daarop dat baie pare gradeerbare antonieme een item met 'n geïmpliseerde waarde en een neutrale term bevat, bv. *oud* x *jonk*, *swaar* x *lig*, *vinnig* x *stadig*, maar baie ander pare bevat weer twee items met geïmpliseerde waardes, bv. *onskuldig* x *skuldig*, *mooi* x *lelik*, *gelukkig* x *hartseer*, ens.

- (c) Gemerktheid is volgens Muehleisen (1997: 8) 'n sambreelterm vir verskeie verwante verskynsels wat die gemerkte lid van 'n antoniempaar van die ongemerkte lid onderskei. Lehrer (1985) bespreek verskeie kriteria wat voorgestel is om gemerktheid te definieer. Geïmpliseerde waarde is een van hulle: die neutrale lid van 'n antoniempaar word as ongemerk beskou en die lid met geïmpliseerde waarde gemerk is, gevolglik is *oud* ongemerk, terwyl *jonk* gemerk is. Indien die naam van die semantiese skaal morfologies verband hou met een van die antonieme, dan is dit verwant aan die ongemerkte lid, bv. die naam van die skaal vir lengte is verwant aan die ongemerkte *lank* eerder as die gemerkte *kort*.

Nog 'n kenmerk van gemerktheid is dat die ongemerkte antoniem oor die algemeen in 'n wyer reeks sintaktiese kontekste kan optree; ongemerkte antonieme veral kan met maatfrases voorkom, maar gemerktes gewoonlik nie. So kan 'n mens byvoorbeeld sê iets is drie meter *lank*, maar nie drie meter *kort* nie. Soortgelyk is verhoudings gewoonlik slegs moontlik met die ongemerkte antoniem; 'n mens kan sê *Karin is twee keer ouer as Petro*, maar nie *Petro is twee keer jonger as Karin* nie. Morfologie speel ook 'n rol: in pare waarin een antoniem afgelei is van die ander, word gesê dat die afgeleide lid gemerk is, dus is *gelukkig* ongemerk en *ongelukkig* gemerk.

- (d) Volgens De Stadler (1989: 83) veronderstel gradeerbaarheid verskillende waardes op 'n skaal, dikwels nie net twee soos wat deur 'n antoniempaar veronderstel word nie. 'n Mens kan jouself dan afvra watter waardes op 'n skaal as teenoorgesteldes van mekaar beskou kan word. Hy bespreek die semantiese veld rondom die antoniempaar *baie* en *min* deur na die volgende woorde te verwys: *oorvloedig*, *volop*, *baie*, *heelwat*, *genoeg*, *bietjie*, *min*, *weinig*, *niks*. Die woord *genoeg* kan waarskynlik as die middelpunt van die skaal beskou word en op grond hiervan kan antonimie gedefinieer word as "twee punte op 'n skaal van gradeerbares op gelyke afstande aan weerskante van die middelpunt". Gevolglik kan die volgende items antoniempare wees: *oorvloedig* x *niks*, *volop* x *weinig*, *baie* x *min* en *heelwat* x *bietjie*. Hierdie voorlopige definisie kan egter nie korrek wees nie, aangesien die volgende pare ook as teenoorgesteldes beskou kan word: *genoeg* x *oorvloedig*, *volop* x *bietjie*, *baie* x *weinig*, *heelwat* x *min*, *oorvloedig* x *bietjie*. De Stadler (1989: 84) kom dan tot die gevolgtrekking dat antonieme nie noodwendig reglynige teenoorgesteldes op 'n skaal is nie, maar op verskillende afstande van die middelpunt af kan lê, en selfs in 'n ver-

houding van antonimie met die middelpunterm self kan tree. Die enigste kriterium is dat die lede van 'n antoniempaar aan weerskante van of op die middelpunt van die skaal moet lê.

- (e) De Stadler (1989: 84) wys daarop dat die feit dat 'n leksikale item verskillende polisemiese waardes (of kombinatoriese waardes) kan hê, ook daartoe lei dat daar vir een woord verskillende teenoorgesteldes aangebied kan word. Vergelyk byvoorbeeld die verskillende polisemiese waardes van die woord *soet*.
- (i) wat 'n aangename suikersmaak het: *soet* water (sout, brak), *soet* patats (sout), *soet* verversings (sout), *soet* koffie (bitter), *soet* druiwe (suur), *soet* wyn (droë)
- (ii) baie aangenaam, veral vir die gehoor en vir die reuk: *soet* tone (wanklinkende), *soet* geur (walglike), *soet* drome (slegte, nare)
- (iii) gehoorsaam, nie stout nie, volgsaam: *soet* kind (stout, ongehoorsame)

Uitdr. Die *soet* en *suur* van die lewe

2.2 Ander tipes teenoorgesteldheid

2.2.1 Binêre opposisies

Die tipe teenstelling wat die grootste mate van ooreenkoms met antonieme het, is komplementêre pare. Cruse (1986: 198) beskryf laasgenoemde soos volg: "The essence of a pair of complementaries is that between them they exhaustively divide some conceptual domain into two mutually exclusive compartments, so that what does not fall into one of the compartments must necessarily fall into the other." De Stadler (1989: 84) wys dan ook tereg daarop dat komplementêre pare nie gradeerbaar is nie en indien die een gestel word, word die ander ontken, en indien die een ontken word, word die ander gestel: *Hy is manlik* impliseer naamlik *Hy is nie vroulik nie* en *Daardie persoon is nie manlik nie* impliseer *Daardie persoon is vroulik*.

Dit is wel so dat *vroulik* soms gradeerbaar is (*vrouliker, vroulikste*), maar hierdie gradeerbaarheid vloei voort uit die verskillende polisemiese waardes van die betrokke leksikale item, in Engels geleksikaliseer as *female* en *feminine* onderskeidelik (De Stadler 1989: 85). Hieruit blyk dat verskillende polisemiese waardes van dieselfde leksikale item selfs in verskillende tipes teenoorgesteldheid gevind kan word.

'n Volgende tipe teenstellingsbetrekking is omgekeerdheid. By woordpare soos *man* x *vrou*, *vader* x *seun*, *vader* x *dogter*, *dokter* x *pasiënt*, *bo* x *onder* en *koop* x *verkoop* kry 'n mens voorbeelde van omgekeerdheid (De Stadler 1989: 85).

Omgekeerdheid geld op sinsvlak. Indien X 'n pasiënt is van Y, dan moet Y die dokter wees van X. Die een lid ontken nie die ander lid nie, maar impliseer juis dat dit bestaan. (Vgl. Cruse 1986: 231.)

'n Laaste tipe teenstelling is ruimtelike opposisie. Ruimtelike opposisie kom voor waar daar 'n verandering van toestand of plek is, bv. *op x af, in x uit, kloksgewys x antikloksgewys*, maar ook woorde soos *vasmaak x losmaak, trou x skei* en *verskyn x verdwyn*.

2.2.2 Opposisie en kontras: verenigbaarheid en onverenigbaarheid

In die voorafgaande afdeling is gewerk met binêre opposisies. Telkens is daar gewys op teenoorgesteldheid van leksikale items se betekenis met verwysing na woordpare.

Daar bestaan egter ook ander verhoudings in die leksikon wat onder die dekterm kontras kan ressorteer en wat nie-binêre kontraste verteenwoordig. De Stadler (1989: 86) verdeel hierdie kontraste in verenigbare en onverenigbare kontraste.

In die geval van verenigbare kontraste kan die leksikale versamelings almal beskou word as ko-hiponieme van dieselfde superordinaat; so dui *noord, suid, oos, wes*, ens. almal op rigtings. Vergelyk ook die dae van die week, maande van die jaar, seisoene, kleure, verskillende tipes diere, bome, vrugte, groente, blomme, ens.

Binne hier versamelings verenigbares is daar woordpare wat as teenoorgesteldes (d.w.s. binêre opposisies) van mekaar beskou kan word. By die kleure kan *swart* byvoorbeeld as die teenoorgestelde van *wit* beskou word.

Teenoor die verenigbares staan die onverenigbares, soos in die volgende voorbeeld waar die items se betekenis onverwant is: *kind, hardloop, vinnig*.

Idiomatiese teenoorgesteldes vind 'n mens by woordpare soos *tee x koffie, sout x peper, mes x vurk, vis x skyfies, koppie x ptering*, wat ook verstenings genoem word.

Hierdie woordpare word tradisioneel as teenoorgesteldes van mekaar beskou, hoewel hulle dit logies gesien nie is nie (De Stadler 1989: 87).

2.3 Kwasi-teenoorgesteldes

Cruse (1986) en Lehrer en Lehrer (1982) bespreek kwasi-teenoorgesteldes. Cruse beskryf onsuiver teenoorgesteldes as teenoorgesteldes wat in hul betekenis 'n meer elementêre teenoorgesteldheid omvat of insluit. Cruse (1986: 198) sê byvoorbeeld: "*Giant: dwarf* can be said to encapsulate the opposition between *large* and *small* (but this opposition does not exhaust their meaning); likewise, *shout* and *whisper* encapsulate *loud* and *soft*, *criticize* and *praise* encapsulate *good* and *bad*."

2.4 Kriteria vir teenoorgesteldes (vs. kwasi-teenoorgesteldes)

Die volgende kriteria vir teenoorgesteldes is deur Cruse (1986) en Lehrer en Lehrer (1982) voorgestel:

- (a) Suiwerheid van die semantiese opposisie: In sommige pare kwasi-teenoorgesteldes (bv. *fluister* x *skreeu*) put die semantiese opposisie nie die betekenis van die woorde uit nie. Die implikasie is dat die semantiese kontras in prototipiese pare teenoorgesteldes in 'n sekere sin die betekenis van die woorde uitput.
- (b) Maklike waarneming van 'n semantiese dimensie: Met kwasi-teenoorgestelde pare soos *dorp* x *platteland*, is dit moeilik om te bepaal wat die relevante semantiese dimensie kan wees. Die implikasie is dat die semantiese dimensie maklik geïdentifiseer kan word vir prototipiese teenoorgesteldes.
- (c) Ooreenkoms ten opsigte van niepropositionele betekenis: Sommige kwasi-teenoorgesteldes (bv. *plomp* x *uitgeteer*) het baie verskillende konnotasies. Die implikasie is dat prototipiese teenoorgesteldes baie eenders is in terme van niepropositionele betekenis.
- (d) Afstand vanaf die middelpunt van 'n semantiese dimensie: In sommige gevalle lyk dit asof een lid van 'n paar van kwasi-teenoorgesteldes (bv. *afskuwelik*) 'n meer ekstreme waarde as die ander lid (*lelik*) het. Die implikasie is dat prototipiese teenoorgesteldes op gelyke afstandpunte van 'n middelpunt lê.
- (e) Ooreenkoms in verspreiding: By beide *big* en sy antoniem *little* kan verwys word na relatiewe ouderdom eerder as fisiese grootte in konstruksies soos *big sister* en *little brother*, maar die kwasi-teenoorgestelde van *little*, dit is *large* kan net verwys na fisiese grootte saam met die s.nwe. *sister* en *brother*.

Opposisiebetrekkinge kan op verskillende vlakke voorkom: leksikaal soos reeds aangedui, subleksikaal soos in die volgende twee paragrawe, asook multi-leksikaal, bv. *Hy/Sy is ouer as gister* x *Hy/Sy het nie buskruit uitgevind nie*. Sodanige betrekkinge kan ook oor die grense van leksikale status heen reik, soos in die geval van *Hy/Sy is slim* x *Hy/Sy het nie buskruit uitgevind nie*.

Teenstellende korrelasies kom ook by morfologiese boustukke voor: *in-* x *uit-*, *hoof-* x *by-*, *-vriendelik* x *-onvriendelik*.

Teenstellende leksikale items kan verder op grond van struktuur in verskillende soorte verdeel word. Hulle kan in leksikale (*open* x *sluit*) en morfoleksikale (*oopmaak* x *toemaak*) tipes verdeel word. By die morfoleksikale lê die

teenstellende elemente in die morfologiese bouelemente: by affikse: *antedateer* x *postdateer*, of met 'n nulplek: *vriendelik* x *onvriendelik*, en by die suffikse: *stemhebbend* x *stemloos*. Dan is daar ook nog die kompositumteenstellings, waar kompositum in die wydste sin bedoel word. Die teenstelling is óf by die eerste element (*blyspel* x *treurspel*) óf by die tweede element (*belastingverhoging* x *belastingverlaging*). In uitsonderlike gevalle kom die teenstelling by beide elemente voor: *afstoot* x *aantrek*. (Vgl. Müller 1989: 2794-2795.)

Een betekeniswaarde van 'n leksikale item het in die meeste gevalle net een antoniem (*sterk* x *swak*); dit is veral waar ten opsigte van die morfoleksikale woorde (*ingaan* x *uitgaan*). Soms is daar egter meer as een antoniem: sowel by leksikale antoniempare (*koud* x *warm*, *heet*) as by morfoleksikale items (gemeng): *mobiel* x *immobiel*, *onbeweeglik*. (Vgl. Müller 1989: 2794-2795.) Hier moet naamlik onderskei word tussen 'n teenstellingswoord, waar daar formele ooreenstemming is, soos by *mobiel* x *immobiel*, teenoor *onbeweeglik* wat nie vormlik ooreenstem nie en slegs 'n teenstellingsveldwoord is.

3. Die gebruik van antonieme in verskillende tipes woordeboeke

Teenstellings kan van groot nut wees in minstens die volgende tipes woordeboeke: eentalige verklarende woordeboeke, tweetalige woordeboeke, en eentalige antoniemwoordeboeke of sinoniemwoordeboeke met antonieme. In die volgende drie afdelings sal die behandeling van teenstellings in elkeen van hierdie tipes woordeboeke bespreek word, toegelig met Afrikaanse voorbeelde in soverre hulle voorhande is.

3.1 Eentalige verklarende woordeboeke

Hierdie tipe woordeboeke, wat onder andere omvattende woordeboeke, standaardwoordeboeke en aanleerderwoordeboeke kan insluit, bied sinonieme en teenstellings aan as deel van hulle semantiese inligting of selfs as die enigste semantiese inligting (vgl. Bergenholtz en Tarp 1995: 126). Die aanbieding van teenstellings sal ondersoek word in die *Woordeboek van die Afrikaanse Taal* (voortaan WAT), die *Handwoordeboek van die Afrikaanse Taal* (voortaan HAT) en die *Basiswoordeboek van Afrikaans* (voortaan BWA) onderskeidelik.

Sedert die verskyning van die eerste deel van die WAT het daar 'n merkbare verbetering in die hantering van inligting oor teenoorgesteldes in die toeligtingsgedeelte ingetree. Raidt (1993: 217) wys reeds ten opsigte van die agtste deel van die WAT daarop dat byvoorbeeld die hantering van sinonieme en antonieme in die toeligtingsgedeelte uiteengesit word. Feinauer (1996: 260) wys ook daarop dat die WAT bedag is op sensitiwiteit rondom geslag. Vroulike vorms in geslagsopposisiepare word dus wel gelemmatiseer en volledig bewerk in WAT IX. Nietemin kom daar nog enkele inkonsekwente gevalle voor (vgl. Feinauer 1996: 262).

In die toeligtingsgedeelte van deel I is teenstellings nie verklaar nie en daar is ook nie aangedui of en hoe hierdie tipe semantiese betrekking in die WAT gehanteer word nie. In die toeligtingsgedeelte van die jongste deel, deel XII, vind 'n mens nie die term *antonimie* nie, maar wel *semantiese opposisie*, wat dus op 'n wyer bestek dui. Dit word soos volg omskryf:

'n Aanduiding van semanties teenstellende en maklik met die lemma verwarbare leksikale items. Die formules wat hier gebruik word, is "teenoor" vir semanties teenstellende leksikale items of "onderskei van" vir leksikale items wat maklik met die lemma verwar kan word, bv. onder **normaalhorend**: teenoor *gehoorgestrem*, en onder **oumawoonstel**: onderskei van *eenmanswoonstel*. Vgl. verder paragraaf 5.4.

In paragraaf 5.4 word dan aangedui waar en hoe teenoorgesteldes opgeneem word:

Kruisverwysing na formele antonieme en verwante of maklik verwarbare leksikale items vind deur middel van vermelding by die onderskeie leksikale items plaas. Hierdie vermelding van die sogenaamde semantiese teenstellings en verwante vorme dien as inligting dat die leksikale item 'n bepaalde antoniem het.

Die semantiese teenstellings en verwante vorme word vermeld na die definisie, maar indien daar sinonieme is, dan na die sinonieme, en word voorafgegaan deur 'n kommapunt. Antonieme en ander semanties teenstellende vorme soos komplementêre leksikale pare word deur die term "teenoor" aangedui, terwyl vorme wat soms met die lemma verwar word deur die term "onderskei van", met of sonder verdere kwalifikasies, voorafgegaan word. Leksikale items wat betekenismatig redelik na aan die lemma is, maar nie as sinoniem daarvan beskou kan word nie, word soms ook met die term "onderskei van" aangedui. Die semantiese verband is egter vaer as by die voorbeelde wat onder 5.5 gehanteer word met "Vgl."

Louw (2000: 135) wys in navolging van Wiegand (1982: 133) daarop dat nie slegs sinonimie nie, maar ook ander semantiese verhoudings in 'n omvattende woordeboek met verskeie volumes, soos die WAT, aangedui behoort te word. Volgens Wiegand (1982: 133) is die aandui van sinonieme, naas die definisie, in eenvolume-handwoordeboeke (soos HAT) voldoende, maar antonimie, hiponimie en ander semantiese verhoudings behoort in omvattende woordeboeke apart van die betekenisverklaring gegee te word en duidelik aangedui te word deur 'n unieke strukturele merker.

Oor die algemeen tree die WAT voorbeeldig op deur konsekwent alle lede van antoniem- en ander teenstellingspare in die dele wat reeds verskyn het, op te neem. By **meerderjarig** is daar byvoorbeeld 'n verwysing na **minderjarig** en omgekeerd. Daar is wel enkele uitsonderings by die steekproef aangetref. Die steekproef is gebaseer op die leksikale items in Müller (1989: 2795) in soverre dit in Afrikaans vertaalbaar is. By **aankoms** is daar nie 'n verwysing na **vertrek** nie. Dit geld ook die werkwoord **aankom**. By **afstoot** is daar 'n verwysing na

aantrek, maar by **aantrek** is daar nie 'n verwysing na **afstoot** nie. By **lag** is daar 'n verwysing na **huil**, maar by **huil** is daar nie 'n verwysing na **lag** nie.

Teenstellings word ook soms in die definisie gehanteer, maar nie op die wyse soos aangedui in die toeligtingsgedeelte van die *WAT deel XII*, naamlik ingelei deur die woord "teenoor", nie. Dit is nie duidelik waarom die teenstelling hier gebruik word nie. Vergelyk in hierdie verband byvoorbeeld **buitekant** ten opsigte van die leksikale teenstelling:

buitekant. I s. Buitesy, die uitwendige van iets: *Verf die kis net a/d buitekant.*

Ook fig.: *Iets net van die buitekant beskou, oppervlakkig.*

II bw. Buite, nie binne nie: *Wag jy solank buitekant tot die deur oopgaan.*

III vs. Buite: *Buitekant die huis is dit koud.*

Soms word die teenstelling ook net in die voorbeeldsin(ne) genoem, bv. by **binnekant** waar dit die teenstelling illustreer:

Buitekant is dit koud, maar binnekant b/d vuur lekker warm.

Laasgenoemde verskynsel kom ook by die komplementêre teenstellende items voor. Vergelyk **broer**:

Ons is twee broers en drie susters.

En in enkele gevalle kom die teenstelling in 'n uitdrukking voor, bv. by **huur**:

Huur en verhuur, ooreenkoms waarby die een party a/d ander die genot van 'n saak gedurende 'n beperkte tyd en teen 'n bepaalde prys afstaan.

In die toeligtingsgedeelte van die HAT word in 'n enkele sin onder andere oor antonieme berig:

Naas die betekenis, maar nie in die plek daarvan nie, word dikwels sinonieme gegee en, waar nodig, ook antonieme.

Hoewel die HAT nie in die toeligtingsgedeelte aandui hoe teenstellings gehanteer word nie, kom 'n mens gou agter dat 'n stelsel soortgelyk aan dié van die WAT, naamlik om "teenoor" te gebruik, ook hier van krag is. In 'n meerdere mate verskyn teenoorgesteldes dikwels slegs in die definisies. Vergelyk byvoorbeeld **jonk, vars, maer, gesond, sag, dood, naby, lelik, passief, buitekant, ongesond**, e.s.m. By **ongesond** kry 'n mens:

ongesond b.nw. en bw. 1 Nie gesond nie; met 'n swak of verswakte gestel: *'n Ongesonde kind, mynwerker.* 2 Skadelik vir die gesondheid en die gees: *'n Ongesonde klimaat. 'n Ongesonde lewe lei. Ongesonde lektuur. 'n Ongesonde belangstelling in seks.* 3 Nie gaaf nie: *Ongesonde bome, vrugte.*

Hier sou 'n verwysing na die spesifieke betekenisonderskeiding, naamlik **gaaf**³ duideliker leiding kon verskaf.

In die toeligtingsgedeelte van die BWA word slegs gemeld dat kruisverwysings aan die einde van 'n paragraaf verskyn en gemerk word met die woord "teenoor". Die voorbeeld wat in die sleutel vir gebruikers gegee word, is: **aanleer** ... Teenoor: **afleer**.

Kruisverwysings kan om verskillende redes in woordeboeke gebruik word. Hulle kan die gebruiker byvoorbeeld waarsku dat daar verwante woorde is wat nie met die betrokke lemma verwar moet word nie. Hulle kan ook verwys na ander woorde wat meer inligting bevat. Duidelikheidshalwe behoort die samestellers te vermeld dat kruisverwysings slegs gebruik word om antonieme aan te dui en dat die kruisverwysings ingelei word deur die woord "teenoor" aan die einde van paragrafe. Alternatiewelik: Indien ander leksikale betrekkinge ook deur kruisverwysings geëkspliseer word, behoort die verwysings dienoreenkomstig gemerk te word.

Aanleerderwoordeboeke bevat gewoonlik 'n gekontroleerde, basiese woorde-skat. Om hierdie rede kan 'n mens verstaan dat bepaalde laefrekwensiewoorde nie in die BWA opgeneem is nie. Vergelyk *meerderjarig* x *minderjarig*, *minimaal* x *maksimaal* en *minimum* x *maksimum*. Indien 'n woord soos *aktief* as lemma opgeneem is, is dit nie duidelik waarom die teenoorgestelde *passief* nie ook as lemma opgeneem is nie. Dit geld ook byvoorbeeld *aantrek*, maar nie *afstoot* nie, *ondertoe*, maar nie *boontoe* nie, *meervoud*, maar nie *enkelvoud* nie, *oopmaak* maar nie *toemaak* nie en *afneem*, maar nie *toeneem* nie.

Soms is beide teenoorgesteldes as lemmas opgeneem, maar die teenoorgestelde word net by een van die twee lede aangedui. So is daar byvoorbeeld by **lelik** 'n aanduiding dat die teenoorgestelde *mooi* is, maar by **mooi** is daar nie 'n aanduiding van *lelik* as teenoorgestelde nie. Vergelyk ook *oud* x *vars*, *siek* x *gesond*, *lewend* x *dood*, *swart* x *wit*, *manlik* x *vroulik*, *amateur* x *beroepspeleer*, *toelaat* x *verbied*, *haan* x *hen* (net in die voorbeeld genoem), en *seun* x *dogter* (net in die voorbeelde genoem). Die voorbeeldmateriaal illustreer hier teenstellings wat slegs op bepaalde betekenisonderskeidings betrekking het.

Dit wil voorkom asof die BWA soms baie duideliker leiding gee as die WAT of die HAT. Vergelyk die lemma **aanstaande**. In sowel die WAT as die HAT word geen teenoorgesteldes aangedui nie. By die BWA word die volgende voorbeeldsinne gegee: "*Hy kom aanstaande Maandag op die lughawe aan. Die aanstaande verkiesing beloof om baie interessant te wees. Aanstaande week gaan ons stad toe.*" Hierop volg dan die kruisverwysing, met 'n aanduiding van tipiese kollokators: "Teenoor: **laas** (Maandag, week), **vorige** (verkiesing)". Dit is net jammer dat daar nie by **laas** en **vorige** kruisverwysings na **aanstaande** en **volgende** is nie.

3.2 Tweetalige woordeboeke

In tweetalige woordeboeke (vgl. Hausmann 1991: 2794-2795) help sinonieme en antonieme met betekenisonderskeiding en teksproduksie en besit hulle ook 'n aanleerfunksie. Die hantering van teenstellings in Pharos se *Afrikaans-Engels*

Woordeboek/English–Afrikaans Dictionary (Du Plessis 2005) (voortaan AEW) sal bespreek word in vergelyking met ander tweetalige Afrikaans–Engelse woordeboeke.

Wat die toeligtingsgedeelte betref, is daar 'n verbetering deurdat die AEW uiteensit hoe manlike en vroulike vorme in dié woordeboek gehanteer word. In vorige/ander tweetalige woordeboeke word dit nie aangedui nie. Daar word genoem dat manlike en vroulike vorme as afsonderlike lemmas opgeneem word en, waar nodig, van die etikette *masc.* (manlik) en *fem.* (vroulik) voorsien word. Die wyse waarop ander tipes teenstellingsbetrekkings gehanteer word, word egter ook nie in die AEW verduidelik nie.

'n Seleksie van twee geslagsopposisiepare (vgl Beyer 1997: 108) met hul onderskeie vertaalekwivalente word vervolgens bespreek soos hulle telkens in AEW, die *Tweetalige Woordeboek* (Bosman, Van der Merwe en Hiemstra 1984) (voortaan TW), die *Groot Woordeboek* (Eksteen 1997) (voortaan GW) en die *Tweetalige Aanleerderswoordeboek* (Du Plessis 1993) (voortaan TA) gehanteer is.

Die eerste lemapaar wat bespreek sal word, is: **koning** x **koningin** teenoor **king** x **queen**. In die bespreking word daar slegs gefokus op dié polisemiese waardes waar daar sprake is van geslagsopposisiepare. So staan *koning* en *koningin* glad nie in 'n geslagsopposisieverhouding ten opsigte van die betekeniswaarde "stuk in die skaakspel" nie, maar in 'n hiërargiese kontrastiewe verhouding (vgl. De Stadler 1989: 86-87).

In AEW word **koning** en **koningin** inderdaad as afsonderlike lemmas opgeneem en in dié twee lemmas se artikels word nie na die teenoorgestelde woord verwys nie. By **king** is daar wel 'n kruisverwysing na **queen** en by **queen** weer 'n kruisverwysing na **king**. Dit gee goeie leiding, aangesien **king** en **queen** alfabeties ver van mekaar verwyder is.

In die Afrikaans–Engelse deel van die TW is daar by die lemma **koning** nêrens 'n verwysing na die vroulike geslagsvorm *koningin* nie, selfs nie eers in die lys van vertaalekwivalente of die sublemmas van **koning** nie. Daar sou geredeneer kon word dat die insluiting van so 'n verwysing in die artikel van **koning** as oorbodig beskou kan word, omdat die lemma **koningin** direk onder die lemma **koning** gelys word. Dieselfde geld vir die lemma **koningin** waar daar ook geen verwysing na **koning** is nie. Dit is egter opvallend dat daar in die Engels–Afrikaanse deel, alhoewel nie eksplisiet nie, tog telkens 'n aanduiding van die komplementêre lid in beide die artikels se inskrywings onder **king** en **queen** verskaf word. Onder **king** word *Queen* en *Queen's* herhaaldelik in die voorbeeldmateriaal as teenstelling opgeneem, maar die aanbod daarvan is problematies, want volgens konvensie dui hakies ook op opsionele aanvullings. Kontekstueel gaan dit in die inskrywings ook oor wie die heerser is, en word die inskrywings dus nie noodwendig so aangebied om die opposisiebetrekking te impliseer nie. Dit kon byvoorbeeld ook "king (prince)" wees, wat 'n ander soort opposisieverhouding kan enkapsuleer.

Ten opsigte van die Afrikaans–Engelse deel bied die GW, anders as die TW, **koning** en **koningin** nie as aparte artikels onder mekaar aan nie. Die mor-

fologiese verbuiging van *koning* na *koningin* word as 'n sublemma onder **koning** ingesluit.

In die Engels–Afrikaanse deel van die GW word daar weer eens, net soos in die geval van die Afrikaans–Engelse deel, geen kruisverwysings by die lemma, of sublemmas van **king** na **queen** gemaak nie. Daar is wel 'n potensieel implisiete verwysing by **queenconsort** na **koning**, mits die gebruiker weet om *vors* met *koning* as sinoniem te vervang. Vergelyk die volgende artikel:

queen, (n), koningin; vrou (kaart); koningin, dame (skaak); homoseksueel; [...] ~ **consort**, vrou v.d. regerende vors; [...].

Daar is 'n bykomende probleem betreffende geslagsaanduiding in hierdie artikel van **queen**. Die vertaalekwivalent *homoseksueel* in die sinoniemparadigma, soos hierbo gelys, behoort geëtiketteer te word met byvoorbeeld (*pej.*, *slegs manlike geslag*). (Vgl. verder Gouws 1989: 243-244 in hierdie verband.) Die AEW tree hier voorbeeldig op deur **queen** soos volg te merk: (*neerh. sl.: homoseksueel*).

Koning en **koningin** word in die TA, soos ook in die mikrostruktuur van die TW, as twee afsonderlike artikels in die Afrikaans–Engelse deel onder mekaar geplaas. Hierdie teenstellingspaar is noodgedwonge in die Engels–Afrikaanse deel alfabeties uiteengeplaas. Kruisverwysings kom egter nie in die TA voor nie, en dit is 'n ongelukkige tekortkoming van hierdie woordeboek, omdat aanleerders van 'n vreemde taal soveel te meer hulp met geslagsaanduiding behoort te ontvang, en waarskynlik hierop staatmaak.

Die volgende lemmapaar wat bespreek sal word, is: **onderwyser** x **onderwyseres** teenoor **teacher** x **Ø**.

By die AEW is daar 'n inkonsekwentheid met die hantering van die geslagsvorme wat deur middel van fleksieverbuiging gevorm word. Terwyl **koning** en **koningin** in die Afrikaans–Engelse deel, soos reeds vermeld, as twee afsonderlike lemmas gehanteer word, is dit nie die geval by **onderwyser** en **onderwyseres** nie. **Onderwyseres** word as 'n sublemma van **onderwyser** by die artikel ingeskryf. Dieselfde gebeur by die TW. Die geslagsopposisiepaar *onderwyser* x *onderwyseres* is 'n komplekse voorbeeld, want daar is ook sprake van die geslagtelik neutrale onderwyser. Die AEW en ander woordeboeke se hantering kan hier gekritiseer word, aangesien geensins vir die geslagtelik neutrale referent van die leksikale item *onderwyser* voorsiening gemaak word nie.

Die GW tree wel konsekwent op: in die Afrikaans–Engelse deel word weer eens 'n fleksieverbuigingsvorm **~es** gebruik om die vroulike geslagsvorm in die artikel onder **onderwyser** as lemma aan te gee.

In die Afrikaans–Engelse deel sien die artikel soos volg daar uit:

onderwyser, (-s), teacher, instructor, educator; dominie; master, preceptor; **~es**, (-se), teacher, mistress; [...]

In die eerste plek is daar geen pertinente aanduiding van die leksikale gaping wat daar in Engels ten opsigte van die vroulike geslagsvorm van *teacher* be-

staan nie. **Onderwyseres** behoort gemerk te word met (*female*) vooraan *teacher*. Te oordeel na die uiteensetting van die woordeboekinskrywing soos hierbo weergegee, sal die gebruiker dalk wonder waarom *master* nie deel vorm van die eerste sinoniemparadigma nie en verder waarom dit nie met (*school*) vooraan gemerk is nie. Dieselfde geld vir *mistress*. Die woord *mistress* op sigself het nie juis positiewe konnotasies nie, gevolglik behoort (*school*) vooraan *mistress* gevoeg te word.

Die AEW hanteer **onderwyser** en **onderwyseres** wel inkonsekwent deur dat daar nie gehou word by die sisteem (volgens die toeligting) om aparte lemmas te gebruik nie, maar andersins is die uiteensetting baie duideliker as byvoorbeeld dié van die GW, hoewel daar nie voorsiening gemaak word vir die neutrale vorm **onderwyser** nie. Vergelyk:

onderwyser *=sers, (masc.)*, **onderwyseres** *=resse, (fem.)* (school)teacher, (*chiefly Br.*) (school)master (*masc.*), schoolmistress (*fem.*); [...]

Die TA lys die twee komplementêre leksikale items **onderwyser** en **onderwyseres** konsekwent onder mekaar as sublemmas in die Afrikaans–Engelse deel van die woordeboek en nie as aparte lemmas, soos deur De Wet (1998: 158) voorgestel word nie. De Wet (1998: 158) beveel aan dat die leksikograaf aan die Afrikaanse kant van die woordeboek dié pare geslagsgemerkte leksikale items waarvoor daar in Engels net een vertaalekwivalent is, as aparte lemmas opneem:

A–E	
onderwyser	teacher
onderwyseres	teacher
E–A	
teacher	onderwyser (n), onderwyser (m), onderwyseres (v)

De Wet (1998: 158) wys verder daarop dat die volgorde waarin die vertaalekwivalente verskaf word, volgens die heersende leksikografiese praktyk, naamlik gebruiksfrekwensie, gedoen behoort te word. Dit beteken dat die woord wat die meeste gebruik word, eerste gegee behoort te word — en nie noodwendig die ongemerkte item nie. Vergelyk verder Beyer 1997: 107-115 vir die hantering van geslagsaanduiding.

3.3 Eentalige antoniemwoordeboeke of sinoniemwoordeboeke met antonieme

Hierdie tipe woordeboeke word gebruik om oplossings vir blokkiesraaisels te vind en vir teksproduksie deur sowel moedertaal- as niemoedertaalsprekers (vgl. Bergenholtz en Tarp 1995: 126). Waar onbekende lemmas verduidelik word deur meer bekende sinonieme of antonieme te gebruik, kan woordeboeke van sinonieme en antonieme ook as definiëringswoordeboeke dien. Dit gaan

hier oor die onomasiologiese vs. die semasiologiese benadering tot die aanbod van betekenisinligting (vgl. Louw 2000).

Aangesien sinonieme en antonieme so nou met mekaar verband hou, is dit waarskynlik beter om beide in een woordeboek te hanteer, eerder as apart. Woordeboeke wat sinonieme (en antonieme) hanteer, se makrostrukture kan op twee verskillende maniere saamgestel word: die eerste metode is die alfabetiese of semasiologiese ordeningsmetode, waarvolgens die lemmas en hul aangebode sinonieme in alfabetiese orde aangebied word, en die tweede metode is dié van tematiese of onomasiologiese ordening, waarvolgens lemmas en hul aangebode sinonieme volgens bepaalde onderwerpe gerangskik word. Die voorbeelde (vir Engels) wat hieronder gebruik sal word, is *Merriam-Webster's Collegiate Thesaurus* wat gekontrasteer word met *The Merriam-Webster Dictionary of Synonyms and Antonyms* en laastens sal daar ook kortliks gewys word op die enigste Afrikaanse ekwivalent, naamlik die *Afrikaanse Sinoniemwoordeboek met Antonieme* deur Eksteen.

Terwyl die tesourus baie meer woordeskat aanbied, is die sinoniemwoordeboeke met antonieme meer gesteld daarop om byvoorbeeld vir 'n intermediêre aanleerder te help om tussen die verskillende sinonieme (en teenstellings) te onderskei. Die volgende voorbeelde van **abandon** (slegs as werkwoord hier aangehaal), eerstens uit *Merriam-Webster's Collegiate Thesaurus* en tweedens uit *The Merriam-Webster Dictionary of Synonyms and Antonyms*, illustreer dit baie duidelik. (Die inskrywing *syn* staan vir "synonym", *idiom* vir "idiomatic equivalent(s)", *ant* vir "antonym(s)", *rel* vir "related word(s)", *con* vir "contrasted word(s)", en // vir "use limited, if in doubt, see a dictionary".)

abandon *vb* 1. to give up without intent to return or reclaim (*abandoned* his family)
syn chuck, desert, forsake, quit, renounce, throw over
rel cast (off), discard, disuse, drop, junk, scrap; reject, repudiate
idiom have done with, leave, flat, quit cold, run out on, turn one's back on (or upon), walk out on
con hold, keep, possess, retain; redeem, rescue, save; acquire, gain, get, procure, win; cherish, foster
ant reclaim
 2. *syn* RELINQUISH, cede, give up, hand over, leave, resign, surrender, // turn up, wave, yield
ant retain

abandon *vb* 1. **Abandon, desert, forsake** mean to give up completely.
Abandon can suggest complete disinterest in the future of what is given up (they *abandoned* their cat at the end of the summer) *ant* reclaim
Desert implies a relationship (as of occupancy or guardianship): it can suggest desolation (deserted farms growing up to brush) or culpability (soldiers who *desert* their posts) *ant* cleave to, stick to
Forsake implies a breaking of a close association by repudiation or renunciation (she *forsook* her husband for a career) *ant* return to, revert to
 2. see RELINQUISH

Vergelyk ook die volgende voorbeeld uit die *Afrikaanse Sinoniemwoordeboek met Antonieme* deur Eksteen, waar daar by 'n groot aantal lemmas ná die sinonieme en hiponieme ook teenstellingswoorde gegee word. Dié teenstellingswoorde verskyn na 'n vierkantige blokkie en is kursief gedruk om hulle van die res van die artikel te onderskei. Vergelyk die volgende voorbeeld:

eggenoot: man, huweliksmaat, huweliksgenoot, lewensgesel, wederhelf, gemaal,
beste, ou beste, ega, gade
□ *eggenote, vrou, gemalin*

4. Slot

Paradis en Willners (2006) het die hantering van antonimie in die *Collins COBUILD Advanced Learners' English Dictionary* (2003) (voortaan CCALED) volledig ondersoek. Hulle wys eers op positiewe aspekte soos die praktiese voordeel wat hierdie woordeboek bo die meeste ander woordeboeke het deurdat leksikale verhoudinge in die kantlyn gespesifiseer word en daarom maklik raakgesien en ontsluit kan word. Verder kom hulle tot die gevolgtrekking dat daar duidelike ooreenkomste is betreffende die betekenis van die lemmas wat gedefinieer word deur antonieme: Hulle is almal inherent binêr en hulle is saamgestel volgens 'n skaal of grens. Die meerderheid lemmas met antonieme is adjektiewe. Dit is egter nie vir hulle duidelik watter beginsels gegeld het by die opname van antonieme nie, aangesien sommige leksikale items nie van antonieme voorsien is nie. Volgens hulle is een van die hoofmerke van 'n aanleerderswoordeboek soos die CCALED om leerders te begelei in hul pogings om 'n goeie greep op die struktuur van die woordeskat van 'n vreemde taal te kry. Aangesien die CCALED 'n korpusgebaseerde woordeboek is, behoort dit gebruik te maak van die inligting wat onttrek kan word uit die korpus. Hulle stel 'n volledige en prinsipiële gebruik van die korpus voor. Een van die maniere waarop die korpus volgens vaste beginsels gebruik kan word, is deur rekenaarprogramme soos *Coco*.

Volgens Justeson en Katz (1991) en Jones (2002) is dit 'n welbekende feit dat antonieme meer gereeld in sinne verskyn as wat voorspel sou kon word. En kanonieke antonieme kom meer dikwels saam voor as kontekstueel beperkte antonieme (vgl. Willners 2001). Paradis en Willners (2006) meen dat hierdie toedrag van sake 'n waardevolle wenk verskaf oor watter antonieme geselekteer en in 'n woordeboek ingesluit behoort te word. Willners (2001: 83) en Holtsberg en Willners (2001) het 'n rekenaarprogram genaamd *Coco* ontwikkel om verwagte en waargenome saamvoorkomende woorde in sinne in 'n gegewe lys in terme van waarskynlikheid te bepaal. *Coco* is ook ontwerp om sinslengtevariasie in ag te neem, wat 'n verbetering was in vergelyking met die studie deur Justeson en Katz (1991). Met behulp van *Coco* het Willners die adjektiewe in die Sweedse korpus ('n miljoenwoordkorpus) ondersoek. Die studie het alle adjektiewe in die korpus wat meer as vyf keer verskyn, ingesluit en het aangetoon dat 357 van die adjektiefpare beduidend meer as voorspel sou word, saam

voorgekom het op 'n beduidendheidsvlak van 10^{-4} . Woordeboeke sou baat kon vind by die bepaling van die mees frekwente antoniempare, aangesien sommige woordeboeke nie die ruimte het om alle teenstellings te akkommodeer nie. Verder kan hierdie instrument help om te verseker dat minstens die mees frekwente teenstellings wel konsekwent aangedui word.

Leksikograwe behoort veel meer aandag te skenk aan teenstellings in verskillende tipes woordeboeke, veral omvattende woordeboeke en aanleerderwoorde. Dit kan gedoen word deur 'n korpusgerigte benadering soos hierbo aangetoon. Teenstellings moet duidelik deur 'n unieke strukturele merker aangedui word. Die metode vir die hantering van teenstellings waarop die leksikograaf besluit, moet dan ook konsekwent gehanteer word in die woordeboek.

Bibliografie

Woordeboeke

- Bosman, D.B., I.W. van der Merwe en L.W. Hiemstra.** 1984^s. *Tweetalige Woordeboek/Bilingual Dictionary*. Kaapstad: Tafelberg.
- Du Plessis, M.** 1993. *Tweetalige Aanleerderswoordeboek/Bilingual Learner's Dictionary*. Kaapstad: Tafelberg.
- Du Plessis, M. (Hoofred.)**. 2005. *Pharos Afrikaans–Engels Woordeboek/English–Afrikaans Dictionary*. Kaapstad: Pharos Woordeboeke.
- Eksteen, Louis.** 1983. *ASA Afrikaanse Sinoniemwoordeboek met Antonieme*. Pretoria: J.L. van Schaik.
- Eksteen, L.C.** 1997. *Groot Woordeboek Afrikaans–Engels/Engels–Afrikaans/Major Dictionary Afrikaans–English/English–Afrikaans*. Kaapstad: Pharos.
- Gouws, Rufus, Ilse Feinauer en Fritz Ponelis.** 1994. *Basiswoordeboek van Afrikaans*. Pretoria: J.L. van Schaik.
- Merriam-Webster's Collegiate Thesaurus*. 1994. Springfield: Merriam-Webster.
- Odendal, F.F. en R.H. Gouws (Reds.)**. 2000. *Verklarende Handwoordeboek van die Afrikaanse Taal*. Midrand: Perskor.
- Sinclair, J. (Red.)**. 2003⁴. *Collins COBUILD Advanced Learner's English Dictionary*. Glasgow: Harper Collins Publishers.
- The Merriam-Webster Dictionary of Synonyms and Antonyms*. 1992. Springfield: Merriam-Webster.
- Woordeboek van die Afrikaanse Taal*. 1950–. Pretoria: Die Staatsdrukker.

Algemeen

- Bergenholtz, H. en S. Tarp (Reds.)**. 1995. *Manual of Specialized Lexicography*. Amsterdam: John Benjamins.
- Beyer, Herman L.** 1997. Aard en leksikografiese hantering van sogenaamde geslagtelik neutrale lede van Afrikaanse geslagsopposisiepare. *S.A. Tydskrif vir Taalkunde* 15(4): 107-115.
- Cruse, D.A.** 1986. *Lexical Semantics*. Cambridge: Cambridge University Press.
- De Stadler, L.G.** 1989. *Afrikaanse semantiek*. Johannesburg: Southern.

- De Wet, Gerda.** 1998. Die makrostrukturele vergestaltung van affikse en tegnostamme in Afrikaanse vertalende woordeboeke. *Lexikos* 8: 154-172.
- Feinauer, A.E.** 1996. Die negende deel van die *Woordeboek van die Afrikaanse Taal*. *Lexikos* 6: 233-271.
- Gouws, R.H.** 1989. *Leksikografie*. Pretoria: Academica.
- Gouws, R.H.** 1990. Information Categories in Dictionaries, with Special Reference to Southern Africa. Hartmann, R.R.K. (Red.). 1990: 52-65.
- Halliday, M.A.K. en R. Hasan.** 1976. *Cohesion in English*. Harlow: Longman.
- Hartmann, R.R.K (Red.).** 1983. *Lexicography: Principles and Practice*. Londen: Academic Press.
- Hartmann, R.R.K. (Red.).** 1990. *Lexicography in Africa. Progress Reports from the Dictionary Research Centre Workshop at Exeter, 24-25 March 1989*. Exeter Linguistic Studies 15. Exeter: Exeter University Press.
- Hausmann, F.J.** 1991. Die Paradigmatik im zweisprachigen Wörterbuch. Hausmann, F.J., O. Reichmann, H.E. Wiegand en L. Zgusta (Reds.). 1989-1991: 2794-2795.
- 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*. Handbücher zur Sprach- und Kommunikationswissenschaft 5.1-5.3. Berlin/New York: Walter de Gruyter.
- Holtsberg, A. en C. Willners.** 2001. Statistics for Sentential Co-occurrence. *Working Papers Lund, Sweden: Department of Linguistics, Lund University* 48: 135-148.
- Jones, S.** 2002. *Antonymy: A Corpus-based Perspective*. Amsterdam: Routledge.
- Jones, S.** 2006. The Discourse Functions of Antonymy in Spoken English. *Text and Talk* 26(1).
- Justeson, J.S. en S.M. Katz.** 1991. Co-occurrences of Antonymous Adjectives and their Contexts. *Computational Linguistics* 17: 1-19.
- Lehrer, A.** 1985. Markedness and Antonymy. *Journal of Linguistics* 31(3): 397-429.
- Lehrer, A. en K. Lehrer.** 1982. Antonymy. *Linguistics and Philosophy* 5: 483-501.
- Louw, Phillip.** 2000. An Integrated Semasiological and Onomasiological Presentation of Semantic Information in General Monolingual Dictionaries as Proposed in H.E. Wiegand's *Semantics and Lexicography*. *Lexikos* 10: 119-137.
- Lyons, J.** 1977. *Semantics*. Cambridge: Cambridge University Press.
- Moulin, A.** 1983. The Pedagogical/Learner's Dictionary. II LSP Dictionaries for EFL Learners. Hartmann, R.R.K. (Red.). 1983: 144-152.
- Muehleisen, V.** 1997. *Antonymy and Semantic Range in English*. Ph.D.-thesis. Evanston, IL: Northwest University.
- Müller, W.** 1989. Die Antonyme im allgemeinen einsprachigen Wörterbuch. Hausmann, F.J., O. Reichmann, H.E. Wiegand en L. Zgusta (Reds.). 1989-1991: 2794-2795.
- Nation, I.S.P.** 2001. *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Palmer, F.R.** 1976. *Semantics*. Cambridge/Londen: Cambridge University Press.
- Paradis, C. en C. Willners.** 2006. *What a Corpus-based Dictionary Tells Us about Antonymy*. Referaat gelewer by die Twaalfde EURALEX Internasionale Kongres, 6-9 September 2006, aangebied deur die Fakulteit Moderne Tale en Literatuur en die Departement Lettere en Filologie van die Universiteit van Turyn, Italië, saam met die Accademia della Crusca.
- Raidt, Edith H.** 1993. Die agtste deel van die *Woordeboek van die Afrikaanse Taal*. *Lexikos* 3: 215-226.

- Wiegand, Herbert E.** 1982. On the Meaning Explanation of Sentence Adverbs in Monolingual Dictionaries. Wiegand, Herbert E. 1999: 113-138.
- Wiegand, Herbert E.** 1999. *Semantics and Lexicography. Selected Studies (1976-1996)*. Redaksioneel versorg deur Antje Immken en Werner Wolski. Tübingen: Max Niemeyer.
- Willners, C.** 2001. *Antonyms in Context. A Corpus-based Semantic Analysis of Swedish Descriptive Adjectives*. Travaux de l'Institut de Linguistique de Lund 40. Lund, Sweden: Department of Linguistics, Lund University.

Die bewerking van idiome in tweetalige woordeboeke*

Liezl Potgieter, *Departement Afrikaans en Nederlands, Universiteit van
Stellenbosch, Stellenbosch, Republiek van Suid-Afrika*
(*liezlpotgieter@gmail.com*)

Opsomming: Woordeboeke is belangrike en noodsaaklike hulpmiddels vir vertalers, maar tog het baie vertalers gemengde gevoelens oor veral tweetalige woordeboeke. Terwyl hulle aan die een kant nie werklik sonder hulle kan werk nie, bied tweetalige woordeboeke aan die ander kant dikwels vir die vertaler min of geen leiding. Dit blyk ook dat tweetalige woordeboeke nie voldoende hulpmiddels vir professionele vertalers is by die vertaling van idiome nie. Teorieë betreffende en benaderings tot vertaling word in die bespreking betrek. Die vertaling van idiome behels meer as net die vertaling van enkelwoorde. In hierdie artikel word aandag gegee aan die huidige bewerking van idiome in tweetalige woordeboeke sowel as die probleme waarmee vertalers gekonfronteer word by die vertaling van idiome.

Sleutelwoorde: IDIOM, TWEETALIGE WOORDEBOEK, VERTAAL-EKWIVALWENT, ABSOLUTE EKWIVALENSIE, GEDEELTELIKE EKWIVALENSIE, DIVERGENSIE, ZERO-EKWIVALENSIE, SURROGAATEKWIVALENSIE, REGISTERVERSILLE, ONVASTHEDE, TAALBOUSELS

Abstract: The Treatment of Idioms in Bilingual Dictionaries. Dictionaries are important and necessary aids for translators, but still many translators have mixed feelings especially regarding bilingual dictionaries. While on the one hand they cannot really work without them, on the other hand bilingual dictionaries often give the translator little or no guidance. It also appears that bilingual dictionaries are not sufficient aids for professional translators in the translation of idioms. Theories regarding and approaches to translation are broached in the discussion. The translation of idioms entails more than just the translation of single words. In this article, attention is given to the present treatment of idioms in bilingual dictionaries, as well as problems with which translators are confronted in the translation of idioms.

Keywords: IDIOM, BILINGUAL DICTIONARY, TRANSLATION EQUIVALENT, ABSOLUTE EQUIVALENCE, PARTIAL EQUIVALENCE, DIVERGENCE, ZERO EQUIVALENCE, SURROGATE EQUIVALENCE, DIFFERENCES IN REGISTER, VARIABLE UNITS, LANGUAGE UNITS

* Hierdie artikel is 'n uittreksel uit 'n ongepubliseerde M.A.-skripsie deur Liezl Gouws *Die bewerking van idiome in tweetalige woordeboeke: 'n probleem vir vertalers?* wat in April 2006 deur die Universiteit van Stellenbosch, Stellenbosch aanvaar is.

1. Inleidend

There is little doubt that dictionaries can be made to better serve their users, among whom translators constitute a very important group. (Roberts en Martin-Rutledge 1999)

Een van die groot tekortkominge in baie tweetalige woordeboeke is die bewerking van idioome. Die bewerkings wat gebruikers in woordeboeke aantref, is dikwels onvoldoende (Gouws 2005: 51) en veroorsaak meermale meer verwarring as duidelikheid. Ook wat betref Afrikaans–Engelse tweetalige woordeboeke, ervaar gebruikers dikwels probleme met idioome wat veroorsaak dat die vertaling van idioome meermale 'n moeilike en tydrowende proses is, bloot weens die gebrekkige bewerking van idioome in tweetalige woordeboeke. In hierdie artikel sal daar gekyk word na verskillende probleme waarmee vertalers en ander woordeboekgebruikers gekonfronteer word met betrekking tot die bewerking van idioome in tweetalige woordeboeke.

2. Aanbieding

Dictionaries are a valuable source of information about multiwords. Unfortunately, only few multiwords are explicitly marked as such in dictionaries: most of them are presented without being distinguished from free combinations of words. (Bentivogli en Pianta 2002)

Die aanbieding van meerwoordige items (insluitende idioome) in tweetalige woordeboeke is een van die redes waarom hierdie woordeboeke vertalers nie genoegsaam help tydens die vertaling van idioome nie. Die algemene tendens in Afrikaans–Engelse en ander tweetalige woordeboeke is dat idioome nie as onafhanklike leksikale items behandel word nie. Idioome word nie in *Groot Woordeboek* (voortaan GW) of in *Pharos Afrikaans–Engels Woordeboek* (voortaan PAEW) aangebied as deel van die makrostruktuur (met ander woorde as lemmas in die sentrale lys of in 'n aparte lys) nie, maar wel as deel van die mikrostruktuur (met ander woorde as deel van die bewerking van 'n betrokke lemma). (Sien ook Roberts en Martin-Rutledge 1999.) Lemmas kan as primêre bewerkingsseenhede beskou word en die aanbieding van idioome sou baie meer geslaagd wees as dit nie net as deel van die mikrostruktuur van die artikel aangebied word nie, maar eerder as sekondêre bewerkingsseenhede met minstens sublemma-status. Deur die aanbieding van idioome in tweetalige woordeboeke sodanig te verbeter, sal die vertaler se soektog na 'n betrokke idioom baie vergemaklik word.

Sowel GW as PAEW gebruik 'n tradisionele leksikografiese prosedure wat behels dat idioome gewoonlik aangebied word as deel van die bewerking van een of meer van die woorde wat in die idioom voorkom (meestal by die naamwoorde of werkwoorde) (Gouws 2005: 51). Tog is hierdie tipe bewerking nie ideaal nie. Dit is belangrik om te onthou dat 'n idioom 'n uitdrukking is waar-

van die betekenis nie uit die betekenis van die dele daarvan afgelei kan word nie. Die feit dat 'n betrokke idioom nie lemmastatus verkry nie, maar in die bewerking van 'n lemma ingesluit word, impliseer dat daar 'n tipe semantiese verhouding tussen die lemma en die idioom bestaan (Gouws 1996: 69). Tog is dit nie altyd die geval nie. As voorbeeld kan gekyk word na die idioom "spill the beans" wat in sowel GW as PAEW as deel van die bewerking van "bean" aangetref word. Die vertaalekwivalente wat in die twee onderskeie woordeboeke vir "bean" verskaf word, is soos volg:

PAEW: boon(tjie); (*i.d. mv.*) bone, boontjies; (*Am. infml.*) kop.

GW: boontjie; (*pl.*) boontjies; pitte (geld).

Alhoewel die idioom "spill the beans" onder die lemmateken "bean" bewerk word, toon die betekenis van die idioom, naamlik "om 'n geheim te verklap", geen ooreenkomste met enige van die betekenisonderskeidings van die leksikale item wat deur die lemmateken "bean" verteenwoordig word nie. Die Afrikaanse vertaalekwivalente wat vir die idioom verskaf word, naamlik "met die hele mandjie patats vorendag kom" (GW en PAEW) en "die aap uit die mou laat" (PAEW) toon ook geen semantiese verhouding met enige van die vertaalekwivalente van die lemma "bean" nie. Beide die brontaal- en die doeltaalvorme van die idioom se semantiese posisie binne die artikel van die lemma "bean" is dus onvanpas. Dit kan aan woordeboekgebruikers baie probleme verskaf in hul soektog na 'n vertaalekwivalent van 'n betrokke idioom omdat die gebruikers nie noodwendig sal weet waar in die woordeboek om na die idioom te gaan soek nie en omdat die idioom nie semanties pas by die lemma waaronder dit bewerk word nie.

'n Verdere probleem met die bewerking van idiome soos wat dit tans in GW en PAEW aangetref word, is dat die idiome as deel van die bewerking van die lemma aangebied word, tussen die ander voorbeeldsinne en kollokasies, maar sonder dat daar vir die woordeboekgebruiker aangedui word in watter gevalle hy/sy te make het met 'n koteksinskrywing en in watter van die gevalle daar sprake is van idiome. Dit is belangrik dat daar 'n onderskeid tussen idiome en voorbeeldmateriaal getref sal word en dat idiome nie as voorbeeldmateriaal aangebied en bewerk sal word nie, maar eerder lemma- of sublemma-status sal kry. Indien idiome bloot as voorbeeldmateriaal bewerk word en daar nie aangedui word in watter gevalle die gebruiker te make het met 'n idioom al dan nie, is daar geen manier om die idiome van ander letterlike voorbeeldsinne te onderskei nie en bemoeilik dit die vertaler se taak om eerstens die idioom in die betrokke artikel op te spoor, en tweedens om te weet in watter gevalle die voorbeeldsin 'n letterlike betekenis het en in watter gevalle die betekenis idiomaties is. Dit is daarom noodsaaklik dat die leksikograaf pertinent vir die gebruiker sal aandui in watter gevalle die woordkombinasies die letterlike gebruik van die lemma as koteksinskrywings demonstreer en in watter gevalle dit 'n idiomatiese waarde het.

'n Gebrek aan ruimte is dikwels een van die hoofredes waarom idioome as deel van die bewerking van 'n lemmateken behandel word, selfs al hou die betekenis van die idioom geensins verband met die betekenis van die leksikale item nie. Alhoewel die aanbieding van idioome as deel van die bewerking van 'n ander lemmateken (soos wat idioome tans in GW en PAEW behandel word) glad nie ideaal is vir woordeboekgebruikers nie, moet in ag geneem word dat die ruimte in 'n woordeboek beperk is en dat leksikograwe dikwels hierdie bewerking gebruik omdat daar soveel data in 'n relatief beperkte ruimte ingepas moet word. Juis daarom is die kans skraal dat leksikograwe in die nabye toekoms van hierdie tipe bewerking van idioome sal wegbeweeg. Om dié rede is dit belangrik dat leksikograwe hierdie minder gewenste werkswyse so goed moontlik moet benut. Indien 'n leksikograaf geen ander uitweg het nie en besluit om 'n idioom se aanbieding en bewerking te koppel aan die artikel van 'n lemma wat ooreenstem met 'n woord uit daardie idioom, is dit nodig dat die leksikograaf nie die idioom of idioome en die koteksinskrywings van die lemma bymekaar sal plaas sonder enige aanduiding van watter gevalle voorbeelde van watter tipe inskrywing is nie. Dit is daarom nodig dat idioome as sublemmas volwaardige bewerkingsenhede moet wees en dat die leksikograaf deur middel van nielemmatiese adressering (byvoorbeeld aanvullende inligting) of met behulp van 'n verbeterde toegangstruktuur (deur die aanbring van byvoorbeeld 'n struktuurmerker) die idioome aan die gebruikers sal uitwys. Deur die aanwending van 'n struktuurmerker soos **UITDR.** (wat in die HAT gebruik word) of **IDM.** om die teksblok te merk waarin idioome aangebied word, maak die leksikograaf dit vir die woordeboekgebruikers moontlik om vinniger die idioome in die artikeltrajek raak te sien en 'n vertaalekwivalent op te spoor, omdat die idioom in 'n aparte soeksone val en dus nie tussen die voorbeeldsinne en kollokasies aangetref word nie.

3. Vertaalekwivalente

Die plek in die woordeboek waar idioome aangebied word, is nie die enigste probleem ten opsigte van die bewerking van idioome in tweetalige woordeboeke nie. Die vertaalekwivalente wat verskaf word en die bewerking daarvan is dikwels 'n nog groter probleem vir vertalers. In die meeste gevalle in GW en PAEW word daar by die bewerking van idioome aan die gebruiker 'n lys vertaalekwivalente verskaf met min of geen inligting rakende die doeltaalkonteks waarbinne die idioome gebruik behoort te word nie. In die volgende gedeelte sal gekyk word na verskeie probleme met betrekking tot die vertaalekwivalente wat vir idioome in tweetalige woordeboeke verskaf word.

3.1 Ekwivalensie

The main aim of the [bilingual] dictionary should not only be the establishment of semantic equivalence between source and target language. Instead a lexicographer has to endeavour to reach communicative equivalence. (Gouws 1996a: 16)

Die tweetalige woordeboek as die vertaler se basiese hulpmiddel dien om intertalige oordrag moontlik te maak. Alhoewel die meeste Suid-Afrikaanse tweetalige woordeboeke daarin slaag om die gebruiker te help om semantiese ekwivalensie te bereik (waar daar een of meer vertaalekwivalente vir 'n betrokke woord verskaf word), is daar min Suid-Afrikaanse tweetalige woordeboeke wat gebruikers voldoende help om ook kommunikatiewe ekwivalensie te bereik.

Wanneer 'n tweetalige woordeboek geraadpleeg word, is dit meestal om 'n vertaalekwivalent te vind om 'n spesifieke woord, uitdrukking of idioom in die brontaal mee te vervang in die doeltaal. In teenstelling met wat dikwels gedink word, kan 'n spesifieke vertaalekwivalent in 'n tweetalige woordeboek nie somer arbitrêr as die betekenis van die leksikale item beskou word nie. 'n Vertaalekwivalent is 'n doeltaalitem wat gebruik kan word om die brontaalitem binne 'n spesifieke situasie te vervang, afhange van spesifieke kontekstuele en kotekstuele beperkings (Gouws 2002). In gevalle waar die woordeboek egter nie aan die gebruiker enige konteks- of koteksleiding verskaf nie, is dit vir die gebruiker dikwels baie moeilik om binne die doeltteks kommunikatiewe ekwivalensie te bereik.

Volgens Gouws (1996a: 16) bestaan daar wel 'n verhouding van ekwivalensie tussen die lemma (of in hierdie geval die idioom) en die onderskeie vertaalekwivalente wat in die betrokke woordeboekartikel aangebied word (die vertaalekwivalentparadigma). Hy noem verder dat leksikograwe moet probeer om vertaalekwivalensie te bereik, wat impliseer dat daar 'n semantiese koördinasie tussen die lemma en sy vertaalekwivalentparadigma moet bestaan. Dit is volgens Gouws (1996a: 17) ook belangrik dat vertaalekwivalente aangevul behoort te word met bykomende inligting en voorbeeldmateriaal wat die tipiese gebruikskonteks illustreer. Die probleem is dat sommige leksikograwe hulle pogings om vertaalekwivalensie te bereik, tot slegs die lys van 'n aantal doeltaalitems beperk. Alhoewel hierdie items wel die semantiese waarde van die lemma verteenwoordig en semantiese ekwivalensie skep, word daar aan die gebruiker geen hulp gebied om die korrekte ekwivalent vir die spesifieke konteks te kies nie. Soos Martin (1962: 156) met reg sê:

Sometimes the uncritical heaping up of near-synonyms is simply an evasion of responsibility on the part of the dictionary-maker: unable (or too little informed) to make up his own mind, he shifts the burden of choice to the user of the dictionary.

Net soos in gevalle waar vertalers tweetalige woordeboeke vir die soek van 'n vertaalekwivalent vir 'n enkelwoord gebruik, is die bereiking van kommunikatiewe ekwivalensie hul hoofdoel wanneer hulle na 'n vertaalekwivalent vir 'n betrokke idioom soek. Dit is daarom noodsaaklik dat leksikograwe ook in die geval van idioome aan die vertaalekwivalentparadigma aandag sal gee en dat vertaalekwivalente van idioome met voorbeeldmateriaal en addisionele inligting aangevul word om die gebruiker te help om vinnig en maklik die korrekte

vertaalekwivalent op te spoor. Deur aanvullende inligting soos konteks- en koteksleiding te gebruik, kan daar verseker word dat die lemma (of in hierdie geval die idioom) en die vertaalekwivalent in 'n verhouding van kommunikatiewe ekwivalensie tot mekaar staan. Semantiese en kommunikatiewe ekwivalensie in woordeboeke lei in die praktyk dikwels ook tot dit waarna Newmark (1982: 39) verwys as "semantiese vertalings" en "kommunikatiewe vertalings". Hy beskryf hierdie twee tipes vertalings soos volg:

Communicative translation attempts to produce on its readers an effect as close as possible to that obtained on the readers of the original. Semantic translation attempts to render, as closely as the semantic and syntactic structures of the second language allow, the exact contextual meaning of the original.

Dit is wenslik dat vertalers eerder kommunikatiewe vertalings as semantiese vertalings sal gebruik, veral ten opsigte van die vertaling van idioome. Juis daarom is dit belangrik dat leksikograwe die lemma en die onderskeie vertaalekwivalente in 'n verhouding van kommunikatiewe ekwivalensie met mekaar sal laat staan, om vertalers te help om 'n kommunikatiewe vertaling as eindproduk te kry.

Daar bestaan verskillende tipes ekwivalensie binne die leksikografie. Daarom moet leksikograwe nie alle vertaalekwivalente eenders bewerk nie, maar die woordeboekgebruikers attent maak op die verskillende tipes ekwivalensie en ook aan hulle uitwys watter tipe ekwivalentverhouding in watter gevalle geld. Sodoende kan leksikograwe die woordeboekgebruikers help om die korrekte vertaalekwivalent vir 'n betrokke idioom binne 'n spesifieke teks en konteks te vind.

Die belangrikste tipes ekwivalensie wat onder vertaalekwivalente aangetref word, is absolute ekwivalensie, gedeeltelike ekwivalensie (wat divergensie insluit) en zero-ekwivalensie (wat surrogaatekwivalensie insluit). Vervolgens sal elk van hierdie tipes ekwivalensie ten opsigte van die vertaling van idioome bespreek word.

3.1.1 Absolute ekwivalensie

Absolute ekwivalensie kom voor wanneer die vertaalekwivalent(e) wat verskaf word in die doeltaal die woord of frase in die brontaal binne alle kontekste en omstandighede kan vervang. Idioome en hul vertaalekwivalente word dikwels in tweetalige woordeboeke as absolute ekwivalente aangegee. In gevalle waar die brontaalidioom en die doeltaalidioom absolute ekwivalente is en dus kongruent aan mekaar is, is daar 'n een-tot-een-verhouding tussen die bron- en die doeltaal. Daar bestaan dus sowel semantiese as kommunikatiewe ekwivalensie tussen die bron- en die doeltaal. In gevalle waar daar absolute ekwivalensie tussen die brontaalidioom en die doeltaalidioom is, bestaan daar ook dikwels 'n verhouding van ekwivalente idiomatisiteit tussen die twee idioome. Sodanige idioome kan dus gewoonlik maklik en probleemloos vertaal word.

Tog gebeur dit ook dikwels dat 'n vertaalekwivalent as 'n absolute ekwivalent vir 'n idioom aangegee word, alhoewel dit nie die geval is dat die doeltaalidioom absoluut ekwivalent aan die brontaalidioom is nie. In sodanige gevalle is dit nodig dat die leksikograaf die gebruiker attent sal maak op die feit dat die twee idiome nie absoluut ekwivalent is nie. Dit is daarom nodig dat die vertaalekwivalent as 'n gedeeltelike ekwivalent bewerk sal word en nie as 'n absolute ekwivalent nie.

3.1.2 Gedeeltelike ekwivalensie en divergensie

Idiome het dikwels meer as een vertaalekwivalent, maar daar word selde aan woordeboekgebruikers enige leiding verskaf ten opsigte van watter ekwivalent om binne 'n bepaalde konteks te kies. Dikwels word die onderskeie ekwivalente as absolute ekwivalente aangedui en alhoewel gevalle wel voorkom waar een of meer vertaalekwivalente absoluut kan wees, gebeur dit meer dikwels dat die vertaalekwivalente slegs gedeeltelik ekwivalent is of dat daar 'n verhouding van divergensie tussen die lemma (of idioom) en die vertaalekwivalente bestaan.

Een van die belangrike aspekte van die bewerking van idiome in tweetaalige woordeboeke waaraan leksikograwe aandag moet gee, is die hantering van divergensie. Divergensie kom in artikels voor waar daar 'n een-tot-meer-as-een-verhouding tussen die bron- en die doeltaal is, en waar die leksikograaf spesiale aandag aan die hantering van die vertaalekwivalentparadigma moet gee. In artikels waar daar 'n verhouding van divergensie voorkom, is dit eersens belangrik dat die leksikograaf tussen leksikale en semantiese divergensie sal onderskei. Volgens Gouws (2002) geld leksikale divergensie in gevalle waar 'n lemma meer as een vertaalekwivalent het, maar waar die vertaalekwivalente in die doeltaal absolute sinonieme of gedeeltelike sinonieme is.

In die geval van leksikale divergensie word die onderskeie vertaalekwivalente in hierdie woordeboeke met behulp van kommas (,) van mekaar geskei. Dit is egter noodsaaklik dat die leksikograaf sal aandui indien die vertaalekwivalente slegs gedeeltelik sinoniem is. In gevalle van absolute sinonimie kan die vertaler die vertaalekwivalent op dieselfde manier hanteer as wat met die behandeling van kongruensie in die woordeboek plaasvind. Indien daar vir die gebruiker geen konteks- of koteksleiding gegee word nie, kan hy/sy van die veronderstelling uitgaan dat die onderskeie vertaalekwivalente almal dieselfde semantiese waarde het. In gevalle van gedeeltelike sinonimie is dit belangrik dat die leksikograaf die gebruiker daarop attent sal maak dat die gedeeltelike sinonieme as vertalings van die brontaalvorm dien, maar dat hulle mekaar nie binne alle kontekste in die doeltaal kan vervang nie. Dit is daarom noodsaaklik dat die leksikograaf konteks- of koteksleiding by die onderskeie vertaalekwivalente sal voeg, sodat die gebruiker sal weet wat die tipiese konteks is waarbinne watter vertaalekwivalent normaalweg voorkom en waar die algemene semantiese waarde van die vertaalekwivalent geaktiveer word (Gouws 2002).

Semantiese divergensie kom voor wanneer 'n lemma meer as een vertaal-ekwivalent het omdat die betrokke brontaalvorm polisemies is. In gevalle van semantiese divergensie word die verskillende polisemiese waardes in GW en PAEW met behulp van kommapunte (;) geskei. Volgens Gouws (1989: 168) is polisemiese lemmas (en dus ook idioome) vanuit 'n semantiese oogpunt een van die vertalende leksikograaf se grootste probleme. Die bewerking van die idioom "keep an eye (up)on" in GW is 'n voorbeeld van 'n artikel waarin daar semantiese divergensie voorkom. Terwyl die eerste vertaalekwivalent "dop-hou" volgens die HAT soos volg gedefinieer word: "in die oog hou; let op — dikwels met onvriendelike bybedoeling", word die tweede vertaalekwivalent wat verskaf word (" 'n oog hou oor") in Prinsloo (2004: 252) so omskryf: "be-waak, sorg vir".

Die probleem is dat daar nie aan die gebruiker gesê word wat die semantiese verskille tussen die verskillende vertaalekwivalente is nie en dat daar ook nie onderskei word watter van die vertaalekwivalente idioome is en watter nie. Indien 'n vertaler nie bewus is van die semantiese verskille tussen die onderskeie vertaalekwivalente nie, kan dit maklik gebeur dat hy/sy die verkeerde ekwivalent binne die betrokke doelteks en -konteks kies. In gevalle soos hierdie is dit belangrik dat die leksikograaf die gebruiker sal help om die regte vertaalekwivalent vir die regte konteks te kies. Die beste manier waarop die leksikograaf kommunikatiewe ekwivalensie kan waarborg, is deur addisionele inskrywings in te voeg om die gebruiker te help sodat hy/sy sal kan aflei watter vertaalekwivalent die korrekte een is om te gebruik.

Dit gebeur ook dikwels dat daar binne een artikel beide leksikale en semantiese divergensie voorkom. In sodanige gevalle van polidivergensie is dit weereens noodsaaklik dat die leksikograaf genoeg konteks- en kontekstleiding sal verskaf en duidelik sal onderskei tussen die sinonieme, gedeeltelike sinonieme en poliseme sodat dit steeds vir die gebruiker duidelik sal wees watter vertaalekwivalent binne watter konteks gebruik moet word om sodoende kommunikatiewe ekwivalensie te kan bewerkstellig.

By die voorkoms van vertaalekwivalente waar daar 'n verhouding van divergensie (hetsy leksikale divergensie, semantiese divergensie of polidivergensie) bestaan, is dit dus nodig dat leksikograwe van aanvullende notas en glosse gebruik sal maak om aan die gebruiker die nodige kontekstleiding te verskaf. Veral vertalers werk dikwels onder groot tyddruk en aanvullende notas en glosse sal hulle help om makliker en vinniger die gepaste ekwivalent op te spoor deur die soektog te vereenvoudig en te verkort.

3.1.3 Zero-ekwivalensie en surrogaatekwivalensie

Dit gebeur ook dat daar idioome in die brontaal bestaan waarvoor daar nie vertaalekwivalente in die doeltaal is nie. Sodanige gevalle is voorbeelde van wat as "zero-ekwivalensie" bekend staan. By gevalle van zero-ekwivalensie is dit belangrik dat die idioome nie net uit die woordeboek weggelaat sal word nie, maar dat hulle inderdaad ook bewerk sal word. Alhoewel dit nie vir die verta-

ler moontlik sal wees om in 'n geval van zero-ekwivalensie die idioom in die brontaal met 'n idioom in die doeltaal te vertaal nie, is dit noodsaaklik dat die vertaler steeds die idioom, tesame met die nodige bewerking daarvan, in die tweetalige woordeboek sal kan vind. Dit is belangrik dat die leksikograaf dit nie aan die vertaler se intuïsie sal oorlaat om af te lei dat, aangesien die brontaalidioom nie 'n geskikte doeltaalidioom het nie, die idioom daarom nie in die woordeboek opgeneem is nie. Dit is gewoonlik juis die idiome waarvoor daar nie 'n doeltaalvorm bestaan nie wat vir vertalers die moeilikste is om suksesvol weer te gee (of op 'n ander manier te bewerk). Dit is dus noodsaaklik dat die leksikograaf die vertaler daarop sal attent maak dat daar nie 'n geskikte doeltaalvorm vir die idioom bestaan nie en voorts ook die alternatiewe maniere waarop die idioom in die doeltaal oorgedra kan word, aan die vertaler sal uitwys.

Die beste en maklikste manier om sodanige idiome te bewerk is deur van surrogaatekwivalensie gebruik te maak. Surrogaatekwivalensie behels dat daar vir die brontaalitem (in hierdie geval 'n idioom) 'n surrogaatvertaalekwivalent verskaf word. Aangesien die surrogaatekwivalent juis benodig word vanwege die afwesigheid van 'n beter vertaalekwivalent, is 'n surrogaatekwivalent dikwels 'n omskrywing van die betrokke woord of idioom. Dit is daarom belangrik dat daar vir die gebruiker aangedui word wanneer die vertaalekwivalent in die doeltaal 'n idioom is en wanneer nie, deur dit byvoorbeeld met 'n asterisk te merk om hom/haar van die nie-idiomatisiteit bewus te maak. Dit gebeur ook dikwels dat, alhoewel daar nie 'n ekwivalente idioom in die doeltaal bestaan nie, daar wel 'n enkelwoord in die doeltaal is om die idioom mee te vervang. Ook in sodanige gevalle is dit noodsaaklik dat die gebruiker daarop attent gemaak sal word dat die vertaalekwivalent nie 'n enkelwoordidioom is nie, maar slegs 'n enkelwoord.

3.2 Register- en ander verskille

In die meeste Suid-Afrikaanse tweetalige woordeboeke word idiome relatief selde van etikette, glosse of konteks- en koteksleiding voorsien. Die woordeboekgebruiker sal meestal slegs 'n idioom en 'n lys vertaalekwivalente aantref. Tog, net soos sommige enkelwoorde, word sekere idiome ook net binne sekere kontekste en sekere registers gebruik. Dit is belangrik dat die leksikograaf die woordeboekgebruiker ook in die bereiking van kommunikatiewe ekwivalensie sal help deur genoegsame leiding te verskaf ten opsigte van wat 'n betrokke idioom se register is en binne watter tipe teks en konteks dit gebruik behoort te word. Die aanwending van etikette, glosse en konteks- en koteksleiding maak ekwivalentonderskeiding makliker vir die woordeboekgebruiker en help hom/haar om te besluit watter doeltaalitem om te kies vir 'n betrokke situasie.

Alhoewel daar registeretikette in beide PAEW en GW aangetref word, is dit meestal net die brontaalitem se register wat aangedui word. Die doeltaalitem word meestal ongeëtiketteer gelaat. Tog is dit belangrik dat leksikograwe in 'n poging om woordeboekgebruikers te help om kommunikatiewe ekwiva-

lensie te bereik, ook aan registerekwivalensie aandag sal gee. Dit beteken dat die leksikograaf daarop moet let om vertaalekwivalente van dieselfde register te verskaf, of om, indien die register van die brontaalitem en die doeltaalitem verskil, etikette te gebruik om dit pertinent onder die woordeboekgebruiker se aandag te bring. Dit is ook noodsaaklik dat die leksikograaf beide die brontaalen en die doeltaalitems sal bewerk om die gebruiker in te lig rakende register en nie net die brontaalitems te bewerk en te aanvaar dat die woordeboekgebruiker self sal weet wat die register van die doeltaalitems is nie. Sodoende kan leksikograwe vertalers help om te besluit watter vertaalekwivalent om binne watter styl en register te gebruik.

4. Onvasthede by idioome

Alhoewel daar talle idioome is wat heeltemal vas en verstar is en waarby daar niks weggeneem, gewysig of bygevoeg kan word nie, is die oorgrote meerderheid idioome op sekere wyses veranderbaar, smeebaar en onvas (Combrink 1989: 57). In die volgende gedeelte sal gekyk word na bepaalde wyses waarop idioome onvas kan wees, sowel as die huidige hantering daarvan in GW en PAEW.

4.1 Opsionele bykomstige taalbousels

Volgens Combrink (1989: 57) is party idioome onvas in dié opsig dat ekstra taalbousels in die idioom ingelas kan word, maar nie noodwendig ingelas hoef te word nie. Moontlike inlasstukke sluit in: opsionele bykomstige woordgroepe, woorde en woorddele.

'n Tweetalige woordeboek behoort, volgens Combrink (1989: 58), die algemeen gangbare, opsionele bykomstige woorde in idioome aan te gee, maar moet hulle konsekwent merk as opsioneel bykomstig. Die algemeenste manier om opsionele bykomstige woordgroepe, woorde of woorddele te merk, is deur hulle tussen hakies te plaas.

Die probleem wat dikwels in tweetalige woordeboeke aangetref word wat betref die aanbieding van idioome met opsionele bykomstige taalbousels, is dat hulle dikwels óf weggelaat word óf verkeerdelik aangedui word asof hulle 'n wesenlike deel van die betrokke idioom is. By die bewerking van die idioom "spill the beans" word die Afrikaanse vertaalekwivalent "met die (hele) mandjie patats vorendag kom" aangetref. In PAEW word die woord "hele" tussen hakies aangedui as 'n opsionele taalbousel, alhoewel dit nie so aangetoon word in GW nie. 'n Vertaler wat van GW gebruik maak om die betrokke idioom te vertaal, sal dus nie weet dat die woord "hele" nie 'n wesenlike deel van die idioom is nie, maar 'n opsionele taalbousel wat wel weggelaat kan word. Vir die vertaler om die idioom korrek te kan vertaal, is dit noodsaaklik dat opsionele bykomstige taalbousels pertinent aangedui sal word. Dit is nodig dat leksikograwe sal let op die bewerking van taalbousels en sal probeer om hulle op 'n konsekwente en verbruikersvriendelike wyse aan te bied.

4.2 Opsionele bykomstige negatiewe woorde

Combrink (1989: 59) noem dat negatiewe woorde 'n addisionele tipe is waaraan die leksikograaf spesiale aandag behoort te skenk by die aangawe van idiome in tweetalige woordeboeke. Daar bestaan idiome soos byvoorbeeld "nothing to write home about" wat oorspronklik slegs in die negatief voorgekom het (alhoewel dit deesdae al hoe meer ook in die positief gebruik word), maar daar is ook verskeie idiome wat in sowel die positief as die negatief gebruik kan word en wat met behulp van woorde soos *geen*, *g'n*, *nie*, *niemand*, *niks* en *nooit* negatief gemaak kan word.

As 'n bepaalde idioom in die positief én die negatief voorkom (met ander woorde as 'n bepaalde idioom opsioneel bykomstig die inlas van 'n negatieweerder toelaat), kan dit volgens Combrink (1989: 60) verwarrend wees as slegs die negatief aangedui word en dit dan nie as opsioneel bykomstig gemerk word nie. Die beste leiding wat 'n tweetalige woordeboek in verband met idiome aan sy gebruikers kan gee, is deur die basiese idioomvorm as sodanig saam met die opsionele bykomstige elemente aan te dui. Die rede waarom dit belangrik is dat die idioom in sy basiese vorm in woordeboeke aangegee sal word, is dat, as die idioom slegs in die negatief aangegee word, gebruikers verkeerdelik die gevolgtrekking kan maak dat die idioom slegs in die negatief gebruik mag word. 'n Leksikograaf kan egter nie by al die idiome wat in die woordeboek voorkom, sowel die positief as die negatief aandui nie omdat dit te veel plek sal opneem. Daarom noem Combrink (1989: 60) dat die negatiewe variant (naas die basiese positiewe) slegs aangegee hoef te word as die negatief een of meer vertalings het wat nie uit die positiewe aangawe afgelei kan word nie. Dit is noodsaaklik dat inligting oor wanneer 'n idioom in sowel die positief as die negatief gebruik kan word en wanneer 'n idioom slegs in die negatief aangewend kan word, pertinent aan gebruikers uitgewys sal word. Juis om hierdie rede ag Combrink (1989: 60) dit so belangrik dat die idioom altyd in die basiese idioomvorm vermeld sal word.

4.3 Alternatiewe bykomstige taalbousels

Volgens Combrink (1989: 62) is daar by idiome ook onvasthede wat hy "alternatiewe woorde en woordgroepe" noem. Hierdie alternatiewe bousels is nóg opsioneel nóg bykomstig. Anders as by opsionele bykomstige taalbousels, gaan dit nie in hierdie geval oor die vraag of die gebruiker die bousel wil gebruik al dan nie, maar oor 'n keuse deur die gebruiker tussen twee of meer verpligte bousels.

In PAEW word alternatiewe bykomstige taalbousels sowel in die brontaal as in die doeltaal aangedui met behulp van óf hakies en die woorde "or" of "of" óf 'n skuinsstreep (/) om die alternatiewe aan die gebruiker uit te wys en hulle terselfdertyd te onderskei van die opsionele bykomstige taalbousels wat ook tussen hakies aangedui word. Hierdie verskillende aanbiedings van die alter-

natiewe bousels kan gebruikers maklik verwar omdat hulle nie noodwendig sal beseef dat daar in beide gevalle sulke bousels gegee word nie. Hierdie inkonsekwente aanbieding is selfs nog meer verwarrend as beide vorme van aanbieding in een artikel voorkom, veral wanneer daar boonop nie net alternatiewe bousels in die artikel aangetref word nie, maar ook 'n verskeidenheid opsionele bousels. Dit is belangrik dat die leksikograaf die data op so 'n wyse sal aanbied dat die gebruiker (of vertaler) maklik sal kan aflei wat dit presies is wat die leksikograaf probeer weergee en hoe hy/sy die inligting wat in die betrokke artikel aangebied word, in sy/haar teks kan of moet gebruik.

Die aanbieding van alternatiewe taalbousels in GW is ook nie juis geslaagd nie, omdat sowel opsionele as alternatiewe taalbousels op presies dieselfde manier behandel word en die gebruiker op geen manier kan weet wanneer dit 'n opsionele taalbousel en wanneer dit 'n alternatiewe taalbousel is nie.

Nog 'n probleem wat ontstaan by die aanbieding van alternatiewe taalbousels is wanneer die woordeboek nie al die verskillende alternatiewe in die artikel aanbied nie, maar slegs een van die alternatiewe, en dit dan boonop laat lyk asof die alternatief nie slegs 'n alternatief is nie, maar 'n wesenlike deel van die betrokke idioom.

'n Voorbeeld van so 'n geval waar die alternatiewe taalbousels nie pertinent aangedui word nie, is die idioom "to see something in a certain light". In hierdie idioom kan die woord "certain" vervang word met een van 'n verskeidenheid byvoeglike naamwoorde, byvoorbeeld "to see something in a *good* light", "to see something in a *bad* light" of "to see something in a *different* light". Tog bly die basiese idioom in die doeltaal deurgaans dieselfde, naamlik "om iets in 'n goeie/slegte/ander lig te beskou". In GW word hierdie idioom nêrens gelys nie en in PAEW word slegs die opsie "to see something in a different light" verskaf.

Dié wyse van aangawe soos dit in PAEW aangetref word, laat dit ten onregte lyk asof "different" 'n wesenlike deel van die betrokke idioom is, terwyl dit juis nie die geval is nie. 'n Vertaler wat byvoorbeeld die idioom "to see something in a bad light" moet vertaal en PAEW raadpleeg, sal nie noodwendig weet dat hierdie idioom op dieselfde wyse vertaal kan word as die idioom "to see something in a different light" wat in PAEW vermeld word nie. Juis daarom is dit belangrik dat ten minste al die algemeenste alternatiewe vir die gebruiker uitgewys sal word en dat die alternatiewe bousels slegs as alternatiewe aangedui sal word en nie as wesenlike dele van die idioom nie.

5. Ten slotte

Dit is duidelik dat daar baie ruimte vir verbetering is ten opsigte van die bewerking van idioome in tweetalige woordeboeke. Dit is daarom noodsaaklik dat daar aandag gegee sal word aan die behandeling van idioome in tweetalige woordeboeke om dit sodanig te verbeter dat dit gebruikersvriendeliker en 'n groter hulp vir vertalers sal wees.

Bronnelys

Woordeboeke

- Eksteen, L.C.** 1997¹⁴. *Groot Woordeboek Afrikaans–Engels/Engels–Afrikaans / Major Dictionary Afrikaans–English/English–Afrikaans*. Kaapstad: Pharos.
- Odendal, F.F. en R.H. Gouws.** 2000⁴. *Verklarende Handwoordeboek van die Afrikaanse Taal*. Midrand: Perskor.
- Du Plessis, M. (Hoofred.)**. 2005. *Pharos 1, Afrikaans–Engels/English–Afrikaans Woordeboek/Dictionary*. 2005. Kaapstad: Pharos Woordeboeke.

Ander bronne

- Bentivogli, L. en E. Pianta.** 2002. Detecting Hidden Multiwords in Bilingual Dictionaries. Braasch, A. en C. Povlsen (Reds.). 2002. *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002, II*: 785-793. Kopenhagen: Center for Sprogteknologi. Intyds beskikbaar by: <<http://tcc.itc.it/people/pianta/publications/euralex-collocations02.pdf>> [24 Februarie 2005].
- Combrink, J.G.H.** 1989. Wesenlike elemente, ekstras en vulsels in die aangawe van idioome in tweetalige woordeboeke. Botha, T.J.R. (Red.). *Leksikografie en leksikologie*: 55-77. Menlopark: Serva Uitgewers.
- Gouws, R.H.** 1989. *Leksikografie*. Kaapstad: Academica.
- Gouws, R.H.** 1996. Idioms and Collocations in Bilingual Dictionaries and their Afrikaans Translation Equivalents. *Lexicographica* 12: 54-88.
- Gouws, R.H.** 1996a Bilingual Dictionaries and Communicative Equivalence for a Multilingual Society. *Lexikos* 6: 14-31.
- Gouws, R.H.** 2002. *Leksikografie: Ekwivalentverhoudinge in vertalende woordeboeke*. Ongepubliseerde Klasnotas Afrikaans 348.
- Gouws, R.H.** 2005. Issues Regarding the Comment on Semantics in Bilingual Dictionaries Dealing with Closely Related Languages. Iglá, B., P. Petkov en H.E. Wiegand (Reds.). 2005. *Kontrastive Lexikologie und zweisprachige Lexikographie. Germanistische Linguistik* 179: 39-56. Hildesheim: Georg Olms Verlag.
- Martin, S.E.** 1962. Selection and Presentation of Ready Equivalents in a Translation Dictionary. Householder, F.W. en S. Saporta (Reds.). *Problems in Lexicography*: 153-159. Bloomington: Indiana University Press.
- Newmark, P.** 1982. *Approaches to Translation*. Oxford: Pergamon Press.
- Prinsloo, A.F.** 2004. *Spreekwoorde en waar hulle vandaan kom*. Kaapstad: Pharos.
- Roberts, R. en V. Martin-Rutledge.** 1999. Dictionary Use Strategies for Translators. Awaiss, H. en J. Hardane (Reds.). 1999. *Traduction: Approches et Théories*: 345-364. Beiroet: Skool van Vertalers en Tolke van Beiroet, Sint Josef Universiteit. Intyds beskikbaar by: <<http://www.dico.uottawa.ca/articles-en.htm>> [30 Mei 2005].

Compiling a Bidirectional Dictionary Bridging English and the Sotho Languages: A Viability Study*

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 investigate the viability of the compilation of a single bidirectional dictionary with a single lemma list for the Sesotho sa Leboa, Setswana and Sesotho → English side and a simultaneous treatment of the three Sotho languages in the articles of the English lemmas in the English → Sesotho sa Leboa, Setswana and Sesotho side of the dictionary. Specific attention will be given to selected macrostructural and microstructural aspects of such a compilation.

Keywords: SOTHO LANGUAGES, NGUNI LANGUAGES, BILINGUAL DICTIONARIES, COMMUNICATIVE EQUIVALENCE, CORPORA, USER PERSPECTIVE

Opsomming: Die samestelling van 'n tweerigtingwoordeboek wat Engels en die Sothotale oorbrug: 'n Lewensvatbaarheidstudie. Die doel van hierdie artikel is om die lewensvatbaarheid van die samestelling van 'n enkele tweerigtingwoordeboek met 'n enkele lemmalys vir die Sesotho sa Leboa, Setswana en Sesotho → Engelse kant en 'n gelyktydige bewerking van die drie Sothotale in die artikels van die Engelse lemmas in die Engels → Sesotho sa Leboa, Setswana en Sesotho kant van die woordeboek te ondersoek. Besondere aandag sal aan geselekteerde makrostrukturele en mikrostrukturele aspekte van so 'n samestelling gegee word.

Sleutelwoorde: SOTHOTALE, NGUNITALE, TWEETALIGE WOORDEBOEKE, KOMMUNIKATIEWE EKWIVALENSIE, KORPUSSE, GEBRUIKERSPERSPEKTIEF

Introduction

The aim of this article is to study the viability of a bidirectional dictionary bridging English and the Sotho languages[†]: Sesotho, Setswana and Sesotho sa Leboa. The focus will be on the advantages and disadvantages of such a single

* This article is based on a paper presented at the Tenth International Conference of the African Association for Lexicography (AFRILEX), organised by the Sesiu sa Sesotho Lexicography Unit, University of the Free State, Bloemfontein, Republic of South Africa, 13–15 July 2005.

dictionary compared to three comparative bidirectional bilinguals: English–Sesotho/Sesotho–English, English–Setswana/Setswana–English and English–Sesotho sa Leboa/Sesotho sa Leboa–English, and the additional value it would have in the absence of bilingual dictionaries bridging African languages with each other. This will be an important achievement since publishers generally do not regard the compilation of separate dictionaries bridging the African languages with each other as economically viable. Compiling a dictionary with a single lemma list for the Sotho languages at this stage can indeed become the forerunner to such an eventual goal, i.e. true bidirectional dictionaries bridging the African languages with each other. It also has the potential to pave the way for an English ↔ Nguni languages dictionary.

The analysis and design of the macrostructure and microstructure will be based on existing bilingual dictionaries bridging English and a Sotho language and will be aimed at the same target users. The viability study will firstly be performed for a combined article for the Sotho languages usable mainly for basic receptive information, i.e. treatment limited to a translation equivalent or two, and secondly for combined articles where a more exhaustive treatment is given. The bilingual dictionaries analysed are *The New English–Northern Sotho Dictionary*, *English–Northern Sotho*, *Northern Sotho–English* (NEN) (Kriel 1976) for Sesotho sa Leboa, *Dikišinare ya Setswana English Afrikaans* (DS) (Snyman et al. 1990) and *Setswana–English–Setswana Dictionary* (SESD) (Matumo 1993) for Setswana and *Southern Sotho–English Dictionary* (SSED) (Mabille and Dieterlen 1988) for Sesotho.

The results of this study will hopefully enable prospective compilers to decide whether it is worthwhile to compile such a dictionary and to provide guidelines and examples for such a compilation. It is not possible to do a detailed analysis of all relevant lexicographic aspects within the limitation of a journal article and the discussion will therefore be limited to a number of key microstructural and macrostructural aspects.

The compilation of such a dictionary will require the combined skills of mother-tongue speakers of all four languages and corpora for these languages.

Impact and range of application for the Sotho and Nguni languages

A bidirectional English → {Sesotho, Setswana and Sesotho sa Leboa}, {Sesotho, Setswana and Sesotho sa Leboa} → English dictionary is comparable to *three* bidirectional bilingual dictionaries, English–Sesotho, Sesotho–English, English–Setswana, Setswana–English, and English–Sesotho sa Leboa, Sesotho sa Leboa–English, thus two directions for the envisaged model versus six sides for separately bridging English and a Sotho language. It could be argued that separate bilinguals bridging English and each of the Sotho languages do exist but in most cases they are out of print or in need of revision. The situation for bridging Sotho languages with each other is much less promising. At this stage in the development of South African lexicography publishers' interest is virtually

non-existent for bridging African languages with each other, thus little hope for bi-directional Sesotho sa Leboa ↔ Sesotho, Sesotho sa Leboa ↔ Setswana and Sesotho ↔ Setswana dictionaries.

A successful bidirectional English ↔ Nguni languages dictionary will be even more significant for the official South African Nguni languages because four languages are involved. A bidirectional English → {siSwati, isiXhosa, isiNdebele and isiZulu}, {siSwati, isiXhosa, isiNdebele and isiZulu} → English dictionary is comparable to *four* bidirectional bilingual dictionaries, English–siSwati, siSwati–English, English–isiXhosa, isiXhosa–English, English–isiZulu, isiZulu–English, and English–isiNdebele, isiNdebele–English. Thus two directions for the envisaged model versus eight sides for separately bridging English and a Nguni language. For bridging Nguni languages with *each other* six bilinguals, thus twelve directions should be compiled, i.e. isiZulu–isiXhosa, isiXhosa–isiZulu, isiZulu–siSwati, siSwati–isiZulu, isiZulu–isiNdebele, isiNdebele–isiZulu, isiXhosa–siSwati, siSwati–isiXhosa, isiXhosa–isiNdebele, isiNdebele–isiXhosa, and siSwati–isiNdebele, isiNdebele–siSwati.

Orthographic words versus lemmas in the Sotho languages

As a result of the disjunctive orthographies and the word-lemmatisation strategy followed in the Sotho languages, the relation between orthographic words versus lemmas is almost 1–1. In corpus-based lexicography this simplifies the compilation significantly since limited lemmatisation has to be performed upon the types generated from the corpora for the Sotho languages. Consider, for example, the 100 most frequently used words in Sesotho sa Leboa in ranking order on high to low frequency count:

- (1) a, moo, le, gago, go, bjale, ka, bolela, ba, tseba, o, mme, ke, dira, e, morena, ya, wena, re, monna, ge, taba, se, kua, wa, lego, gore, tšona, ga, fao, sa, mola, di, rena, mo, kgoši, be, bana, tša, mongwe, la, gwa, bona, leo, ye, bao, tla, ao, gona, bile, tše, ra, gagwe, šetše, ile, eng, yo, tsebe, na, woo, yena, tloga, yeo, kwa, fela, no, motho, mang, gomme, lena, goba, gape, bjalo, morago, tlo, swanetše, batho, mokgwa, bja, selo, nna, ngwana, yona, fihla, tšeo, mosadi, pele, banna, bego, thoma, wo, pelo, gobane, modimo, moka, eupša, bo, nako, seo, tee

A lemma list for the words in (1) will consist of between 95 and 100 lemmas depending on the lemmatisation strategy.

Using the Dutch–Afrikaans dictionary as a design model

The envisaged English ↔ Sotho languages dictionary is a multifunctional dictionary where a dictionary consultation environment is created in which, in

terms of Martin and Gouws (2000: 788), 'both differences and similarities become apparent in an efficient and contrastive way'. Reflecting differences and similarities will indeed be the key factor in simultaneous treatment of the Sotho languages.

A second important observation made by Martin and Gouws (2000: 790) is that 'the combinatory data represents the core of the lexicographic presentation'. Compare in this regard the following example of the approach of Martin and Gouws (2000: 790) where these principles are honoured for non-contrastive combinations, contrastive combinations and idiomatic expressions in the article of the lemma *bril* 'spectacles':

<p>bril 1 [om te kijken] # – een <u>bril</u> hebben/dragen; hij heeft zijn <u>bril</u> niet op; ... • <inf.> een <u>bril</u> moeten 'n <u>bril</u> moet kry ... ♦ elk ziet door zijn eigen <u>bril</u> <i>elkeen kyk deur sy eie <u>bril</u></i> door 'n roze <u>bril</u> kijken <i>deur 'n rooskleurige <u>bril</u> kyk</i> ‣ <i>iemand 'n <u>bril</u> op die neus sit</i> iemand te grazen nemen</p>
--

Non-contrastive combinations are marked by '-' and only a Dutch example is given. Contrastive combinations are marked with '•' and '♦' marks idiomatic expressions. In this way different search zones (in a fixed order) are clearly marked in a user-friendly way and differences and similarities are clearly illustrated.

The envisaged English ↔ Sotho languages dictionary reflects a striking resemblance to, but also clear differences with, the Dutch–Afrikaans dictionary described by Martin and Gouws (2000). Among the similarities on macro-structural level count the compilation of a single central list, i.e. a single access structure, and consideration of different lemma types. The study for the Sotho languages differs from the model of Martin and Gouws in that the lemma list is for three and not for only two languages and that a full bridging with English is done.

The macrostructure

Size and impact of a single lemma list

The prospective compiler of a dictionary with a single lemma list should firstly decide on the size of the lemma lists for both sides of the dictionary in order to compile a dictionary that would cover a reasonable percentage of use of the four languages in question. As a point of departure an assessment was made of the size of the lemma lists of dictionaries bridging English and Sotho languages as well as the top frequencies in English dictionaries such as *Collins COBUILD English Dictionary* (COBUILD2), *Macmillan English Dictionary for Advanced Learners* (MED) and *Longman Dictionary of Contemporary English* (LDOCE).

For English, the data given in COBUILD2 regarding the impact of the frequency bands give useful guidelines to the size and nature of an English lemma list for the envisaged dictionary.

Table 1: Summary of frequency band values in COBUILD2

Number of filled diamonds	Lemmas per category	Totals	% of all written and spoken English
5	700		
4	1 200		
(Total 5+4)		1 900	75
3	1 500		
2	3 200		
1	8 100		
(Total 3+2+1)		12 800	20
(Total 5+4+3+2+1)		14 700	95

From Table 1 it is clear that the top 14 700 lemmas represent an astonishing 95% of the tokens or running words in a given English text.

The words in the five frequency bands are of immense importance to learners because they make up 95% of all spoken and written English. (COBUILD2 1995: xiii)

The sizes of the lemma lists for the Sotho languages are reflected in Table 2.

Table 2: Number of lemmas in NEN, DS, SESD, SSED

Dictionary	Lemmas: English → Sotho languages	Lemmas: Sotho languages → English
Sesotho sa Leboa: NEN	11 640	13 900
Setswana: DS	10 600	15 600
Setswana: SESD	7 258	21 700
Sesotho: SSED	—	11 900

The impact of a single lemma list for the Sotho languages will be studied taking the top 10 000 words in the Pretoria corpora for each of the three languages as a point of departure.

Table 3: Sources, types and tokens in the Pretoria Sotho languages corpora

	Sesotho sa Leboa	Setswana	Sesotho
Sources	327	235	77
Tokens	5 957 553	5 466 438	3 159 568
Types	157 391	151 701	

The number of types given in the final row of Table 3 reflects 100% of the use of the languages given in terms of tokens in the second row. This simply means, for example, that a Sesotho sa Leboa dictionary containing 150 000 lemmas would account for each of the 5.9 million words in the 327 texts that make up this corpus. The same analysis is applicable to Setswana and Sesotho from the figures given in Table 3. The question is what the impact of a lemma list consisting of only 10 000 lemmas for each of the languages will be in terms of token coverage. The comparable statistics for the Sotho languages are given in Table 4.

Table 4: Percentage of tokens represented by the top 10 000 types

Language	Tokens for the top 10 000 types	Tokens in the entire corpus	% of words covered by the top 10 000 types
Sesotho sa Leboa	5 462 558	5 957 553	91.7%
Setswana	4 948 959	5 466 438	90.5%
Sesotho	2 868 451	3 159 568	90.8%

From Table 4 it is clear that, as in the case of English, lemma lists compiled for the top 10 000 tokens in each of the Sotho languages represent more than 90% of the use of the language. It can therefore be argued that the selection of 10 000 lemmas for English, Sesotho sa Leboa, Setswana and Sesotho is viable in terms of considerable coverage of all four languages.

Lexical overlap in the Sotho languages

The second aspect studied on macrostructural level in the consideration of a consolidated lemma list is the percentage of words that the languages have *in common*, simply referred to as *overlap*. It stands to reason that the greater the overlap the better the chances of success for such a dictionary will be.

As a point of departure the overlap between Dutch and Afrikaans was studied since, as reported above, that project is regarded as a viable one. A comparison between Dutch and Afrikaans corpora reveal an overlap of 20%. Consider in this regard a selection of such mutual lexical items with high occurrence frequencies per million running words in Table 5.

Table 5: Afrikaans compared to Dutch: mutual lexical items, with frequencies per million running words (Gouws et al. 2004: 798)

Item	Afrikaans	Dutch	Item	Afrikaans	Dutch	Item	Afrikaans	Dutch
<i>van</i>	19 688	18 091	<i>niet</i>	46	7 358	<i>als</i>	26	4 011
<i>het</i>	15 111	16 655	<i>te</i>	8 876	7 025	<i>aan</i>	4 338	3 846
<i>een</i>	1 892	16 223	<i>die</i>	58 733	6 520	<i>of</i>	6 653	3 811

<i>en</i>	22 516	13 663	<i>met</i>	7 139	6 268	<i>nog</i>	1 930	3 507
<i>in</i>	15 619	13 483	<i>voor</i>	1 316	6 220	<i>wel</i>	276	3 146
<i>is</i>	15 628	11 539	<i>maar</i>	3 862	5 121	<i>om</i>	7 323	3 074
<i>dat</i>	4 841	9 565	<i>ook</i>	3 068	4 414	<i>wat</i>	5 268	2 918
<i>op</i>	7 132	8 125	<i>dan</i>	907	4 032	<i>kan</i>	2 069	2 488

For the English → Sotho languages side of the dictionary a flying start exists, (technically speaking a 100% overlap), since the lemma list will only be 10 000 English lemmas and not three times 10 000 lemmas as for three separate dictionaries.

For the Sotho languages a comparison of the top 10 000 words in Sesotho, Setswana and Sesotho sa Leboa reveals that the three languages have 1 943 (19.4%) words in common. Sesotho sa Leboa and Setswana share 3 276 (32.7%) words. Sesotho sa Leboa and Sesotho have 2 689 (26.9%) words in common and Setswana and Sesotho share 3 441 (34.4%) words. This results in a single lemma list of 22 537 compared to a 30 000 lemma list in three separate dictionaries, thus a reduction of almost 30%.

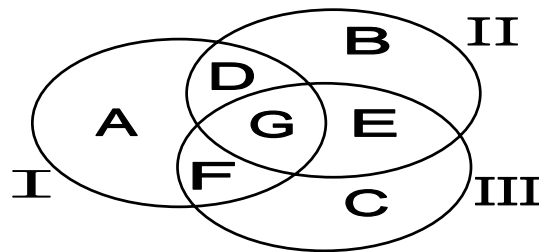


Figure 1: I = Sesotho sa Leboa; II = Setswana; III = Sesotho

(2) Formula and values for I, II and III:

Single lemma list =
 [top 10 000 Sesotho sa Leboa lemmas (A+F+D+G) +
 top 10 000 Setswana lemmas (B+E+D+G) +
 top 10 000 Sesotho lemmas (F+E+C+G)] –
 [[two times the Sesotho sa Leboa/Setswana/Sesotho lemmas in common
 (2xG)] +
 [Sesotho sa Leboa/Setswana lemmas in common (D)] +
 [Sesotho sa Leboa/Sesotho lemmas in common (F)] +
 [Setswana/Sesotho lemmas in common (E)]
 A = 5 978; B = 5 226; C = 5 813; D = 1 333; E = 1 498; F = 746; G = 1 943

Consider as a concrete example a mini-lemma list compiled on the basis of the overlap between the top 100 types in Sesotho sa Leboa, Sesotho and Setswana. The trilingual overlap, i.e. words that all three languages share within the top 100 is 32%, i.e. the 32 words in (3).

- (3) motho, bo, bona, se, sa, re, di, pele, e, nna, ngwana, tla, na, o, mosadi, morena, monna, mo, mme, yona, a, le, la, ke, ka, nako, bana, wa, wena, batho, ba, ya

The bilingual overlap between Sesotho sa Leboa and Setswana within the top 100 is 47%, i.e. the words in (3) plus the 15 words in (4).

- (4) ga, dira, eng, pelo, morago, mongwe, fela, gago, kwa, gore, yo, go, gape, bile, gagwe

The bilingual overlap between Sesotho sa Leboa and Sesotho within the top 100 is 45% or 45, i.e. the words in (3) plus the 13 words in (5).

- (5) ile, ra, tseba, fihla, moo, yena, lena, modimo, leo, seo, mang, taba, be

Finally, the bilingual overlap between Sesotho and Setswana within the top 100 is 46%, i.e. the words in (3) plus the 14 words in (6).

- (6) ntse, nngwe, tswa, ne, neng, teng, ena, utlwa, tsa, tsena, tse, tle, bua, rona

Sesotho sa Leboa has 40 unique words and Setswana and Sesotho 39 and 41 unique words respectively. This renders a single lemma list of 194 lemmas. For this section it can be concluded that lemma lists based upon the top 10 000 tokens in English and the Sotho languages will render sufficient coverage of these languages and that the amount of overlap in the Sotho languages and the resulting single lemma list suggest that the compilation of an English ↔ Sotho languages dictionary is viable on macrostructural level. The prospective compiler, however, has to keep in mind that words which have the same orthographic form in the Sotho languages but different grammatical functions will have to be entered as more than one lemma depending on the lexicographic approach. For example, *nna* in (3) as a pronoun of the first person singular in the Sotho languages but also as a verb in Setswana.

The microstructure

On microstructural level preliminary tests indicate that the average article length in the envisaged English → Sotho languages side of the dictionary would vary between one-third and two-thirds of the combined article length of the English → Sesotho/Setswana/Sesotho sa Leboa sides of three separate dictionaries, thus a 30%–60% reduction. Compare the following randomly selected lemmas where unmarked forms such as *pula* 'rain' and *motho* 'a person' reflect a complete overlap between the three Sotho languages while double subscripts, e.g. *gagwe* 'his/her' and *phela* 'live', mark similarities between two languages and single subscripts, such as for *hae* 'his/her' and *jang* 'how', uniqueness in one language only.

- (7) pula
- motho
- gagwe [SsL/Set]
- phela [Ses/SsL]
- hae [Ses]
- jang [Set]

The real challenge lies in the successful compilation of the Sotho articles in the English → Sotho languages side of the dictionary. Failure to do so will simply result in articles reflecting the mere stacking of translation equivalents of English lemmas in Sesotho sa Leboa, Setswana and Sesotho, without consideration of crucial aspects of *differences, similarities and combinatory data* as highlighted in terms of Martin and Gouws (2000) above. There will thus be no gain in reduction and comparison. Consider Table 6 as an extract from the *Concise Multilingual Dictionary* (CMD) as a case in point.

Table 6: *Concise Multilingual Dictionary*

English	Xhosa	Zulu	Northern Sotho	Southern Sotho	Tswana	Afrikaans
carry	-thwala	-thwala	-rwala	-rwala	-rwala	dra
close (v.)	-vala	-vala	-tswalela	-kwala	-tswala	toemaak (ww.)
hot	-shushu	-shisa	-borutho (nom. rel.)	-tjhesa	bolelo	warm
pretty (water)	amahle	amanzi amahle	-bose (adj.)	metsi a matle	monate	lekker (water)

Firstly, translation equivalents are simply chronologically stacked for each language with considerable repetition in both the Sotho and Nguni languages without any attempt towards reduction. Secondly, a complete lack of communicative equivalence poses a great risk to the user to incorrectly use the equivalent(s). For example, *tswalela* has a limited range of application in Sesotho sa Leboa and cannot be used in all contexts as an equivalent of 'close', thus misleading the user. He/She is further misled by gross inconsistencies, e.g. in the final row where the compilers failed to add the word for *water* in isiXhosa, Sesotho sa Leboa and Setswana. The user would conclude that *monate* means 'pretty water' in Setswana while it only means 'nice, pretty'.

In spite of its shortcomings, it could be argued that an English → Sotho languages/Nguni languages/Afrikaans dictionary of this magnitude is a useful contribution in the complete absence of dictionaries bridging African languages with each other and could be improved by simultaneous treatment of the target languages. Consider the following attempt to improve CMD's articles for the lemmas *coffee, page* and *verb* in Table 7 versus example (8).

Table 7: *Concise Multilingual Dictionary*

English	Northern Sotho	Southern Sotho	Setswana
coffee	kofi	kofi	kofi
page	letlakala	leqephe	tsebe
verb	lediri	leetsi	lediri

- (8) **coffee** kofi
page letlakala _[SsL], leqephe _[Ses], tsebe _[Set]
verb lediri _[SsL/Set], leetsi _[Ses]

The aim of this viability study, however, is to compile more comprehensive articles, at least of the same size as existing bilingual dictionaries such as NEN, DS, SESD and SSED. The articles compiled in (9) for the English → Sotho languages and Sotho languages → English sides of the dictionary are still relatively speaking restricted to receptive use but they do go some way towards productive use and communicative equivalence.

- (9) **rain** *n.* pula
motho *n.* a person, a human being
his *pp.* gagwe _[SsL/Set], hae _[Ses]
live *v.* phela _[Ses/SsL], tshela _[Set] ~ **(survive)** phela _[Set]
how? *adj.* jwang? _[Ses] o phela jwang?; jang? _[Set] o tshela jang?; bjang? _[SsL]
o phela bjang? **how are you?**
pelo (peló); N Cl. 9/10; **dipelo** heart; ~**ng** in the heart; **pelwana** little heart; **fela** ~ give up; **beta** ~ be brave; ~**bohloko** _[SsL/Ses]; ~**botlhoko** _[Set]; ~**mpe**, ~**nolo**; ~**telele**; ~**tlhomogi** _[SsL/Set]; ~**tlhomohi** _[Ses]

Sensible reduction in these examples is achieved in terms of, among others, tonal indication, grammatical information, and translation equivalents. They should, however, be submitted to target users and the feedback obtained should be carefully studied.

In the Sotho languages → English side, the mediostructure (system of cross-referencing) can be fruitfully utilised to link the lemma with its equivalent in the other Sotho language(s), thus further strengthening the aspect of bridging Sesotho sa Leboa, Setswana and Sesotho with each other.

- (10) **hore** _[Ses] *conj.* that, in order that ... cf. gore _[SsL/Set]

Selecting suitable dictionary conventions will be a crucial aspect in order to present user-friendly search zones. In the examples above, subscripts were used to mark the distinctions between Sesotho, Setswana and Sesotho sa Leboa. Similar layouts should be tested utilising a combination of different colours and standard conventions such as bold, underline and italics as in Figure 2

where the different languages are marked in the dictionary in colour and coloured shadings (which can unfortunately not be reproduced in this article) with the aid of running footers:

rain *n.* pula
 motho *n.* a person, a human being
 his *pp.* gagwe, hae
 live *v.* phela, tshela; ~ (survive) phela
 how? *adj.* jwang? o phela jwang?; jang? o tshela jang?; bjang? o phela bjang?
how are you?
 {Running footer:}
 Ses; Set; SsL; SsL/Set; Ses/SsL; Ses/Set

All three languages in common = black
 SsL (Sesotho sa Leboa) = red
 Set (Setswana) = green
 Ses (Sesotho) = blue

Figure 2: Using colour and colour shading in Sotho languages articles of English lemmas

Conclusion

In this article macrostructural and microstructural aspects were analysed in terms of the viability of an English ↔ Sotho languages dictionary with a single lemma list for the Sotho languages and simultaneous treatment of the Sotho languages in the English → Sotho languages side. It can be concluded that such a compilation will be successful (a) for lemma lists of a reasonable size taken from all four languages as a point of departure, (b) because substantial lexical overlap exists between Sesotho sa Leboa, Setswana and Sesotho, (c) because treatment in terms of similarities, differences and combinatory data is possible, and (d) because user-friendly articles comprehensive enough for receptive and limited productive use can be compiled.

Endnote

† The term *Sotho languages* refers to Sesotho sa Leboa (Northern Sotho), Setswana (Tswana), Sesotho (Southern Sotho), and the term *Nguni languages* refers to isiZulu (Zulu), siSwati (Swazi), isiXhosa (Xhosa) and isiNdebele (Ndebele), cf. Van Wyk (1966: 17, 21, 36-37), Guthrie (1970), etc.

References

Dictionaries

- CMD = Jennings, L.E. et al. 1995. *The Concise Multilingual Dictionary*. Johannesburg: Ad Donker.
 COBUILD2 = Sinclair, J. (Ed.). 1995. *Collins COBUILD English Dictionary*. London: HarperCollins.

- DS = Snyman, J.W. et al. 1990. *Dikišinare ya Setswana English Afrikaans Dictionary Woordeboek*. Pretoria: Via Afrika.
- LDOCE = Procter, P. (Ed.). 1978. *Longman Dictionary of Contemporary English*. Harlow: Longman.
- MED = *Macmillan English Dictionary for Advanced Learners*. 2002. Macmillan Publishers Limited.
- NEN = Kriel, Theunis J. 1976⁴. *The New English–Northern Sotho Dictionary, English–Northern Sotho, Northern Sotho–English*. Johannesburg: Educum Publishers.
- SESD = Matumo, Z.I. 1993 *Setswana–English–Setswana Dictionary*. Gaborone: Macmillan.
- SSED = Mabile, A. and H. Dieterlen. 1988. *Southern Sotho–English Dictionary*. Revised by R.A. Paroz. Morija: Morija Sesotho Book Depot.

Other literature

- Gouws, R.H., D.J. Prinsloo and G.-M. de Schryver.** 2004. Friends will be Friends — True or False. Lexicographic Approaches to the Treatment of False Friends. Williams, G. and S. Vessier (Eds.). 2004. *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004, Lorient, France, July 6–10, 2004*: 797-805. Lorient: Faculté des Lettres et des Sciences Humaines, Université de Bretagne Sud.
- Guthrie, M.** 1970. *Comparative Bantu Vol. III. An Introduction to the Comparative Linguistics and Prehistory of the Bantu Languages*. Farnborough: Gregg Press.
- Martin, Willy and Rufus Gouws.** 2000. A New Dictionary Model for Closely Related Languages: The Dutch–Afrikaans Dictionary Project as a Case in Point. Heid, Ulrich, Stefan Evert, Egbert Lehmann and Christian Rohrer (Eds.). 2000. *Proceedings of the Ninth EURALEX International Congress, 8–12 August 2000*: 783-792. Stuttgart: Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.
- Van Wyk, E.B.** 1966. *Die Bantoetale: 'n Beknopte algemene oorsig*. Pretoria: Van Schaik.

Das Lern- und Konsultationswörterbuch. Ein neuer Fachwörterbuchtyp am Beispiel der *Wörterbücher zur Sprach- und Kommunikationswissenschaft (WSK)*

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

Anlässlich meiner Ehrenpromotion an der Universität Stellenbosch am 20. April 2006 dieser dankbar gewidmet.

Zusammenfassung: Nach kurzen Hinweisen zum Begriff des Wörterbuchtyps werden die zentralen Probleme der lexikographischen Versorgung bezüglich der deutschen Fachsprache im Fachgebietsverbund „Sprach- und Kommunikationswissenschaft“ kurz charakterisiert. Sie zeigen sich besonders in der nicht ausreichenden fachlexikographischen Abdeckung, in der idiosynkratischen Lemmaselektion und in der Nichtberücksichtigung metalexikographischen Wissens. Dann wird die Lern- und die Konsultationskomponente der WSK-Bände beschrieben. Dabei wird u. a. auf die Datendistribution, die Mediostruktur und die typkonstitutiven Eigenschaften der Wörterbuchartikel eingegangen, so dass insgesamt ein Eindruck entsteht, welches die zentralen Eigenschaften eines Lern- und Konsultationswörterbuchs sind.

Stichwörter: ÄQUIVALENTREGISTER, FACHLEXIKOGRAPHISCHE ABDECKUNG, LEMMASELEKTION, LERN- UND KONSULTATIONSWÖRTERBUCH, TEXTUELLE ARTIKELPOSITION, WISSENSRECHERCHESITUATION, WÖRTERBUCHKLASSE, WÖRTERBUCHTYP

Abstract: *The Teaching and Consultation Dictionary. A New Type of Special-field Dictionary Exemplified by the Wörterbücher zur Sprach- und Kommunikationswissenschaft (WSK).* Following a short exposition regarding the understanding of dictionary types, a brief characterisation is given of the central problems of lexicographic treatment with regard to German technical language in the specialised field of "language and communication science". These problems become especially clear from the insufficient special-field lexicographic coverage, the idiosyncratic lemma selection and the non-cognisance of metalexigraphic knowledge. Subsequently the teaching and consultation components of the WSK volumes are discussed. In doing so the emphasis is a.o. on the data distribution, the mediostructure and the type-constitutive features of the dictionary articles, leading to a notion of what the central features of a teaching and consultation dictionary are.

Keywords: DICTIONARY CLASS, DICTIONARY TYPE, EQUIVALENT REGISTER, KNOWLEDGE RESEARCH SITUATION, LEMMA SELECTION, SPECIAL-FIELD LEXICOGRAPHIC COVERAGE, TEACHING AND CONSULTATION DICTIONARY, TEXTUAL ARTICLE POSITION

1. Kurze Vorbemerkung über Wörterbuchtypen

Bevor wir uns mit einem speziellen Fachwörterbuchtyp näher befassen, ist es zweckmäßig, kurz eine Vorverständigung über die Frage herbeizuführen, was nachfolgend unter einem Wörterbuchtyp verstanden werden soll (vgl. dazu Wiegand 1988). Ganz allgemein gilt: Ein Typ lässt sich als eine Menge von Eigenschaften auffassen, die miteinander verträglich und typkonstitutiv sind. Entsprechend wird unter einem Wörterbuchtyp eine Menge von solchen miteinander verträglichen lexikographischen Eigenschaften verstanden, die als typkonstitutiv gelten. Dies sei an einem einfachen Beispiel veranschaulicht. Gegeben sei der Typ des Abkürzungswörterbuchs, der in allen verschrifteten Kultursprachen als Printwörterbuch realisiert ist. Die Menge der typkonstitutiven Eigenschaften weist lediglich zwei Elemente (E) auf.

- E₁ besteht in der Eigenschaft, dass zugriffsbereite Abkürzungen genannt werden.
- E₂ besteht in der Eigenschaft, dass den Abkürzungen Abkürzungsauflösungsangaben zugeordnet werden.

Es ist zu beachten, dass die jeweilige makrostrukturelle Anordnung der Abkürzungen, die ihre Zugriffsbereitschaft sicherstellt, nicht zu den typkonstitutiven Eigenschaften für ein Abkürzungswörterbuch gehört. Sind die Lemmata, mit denen die Abkürzungen genannt werden, alphabetisch geordnet, liegt ein alphabetisches Abkürzungswörterbuch vor. Der Typ des alphabetischen Abkürzungswörterbuchs ist ein Untertyp des Abkürzungswörterbuchs, der erhältlich ist, wenn man das Typologiekriterium „makrostrukturelle Anordnungsmethode“ zur Unterteilung des Ausgangstyps anwendet. Ich betone ausdrücklich, dass es in diesem Beitrag nicht um Wörterbuchtypologien geht, sondern nur um einen bestimmten Fachwörterbuchtyp. Zu Wörterbuchtypen lassen sich Typvarianten angeben. So wäre z.B. eine Typvariante des Typs des alphabetischen Abkürzungswörterbuchs dann gegeben, wenn auf die Abkürzungsauflösungsangabe in manchen Wörterbuchartikeln ein kurzer Angabetext zu demjenigen Bezugsgegenstand steht, der mit der Abkürzung üblicherweise bezeichnet wird.

Schließlich sei zum Schluss der Vorbemerkung noch darauf hingewiesen, dass man Wörterbuchtypen von Wörterbuchklassen unterscheiden sollte.

Der Terminus *Wörterbuchtyp* ist intensional zu verstehen; das bedeutet u.a.: Um ihn angemessen verwenden zu können, muss man nicht angeben

können, wie viele Wörterbuchexemplare zum jeweiligen Wörterbuchttyp gehören. Weiterhin gilt: Um zu wissen, was ein Abkürzungswörterbuch ist, braucht man nicht alle Abkürzungswörterbücher zu kennen.

Der Terminus *Wörterbuchklasse* ist dagegen extensional zu verstehen; das bedeutet u.a.: Um ihn angemessen verwenden zu können, muss man angeben können, wie viele Wörterbücher zur jeweiligen Wörterbuchklasse gehören.

2. Zwischenbemerkung zur lexikographischen Versorgung bezüglich der deutschen Fachsprache im Fachgebietsverbund Sprach- und Kommunikationswissenschaft

Diese Zwischenbemerkung soll dazu dienen, kurz die Gründe zu erläutern, die dazu geführt haben, dass Stefan J. Schierholz und ich die Reihe *Wörterbücher zur Sprach- und Kommunikationswissenschaft* (WSK) gegründet haben (vgl. Schierholz und Wiegand 2005).

Die Reihe WSK umfasst derzeit die nachfolgend genannten 23 Bände. In Klammern stehen die Bandherausgeber, soweit sie bereits feststehen. Weiteres zur Organisation der Reihe findet man unter <www.wsk.uni-erlangen.de>.

1. *Grammatik* (2 Teilbände: 1.1 *Formenlehre*; 1.2 *Syntax*) (Christa Dürscheid und Stefan J. Schierholz)
2. *Wortbildung* (Peter O. Müller und Susan Olson)
3. *Historische Sprachwissenschaft* (Mechthild Habermann und Markus Hundt)
4. *Phonetik und Phonologie* (Tracy Alan Hall und Bernd Pompino-Marschall)
5. *Schriftlinguistik* (Martin Neef und Rüdiger Weingarten)
6. *Textlinguistik und Stilistik* (Jannis Androutsopoulos und Stephan Habscheid)
7. *Sprachtechnologie und Computerlinguistik* (Dafydd Gibbon und Ulrich Schmitz)
8. *Dialektologie* (Heiko Girnth und Franz Patocka)
9. *Quantitative und Formale Linguistik* (Peter Grzybek und Reinhard Köhler)
10. *Semantik und Pragmatik* (Ralf Klabunde und Beatrice Primus)
11. *Sprachtheorie und Methodenlehre* (Johannes Kabatek und Bernd Kortmann)
12. *Lexikologie und Phraseologie* (Christiane Fellbaum und Ingo Warnke)
13. *Medienwissenschaften* (Irmela Schneider und Peter M. Spangenberg)
14. *Kognitive Grammatik* (Susanne Niemeier und Doris Schönefeld)
15. *Soziolinguistik* (Stephan Elspaß und Ekkehard Felder)

16. *Sprachphilosophie* (Christoph Demmerling und Pirmin Stekeler-Weithofer)
17. *Sprachdidaktik: Fremd- und Muttersprache* (Jörg Kilian und Jutta Rymarczyk)
18. *Sprachtypologie/Allgemeine Sprachwissenschaft* (Johannes Helmbrecht und N.N.)
19. *Semiotik* (Guido Ipsen und N.N.)
20. *Gesprächslinguistik*
21. *Lexikographie und Wörterbuchforschung*
22. *Psycho- und Neurolinguistik* (Christina Kauschke und Prisca Stenneken)
23. *Sprachen- und Varietätennamen*

Alle Bände erhalten den gleichen Untertitel, und zwar den Folgenden: *Ein Lern- und Konsultationswörterbuch. Mit einer Systematischen Einführung und englischen Übersetzungen.*

Die lexikographische Versorgung des Fachgebietverbundes „Sprach- und Kommunikationswissenschaft“ ist bezüglich der deutschen Fachsprache derzeit nicht ausreichend, und zwar besonders in den drei folgenden Hinsichten:

- (1) Die erste Hinsicht betrifft die fachlexikographische Abdeckung:
Von hochgerechnet ca. 35 000 bis 40 000 deutschen Termini sind ca. 5 000 lexikographisch bearbeitet. Wenn die geplanten 23 WSK-Bände fertig sind, sollten ca. 40 000 Termini in moderner Art und Weise und möglichst auf einem aktuellen Stand ihres wissenschaftlichen Gebrauchs lexikographisch bearbeitet sein.
- (2) Die zweite Hinsicht, in der die lexikographische Versorgung nicht ausreichend ist, betrifft die Lemmaselektion:
Die Lemmaselektion in den vier derzeit maßgeblichen deutschen linguistischen Fachwörterbüchern — nämlich Abraham (1988), Bußmann (2002), Glück und Schmoe (Hrsg.) (2005) und Lewandowski (1990) — ist mehr oder weniger idiosynkratisch. Ja, man kann sogar nachweisen: Die Lemmaselektion ist teilweise ideologisch, weil sie schulendeterminiert ist, was aus den gewählten Wörterbuchtiteln, wie z.B. *Lexikon der Sprachwissenschaft* (Bußmann 2002), nicht ersichtlich ist. Auf diese Weise wird Fachlexikographie zu einem Instrument von hilflosen Versuchen zur Durchsetzung von Theorien und Konzepten im Bereich der wechselnden Mainstream-Perspektiven; kleinere Teilgebiete und nicht gewünschte Neuerungen werden nicht in die Lemmaselektion einbezogen und damit unterdrückt. Das ist aber gerade nicht die Aufgabe der linguistischen Fachlexikographie.

- (3) Die dritte Hinsicht betrifft das metalexikographische Wissenskorpus:
Das in den drei letzten Jahrzehnten erarbeitete Wissen der Wörterbuchforschung hat bei der Konzeption und Erarbeitung der linguistischen Fachwörterbücher so gut wie keine Rolle gespielt. Entsprechend liegen traditionelle Fachwörterbücher mit 0-8-15-Konzeption und Schema-F-Kodifikation vor, die weit hinter dem zurückbleiben, was in anderen Fächern geboten wird. Insbesondere die Wörterbuchform ist unzureichend und keineswegs auf die Benutzerbedürfnisse zugeschnitten.

Stefan J. Schierholz und ich haben aus der soeben nur sehr grob charakterisierten Situation folgende Konsequenz gezogen:

Die Fachlexikographie für den Fachgebietsverbund „Sprach- und Kommunikationswissenschaft“ muss insgesamt neu konzipiert werden. Die wichtigsten Eckpfeiler einer solchen Neukonzeption sind die Folgenden:

- (a) Trennung der Verantwortlichkeiten bei der Erarbeitung der Wörterbücher der WSK-Reihe in folgender Form:
- Die Metalexikographen erarbeiten die Wörterbuchform und sorgen für ihre Implementierung in einem webbasierten Redaktionssystem, so dass das metalexikographische Wissenskorpus angemessene Berücksichtigung findet.
 - Die Linguisten in ihrer Rolle als Fachlexikographen sind für den jeweiligen bandspezifischen Ausschnitt des Wörterbuchgegenstandes zuständig, und zwar als Bandherausgeber und als Artikelautoren, so dass eine angemessene teilgebietspezifische Fachkompetenz gegeben ist.
- (b) Es wird kein Gesamtfachwörterbuch erarbeitet, dessen Titel nur ein unfreiwilliger Ausdruck von Hybris sein kann, sondern eine neue Fachwörterbuchserie hinsichtlich der Fachwörterbuchform gleichartiger, gedruckter Teilgebietsfachwörterbücher mit spezifischen und treffenden Titeln, die in einem Online-Fachwörterbuch, dessen Konzeption bereits in Arbeit ist, zusammengeführt werden können, aber nicht zusammengeführt werden müssen.
- (c) Alle WSK-Bände werden so realisiert, dass sie zu einem neuartigen Fachwörterbuchtyp gehören, nämlich zum Typ des Lern- und Konsultationswörterbuchs, der in den naturwissenschaftlichen Fächern z.T. in bestimmten Typvarianten bereits ausgeprägt ist.¹
- (d) Die Rolle der englischen Sprache als wissenschaftliche *lingua franca* wird ernst genommen. Entsprechend sind die WSK-Bände bilingualisierte Fachwörterbücher, so dass es in den Einzel- und Synopseartikeln englische Terminusäquivalentangaben, Übersetzungen der deutschen Defi-

niensangaben sowie im Nachspann ein Englisch–Deutsches Äquivalent-zugriffsregister gibt.

- (e) Die computergestützte fachlexikographische Arbeit wird als Teamarbeit organisiert, so dass möglichst viele kompetente Fachkollegen mitarbeiten.

Um einer größeren Anschaulichkeit willen wird bereits zum Ende dieser Zwischenbemerkung ein WSK-Artikel aus dem Band 21 *Lexikographie und Wörterbuchforschung* mit Markierungen für die textuellen Artikelpositionen abgebildet. Erläuterungen zum Artikel finden sich unter 4.

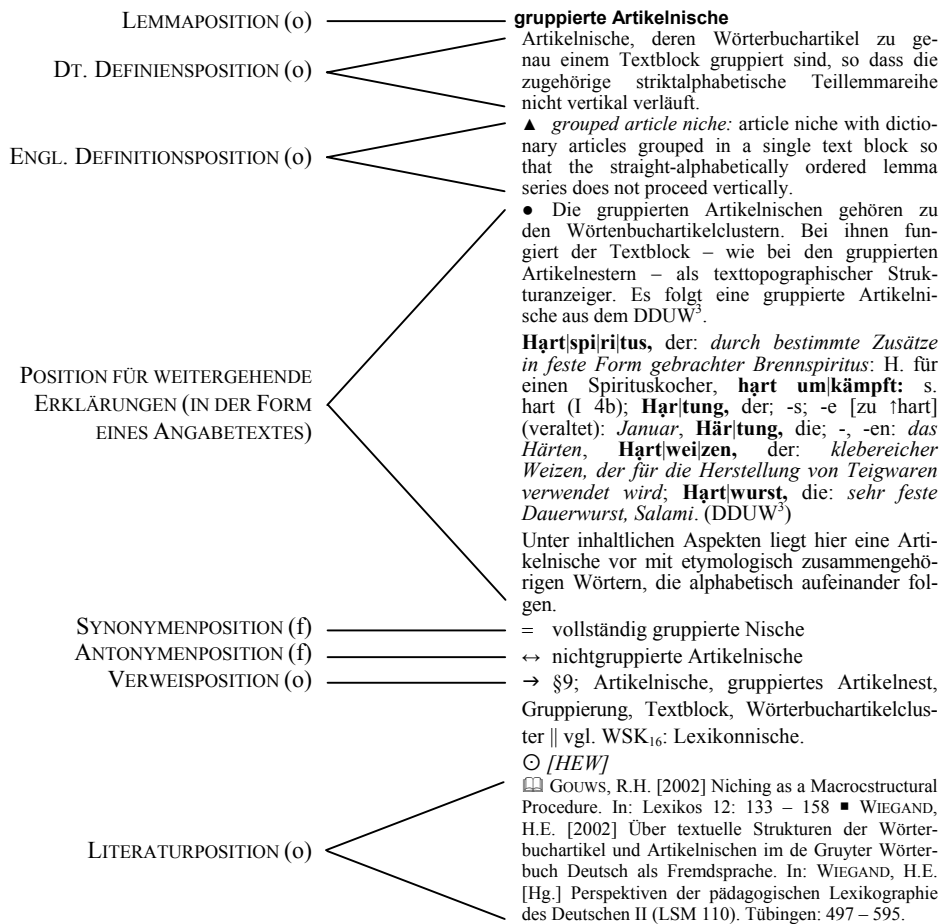


Abb.: Veranschaulichung zu den textuellen Artikelpositionen eines WSK-Artikels, der ein Einzelartikel zu einem monosemen Lemmazeichen ohne Positionsreduzierung ist; Abkürzungen: o = obligatorisch, f = fakultativ

3. Die WSK-Bände als fachliche Lern- und Konsultationswörterbücher

3.1 Wissenschaftliches Fachwissen und Fachlexikographie

Der Typ des fachlichen Lern- und Konsultationswörterbuchs wurde von mir im Anschluss an meine theoretischen und empirischen Untersuchungen zur Wörterbuchbenutzungsforschung konzipiert. Der Typ des fachlichen Lern- und Konsultationswörterbuchs kann verstanden werden als eine fachlexikographische Antwort auf zentrale Fragen zum Erwerb, zur Struktur und zur handlungsermöglichenden Funktion von wissenschaftlichem Fachwissen. Die Antwort ist lexikographietheoretisch fundiert und berücksichtigt die neuere Fachsprachenforschung. Für den Fachgebietsverbund „Sprach- und Kommunikationswissenschaft“ gilt demnach u.a. Folgendes:

- Wissenschaftliches Wissen wird in erster Linie zeichenvermittelt, also sprach-, bild- und symbolvermittelt erworben; dies geschieht in unterschiedlichen Lehr-, Studier- und Handlungssituationen. Die Wissensvermittlung durch fachliche Texte in natürlichen Sprachen spielt hierbei die zentrale Rolle. Sieht man von den naturwissenschaftlichen Kernfächern Physik und Chemie sowie der Mathematik ab, dann gilt: Ein wissenschaftlicher Gegenstand ist ein bestimmter nur unter mindestens einer bestimmten natürlichsprachlichen wissenschaftlichen Benennung und ihrem definitionsgemäßen Gebrauch. Termini lassen sich entsprechend nicht angemessen als äußerliche sprachliche Etiketten verstehen, die nachträglich vorgegebenen Begriffen nur angeheftet werden.
- Jede Art von wissenschaftlichem Wissen ist immer mehr oder weniger zusammenhängendes Wissen zu einem bestimmten Gegenstandsreich.
- Punktueller wissenschaftliches Wissen, das einer Person fehlt, benötigt sie bei der gestörten Rezeption wissenschaftlicher Texte, weiterhin bei der gestörten Fachtextproduktion sowie bei der gestörten Übersetzung wissenschaftlicher Texte, also insgesamt in konfliktären Situationen. Zusammenhängende Wissensausschnitte werden dagegen außerhalb von konfliktären Situationen benötigt, so dass zusätzliches Wissen zur Lösung wissenschaftlicher Aufgaben systematisch recherchiert werden muss. Will man auf den unterschiedlichen Wissensbedarf potenzieller Fachwörterbuchbenutzer lexikographisch angemessen reagieren, dann werden Fachwörterbücher benötigt, die mindestens die drei nun zu nennenden Aufgaben erfüllen können.

(1) Das Fachwörterbuch muss in konfliktbedingten Konsultationssituationen, also immer dann einfach und schnell benutzbar sein, wenn bei

der Arbeit mit wissenschaftlichen Texten aufgrund von fachbezogenen punktuellen Wissenslücken der verschiedensten Art aktuelle Kommunikationsstörungen auftreten. Diese können entweder die Form von Textrezeptionsstörungen aufweisen oder als Textproduktionsstörungen oder auch als Störungen bei der Übersetzungstätigkeit gegeben sein. In solchen konfliktären Situationen können konsultative Suchfragen nach einzelnen wissenschaftlichen Termini an das Fachwörterbuch gerichtet werden, die meistens auch Fragen nach dem Fachwissen sind, das an die jeweiligen Termini gebunden ist (vgl. Wiegand 1998).

(2) Das Fachwörterbuch muss aber gleichermaßen in den ganz anders gearteten nichtkonfliktbedingten Benutzungssituationen erfolgreich benutzbar sein, die zum Typ der wörterbuchbezogenen Wissensrechersituationen gehören (vgl. Wiegand 2006); das gilt besonders für solche Recheresituationen, die während des akademischen Studiums in verschiedenen Ausprägungen auftreten oder während der Einarbeitung von Dozenten in ein Teilgebiet ihres Faches, das ihnen weniger vertraut ist. Sowohl Studierende als auch Dozenten werden eine wörterbuchinterne Fachwissensrecherche, die aus einer zeitlich geordneten Folge von wörterbuchbezogenen Recheresituationen besteht, welche vor allem durch die Ausführung von verweismotivierten Konsultationshandlungen verknüpft sind, gerade dann vornehmen, wenn unter bestimmten vorgegebenen Frage- oder Aufgabestellungen begriffliche Zusammenhänge neu erarbeitet werden müssen.

(3) Das Fachwörterbuch muss schließlich den Übergang von konfliktbedingten Konsultationssituationen zu Recheresituationen problemlos zulassen sowie auch den Übergang in der umgekehrten Richtung. Fachwörterbücher, die so konzipiert und deren fachlexikographische Daten (einschließlich der bibliographischen Daten) so präsentiert sind, dass sie die drei genannten Aufgaben erfüllen, gehören zum Typ des fachlichen Lern- und Konsultationswörterbuchs. Obwohl Fachwörterbücher dieses Typs unterschiedlich ausgeprägt sein können, gilt für alle Typvarianten grundsätzlich dies: Jedes fachliche Lern- und Konsultationswörterbuch weist zwei Komponenten auf: Eine Konsultationskomponente und eine Lernkomponente. Diese beiden Komponenten treten in einem Fachwörterbuch nicht — wie z.B. Textkonstituenten eines kondensierten Wörterbuchartikels — als isolierbare und zusammenhängende textuelle Teile auf. Vielmehr handelt es sich bei diesen Komponenten um Mengen von Eigenschaften, deren Ausprägungen bei den Wörterbüchern auftreten, und zwar um solche, die z.T. bei der Wörterbuchform und z.T. beim Wörterbuchgegenstand ausgeprägt sind. Die beiden Eigenschaftsmengen haben eine Schnittmenge, so dass es Eigenschaften gibt, die sowohl zur Konsultationskomponente als auch zur Lernkomponente gehören.

3.2 Zur Konsultationskomponente von WSK-Bänden

Die Konsultationskomponente besteht aus solchen Eigenschaften, welche die Voraussetzungen dafür bilden, dass das jeweilige Wörterbuch die unter (1) genannten Aufgaben erfüllen kann. Die wichtigsten dieser Eigenschaften und Eigenschaftskomplexe sind in den WSK-Bänden wie folgt ausgeprägt:

- (i) In jedem WSK-Band existiert eine glattalphabetische makrostrukturelle äußere Zugriffsstruktur.
- (ii) Jeder WSK-Band soll einen möglichst hohen Grad der lexikographischen Abdeckung aufweisen und damit eine so genannte „reichhaltige Makrostruktur“.
- (iii) In jedem WSK-Band wird eine relativ engmaschige, jeweils terminusdeterminierte und damit durch die sprachlichen Besonderheiten des jeweiligen Fachsprachenausschnittes bestimmte Parzellierung des wissenschaftlichen Fachwissens in partiell kondensierten Einzelartikeln vorgenommen, die präzises fachliches Detailwissen, und zwar einschließlich bibliographischen Wissen präsentieren. Ein Beispiel für einen solchen Artikel findet sich in der Abb. Neben Einzel- gibt es Synopseartikel.
- (iv) Alle Einzel- und Synopseartikel in allen WSK-Bänden weisen eine textuelle Struktur auf, die es den Artikelautoren ermöglicht, das zusammenhängende und nur mit unscharfen Grenzen unterscheidbare gegenstandskonstitutive fachzyklopädische Bedeutungswissen sowie das fachzyklopädische Sachwissen (i.S.v. Wiegand 1988a) angemessen lexikographisch zu bearbeiten.
- (v) Alle Einzel- und Synopseartikel in allen WSK-Bänden weisen ein artikelinternes positionsgebundenes Verweisungsangebot auf, das bei Verständnisschwierigkeiten, die bei der Artikellectüre auftreten, weiterhilft.

3.3 Zur Lernkomponente von WSK-Bänden

Im Folgenden werfen wir einen kurzen Blick auf die Lernkomponente von WSK-Bänden. Diese Komponente besteht aus solchen Eigenschaften, welche die Voraussetzungen dafür bilden, dass das jeweilige Wörterbuch die unter (2) genannten Aufgaben erfüllen kann. Die wichtigsten dieser Eigenschaften und Eigenschaftskomplexe sind wie folgt ausgeprägt:

- (vi) In jedem WSK-Band wird die hochgradige Parzellierung des Fachwissens, die in den Einzelartikel gegeben ist, in Synopseartikeln partiell aufgehoben.

- (vii) Das in den Wörterbuchartikeln von WSK-Bänden vermittelte Fachwissen wird über Verweisungen paragraphenspezifisch mit der Systematischen Einführung verknüpft, die jeder WSK-Band aufweist, und erhält dort einen weiter angelegten, fachlichen Verständnisrahmen.
- (viii) Es existiert eine bidirektionale Verweisung zwischen jedem Paragraphen der Systematischen Einführung und den fachlich zugehörigen Synopseartikeln, die ihrerseits mit vielen Einzelartikeln mediostrukturell uni- und z.T. bidirektional vernetzt sind.
- (ix) Schließlich weisen alle WSK-Bände eine mehrdimensionale Datenakzessivität auf.

Die beiden zuletzt genannten Eigenschaften (viii) und (ix) bilden wichtige Voraussetzungen dafür, dass die unter (3) genannte Aufgaben erfüllt werden: Sie garantieren dem kundigen Benutzer einen problemlosen Übergang von Konsultationssituationen zu Wissensrecherchesituationen und umgekehrt.

3.4 Typkonstitutive Aspekte der Wörterbuchform

Durch die bisher gegebene Darstellung von WSK-Bänden mit Blick auf die Lern- und Konsultationskomponente ist das charakteristische Profil des Wörterbuchtyps „fachliches Lern- und Konsultationswörterbuch“, wie er in allen WSK-Bänden gleichartig ausgeprägt ist, noch nicht scharf genug herausgearbeitet. Um dies zu erreichen, werfen wir im Folgenden noch einen Blick auf solche ausgewählten Teile und Eigenschaften der Wörterbuchform, die als typkonstitutiv und damit als besonders charakteristisch gelten können. Einige bereits erwähnte Sachverhalte kommen dabei erneut ins Blickfeld, jedoch in einer anderen Perspektive. Insgesamt werden vier Aspekte betrachtet.

(a) Typspezifische Datendistribution

Alle WSK-Bände weisen eine typspezifische fachwörterbuchinterne Datendistribution auf. Das Datendistributionsprogramm sieht für alle WSK-Bände hinsichtlich der zentralen Daten folgende Verteilung vor.

Erstens die lexikographischen Daten werden auf die nach Paragraphen gegliederte Systematische Einführung in das jeweilige Teilgebiet eines WSK-Bandes im Wörterbuchvorspann verteilt. *Zweitens* werden sie auf ein glatt-alphabetisches zweiseitig präsentiertes Verzeichnis der Termini (das so genannte Wörterverzeichnis) und schließlich *drittens* auf ein alphabetisches Englisch-Deutsches Äquivalenzzugsregister im Wörterbuchnachspann verteilt. Im Folgenden werfen wir kurz einen Blick auf die drei zugriffsverschiedenen Speicherplätze für Daten in WSK-Bänden.

Die Systematische Einführung soll nicht mehr als 64 Druckseiten umfassen. Am Ende jedes ihrer Paragraphen findet sich eine feste Verweisposition; in dieser wird auf alle Synopseartikel verwiesen, die inhaltlich zum jeweiligen Paragraphen gehören. Eine solche Verweisposition könnte z.B. im Paragraphen 14 mit dem Paragraphentitel „Die Bauteile von Wörterbuchartikeln“ der Systematischen Einführung zum WSK-Band 21 *Lexikographie und Wörterbuchforschung* folgende Form aufweisen und damit aus einer Verweisbeziehungsangabe, realisiert durch einen Schrägpfeil „↗“, und sechs Verweisadressenangaben bestehen:

„↗ Angabe, Angabeklasse, Angabestruktur, Angabetext, Kommentar, Strukturanzeiger“.

Im Verzeichnis der Termini handelt es sich um ein artikelheterogenes Wörterverzeichnis, weil es sowohl Synopseartikel als auch Einzel- und Verweisartikel aufweist. Da die Lemmata und damit alle Wörterbuchartikel glattalphabetisch geordnet sind, weisen die WSK-Bände ein Wörterverzeichnis mit homogener Wörterverzeichnisarchitektur auf, weil eine *oberhalb*-Relation auf eine Menge definiert ist, deren Elemente ausschließlich Wörterbuchartikel sind. Das alphabetische Englisch-Deutsche Äquivalenzzugriffsregister im Wörterbuchnachspann weist z.B. im WSK-Band 21 folgende Registereinträge auf:

[REGISTEREINGÄNGE]	[REGISTERANGABEN]
active-passive principle	Aktiv-Passiv-Prinzip
dictionary of anecdotes	Anekdotenwörterbuch
item	Angabe
grouped article niche	gruppierte Artikelnische

Wie jedes Wörterbuchregister, so sind auch die bilingualen Register in den WSK-Bänden geordnete Mengen von Registereinträgen. Jeder Registereintrag in den Äquivalenzzugriffsregistern besteht aus einem Registerzugang (z.B. „item“) und mindestens einer auf diesen folgende Registerangabe (z.B. „**Angabe**“). Mit den Registereingängen werden die englischen Äquivalente genannt, die in der Äquivalentposition der Fachwörterbuchartikel mit der Äquivalentangabe genannt werden. Mit den Registerangaben werden die Termini genannt, die in den Fachwörterbuchartikeln mit dem Lemma genannt werden. Die Menge aller derjenigen Registerangaben, die auf die Registereingänge folgen, bilden den direkten Zugriffsbereich der alphabetischen Registerzugriffsstruktur des Äquivalenzzugriffsregisters. Eine genuine Funktion des bifunktionalen Registers besteht daher darin, Benutzerfragen direkt (d.h. ohne weiteres Nachschlagen) zu beantworten, welche zu folgendem Typ von Suchfragen gehören: WIE HEISST DAS DEUTSCHE ÄQUIVALENT ZU EINEM ENGLISCHEN

TERMINUS X? Jede Registerangabe, die auf einen Registerzugang folgt, fungiert weiterhin als reduzierte Verweisangabe. Entsprechend kann das Äquivalenzzugriffsregister eine weitere Funktion erfüllen: Es liefert dem Benutzer die Fundortdaten im Wörterverzeichnis. Demgemäß ist das Wörterverzeichnis der indirekte Zugriffsbereich der alphabetischen Registerzugriffsstruktur des Äquivalenzzugriffsregisters. Für bilingualisierte Fachwörterbücher, wie die WSK-Bände, die ein Äquivalenzzugriffsregister aufweisen, gilt, dass die Fachwörterbuchartikel zum Typ der direkt und indirekt akzessiven Wörterbuchartikel gehören. Entsprechend sind alle WSK-Bände biakzessiv.

(b) Textstrukturelle Besonderheiten von WSK-Artikeln

Alle Einzel- und Synopseartikel von WSK-Bänden weisen bestimmte strukturelle Besonderheiten auf, die bei der lexikographischen Bearbeitung der Termini zu beachten sind. Die typkonstitutiven textstrukturellen Besonderheiten der Wörterbuchartikel bestehen u.a. darin, dass zwischen den Definitionspositionen und der auf diese folgende Position für weitergehende Erklärungen unterschieden wird. Diese textuelle Zweiteilung geht auf die Auffassung zurück, dass man das fachenzyklopädische Bedeutungswissen von fachenzyklopädischen Sachwissen zwar deutlich unterscheiden kann, dass es aber als nicht strikt getrenntes und daher nicht immer eindeutig isolierbares Wissen gelten muss (vgl. Wiegand 1988a). Demgemäß wird in jeden Wörterbuchartikel eines WSK-Bandes, der kein Verweisartikel ist, zunächst mittels der lexikographischen Fachdefinition (und zwar zuerst in Deutsch und danach in Englisch) ein gegenstandskonstitutives fachenzyklopädisches Bedeutungswissen vermittelt; dadurch hat der Benutzer gleich zu Beginn der Artikellectüre die Chance zu erfahren, wovon im weiteren Artikel die Rede ist, so dass er über einen fachlichen Gegenstand als Konzept verfügt, zu welchem erst später in der textuellen Position für weitergehende Erklärungen fachenzyklopädisches Sachwissen mitgeteilt wird; es wird mithin zuerst ein bestimmter fachlicher Gegenstand nicht nur durch seine Nennung mit der Terminusgestaltangabe, sondern auch definitorisch eingeführt, und dann wird über diesen geschrieben. Letzteres geschieht in der Form eines Angabetextes, in dem keine Textkondensationen auftreten, die sich auf die Syntax auswirken. Alle Artikel in den WSK-Bänden sind entsprechend partiell kondensierte Fachwörterbuchartikel (i.S.v. Wiegand 2003), so dass sie gut lesbar sind.

(c) Datenakzessivität der WSK-Bände

Alle WSK-Bände weisen einen extern und intern mehrdimensionale Datenakzessivität (oder: Datenzugriffsbereitschaft) auf. Die Mehrdimensionalität der Datenakzessivität zeigt sich *erstens* darin, dass alle WSK-Bände extern biakzessiv sind; dies bedeutet: *Erstens* ist direkter externer Zugriff auf jedes Lemma als Element der makrostrukturellen äußeren Zugriffsstruktur möglich. *Zweitens*

ist ein indirekter externer Zugriff auf die Lemmata des Wörterverzeichnis möglich. Dieser wird vermittelt vom Englisch–Deutschen Äquivalenzzugriffsregister. *Drittens* ist das Englisch–Deutsche Äquivalenzzugriffsregister, wenn mit konsultativen Suchfragen vom Typ WIE LAUTET DER DEUTSCHE ÄQUIVALENTE TERMINUS Y ZUM ENGLISCHEN TERMINUS X? auf seine Registereingänge zugegriffen wird, extern monoakzessiv.

Die Mehrdimensionalität der Datenakzessivität zeigt sich *zweitens* darin, dass alle WSK-Bände intern polyakzessiv sind; dies bedeutet: Der mit den verweisvermittelnden Textsegmenten in der jeweiligen Verweisausgangsposition beginnende und im jeweiligen Verweiszielbereich endende, vom Benutzer-in-actu etablierte, binnenkurrente äußere Zugriffspfad, dessen Pfadende entweder in einer anderen Artikelposition oder in einer anderen Textverbundkonstituente liegt, führt meistens über ein innere Zugriffsstruktur oder über eine innere Schnellzugriffsstruktur, die Zugriff auf den jeweiligen Verweiszielbereich eröffnen. Damit sind die textuellen Voraussetzungen für eine optimale wörterbuchinterne Orientierung gegeben.

(d) Mediostrukturelle Vernetzung

Alle WSK-Bände weisen eine wörterverzeichnisübergreifende mediostrukturelle Vernetzung der auf zugriffsverschiedene bandinterne Positionen verteilten lexikographischen Daten auf. Ohne dass eine angemessene Menge von mediostrukturellen Netzen textuell verankert ist, kann ein gedrucktes Fachwörterbuch nicht beanspruchen, eine Lernkomponente aufzuweisen. Denn gelernt werden sollen nicht isolierte Termini und das relativ isolierte zugehörige Wissen, sondern größere zusammenhängende Wissensausschnitte. Bei jedem alphabetischen Wörterbuch ist die externe Datenakzessivität primär. Die größtmögliche Zugriffsbereitschaft und damit auch eine rasche Auffindbarkeit der artikelinternen Daten rangiert daher im alphabetischen Wörterbuch stets vor der lexikographischen Darstellung terminusübergreifender Wissenszusammenhänge. Daher müssen durch funktionsgerechte lexikographische Verweisungen die textuellen Voraussetzungen dafür geschaffen werden, dass ein Benutzer-in-actu diejenigen Wissenszusammenhänge durch Verweisfolgung wieder herstellen kann, die er kennen muss, um das Fachwörterbuch nicht nur in konfliktbedingten Konsultationssituationen punktuell erfolgreich benutzen zu können, sondern auch in Wissensrecherchesituationen, die er durch die Ausführung von verweismotivierten Konsultationshandlungen zu einer — vom Rechercheziel bestimmten — Folge von Wissensrecherchesituationen verknüpfen kann, die dann als eine wörterbuchinterne Fachwissensrecherche gelten (vgl. Wiegand 2006).

Die für den Typ des Lern- und Konsultationswörterbuchs am meisten charakteristischen mediostrukturellen wörterverzeichnisübergreifenden Netze entstehen durch die bidirektionale Verweisung zwischen den festen finalen Verweispositionen in den Paragraphen der Systematischen Einführung und

den binnensituierten Verweispositionen in den Synopseartikeln; durch diese werden die wörterverzeichnisinternen mediostrukturellen Artikelnetze wirksam ergänzt, die u.a. durch die unidirektionale Verweisung von Einzel- auf Synopseartikeln gegeben sind.

Alle WSK-Bände weisen weiterhin eine bandübergreifende Verweisung zwischen denjenigen Wörterbuchartikeln auf, zu denen es gleichlautende Lemmata gibt.

4. Die Ausprägungen typkonstitutiver Eigenschaften bei der Artikelform am Beispiel

Im folgendem wird die Artikelform unter besonderer Berücksichtigung typkonstitutiver Eigenschaften am Beispiel des Artikels zum Mehrwortterminus *gruppierte Artikelnische* charakterisiert (vgl. Abb.). Es handelt sich um einen monosemen Terminus, der nicht positionenreduziert ist, so dass die beiden fakultativen textuellen Artikelpositionen, die Synonymen- und die Antonymenposition, da sowohl ein synonyme als auch ein antonyme Terminus existiert, vorhanden sind. Alle Artikel in WSK-Bänden, außer den Verweisartikeln, sind partiell kondensierte Wörterbuchartikel (i.S.v. Wiegand 2003: 203ff): Denn die Position für weitergehende Erklärung muss als Angabetext gestaltet sein, so dass die natürlichsprachliche Syntax vollständig ausgeprägt ist und keine Kondensationsformen auftreten, die sich auf die Syntax auswirken. Einzelartikel sind solche Wörterbuchartikel, in denen sich die lexikographische Bearbeitung nur auf das Lemmazeichen und dessen usuelles Bezugsobjekt bezieht. In dem Beispielartikel dienen mithin alle lexikographischen Aussagen dazu, dem Benutzer den Terminus *gruppierte Artikelnische* zu erklären und damit auch, was eine Artikelnische ist. Dies geschieht so, dass mit dem Lemma der Terminus als Definiendum genannt wird. Daraufhin wird in der deutschen Definiensposition das Definiens genannt, und zwar mit einer adjazent linksadressierten Definiensangabe, so dass in den beiden ersten obligatorischen textuellen Positionen des Artikels insgesamt eine verdichtete Fachdefinition präsentiert wird; die Verdichtung besteht darin, dass der Definitor fehlt. Für einen Benutzer-in-actu, der diese Definition versteht, ist damit ein lexikographischer Gegenstand auf metalexikographischer Grundlage fachlexikographisch konstituiert, so dass der Benutzer über das fachgegenstandskonstitutive Bedeutungswissen verfügt. Eine besondere typkonstitutive Eigenschaft der Artikelform besteht darin, dass auf die deutsche Definiensposition die obligatorische englische Definitionsposition folgt. Sie beginnt mit der englischen Terminusäquivalentangabe (hier „*grouped article niche*“), mit der das Äquivalent zum deutschen Terminus genannt wird, das zugleich als Definiendum für die englische Definiensangabe fungiert, die eine Übersetzung der deutschen ist: Die Gegenstandskonstitution kann somit auch von solchen Benutzern vorgenommen werden, die kein oder nicht genügend Deutsch aber ausreichend Englisch können.

Ein Benutzer, der ein konsultative Suchfrage vom Typ WAS BEDEUTET DER TERMINUS X? oder eine vom Typ WAS IST EIN Y? im Rahmen einer Konsultationshandlung an einen WSK-Band gerichtet hat, kann nach der erfolgreichen Lektüre einer der beiden definitiven Positionen bereits über eine für ihn ausreichende Antwort verfügen. Er braucht den Angabetext in der Position für weitergehende Erklärungen nicht zu lesen. Nur Benutzer, die die Definitionen nicht verstehen oder solche, die mehr wissen wollen, z.B. weil sie den WSK-Band nicht im Rahmen einer konfliktbedingten Konsultationssituation, sondern im Rahmen einer nichtkonfliktbedingten Wissensrecherchesituation benutzt haben, müssen die weiterführende Erklärung lesen. Die Position für weitergehende Erklärung kann als einer der zentralen Teile der Lernkomponente von WSK-Bänden gelten. Aus der Perspektive der Artikelautoren ist in dieser textuellen Position der größtmögliche Formulierungs- und Gestaltungsspielraum gegeben: Eine große Zahl von Schrifttypen steht zur Verfügung, Abbildungen aller Art sind möglich usw. Damit können je nach dem Typ des Terminus unterschiedliche Aspekte lexikographisch bearbeitet werden.

Auf die Position für weitergehende Erklärungen folgt die Synonymenposition. In dieser können mehrere Synonyme genannt werden. Zu jeder Synonymangabe in einem Einzel- oder Synopseartikel gehört ein Verweisartikel; in diesem wird das Synonym mit dem Verweislemma genannt. Zu dem Beispielartikel gehört demgemäß folgender Verweisartikel:

vollständig gruppierte Nische

→ gruppierte Artikelnische

Auf die Synonymenposition folgt die Antonymenposition; in dieser können mehrere Antonyme genannt werden.

Eine weitere typkonstitutive Eigenschaft der Artikelform besteht darin, dass alle Artikel als vorletzte Position eine feste Verweisposition aufweisen. Damit wird eine genuine Eigenschaft fachlichen Wissens, nämlich theorievermittelt zusammenhängendes Wissen zu sein, so berücksichtigt, dass es sich auch in der Artikelform zeigt. Die festgelegte Verweisposition zwingt im höheren gerade dazu, die Verweisung systematisch zu konzipieren, und zwar besonders durch eine Kontrolle der mediostrukturellen Selektion. Bei dieser ist vor allem darauf zu achten, dass die Verweisung die Gegenstandskonstitution unterstützt. In der Verweisposition wird bandintern auf die Paragraphen der Systematischen Einführung und auf andere Artikel, wörterbuchserienintern auf Artikel in anderen Bänden sowie auf die CD-ROM und weiterhin mit der Autorensiglenangabe auf das bandinterne Autorenverzeichnis verwiesen. Die wörterbuchserieninterne Verweisung und die auf die CD-ROM sind fakultativ.

Die letzte textuelle Position ist die Literaturposition; sie ist obligatorisch. Die Titel sind alphabetisch geordnet, so dass eine positionenspezifische alphabetische Zugriffsstruktur gegeben ist. In der Position für weitergehende Erklärung darf auf die Titel verwiesen werden.

Alle textuellen Positionen beginnen mit einer Identifizierungsangabe. Bei der deutschen Definiensangabe ist sie als Nullangabe realisiert, bei allen anderen Positionen durch ein Angabesymbol. Jede textuelle Position beginnt mit einer neuen Zeile, so dass jede textuelle Position einen eigenen Textblock bildet. Die WSK-Artikel weisen dadurch eine übersichtliche vertikale Textarchitektur auf.

Die Synopseartikel unterscheiden sich von den Einzelartikeln nur geringfügig, und zwar nur positionenintern: Die Anzahl der erlaubten Verweisadressenangaben und die der Literaturangaben ist größer. Damit sind die wichtigsten Eigenschaften von WSK-Artikeln am Beispiel erläutert. Eine genaue Darstellung der Artikelmikrostrukturen war hier nicht beabsichtigt. Sie findet sich in Schierholz und Wiegand (2005).

5. Schlussbemerkung

Es besteht die Absicht, jeweils dann eine englische Version eines WSK-Bandes zu bearbeiten, wenn der deutsche Band erschienen ist.

Anmerkung

1. Auch das in Arbeit befindliche *Wörterbuch zur Lexikographie und Wörterbuchforschung* (WLWF) gehört zum Typ des Lern- und Konsultationswörterbuchs. Allerdings sind seine Adressaten vor allem Experten, während die Adressaten der WSK-Bände in erster Linie Semixperten sind. Zum WLWF vgl. man: <www.fabulex.de> und die dort genannte Literatur.

Literatur

- Abraham, Werner.** 1988. *Terminologie zur neueren Linguistik*. 2., völlig neu bearb. u. erw. Aufl. Tübingen: Max Niemeyer Verlag.
- Bußmann, Hadumod.** 2002. *Lexikon der Sprachwissenschaft*. 3., aktual. u. erw. Aufl. Stuttgart: Alfred Kröner Verlag.
- Glück, Helmut und Friederike Schmöe (Hrsg.).** 2005. *Metzler Lexikon Sprache*. 3., Neubearb. Aufl. Mit 40 Abbildungen u. 12 vierfarbigen Karten. Stuttgart/Weimar: Verlag J.B. Metzler.
- Lewandowski, Theodor.** 1990. *Linguistisches Wörterbuch*. 5. überarb. Aufl. Uni-Taschenbücher 1518. Heidelberg: Quelle und Meyer.
- 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.
- Wiegand, Herbert Ernst.** 1988. Vorüberlegungen zur Wörterbuchtypologie. Teil I. Hyldgaard-Jensen, Karl und Arne Zettersten (Hrsg.). *Symposium on Lexicography III. Proceedings of the Third International Symposium on Lexicography. May 14-16, 1986 at the University of Copenhagen*: 3-105. *Lexicographica. Series Maior* 19. Tübingen: Max Niemeyer Verlag.

- Wiegand, Herbert Ernst.** 1988a. Was eigentlich ist Fachlexikographie? Mit Hinweisen zum Verhältnis von sprachlichem und enzyklopädischen Wissen. Munske, Horst Haider, Peter von Polenz, Oskar Reichmann und Reiner Hildebrandt (Hrsg.). 1988. *Deutscher Wortschatz. Lexikologische Studien. Ludwig Erich Schmitt zum 80. Geburtstag v. seinen Marburger Schülern*: 729-790. Berlin/New York: Walter de Gruyter.
- Wiegand, Herbert Ernst.** 1998. *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie. 1. Teilband.* Mit 159 Abbildungen im Text. Berlin/New York: Walter de Gruyter.
- Wiegand, Herbert Ernst.** 2003. Überlegungen zur Typologie von Wörterbuchartikeln in Printwörterbüchern. Ein Beitrag zur Theorie der Wörterbuchform. *Lexicographica* 19: 169-313.
- Wiegand, Herbert Ernst.** 2004. *Die Wörterbücher zur Sprach- und Kommunikationswissenschaft (WSK) und ihre Benutzungsmöglichkeiten im Fach Deutsch als Fremdsprache.* Vortrag gehalten auf dem 3. Internationalen Kolloquium zur Wörterbuchforschung, angeboten in Varna, Bulgare, Oktober 2004. [Erscheint 2006]

Internationalisation, Localisation and Customisation Aspects of the Dictionary Application *TshwaneLex**

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

Abstract: TshwaneLex is the world's only lexicography software suite with which the entire lexicographic process, from initial compilation all the way to final product, may be conducted in the language of one's choice. This is possible thanks to various aspects of internationalisation, localisation and customisation that are built into TshwaneLex. These are discussed by means of examples drawn from a wide variety of projects and languages.

Keywords: INTERNATIONALISATION, LOCALISATION, MAINSTREAM LOCALISATION, DEVELOPMENT LOCALISATION, BLOWBACK LOCALISATION, CUSTOMISATION, TSHWANELEX, LEXICOGRAPHY, DICTIONARY, SOFTWARE, LANGUAGE INTERFACE PACK (LIP), CILUBÀ, ISIZULU, SESOTHO SA LEOA, SETSWANA, SWAHILI, WELSH

Samenvatting: **Aspekte van internasionaaliserings, lokaalisering en aanpasbaarheid in die woordeboektoepassing *TshwaneLex*.** TshwaneLex is 's werelds enige lexicografiese software suite waarmee het volledige lexicografiese proses, van initiële samestelling tot en met het eindprodukt, in een taal naars keuse kan word uitgevoer. Dit is moegelik dankzij verskillende aspekte van internasionaaliserings, lokaalisering en aanpasbaarheid die in TshwaneLex word ingebou. Die woorde bespreek met behulp van voorbeelde uit een breed gamma van projekte en tale.

Sleutelwoorde: INTERNASIONALISATIE, LOKALISATIE, HOOFDSTROOMLOKALISATIE, ONTWIKKELINGSLOKALISATIE, TERUGKOPPELINGSLOKALISATIE, AANPASBAARHEID, TSHWANELEX, LEXICOGRAFIE, WOORDENBOEK, SOFTWARE, LANGUAGE INTERFACE PACK (LIP), CILUBÀ, ISIZULU, SESOTHO SA LEOA, SETSWANA, SWAHILI, WELSH

* This article combines material that was presented by the author at the Eleventh International Conference of the African Association for Lexicography (organised by the Tshivenda National Lexicography Unit, University of Venda for Science and Technology, Thohoyandou, Republic of South Africa, 5–7 July 2006), and the Second Conference of the International Association for Translation and Intercultural Studies (hosted by the Centre for Humanities Research, University of the Western Cape, Bellville, Republic of South Africa, 11–14 July 2006).

1. Localising Software in Today's Global Village

It should not come as a surprise that in today's global village, the localisation of software is picking up momentum. Just this April, for example, the software giant *Microsoft* launched its isiZulu LIP (Language Interface Pack) for Windows XP, which was followed in June by the LIP for Setswana. By the time this article will be in press, it is expected that the Afrikaans version will also have been released. What is true for these generic mass-produced applications has now also become a reality for the highly specialised field of lexicography. Indeed, *TshwaneLex* — the world's only truly off-the-shelf dictionary compilation software, and the flagship product of *TshwaneDJe HLT* — is now not only fully customisable, but also fully localisable. This thus means that the entire lexicographic process, from initial compilation all the way to final product, may henceforth be conducted in any language of one's choice. The purpose of this article is to highlight the main aspects that make this possible.

2. Internationalisation and Localisation: What are these?

The two key concepts 'internationalisation' and 'localisation' are summarised as follows by Schmitz (2006: slides 3 and 4):

What is internationalization (I18N)?

Developing a product or service in such a way that it will be easy to adapt to other markets (languages and cultures)

- Goal: Eliminate the need to redevelop, reprogram or recompile the original product / content
- Carried out by the development group in conjunction with the localisation people

What is localisation (L10N)?

Adapting a product or service to a local or regional market

- Goal: Appropriate linguistic and cultural aspects
- Performed by translators, localisers, language engineers

Note the crucial difference between the first two words in each definition: 'developing' versus 'adapting'. Internationalisation (often abbreviated to just I18N — being the first and last letters with '18' representing the number of letters in-between, and pronounced 'I-eighteen') is thus what developers do *once*; while localisation (analogously abbreviated to L10N and pronounced 'L-ten') is what one has to do *over and over*. This brings Schmitz (2006: slide 6) to conclude: "the more stuff you push into I18N out of L10N, the less complicated and expensive the process becomes".

Although Schmitz's conclusion is of course true, one may deplore the focus on the monetary aspect. Actually, this focus is the result of current 'mainstream

localisation' efforts, summarised by Schäler (2006: slide 8) with the following notion:

Increase return on investment (ROI)

- IF there are markets rich enough to buy our products
- THEN adapt our already developed products superficially to the requirements of these markets (with a minimum effort)
- AND sell them into these new markets for a similar price as the original product (there is no easier way to make money)

Schäler (2006: slide 10) is indeed correct when he claims that "mainstream localisation efforts focus on short-term return on investment and ignore currently non-profitable regions".

3. Brief Internationalisation and Localisation Case Study: *Microsoft's LIPs (Language Interface Packs) for South Africa*

If one focuses on the South African market, and as mentioned at the outset, one sees that work has been done on three localised versions of the Windows XP operating system, by means of the creation of three LIPs. In the words of Rolfe (2002: abstract):

A Language Interface Pack (LIP) is a package that allows users to install a particular language skin on top of their English operating system to provide them with an almost fully localized operating system (OS) User Interface (UI) for the chosen language. This is accomplished via Microsoft's Unicode-based Multilanguage User Interface (MUI) technology which allows the localization of resources for the most visible and most commonly used features of the operating system.

Although seemingly very impressive, what exactly are those "most visible and most commonly used features"? According to Rolfe (2002: slide 24) one needs to answer the following question:

Which 20% of the UI is used 80% of the time?

- Desktop Features
 - Start Menu & Submenus
 - Windows Explorer
- Components
 - Internet Explorer
 - Windows Media Player
- Other Features
 - Program Manager
 - Update Driver Wizard
 - Screensavers
 - Taskbar & Properties
 - Control Panel
 - Outlook Express
 - Windows Address Book
 - Task Manager
 - Remote Desktop
 - Games

- Help & Support Center content
 - Main content remains unlocalized.

In other words, for LIPs the '20-80 rule' is used, meaning that *only* the 20% of the user interface that is used 80% of the time is being localised, and the above shows a partial list of the top-level elements defined to be localised.

In order to visualise — literally — what this means during actual use, Addendum 1 shows the start page of the Outlook Express e-mail client, supposedly in Setswana. Although this application was chosen by Ntaoleng Motaung when she proudly showcased the LIP for Setswana at AFRILEX 2006, one is immediately struck by the amount of untranslated strings. One does not even need to "go deep into the application" (Ntaoleng 2006) to find parts that are not in Setswana, as one is merely looking at the start page. Basics such as 'There is 1 unread mail message in your inbox' have only been partially translated, here to *There is 1 molaetsa wa poso e e sa buisiwang mo bokosokgorogelon** [sic — should be *bokosokgorogelong*] *ya gago*. The reader will see that there are many more untranslated sections that are visible in the screenshot. Further note that even basic button texts such as 'Previous' and 'Next' are also displayed in English. Of course, there are translations in Setswana for these concepts, and as a matter of fact, in various other places of Windows XP these have indeed been translated. See in this regard for example Addendum 2, a screenshot of one of the first questions one has to answer in order to install the Setswana LIP. Here *Morago* stands for 'Previous' and *E e latelang* for 'Next'.

While the above observations could be viewed as 'unfortunate', one really starts to be concerned when one notices the high degree of inconsistencies. Simply clicking on the *Simolola* 'Start' button in Setswana reveals the pop-up window shown in Addendum 3, left. There is no reason why the capitalisation of *Me* 'My' should vary in the top-right corner:

- | | |
|----------------------------------|---------------------|
| • <i>Ditokumente tsa Me</i> | 'My documents' |
| • <i>Ditshwantsho tsa Me</i> | 'My Pictures' |
| • <i>Khomphiutara ya Me</i> | 'My Computer' |
| • <i>Mafelofaratlhatlha a me</i> | 'my Network Places' |

Ironically, the same problem reappears in the isiZulu version when clicking on the *Qala* 'Start' button, as can be seen in Addendum 3, right:

- | | |
|----------------------------------|---------------------|
| • <i>Imibhalo yami</i> | 'my Documents' |
| • <i>Izithombe Zami</i> | 'My Pictures' |
| • <i>Ikhompyutha yami</i> | 'my Computer' |
| • <i>Izindawo zomphambo wami</i> | 'my network Places' |

Regrettably, such inconsistencies are far too numerous throughout Windows XP, as is exemplified in Addendum 4 for the isiZulu LIP, where one notices as

many as three different spellings of 'Internet': *intanethi* in the top-left corner, *Internet* in the middle, and *Intanethi* (as well as again *Internet* on the same line!) in the bottom-right.

According to Mariëtta Alberts (personal communication at AFRILEX 2006, 7 July 2006), the three Technical Committees: Terminology Development of PanSALB's National Language Bodies for isiZulu, Setswana and Afrikaans were requested to verify the terminology of the LIP glossaries for the said languages. All three committees complained about inconsistencies in the translations as well as the strange usage of capitalisation. PanSALB therefore stated clearly that its National Language Bodies cannot authenticate the LIPs prior to proper feedback from stakeholders.

It is highly surprising that such basic errors and inconsistencies could have crept into the localisation effort as, with the right localisation tools, these could all easily have been avoided. In addition to this, however, it is clear that the translators and proofreaders did a very sloppy overall job, as exemplified in Addendum 5. In the string 'Add or Remove Programs', for example, the English 'or' ended up being glued to the Setswana translation *gongwe* 'or', resulting in the erroneous *Fetola gongweor* tlosa Diporokeramo*. (Also, and again, note the inconsistency in word-initial capitalisation here.)

At the IATIS 2006 conference, Dwayne Bailey, Director of the *Zuza Software Foundation*, referred to Microsoft's half-baked LIPs as "merely paying LIP service to the South African languages" (Bailey 2006). Given Bailey supervised the creation of the complete, freely available OpenOffice.org 2.0 suite — which includes the word processor Writer, the spreadsheet component Calc, the presentation application Impress, the graphics package Draw, the database program Base, and the equations and formulae tool Math — and this in *all eleven* of South Africa's official languages, he of course more than anyone else has reason to speak. The current South African LIPs are an affront to the local languages, and in no way does this type of localisation show good practice for others to copy.

4. Mainstream, Development and Blowback Localisation

Against the backdrop just sketched, it is not surprising that localisation professionals like Schäler (2006: slide 27) conclude that the "short-term return-on-investment cannot remain the only driving force behind the localisation effort". In this respect, the 'Global Initiative for Local Computing' (GILC) — with as motto *Localisation is not an option — it is a fundamental right* — was launched in September 2005, to move away from 'mainstream localisation' towards 'development localisation'. In the words of Schäler (2006: slide 13):

The main driver of the development localisation efforts is the belief that a more inspirational, visionary and innovative perspective on localisation includes a variety of reasons to localise. Among them are:

- Social reasons — the intent to bridge social divides;
- Political reasons — the effort to provide equal access to electronic information;
- Cultural reasons — the belief in the need for linguistic and cultural diversity;
- Long-term investment reasons — the conviction that business needs to take a long-term approach and invest sensibly in new and emerging markets.

Schäler (2006: slide 27) further envisages the future creation of a 'blowback effect', whereby custom solutions and technologies that are being launched in developing countries are being picked up by, or fed back into, the developed world. This future 'blowback localisation' would then allow for a genuine exchange or two-way traffic.

In hindsight, this future might already be with us when it comes to the field of lexicography software. When one studies the highly innovative internationalisation and localisation aspects of *TshwaneLex*, one realises that various useful principles developed in *TshwaneLex* could and should indeed be adopted elsewhere. In order to fully appreciate these aspects, however, it is also necessary to look at some customisation aspects of *TshwaneLex*. As such, the remainder of this article will proceed in reverse, from final product back to the heart of the software environment itself, using a variety of languages.

5. *TshwaneLex* in a Nutshell — Examples of Final Products

TshwaneLex is a professional lexicography software suite for the compilation of monolingual, bilingual and multilingual dictionaries, and for the publication of dictionaries in hardcopy, online and electronic formats. The application is currently in use by over two hundred users worldwide, from individuals to large organisations, and for over one hundred different languages. To date, over eighty people have received formal *TshwaneLex* training.

Examples of published hardcopy reference works that are currently stored in *TshwaneLex* are the *Pharos Afrikaans–Engels/English–Afrikaans Woordeboek/Dictionary* (Du Plessis et al. 2005), the *IsiNdebele–English Bilingual and Explanatory Dictionary* (Mahlangu et al. 2006), and the *Klein uitleenwoordenboek* (Van der Sijts 2006). Online examples include Gaebler's (2004–2006) *Zulu–English/English–Zulu Online Dictionary*, Randi's (2005–2006) *An Encyclopedia of Claims, Frauds, and Hoaxes of the Occult and Supernatural*, and Kabuta et al.'s (2006) *Nkòngamyakù wa Cilubà–Mfwàlànsa/Dictionnaire Cilubà–Français*.

6. Customisation Aspects of *TshwaneLex* — From Final Products to the Compilation Environment

In Addendum 6 an example is shown of an existing monolingual dictionary, produced and placed online with *TshwaneLex*, for which both the dictionary

data and the entire dictionary interface text are in the same language. In this case, a mother-tongue speaker of Sesotho sa Leboa is looking up the word *matšulela* '(traditional) grinding stones to grind grains' in the *Pukuntšutlhaloši ya Sesotho sa Leboa ka Inthanete* 'Explanatory Sesotho sa Leboa Dictionary on the Internet' (Mojela et al. 2004–2006). To date, and probably surprisingly, this work remains the only fully monolingual African-language dictionary that is available on the Internet, allowing the user to consult a reference work in a digital dictionary environment that is truly and entirely in his/her language.

Although seemingly straightforward, what a monolingual dictionary does not reveal is the separation between actual, unique dictionary contents and repetitive metalanguage (such as parts of speech, domain labels, etc.). If this separation is done properly, dynamic metalanguage customisation can, for example, instantly 'kick in' while a user consults an electronic bilingual (or multilingual) dictionary. An in-depth discussion of the notion of 'dynamic metalanguage customisation' may be found in De Schryver and Joffe (2005); suffice it here to point out that this implies, amongst others, that a dictionary's metalanguage is 'generated' (or customised, adapted) in real time in the language of the dictionary user. This, again, remains a world's first, and is illustrated in Addendum 7 for Hillewaert et al.'s (2004–2006) *Kamusi ya Kiswahili–Kiingereza Katika Mtandao/Online Swahili–English Dictionary*, a reference work in progress compiled with TshwaneLex. In the screenshot shown in the background, a person browsing the dictionary in Swahili will (for the example shown in this case) see the parts of speech labelled as *kimilikishi nomino* and *kivumishi*, while the cross-reference marker text will be displayed as *Mzizi*. For a person who browses the same dictionary, but now in English, as shown in the screenshot in the foreground, these same metalanguage strings will automatically be adapted to 'possessive pronoun', respectively 'adjective', for the parts of speech, and 'Root' for the cross-reference marker text.

Apart from separating dictionary contents from metalanguage, good lexicography software should also clearly separate these data from the structure of the articles. In technical terms this means that the contents must be 'governed by' and thus 'conform to' a DTD (Document Type Definition, or dictionary grammar), with this DTD again markedly distinct from the actual formatting (or style) of the data. An in-depth discussion of the fully customisable and built-in TshwaneLex DTD and linked Styles system may be found in Joffe and De Schryver (2005); suffice it to point out for the present discussion that all aspects of the grammar and formatting may be created in the language of the user's choice. This is illustrated in Addendum 8, where a Cilubà–French Dictionary (Kabuta et al. 2006) is being compiled within TshwaneLex. The Tree View on the left is 'converted' into one of any number of displays in the Preview Area on the right. What is important to note here is that all the labels for the components of the Tree View are entirely customisable (and have been customised) by the lexicographer: *Ngumvowikilà* stands for 'Sense', *Munyàku* for 'Combination', *Dikù-dimuna* for 'Translation equivalent', *Cileejilu* for 'Example', etc.

This tripartite division of the various levels of the data is also reflected in the layout of the TshwaneLex GUI (graphical user interface). As may be seen in Addendum 9, (1) the bottom-left quadrant is the area where the unique dictionary contents are being typed in (into input boxes) and where the repetitive metalanguage is being selected (from, for instance, drop-down menus), (2) the top-left quadrant is where the Tree View is being constructed in conformance with the DTD, and (3) the right half is the Preview Area displaying a possible output with any of a number of styles applied. Each of these levels is not only seamlessly linked to and interacts with the other levels, each is also fully customisable, including with regard to the language used. In Addendum 9, all customisation aspects were created in Cilubà. Also (and already) visible in this same screenshot, is the fact that all the text strings of the GUI itself have been localised. This, however, is the topic of the next section.

7. Localisation Aspects of TshwaneLex — Manipulating the Compilation Environment

Typical tools for software localisation include terminology management systems, translation memory applications, localisation software, and project management tools (Schmitz 2006: slide 24). In TshwaneLex, however, all of these were brought together into a *single powerful built-in Localisation editor*. All relevant text strings that appear throughout the various menus, dialogue boxes, messages, tabs, buttons, the status bar, etc. are automatically presented in the Localisation editor. This can be seen in Addendum 10, where localisation of TshwaneLex is ongoing for Welsh.

Also note that all strings where the English term 'View' appears were brought together in this example, which enables the translator to make sure that the *translated terminology is consistent*. One simply has to recall the haphazard translation of the word 'Internet' throughout the isiZulu LIP for Windows XP, for example, to realise that being able to bring related material together while translating is indeed an added advantage.

Furthermore, another powerful feature that helps ensure consistency and *contextual correctness* is an indication of where in the GUI each particular text string appears. This is shown in the column headed by 'Key' in Addendum 10. This first column, as well as the columns with the original and translated strings, can also be sorted by simply clicking on the respective headers, which again enables one to bring related material together.

Any number of localised versions of the GUI can be prepared and each immediately becomes available within TshwaneLex. Further note that all the text strings are kept in a single file per language, which for instance means that all the text of each localised version can easily be spellchecked in one pass. With the latter, non-words such as for example *gongweor**, which, as discussed above, found its way into the Setswana LIP for Windows XP, will most definitely be instantly picked up.

Particularly handy, finally, is the fact that the results of the localisation can be seen *in real time* within TshwaneLex itself by simply clicking 'Apply [Partial]'. As an illustration, random examples of localised sections of TshwaneLex are shown in Addendum 11, for both Welsh and Cilubà.

8. Internationalisation Aspects of TshwaneLex — From the Compilation Environment to the Program Code

In order to achieve the localisation aspects discussed in the previous section, TshwaneLex of course had to be internationalised. Simply speaking, this internationalisation was achieved through a strict separation between the actual (C++) program code on the one hand, and all the textual data that appears in the GUI of TshwaneLex on the other. This internationalisation, in combination with the fact that Unicode is supported on all levels throughout the application, means that TshwaneLex can now easily be adapted to other markets, and this by the software *users themselves*. Users can even select/create localised terms suitable to their purposes.

There is, however, an additional advantage to the internationalisation of TshwaneLex from a developer's angle, namely the fact that it was an important step towards being able to generate *multiple applications* (TshwaneLex 1.0, TshwaneLex 2.0, TshwaneTerm 1.0, TshwaneTerm 2.0, etc.) from the same code base.

9. Conclusion: Candidates for 'Blowback Localisation'

Implicit in this article was the conviction that localisation in lexicography can only be truly successful if, in addition to the localisability of the graphical user interface, the dictionary contents on the one hand, and the building blocks to bring those contents together on the other, have such a degree of customisability that the (meta)language can also be adapted on those levels.

With specific regard to the internationalisation and localisation study of the lexicography software TshwaneLex, the following elements appear to be good candidates for 'blowback localisation':

- put the localisation in the hands of the users themselves, through a *built-in localisation editor*, without the need for any other extra software;
- enable all the text strings to be searched for, brought together, and sorted in various ways, so as to *ensure consistency* in the translations;
- make sure all the translated textual material is always seen *in context*;
- ensure *easy spellchecking* of all the translated material, and this by means of just one pass;
- allow users to add and change *any number of languages*, in addition to the default one(s);
- consider the *instant visualisation* of the result.

Summarising, and returning to the software under discussion: While one is using *TshwaneLex* to localise *TshwaneLex*, the emerging localised version of *TshwaneLex* appears in front of the translator's eyes — arguably the ultimate in localisation.

Acknowledgements

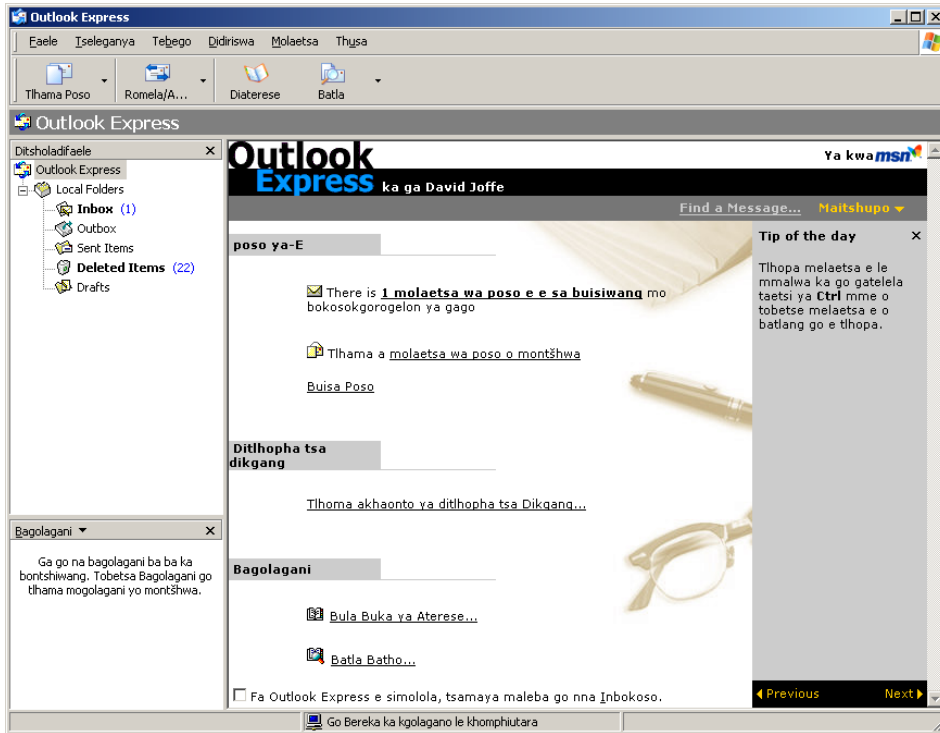
I would firstly like to extend my great appreciation for David Joffe, who not only is the brilliant software engineer behind *TshwaneLex*, turning the wildest dreams into workable products, but also the chief conceptualiser of the material presented here. I would further like to thank the people who were responsible for the creation of the multilingual data, viz. Maropeng V. Mojela and his team for Sesotho sa Leboa (Data © 2004–2006), Sarah Hillewaert and her team for Swahili (Data © 2004–2006), Ngo S. Kabuta and his team for Cilubà (Data © 2006), and Dewi Evans and his team for Welsh (Data © 2006).

References

- Bailey, Dwayne.** 2006. Stories of an African Localiser. Paper presented at the *Second Conference of the International Association for Translation and Intercultural Studies* (University of the Western Cape, Bellville, Republic of South Africa, 11–14 July 2006).
- De Schryver, Gilles-Maurice and David Joffe.** 2005. Dynamic Metalanguage Customisation with the Dictionary Application *TshwaneLex*. Kiefer, F., G. Kiss and J. Pajzs (Eds.). 2005. *Papers in Computational Lexicography, COMPLEX 2005*: 190–199. Budapest: Linguistics Institute, Hungarian Academy of Sciences.
- Du Plessis, Madaleine et al.** 2005. *Pharos Afrikaans–Engels/English–Afrikaans Woordeboek/Dictionnaire*. Cape Town: Pharos.
- Gaebler, Carsten.** 2004–2006. *Zulu–English/English–Zulu Online Dictionary* [online]. Available: <<http://isizulu.net/>>.
- Hillewaert, Sarah, Pitta Joffe and Gilles-Maurice de Schryver.** 2004–2006. *Kamusi ya Kiswahili–Kiingereza Katika Mtandao/Online Swahili–English Dictionary* [online]. Available: <<http://africanlanguages.com/swahili/>>.
- Joffe, David and Gilles-Maurice de Schryver.** 2005. Representing and Describing Words Flexibly with the Dictionary Application *TshwaneLex*. Ooi, V.B.Y., A. Pakir, I. Talib, L. Tan, P.K.W. Tan and Y.Y. Tan (Eds.). 2005. *Words in Asian Cultural Contexts, Proceedings of the 4th Asialex Conference, 1–3 June 2005, M Hotel, Singapore*: 108–114. Singapore: Department of English Language and Literature & Asia Research Institute, National University of Singapore.
- Kabuta, Ngo S. et al.** 2006. *Nkòngamyakù wa Cilubà–Mfwàlànsa/Dictionnaire Cilubà–Français* [online]. Available: <<http://www.ciyem.ugent.be/>>.
- Mahlangu, Sponono K. et al.** 2006. *IsiNdebele–English Bilingual and Explanatory Dictionary*. Cape Town: Phumelela.
- Microsoft.** 1975–2006. Microsoft Corporation [online]. Available: <<http://www.microsoft.com/>>.

- Mojela, Maropeng V. et al.** 2004–2006. *Pukuntšutlhaloši ya Sesotho sa Leboa ka Inthanete (Explanatory Sesotho sa Leboa Dictionary on the Internet)* [online]. Available: <<http://africanlanguages.com/psl/>>.
- Motaung, Ntaoleng.** 2006. Microsoft Language Interface Packs (LIPs) for South African Languages. Paper presented at the *Eleventh International Conference of the African Association for Lexicography* (University of Venda for Science and Technology, Thohoyandou, Republic of South Africa, 5–7 July 2006).
- OpenOffice.org.* 2000–2006. Free Office Suite [online]. Available: <<http://www.openoffice.org/>>.
- Randi, James.** 2005–2006. *An Encyclopedia of Claims, Frauds, and Hoaxes of the Occult and Supernatural* [online]. Available: <<http://randi.org/encyclopedia/>>.
- Rolfe, Russ.** 2002. Windows XP Language Interface Packs (LIPs) — Localized OSs for the Masses. Paper presented at the *Twenty-second International Unicode Conference* (San Jose, California, USA, 9–13 September 2002).
- Schäler, Reinhard.** 2006. The Global Initiative for Local Computing (GILC) — Working Together, or: The Genesis of Localisation. Paper presented at the *Second Conference of the International Association for Translation and Intercultural Studies* (University of the Western Cape, Bellville, Republic of South Africa, 11–14 July 2006).
- Schmitz, Klaus-Dirk.** 2006. Terminology and Localization: Principles and Tools. Paper presented at the *UNESCO TermTrain Training Week* (South African Bureau of Standards, Pretoria (Tshwane), Republic of South Africa, 27–31 March 2006).
- TshwaneDJe HLT.* 2002–2006. TshwaneDJe Human Language Technology [online]. Available: <<http://tshwanedje.com/>>.
- Van der Sijs, Noline.** 2006. *Klein uitleenwoordenboek*. The Hague: Sdu Uitgevers.
- Zuza Software Foundation.* 2003–2006. OpenOffice.org 2.0 in all 11 official South African languages [online]. Available: <<http://translate.org.za/content/view/17/54/>>.

Addendum 1: Microsoft's Outlook Express with the Setswana LIP (Note the large amount of untranslated material)



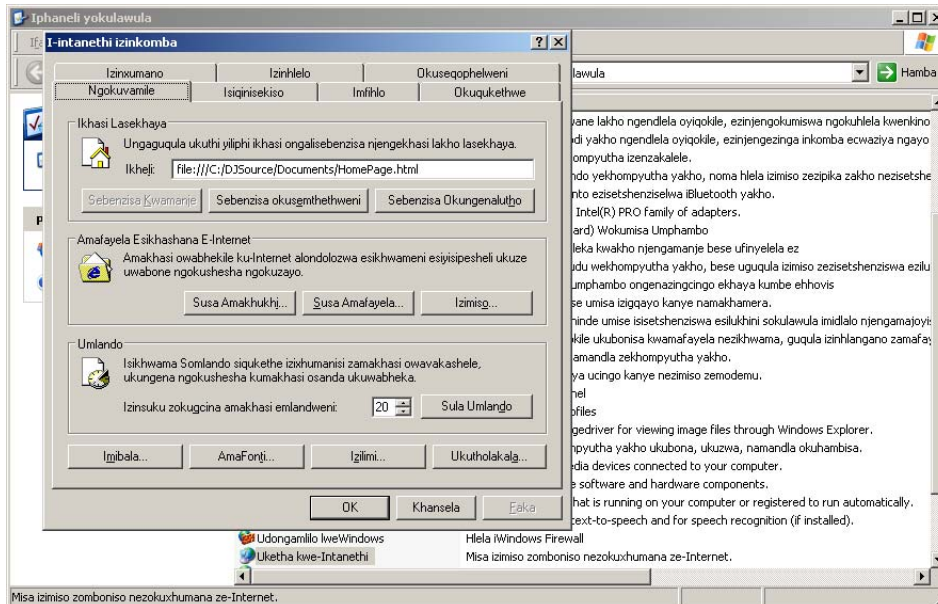
Addendum 2: Installing Microsoft's LIP for Setswana (Note the *Morago* 'Previous' and *E e latelang* 'Next' buttons, and compare with bottom-right of Addendum 1)



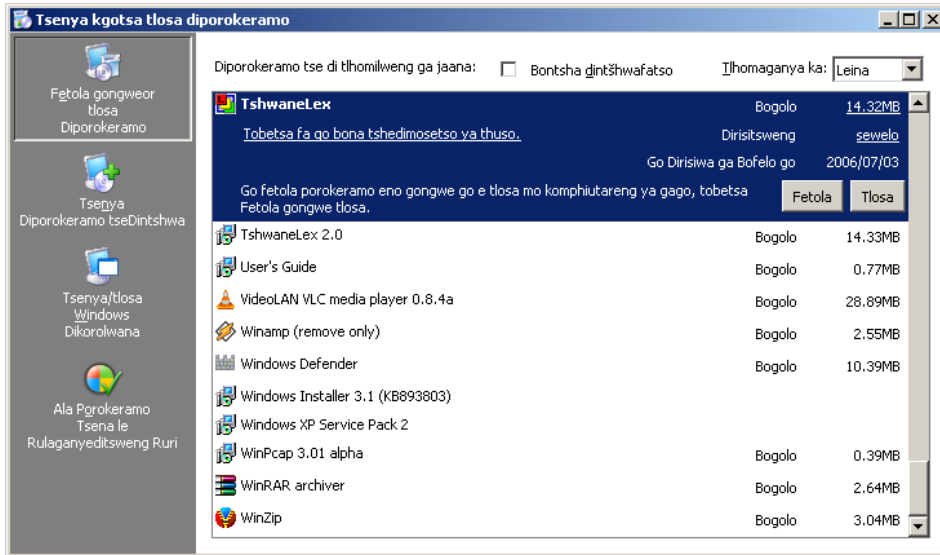
Addendum 3: Microsoft's 'Start' options with the Setswana (left) and isiZulu (right) LIPs (Note the inconsistencies in the top-right corners)



Addendum 4: Microsoft's Internet Explorer 'Internet Options' with the isiZulu LIP (Note the various spellings for Internet as *intanethi*, *Internet*, and *Intanethi*)



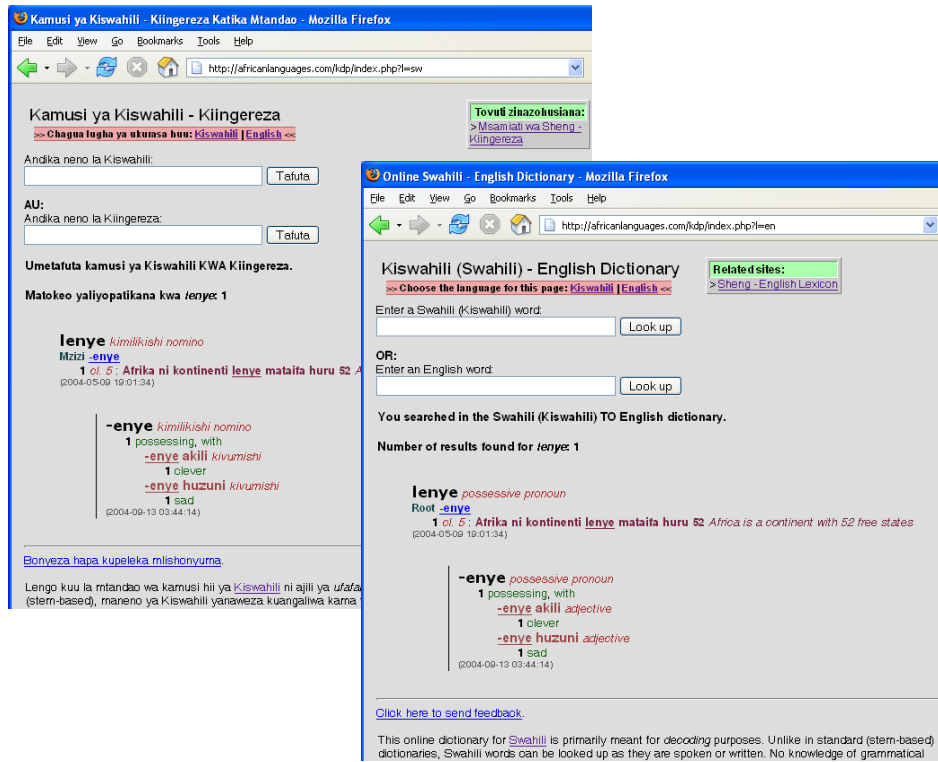
Addendum 5: 'Add or Remove Programs' in *Microsoft's* Control Panel with the Setswana LIP (Note, for instance, that the Setswana *gongwe* 'or' was simply glued to the English 'or' in the top-left corner)



Addendum 6: 'Explanatory Sesotho sa Leboa Dictionary on the Internet', created and maintained with *TshwaneLex*



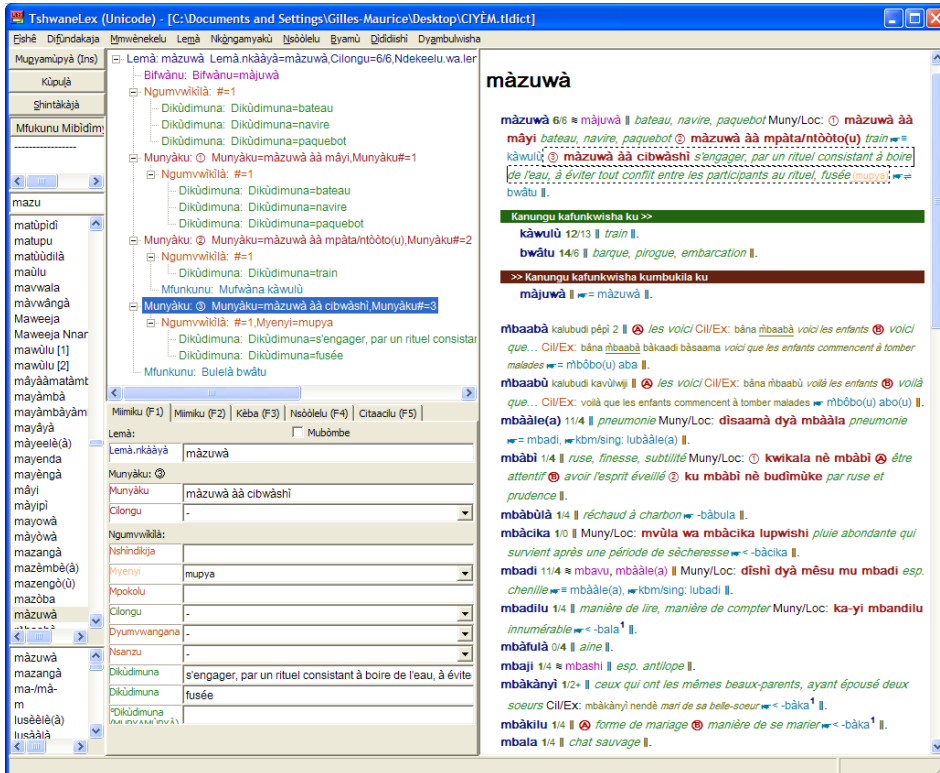
Addendum 7: Dynamic metalanguage customisation in real time on the Internet for an 'Online Swahili–English Dictionary', achieved with TshwaneLex



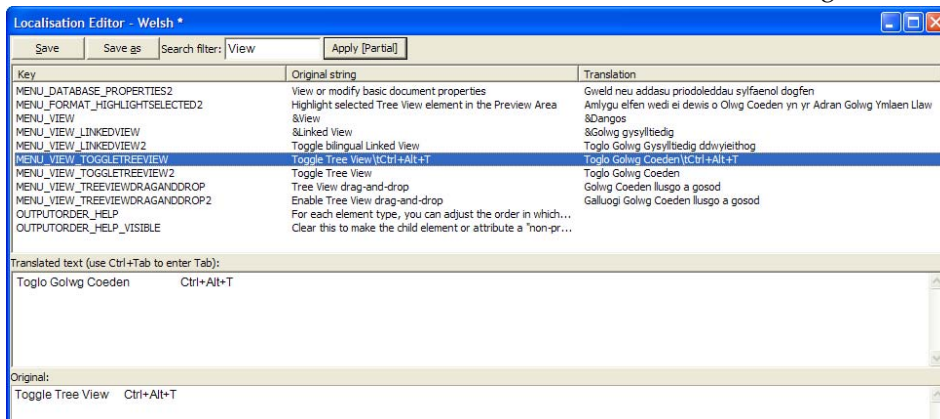
Addendum 8: The fully customisable DTD-driven structure for one article (left), and one possible output/display (right), as seen for a 'Cilubà–French Dictionary' within TshwaneLex



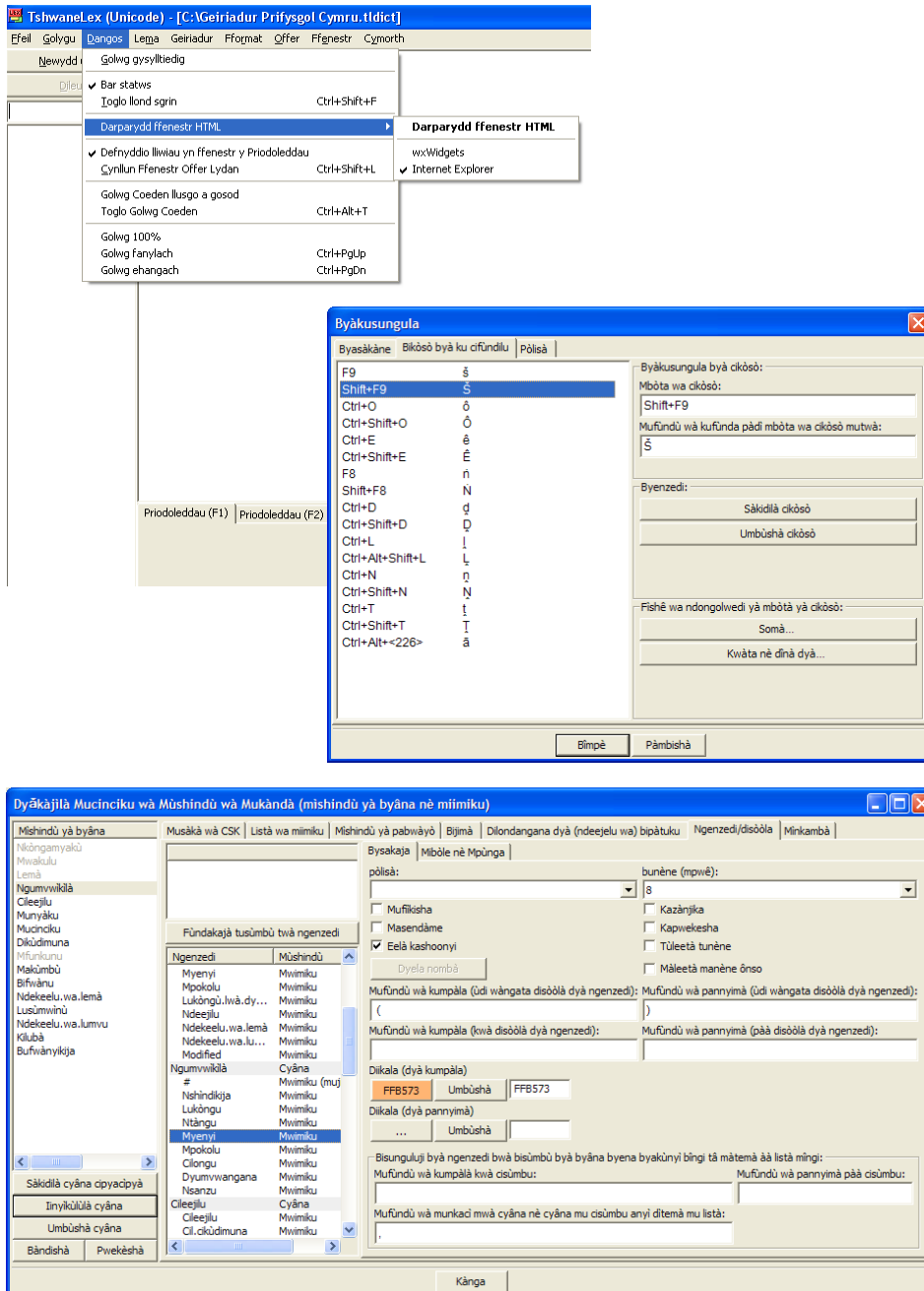
Addendum 9: Different levels of data in TshwaneLex, here for a 'Cilubà-French Dictionary': (1) dictionary contents versus metalanguage in the bottom-left, (2) article structure which conforms to the DTD in the top-left, and (3) output with any of a number of styles applied on the right



Addendum 10: The built-in TshwaneLex Localisation editor, showing Welsh



Addendum 11: Translated GUI sections of TshwaneLex: Main menu (top-left quarter) in Welsh, and Options editor as well as complete DTD-Styles editor in Cilubà



Des considérations sur la lexicographie et la linguistique au Gabon*

Guy-Modeste Ekwa Ebanéga (13751735@sun.ac.za) et

Fatima Tomba Moussavou (13751743@sun.ac.za), *Département d'afrikaans et de néerlandais, Université de Stellenbosch, Stellenbosch, République d'Afrique du Sud*

Résumé: Le Gabon est un pays à diversité linguistique. La constitution gabonaise reconnaît le français comme l'unique langue officielle. En outre, elle s'efforce de promouvoir et de protéger les langues nationales. La lexicographie, discipline de la production des dictionnaires, et la linguistique, l'étude scientifique du langage, ont toutes les deux un rôle à jouer dans la promotion et la protection des langues nationales. L'objectif de cet article est de montrer l'importance d'une collaboration entre lexicographes et linguistes pour un développement effectif des langues du Gabon.

Pour rendre compte de la nécessité de cette collaboration entre les lexicographes et les linguistes au Gabon, cet article commence par la présentation de la situation linguistique du Gabon. Ensuite des aspects tels que les contributions des deux disciplines de la linguistique et de la lexicographie au développement des langues gabonaises, la collaboration entre les linguistes et les lexicographes et la question des unités lexicographiques gabonaises seront discutés.

Mots-clés: LEXICOGRAPHIE, LINGUISTIQUE, COLLABORATION, LEXICOGRAPHERS, LINGUISTES, SCIENCES INTERDISCIPLINAIRES, LANGUES GABONAISES, DICTIONNAIRE, LANGUE, CORPUS, UNITÉS LEXICOGRAPHIQUES, DÉVELOPPEMENT

Abstract: Considerations on Lexicography and Linguistics in Gabon. Gabon is a country with a linguistic diversity. The Gabonese constitution recognizes French as the only official language. Moreover, it endeavours to promote and protect the national languages. Lexicography, the discipline for the producing of dictionaries, and linguistics, the scientific study of language, both have a role to play in the promotion and protection of the national languages.

To account for the necessity of this collaboration between lexicographers and linguists in Gabon, this article starts by presenting the linguistic situation in Gabon. Then aspects such as the contribution of the two disciplines of linguistics and lexicography to the development of the Gabonese languages, the collaboration between linguists and lexicographers and the question of Gabonese lexicographic units will be discussed.

Keywords: LEXICOGRAPHY, LINGUISTICS, COLLABORATION, LEXICOGRAPHERS, LINGUISTS, INTERDISCIPLINARY SCIENCES, GABONESE LANGUAGES, DICTIONARY, LANGUAGE, CORPUS, LEXICOGRAPHIC UNITS, DEVELOPMENT

* Cet article est une version adaptée de la communication présentée à la Neuvième Conférence Internationale Annuelle de l'Association Africaine pour la Lexicographie (AFRILEX), organisée par le Groupe de Recherche en Langues et Cultures Orales (GRELACO), Université Omar Bongo, Libreville, Gabon, 18–23 juillet 2004.

1. Introduction

Le Gabon, comme la majorité des pays africains, est un pays à diversité linguistique. À côté des langues locales, coexistent les langues parlées par les différentes communautés linguistiques d'immigrants ou de migrants de différents pays. La majorité des langues gabonaises est bantoue. Le français jouit d'un statut particulier en tant que langue introduite par la colonisation; il jouit également d'un statut privilégié parmi toutes les langues présentes au Gabon. La constitution gabonaise reconnaît le français comme l'unique langue officielle. En outre, elle s'efforce de promouvoir et de protéger les langues nationales. La lexicographie, discipline de la production des dictionnaires, et la linguistique, l'étude scientifique de la langue, ont toutes les deux un rôle à jouer dans la promotion et la protection des langues nationales. L'objectif de cet article est de montrer l'importance d'une collaboration entre lexicographes et linguistes pour un développement effectif des langues du Gabon.

Pour rendre compte de la nécessité de cette collaboration entre les lexicographes et les linguistes au Gabon, notre propos commencera par la présentation de la situation linguistique au Gabon. Des aspects tels que les contributions des deux disciplines linguistiques et lexicographiques au développement des langues gabonaises, la collaboration entre les linguistes et les lexicographes et la question des unités lexicographiques gabonaises seront également discutés.

2. Situation linguistique

Le Gabon est un pays multilingue avec plus de 40 langues gabonaises; en outre le français est la langue officielle (cf. Grimes 1996: 1). Il est également la langue véhiculaire et la langue d'ascension sociale (cf. Mavoungou 2002: 230), langue de la communication individualisée et institutionnalisée¹ (cf. Mba-Nkoghe 1991: 20-23), ainsi que la langue d'immigrants africains syro-libanais d'introduction récente. La majorité des langues gabonaises est bantoue selon les travaux de Guthrie (1953: 55-73), Doke et Cole (1963: 63-76), Jacquot (1978) et Kwenzi-Mikala (1990). Parmi les langues gabonaises, on peut ajouter celles parlées par les pygmées; elles ne sont pas très bien connues. Les langues gabonaises peuvent être divisées en dix groupes de parlers ou unités-langues. Une unité-langue est l'ensemble de différents parlers, tous mutuellement compréhensibles.

Bon nombre de recherches ont été faites sur les classifications en langues gabonaises; d'après les données dont nous disposons, on peut citer les travaux de Grimes (1996), Jacquot (1978), Raponda-Walker (1960) et Guthrie (1953). La classification la plus récente est celle de Kwenzi-Mikala (1987, 1998). Selon Kwenzi-Mikala, on relève dix groupes de parlers ou unités-langues à savoir les groupes mazuna, myene, mekana-menaa, mekona-mangote, membe (ou okandetsogo), merye, metye, membre, makena et baka. Ces groupes de langues sont établis sur le critère d'intercompréhension mutuelle et la formule de conversation «je dis que» (à l'exception du groupe baka).

Malgré l'article de la constitution qui reconnaît le français comme langue officielle, en même temps s'efforce de promouvoir et protéger les langues nationales, aucune langue gabonaise n'a le statut de langue officielle. Emejulu et Nzang-Bié (1999) ont utilisé la grille théorique de Fishman (1991) pour évaluer le niveau de vie et la dynamique des langues nationales. Il ressort de cette étude qu'au mieux toutes ces langues nationales se classent invariablement des stades 5 à 8 selon leur degré de valence sociale.

Cependant toutes ces langues gabonaises ne sont ni développées, ni standardisées, ni codifiées, ni utilisées dans aucune structure officielle sauf pour quelques unes (et ce d'une manière *ad hoc*) à la Radio Nationale ou Provinciale selon les besoins politiques. De récents travaux sur la vitalité des langues gabonaises montrent d'ailleurs des résultats très alarmants (Idiata 2003a). Parmi ces langues, certaines d'entre elles dont les seuls locuteurs sont quelques vieillards dans les villages, sont simplement menacées d'extinction (cf. Emejulu 2001: 39).

Quand le gouvernement a pris conscience de l'importance des langues nationales dans le développement du pays, il a commencé à les promouvoir. À cet effet, un Séminaire sur les États Généraux de l'Éducation et de la Formation s'est réuni du 12 au 13 décembre 1983. L'une des recommandations du séminaire était l'introduction des langues d'héritage culturel dans le système éducatif et national. Cependant, aucune mesure concrète ne découle de cette recommandation. Quatorze ans plus tard, cette recommandation fut encore introduite pendant la Table Ronde sur les Recherches Linguistiques et l'Enseignement des Langues au Gabon qui s'est tenue du 9 au 11 décembre 1997.

Recentement en avril 1997, le Séminaire sur la Standardisation de l'Orthographe des Langues Gabonaises était organisé par le Ministre de l'Éducation Nationale.

Les langues nationales sont utilisées comme médium d'instruction dans le programme littéraire pour adultes et à la télévision. Une section sur les langues nationales a vu le jour pendant l'année académique 1998–1999 à l'École Normale Supérieure (ENS) de Libreville. Malheureusement, ce projet n'aboutit toujours pas. Trois raisons expliquent cette situation:

- (a) Les gens pensent que la promotion et le développement des langues sont l'affaire des linguistes.
- (b) Les linguistes ne sont pas nombreux au Gabon; on y dénombre environ une quinzaine.
- (c) Le travail des linguistes n'est pas accessible au public puisqu'il est très technique. Ils se perdent aussi dans des descriptions phonologiques.

Pour cela, il nous faut faire une projection de l'avenir, c'est-à-dire compter sur le soutien de quelques lexicographes gabonais formés à l'Université de Stellenbosch (Afrique du Sud) pour que l'étude des langues gabonaises soit effective.

3. Linguistique, lexicographie et développement des langues gabonaises

La lexicographie et la linguistique contribuent au développement des langues gabonaises. La linguistique traite les classifications des langues et des descriptions linguistiques. Sur le domaine des classifications des langues gabonaises, nous pouvons mentionner les recherches de Kwenzi-Mikala (1990) qui a énuméré les langues gabonaises dans 62 parlers, catégorisés dans dix groupes de langues selon leur intelligibilité mutuelle.

Quant aux descriptions linguistiques, elles regroupent les domaines de la phonétique, la phonologie et la morphologie. Les descriptions du domaine de la syntaxe, de la sémantique, de la psycholinguistique et de la sociolinguistique sont rares et sont limitées à quelques langues (Idiata 2003: 37). Les descriptions linguistiques sont entreprises au Département des Sciences du Langage de l'Université Omar Bongo, Libreville, créé en 1994. L'étude des langues gabonaises demeure un vaste champ d'investigation à explorer.

Les contributions lexicographiques commencent vers les années 1800 par les missionnaires et les administrateurs coloniaux. Ces derniers, dans leur souci d'évangélisation et de communication avec les populations gabonaises, ont produit des ouvrages à caractère lexicographique. Bien que perfectibles, ces travaux répondent aux critères métalexigraphiques, notamment pour les méthodes utilisées pour la compilation des dictionnaires. La sémantique, la syntaxe et la pragmatique sont négligées dans ces dictionnaires. En outre, les dictionnaires compilés par les missionnaires et administrateurs coloniaux étaient des dictionnaires bilingues et étaient produits pour donner aux non-locuteurs natifs des langues gabonaises accès à la langue étrangère.

Par ailleurs, le Gabon a connu des développements importants dans le domaine lexicographique. Un séminaire sur la lexicographie bantoue tenu au Centre International de Civilisation Bantoue (CICIBA) a eu lieu en 1997 à Libreville par les membres venus de six pays africains (Cameroun, Centrafrique, Gabon, Guinée Equatoriale, Rwanda et São Tomé et Príncipe). Au cours de ce séminaire, le Dr. D.J. van Schalkwyk et le Dr. F.J. Lombard, éditeur et éditeur adjoint précédents du *Woordeboek van die Afrikaanse Taal* (WAT), ont partagé avec leurs homologues africains leur savoir sur la métalexigraphie et la pratique lexicographique en Afrique et dans le monde. Un accord de coopération liant le Groupe de Recherches en Langues et Cultures Orales (GRELACO) de l'Université Omar Bongo et le Bureau du *Woordeboek van die Afrikaanse Taal* (WAT) fut signé le 6 décembre 1999. Toujours en 1999, deux étudiants gabonais sont venus au Bureau du WAT pour une formation continue et pour leur étude doctorale en lexicographie à l'Université de Stellenbosch (République d'Afrique du Sud) sous la direction du professeur R.H. Gouws. En 2000, trois autres étudiants sont venus en Afrique du Sud et en 2001 ils ont été suivis par cinq autres. Le Bureau du WAT leur a donné une formation pilote en lexicographie générale, en lexicographie informatique et en planification des projets lexicographiques. Cinq étudiants ont soutenu leur thèse de docteur en lexicographie.

Trois ans plus tard s'est tenue à Libreville du 18 au 23 juillet 2004, la Neuvième Conférence Internationale de l'Association Africaine pour la Lexicographie (AFRILEX). Le sujet de cette conférence était la problématique du dictionnaire, le développement des langues et l'organisation des centres lexicographiques. Son objet était la recherche des zones de convergences entre le dictionnaire et la langue, afin de proposer une stratégie dynamique pour le développement de la lexicographie gabonaise, appelée à rendre accessibles les connaissances du monde contemporain pour la pluralité des langues et cultures du Gabon.

Il semble toutefois que le dictionnaire est très peu utilisé au Gabon. On consulte généralement un dictionnaire pour chercher le sens d'un mot ou pour vérifier son orthographe. Il ne fait pas partie des ouvrages courants à consulter, et est donc rangé dans la catégorie des ouvrages occasionnels. Les dictionnaires dans les langues gabonaises ne se trouvent pas dans les librairies, les rayons des bibliothèques générales, et dans les établissements scolaires et universitaires. Il est donc primordial qu'une culture du dictionnaire s'établisse au Gabon. On entend, par «culture du dictionnaire», accoutumer l'utilisateur à l'emploi du dictionnaire. Au Gabon, cette culture du dictionnaire peut s'établir au moyen du système éducatif. Il est important que la culture du dictionnaire s'installe chez les enseignants, qui à leur tour transmettront leur connaissance lexicographique aux élèves et étudiants. Il incombe aux lexicographes de véhiculer cette culture.

Les structures telles que l'École Normale des Instituteurs (ENI), l'École Normale Supérieure (ENS) et l'Université Omar Bongo (UOB) sont appropriées pour transmettre des connaissances lexicographiques aux éducateurs qui, à leur tour, les inculqueront aux élèves et étudiants.

4. Collaboration entre linguistes et lexicographes

Il importe qu'une collaboration entre linguistes et lexicographes s'établisse au Gabon. Cette collaboration peut être effective si l'on considère le rôle joué par le dictionnaire comme miroir de la société, le réservoir du savoir linguistique et culturel. Dans le but de préserver la riche variété des langues gabonaises, les lexicographes et linguistes devraient travailler ensemble. Les premiers, les lexicographes, devraient réviser ou améliorer les dictionnaires bilingues existants² en se penchant sur les types de dictionnaires tels les dictionnaires monolingues basés sur des descriptions linguistiques. Selon Tomba Moussavou (2002: 159-160), ces dictionnaires monolingues seront destinés aux usagers natifs; ils seront également destinés aux usagers qui apprennent une seconde langue. L'avènement de ces dictionnaires en langues gabonaises aura plusieurs incidences sur la standardisation, la terminologie, les recherches linguistiques ainsi que l'enseignement des dites langues. En outre, ces dictionnaires se proposeront, d'une part, d'améliorer et de compléter les carences et insuffisances aussi bien macro- que microstructurelles des dictionnaires qui ont été compilés par

les missionnaires, et d'autre part les domaines que ces dictionnaires bilingues n'ont pas exploités. Quant aux dictionnaires bilingues déjà existants, nous sommes tout à fait du même avis que Afane Otsaga (2001: 138, 157) optant pour une nouvelle approche et un modèle théorique de compilation des dictionnaires bilingues, en tenant compte du contexte spécifique dans lequel se trouvent les langues gabonaises. Selon l'auteur, cette nouvelle approche devrait partir de la collection des données, au contenu même des articles du dictionnaire bilingue: la traduction, l'équivalence, les synonymes, les exemples et la forme des entrées. Quelques aspects généraux sur le dictionnaire bilingue tels que la sélection des entrées et les relations sémantiques devraient aussi être revisités. Les deuxièmes, les linguistes, devraient orienter leurs travaux dans des descriptions sémantiques et syntaxiques, qui sont encore jusque-là inexplorées; il faut noter que les travaux de syntaxe sont assez rares et limités à quelques langues qui ont fait l'objet de recherches doctorales en morphosyntaxe, comme le isangu (groupe merye) (cf. Idiata 2003: 37).

Il existe aujourd'hui beaucoup de descriptions linguistiques sur les langues gabonaises — même si certains domaines précédemment cités restent inexplorés — depuis la création du Département des Sciences du Langage au sein de la Faculté des Lettres et Sciences Humaines (FLSH) de l'Université Omar Bongo. Ces descriptions ne sont pas accessibles au public, puisque très techniques et hermétiques, comme énoncé dans les lignes précédentes. Dans cette situation, une collaboration entre lexicographes et linguistes s'impose, pour que la connaissance de la langue soit accessible au public. En effet, les lexicographes pourraient se servir des travaux linguistiques et intégrer les aspects qui les intéressent dans les dictionnaires, sans avoir besoin eux-mêmes de refaire le travail d'analyse linguistique. En outre, le dictionnaire reste l'élément le plus approprié pour rendre effectif la connaissance de la langue. Il donne à la communauté la possibilité d'évaluer la langue: «donne moi ton dictionnaire et j'évaluerai ta langue». Le dictionnaire va développer l'essence même de la langue.

Les lexicographes et les linguistes doivent orienter leurs ressources et leurs efforts dans la compilation des corpora en langues gabonaises. Le corpus se définit comme un corps de textes écrits ou un discours transcrit qui peut servir de base à une analyse et description linguistique (Kennedy 1998: 1; Ekwa Ebanéga 2002: 188). L'objectif du corpus est de permettre aux lexicographes et linguistes de mener des recherches en tenant compte des réalités linguistiques issues du langage quotidien des usagers-cibles (Ekwa Ebanéga 2002: 190). Au Gabon, le corpus serait une importante source permettant aux chercheurs de travailler de façon permanente sur les langues gabonaises. Selon Emejulu (2001: 45), une rupture épistémologique franche s'impose pour que les études sur les langues gabonaises aient une utilité pratique et connaissent un essor que l'on attend d'elles. Pour atteindre ce but, les lexicographes et les linguistes gabonais doivent désormais orienter leurs efforts prioritairement vers le développement des corpora électroniques. Ainsi l'existence d'un tel rassemblement de textes en

langues gabonaises fournira également aux lexicographes et linguistes un accès à des données actuellement dispersées ou non disponibles, quelque soit leurs domaines scientifiques respectifs (lexicographie, linguistique, etc.). Nous sommes d'accord avec Emejulu (2001: 40) lorsqu'il affirme que la création et l'analyse des corpora stockés dans les banques de données informatisées ont non seulement révolutionné la linguistique, mais aussi contribué à un progrès considérable de la lexicographie. Selon lui, il est alors bien évident que si la linguistique gabonaise veut être légitime, au diapason de la science du nouveau millenium, il lui faut marquer une rupture épistémologique et donner la priorité au développement des corpora informatisées pour les langues au Gabon. Au niveau de l'analyse de données, nombreux sont les outils qui permettent d'analyser des corpus électroniques. Un logiciel récent qui est à la fois performant et convivial est le WordSmith Tools de Mike Scott (Oxford University Press). WordSmith effectue plusieurs types de recherches à savoir:

- (a) Listes de fréquence. On peut facilement obtenir des listes de fréquence de tous les mots dans le corpus. Les listes de fréquence se présentent de deux manières: par ordre alphabétique et par ordre de fréquence.
- (b) Listes comparatives. Cette fonctionnalité permet de comparer automatiquement la fréquence des mots dans des corpora différents et de faire ressortir les mots, soit beaucoup plus employés, soit beaucoup moins employés dans un des corpora.
- (c) Concordances. Cette fonctionnalité permet de visualiser toutes les occurrences d'un même mot en contexte et de manipuler ce même contexte.

Grâce à cette facilité d'accès aux données et aux nouvelles possibilités de traitement et de manipulation, ce corpus, mis sous forme électronique, permettra des approches diversifiées pour les recherches dans les langues gabonaises. Il sera question ici de légitimer la construction des corpora électroniques pour toutes les langues du Gabon et de présenter les outils statistiques et informatiques pour l'analyse (qualitative et quantitative) de corpus de telle sorte que les linguistes et lexicographes gabonais soient en position de commencer dans l'immédiat à tirer profit des corpora électroniques pour les rendements meilleurs (cf. Emejulu 2001: 40). Le but final d'un tel corpus est de donner une place aux langues gabonaises dans la recherche lexicographique et linguistique du 21ème siècle.

Dans le cadre des langues africaines en général et des langues gabonaises en particulier, Ekwa Ebanéga (2002: 190) a souligné l'exemple du projet ALLEX (African Language Lexicon), à présent ALRI (African Languages Research Institute). Commencé autour de 1990, le projet ALLEX est né d'un partenariat entre les chercheurs du Département des Langues et Littérature Africaine de l'Université du Zimbabwe et ceux des Départements de Lexicographie des Universités de Göteborg (Suède) et d'Oslo (Norvège), ainsi que les hommes

d'affaires suédois et norvégiens. Le projet ALLEX a essentiellement pour objectif la mise en place d'une base de données lexicographiques pour la production à terme des dictionnaires monolingues en mashona (Mavoungou 2001: 176). Les projets du type ALLEX pour les langues gabonaises auront pour résultat le développement de celles-ci. En termes plus concrets, un projet de l'envergure de l'ALLEX pour les langues gabonaises, aura des implications dans nombre de disciplines, notamment la lexicographie et la linguistique. D'un point de vue lexicographique et linguistique, la manipulation des vastes données est susceptible de permettre les opérations suivantes:

- (a) Une analyse linguistique relativement exhaustive de chaque mot du corpus,
- (b) L'identification des formes macrostructurelles (lemmes) de chaque mot, et
- (c) L'identification des exemples d'emploi pour chaque mot du corpus.

L'utilisation d'un corpus peut offrir au lexicographe et au linguiste tous les contextes d'apparition du mot à l'intérieur de la phrase, ce qui facilite énormément le travail du chercheur. Par exemple, le terme yilumbu *bwâla* (village), compte pour 1,10% (soit 394 occurrences) dans le corpus yilumbu (groupe merye); il est majoritairement précédé dans les phrases par le préfixe locatif de classe 17, *o* (0,67%, soit 237 occurrences dans le corpus). *Bwâla* s'emploie dans une moindre mesure avec des pronoms possessifs utilisés avec des noms. Les données sus-mentionnées peuvent intéresser à la fois le lexicographe et le linguiste. Pour le linguiste, l'occurrence quasi régulière du connectif *o* en combinaison avec *bwâla* peut permettre de prévoir la configuration syntaxique des phrases du yilumbu dans lesquelles *bwâla* apparaît. Pour le lexicographe, les données présentées ci-dessus donnent déjà de bonnes indications quant aux exemples appropriés pour illustrer le lemme *bwâla* dans le dictionnaire.

La collaboration entre les lexicographes et les linguistes s'inscrit dans le cadre du GRELACO. Ces activités qui s'étalent dans l'espace et dans le temps sont d'intérêt national puisqu'elles visent la promotion et la valorisation du patrimoine linguistique du Gabon. Elles sont réunies autour de cinq programmes majeurs, à savoir:

- (a) Programme d'Alphabétisation en Langue Maternelle (PALM),
- (b) Programme de Développement et Standardisation des Langues (PDSL),
- (c) Programme de Dictionnaire Pédagogique et Scientifique en Langues Gabonaises (PDPSL), support majeur et permanent des PALM et PDSL,
- (d) Programme d'Atlas Linguistique du Gabon (PALG), et
- (e) Programme d'Acquisition et Trouble du Langage (PATL).

Dans cette politique d'une collaboration entre lexicographes et linguistes, le GRELACO est devenu membre actif de l'Association Africaine pour la Lexicographie (AFRILEX); il a signifié son désir de collaborer d'une part avec l'Aarhus École des Affaires (Danemark), et d'autre part avec l'African Language Association for Southern Africa (ALASA).

5. Unités lexicographiques

La question de l'établissement des unités lexicographiques est un facteur déterminant pour la collaboration entre lexicographes et linguistes au Gabon. Une unité lexicographique est une structure de recherche dont l'objectif principal est la production pratique des dictionnaires. Dans le cadre de la politique linguistique des langues africaines, le projet de type PanSALB (Pan South African Language Board) pour les langues sudafricaines a été établi par le gouvernement de l'Afrique du Sud en 1995, afin de favoriser le multilinguisme et développer les langues qui étaient autrefois marginalisées. Grâce à ce projet l'unité lexicographique pour chacune des onze langues officielles de l'Afrique du Sud, à savoir: isizulu, isixhosa, afrikaans, sepedi, anglais, setswana, sesotho, xitsonga, siswati, tshivenda et isindebele a vu le jour. Pour le cas du Gabon, un projet de type PanSALB aurait pour résultat la préservation de la diversité linguistique et l'établissement des onze unités lexicographiques telles que les unités françaises et mazuna, myene, mekana-mena, mekana-mangote, membe, merye, metye, membre, makena, baka, et selon les dix groupes de langues préalablement cités. Celles-ci serviront d'interface pour la collaboration entre les lexicographes et les linguistes. En établissant les unités lexicographiques au Gabon, ceci aurait comme résultat la production des dictionnaires et la préservation de la diversité linguistique sous toutes ses formes variées.

La création d'une unité lexicographique est de rigueur compte tenu des besoins lexicographiques des langues gabonaises (Nyangone Assam et Mavoungou 2000: 252 et Nyangone Assam 2001: 188). Cette structure est en train de se mettre en place et sera fonctionnelle dans le cadre du GRELACO au sein de l'Université Omar Bongo (Emejulu et Nzang-Bié 1999). Le GRELACO œuvre donc dans le sens de la création des unités lexicographiques avec le concours des communautés linguistiques. Pour satisfaire les besoins en matière des dictionnaires et produits dérivés, supports indispensables pour le développement et la standardisation des langues, il est d'une nécessité absolue de créer des unités lexicographiques à travers l'ensemble du territoire gabonais, et ce en fonction des besoins spécifiques des langues et des communautés linguistiques. Selon Emejulu (2001: 50), il est plus pratique et économique d'établir les unités lexicographiques dans les localités où se trouvent les unités-langues répertoriées. Les unités lexicographiques seront installées auprès des Académies Provinciales dans d'autres provinces que la province du Haut-Ogooué dont l'unité lexicographique sera installée au sein de l'Université des Sciences et Techniques de Masuku à Franceville. L'unité lexicographique du français sera établie

à Libreville, la capitale du Gabon. Pour le cas de l'unité lexicographique du français, elle prendra en compte la variété du français du Gabon³ (Mavoungou 2002)

Toutes ces unités lexicographiques seront animées par des agents lexicographiques formés par le GRELACO. La première tâche des unités lexicographiques sera l'identification et la planification des projets lexicographiques à court, moyen et long termes. La deuxième et principale tâche est d'assumer la responsabilité du développement des corpora des différentes langues.

6. Conclusion

Dans ce travail, il a été question de l'importance d'une collaboration entre lexicographes et linguistes au Gabon. L'esprit qui a animé cette contribution était d'aller à l'essentiel et de présenter la situation linguistique du Gabon, de montrer que la lexicographie et la linguistique sont deux disciplines autonomes, avec des objets d'études bien spécifiques. La lexicographie a également une vocation interdisciplinaire, elle entretient des relations avec la linguistique et bien d'autres sciences.

Parmi les contributions de la linguistique et de la lexicographie pour le développement des langues gabonaises, il faut retenir que pour la lexicographie, la pratique des dictionnaires a bien commencé par les travaux des missionnaires et les administrateurs coloniaux; pour la linguistique, les contributions traitent de la classifications des langues et des descriptions linguistiques. La collaboration des lexicographes et des linguistes est nécessaire dans la mesure où la lexicographie permet de contenir le savoir linguistique et culturel des usagers. Les lexicographes et les linguistes doivent orienter leurs ressources et leurs efforts dans des domaines jusque-là peu explorés comme la syntaxe et la sémantique. Le corpus permet également de faire un pont entre ces deux disciplines séparées, mais qui sont amenées à interagir intimement dans l'analyse des données par l'entremise des outils performants comme le WordSmith. La création des unités lexicographiques est aussi un facteur déterminant pour la collaboration entre les lexicographes et les linguistes au Gabon. Les unités lexicographiques qui seront installées dans toutes les Académies Provinciales auront comme résultat la production des dictionnaires et la préservation de la diversité linguistique sous toutes ses formes variées. Le GRELACO s'inscrit dans le cadre d'une collaboration effective de ces deux disciplines scientifiques, compte tenu de ses activités réunies autour des programmes majeurs.

Notes

1. Le français est la seule langue dont le statut est clairement défini dans la constitution gabonaise dans l'article 2 du titre 1: "La République Gabonaise adopte le français comme langue officielle." Le français est de ce fait la langue des institutions. Le français est la langue de

l'école, la langue du savoir, la langue de l'administration, la langue des institutions juridiques et économiques et des media.

2. Selon Kavanagh (2002: 276), tous les dictionnaires sont pratiquement basés dans une certaine mesure sur un dictionnaire déjà existant.
3. Mavoungou (2002: 230) propose des différents critères métalexigraphiques pour la planification et la compilation du Dictionnaire du français du Gabon. L'auteur tient compte du fait que le français, langue officielle, est bien établi et considère le truisme selon lequel une langue ne peut exister qu'avec un seul usage. Le moment est venu de se demander si le Gabon ne doit pas assumer son propre usage du français. Il s'agit donc de poser les jalons susceptibles de nous conduire "vers un dictionnaire du français du Gabon".

Bibliographie

- Afane Otsaga, T. 2001. Les dictionnaires bilingues dans les langues gabonaises: Approche méthodologique. Emejulu, J.D. (Éd.). 2001: 137-159.
- Doke, C.M. et D.T. Cole. 1963. *Contributions to the History of Bantu Linguistics*. Johannesburg: Witwatersrand University Press.
- Ekwa Ebanéga, G.-M. 2002. Le corpus lexicographique dans le développement des langues gabonaises. Emejulu, J.D. (Éd.). 2002: 188-205.
- Emejulu, J.D. 2001. Lexicographie multilingue et multisectorielle au Gabon: planification, stratégie et enjeux. Emejulu, J.D. (Éd.). 2001: 38-57.
- Emejulu, J.D. (Éd.). 2001. *Éléments de lexicographie gabonaise. Tome I*. New York: Jimacs-Hillman.
- Emejulu, J.D. (Éd.). 2002. *Éléments de lexicographie gabonaise. Tome II*. New York: Jimacs-Hillman.
- Emejulu, J.D. et Y. Nzang-Bié. 1999. *Linguistic Perspectives in Gabon*. Communication présentée à un colloque organisé par le Summer Institute of Linguistics, Université de Dakota du Nord, Grand Forks, États-Unis, 20 juillet 1999.
- Fishman, J.A. 1991. *Reversing Language Shift: Theoretical and Empirical Foundations of Assistance to Threatened Languages*. Cleveland: Multilingual Matters.
- Grimes, B.F. 1996¹³. *Ethonologue*. Dallas: Summer Institute of Linguistics, Inc.
- Guthrie, M. 1953. *The Bantu Languages of Western Equatorial Africa*. Oxford: Oxford University Press.
- Idiata, D.F. 2003. *Pourquoi le Gabon doit investir sur ses langues vernaculaires*. Le Cap: Centre for Advanced Studies of African Society.
- Idiata, D.F. 2003a. Diversité ethnolinguistique et vitalité des langues au Gabon. Bearth, T. (Éd.). 2003. *African Languages in Global Society/Les langues africaines à l'heure de la mondialisation*. Cologne: Koppe Verlag.
- Jacquot, A. 1978. Le Gabon. Barreteau, D. (Éd.). 1978. *Inventaire des études linguistiques sur les pays d'Afrique noire d'expression française et sur le Madagascar*: 493-503. Paris: CILF.
- Kavanagh, K.P. 2002. Adapter un dictionnaire monolingue pour un usage local. Emejulu, J.D. (Éd.). 2002: 276-290.
- Kennedy, G. 1998. *An Introduction to Corpus Linguistics*. Londres: Longman.
- Kwenzi-Mikala, J.T. 1987. Contribution à l'inventaire des parlers bantu du Gabon. *Pholia. Revue du Laboratoire de Phonétique et Linguistique Africaine* 2: 103-110.
- Kwenzi-Mikala, J.T. 1990. Quel avenir pour les langues gabonaises? *Revue Gabonaise des Sciences de l'Homme* 2: 121-124.

- Kwenzi-Mikala, J.T.** 1998. Parlers du Gabon: classification du 11.12.97. Raponda-Walker, A. (Éd.). 1998. *Les langues du Gabon*: 217. Libreville: Éditions Raponda-Walker.
- Mavoungou, P.A.** 2001. La mondialisation et la lexicographie trilingue ou plurilingue au Gabon. Emejulu, J.D. (Éd.). 2001: 160-183.
- Mavoungou, P.A.** 2002. Vers un dictionnaire du français du Gabon. Emejulu, J.D. (Éd.). 2002: 230-262.
- Mba-Nkoghe, J.** 1991. Le Gabon linguistique. *Notre Librairie. Littérature gabonaise* 105: 20-23.
- 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. (Éd.). 2001: 7-37.
- Nyangone Assam, B.** 2001. La lexicographie dans l'apprentissage des langues gabonaises. Emejulu, J.D. (Éd.). 2001: 187-205.
- Nyangone Assam, B. et P.A. Mavoungou.** 2000. Lexicography in Gabon: A Survey. *Lexikos* 10: 252-274.
- Raponda-Walker, A.** 1960. *Notes d'histoire du Gabon*. Montpellier: Imprimerie Charité.
- Tomba Moussavou, F.** 2002. Quelques aspects de la lexicographie monolingue au Gabon. Emejulu, J.D. (Éd.). 2002: 159-173.

Thoughts and Views on the Compilation of Monolingual Dictionaries in South Africa*

N.C.P. Golele, *Xitsonga Language Research and Development Centre, Letaba,
Republic of South Africa (xitsolrdc@telkomsa.net)*

Abstract: Developing and documenting the eleven official languages of South Africa on all levels of communication in order to fulfil all the roles and uses characteristic of truly official languages is a great challenge. To meet this need various bodies such as the National Lexicography Units have been established by the Pan South African Language Board (PanSALB). As far as dictionary compilation is concerned, acquaintance with the state-of-the-art developments in the theory and practice of lexicography is necessary.

The focus of the African languages should be directed onto the compilation of monolingual dictionaries. It is important that these monolingual dictionaries should be usable right from the start on a continuous basis. Continued attention should be given to enlarging the corpora and actual consultation of these corpora on the macro- and microstructural levels. The end-products should be of a high lexicographic standard, well-balanced in terms of lemma selection, length of the articles, maximum utilisation of available dictionary space etc. They should also be planned and compiled in such a way that the transition from paper dictionaries to electronic dictionaries could be easily and naturally accomplished. Advanced and continued training in the compilation of monolingual dictionaries should be presented.

Keywords: MONOLINGUAL DICTIONARIES, OFFICIAL LANGUAGES, DICTIONARY COMPILATION, CORPORA, NATIONAL LEXICOGRAPHY UNITS, TARGET USERS, DICTIONARY USE, DICTIONARY CULTURE, CORE TERMS

Opsomming: Gedagtes en beskouings oor die samestelling van eentalige woordeboeke in Suid-Afrika. Die ontwikkeling en dokumentering van die elf amptelike tale van Suid-Afrika op alle vlakke van kommunikasie om alle rolle en gebruike van werklik amptelike tale te vervul, is 'n groot uitdaging. Om in hierdie behoefte te voorsien, is liggame soos die Nasionale Leksikografie-eenhede deur die Pan Suid-Afrikaanse Taalraad (PanSAT) tot stand gebring. Wat woordeboeksamestelling betref, is bekendheid met die heel nuutste ontwikkelinge in die teorie en praktyk van die leksikografie noodsaaklik.

Die fokus van die Afrikatale moet op die samestelling van eentalige woordeboeke gerig wees. Dit is belangrik dat hierdie eentalige woordeboeke reg van die begin af op 'n deurlopende basis bruikbaar moet wees. Voortdurende aandag moet aan die uitbreiding van die korpusse en die

* This article is a slightly adapted version of a keynote address presented at the Tenth International Conference of the African Association for Lexicography, organised by the Sesuisa Sesotho Lexicography Unit, University of the Free State, Bloemfontein, Republic of South Africa, 13–15 July 2005.

werklike raadpleging van hierdie korpusse op makro- en mikrostrukturele vlak gegee word. Die eindprodukte moet van 'n hoë leksikografiese gehalte wees, goed gebalanseerd betreffende lem-makeuse, lengte van artikels, maksimum gebruik van beskikbare woordeboekruimte, e.s.m. Hulle moet ook op so 'n manier beplan en saamgestel word dat die oorgang van papierwoordeboeke na elektroniese woordeboeke maklik en natuurlik bewerkstellig kan word. Gevorderde en voortduurende opleiding in die samestelling van eentalige woordeboeke moet aangebied word.

Slutelwoorde: EENTALIGE WOORDEBOEKE, AMPTELIKE TALE, WOORDEBOEK-SAMESTELLING, KORPUSSE, NASIONALE LEKSIKOGRAFIE-EENHEDE, TEIKENGEBRUIKERS, WOORDEBOEKGEBRUIK, WOORDEBOEKKULTUUR, KERNTERME

A challenging new era has dawned on dictionary compilation in South Africa. In the dispensation prior to democracy in 1994, Afrikaans and English were privileged as the only two official languages. African languages were "official" in their respective homelands. In the new democratic South Africa, the African languages too enjoy official status throughout the country. Developing and documenting eleven official languages on all levels of communication in order to fulfil all roles and uses characteristic of truly official languages is a great challenge. In the field of lexicography, the Pan South African Language Board (PanSALB) has been tasked with the responsibility of establishing lexicography units for the eleven official languages to assist the processes of language preservation and development.

The Woordeboek van die Afrikaanse Taal (WAT) on which work started in 1926, can serve as an example of the slow and arduous process often characteristic of the production of a monolingual dictionary. Two decades later, in 1946, work had been produced on only the first ten letters of the alphabet. The manuscript was never published for reasons among which were that Afrikaans had not been properly standardized when the work began. A general lack of terminology was another factor. Furthermore there was uncertainty whether Afrikaans should move away from or closer to Dutch. Because of this, some Dutch words were included which never became part of Afrikaans. It is to obviate such problems that the Pan South African Language Board has established National Language Bodies, National Lexicography Units and Provincial Language Committees to expedite solutions to problems in areas such as lexicography, terminology, translation and interpreting, grammar and literature, to mention but a few challenges to be met. The African Association for Lexicography (AFRILEX) has to be commended for its role in providing guidance in the production of quality dictionaries.

On matters of standardization, the National Language Bodies met to authenticate the terminology lists of the Department of Arts and Culture, produced in various fields. There was also a meeting at the University of Pretoria to attend to matters of standardization of the African languages of South Africa. These activities point to the extent to which language matters receive attention in South Africa, and how lexicographic work could benefit.

The National Lexicography Units are funded by Parliament through PanSALB with strict conditions attached, namely the production of publications which have actually to be exhibited to Parliament. This makes it clear that there is no time to waste as far as lexicography development is concerned. The units can therefore not afford the luxury of dialectics such as the prescriptive versus descriptive approach or issues of word versus stem traditions. There is also no time to leisurely and gradually develop and contrive sound lexicographic principles and practices. The units have to start producing immediately. This means that they must acquaint themselves with state-of-the-art knowledge in the theory and practice of lexicography. In the field of theoretical and practical lexicography, the combination of the contributions of foreign and local experts who show an interest in the development of lexicography in South Africa have to be embraced and optimally used.

The mandate of the National Lexicography Units is to produce comprehensive monolingual dictionaries and related products for the respective languages. Most units fruitfully utilized this opportunity to provide basic bridging dictionaries between the African languages and English. Some publishers are keen to assist with the publication of these dictionaries.

Needless to say, the value of bilingual dictionaries for mother-tongue speakers of African languages is not disregarded or disputed. If proper studies of users' requirements in the African languages reflect an urgent need for bilingual dictionaries, naturally such dictionaries should be compiled to satisfy these requirements. However, it is imperative that long-term preference be given to the compilation of comprehensive monolingual dictionaries for all the official languages, including Afrikaans and English. PanSALB, in terms of its vision and mission, should play an important part, and PanSALB and its structures, together with government departments and language associations and bodies, should collaborate to perform a decisive role in the compilation of such dictionaries. Even as bilingual dictionaries are produced, it is evident that they will need to be revised and updated from time to time in order to serve the lexicographic needs of the respective language communities. However, such revisions and updates should be secondary objectives not distracting attention from the main object, namely the compilation of monolingual dictionaries.

Monolingual dictionaries should be prepared in such a way that they are usable right from the start, and their production maintained. The situation of African languages makes it impossible to wait decades for the completion of a dictionary. Therefore, monolingual dictionaries must be compiled as complete units, i.e. dictionaries comprising lemmas for all the letters of the alphabet, namely from A to Z, continuously being increased in comprehensiveness and released as often as possible. Apart from serving the language communities, completed products also give the units a gratifying sense of accomplishment. Eventually such volumes could be extended to form a multi-volume dictionary.

Continued attention has to be given to enlarging the corpora and actual consultation or querying of these corpora on the macro- and microstructural levels. On the macrostructural level, specific guidance should be given to lexicographers in respect of the compilation of lemma lists for these dictionaries. On the microstructural level, lexicographers should be continuously assisted to maximally utilize concordance lines for enhancing the quality of the articles to be compiled. The end products should be of a high lexicographic standard, well-balanced in terms of lemma selection, length of articles, maximum utilization of available dictionary space and other factors. These products should be planned and compiled in such a way that transition from paper dictionaries to electronic dictionaries, whenever the right time for this arrives, will be easily and naturally accomplished. This statement is made bearing in mind the less privileged circumstances in which some potential users find themselves. As in other spheres of life, the ideal aimed at has to be balanced with the realities of the circumstances of the less privileged. Advanced and continued training in the compilation of monolingual dictionaries should be presented by means of centralized and/or decentralized initiatives.

Frequency of release of products could depend also on strategies employed to reach desired goals. Experienced language practitioners in the community can be tasked with the function of providing definitions for lemmas for the monolingual dictionary. The task of the unit can therefore be reduced to that of formatting of the text. This strategy has the further advantage that more potential lemmas are generated by members of the community through the definitions. This contributes much to dictionary production. As cited by Dr P.A. Louw, formerly of the WAT, words or terms used in definitions should themselves be incorporated into the dictionary. This procedure allows for the broadening of the scope of the contribution to the lexicographic process.

Until recently the area of lexicography has been sadly neglected. It could be argued that the missionaries contributed to lexicographic development through the compilation of a number of bilingual dictionaries for the African languages. Although the value of their contribution is not disputed, it is true that these dictionaries were not compiled with the mother-tongue speakers in mind, but rather for the benefit of the missionaries to understand the relevant languages for their evangelization mission. Defunct departments of education of the past dispensation also produced some kind of dictionaries for the African languages in support of policies in which the intended users had no share or contribution. What is now required for the African languages in South Africa is a discontinuation of such Eurocentric approaches, and the adoption of a system where the lexicographic needs of the mother-tongue speakers of African languages take centre-stage.

Many challenges lie on the way of dictionary production in South Africa, such as staff recruitment and continuous training, motivation and guidance, retention of lexicographic skills, infrastructure, financial constraints and other factors. As far as staff composition is concerned, consideration should perhaps

be given to an idea mooted by Dr M. Alberts of PanSALB at a workshop on terminology, that the lexicography units should incorporate terminologists in their staff structure to enhance their work.

At this stage of the development of the African languages, dictionary units should assist communities by placing a strong focus on the identification of the target users, their dictionary consultation skills, and their general needs. It should not be forgotten that modern dictionaries are compiled with a strong emphasis on the user perspective. In most cases, it amounts to the compilation of very practical, user-friendly dictionaries. In the end dictionaries are judged as good or bad by their target users.

Lemmas should be defined in such a way that users will find all the answers to the questions that cause them to consult the dictionary. Laufer (1992: 71) calls this objective *knowing* a word:

Knowing a word would ideally imply familiarities with all its properties ... When a person "knows" a word, he/she knows the following: the word's pronunciation, its spelling, its morphological components, if any, the words that are morphologically related to it, the word's syntactic behaviour in a sentence, the full range of the word's meaning, the appropriate situations for using the word, its collocational restrictions, its distribution and the relation between the word and other words within a lexical set ...

Lombard (1991: 166) identifies a number of defining criteria that would result in good definitions namely completeness, clarity, accuracy, consistency, independence, objectivity and neutrality.

Emphasis should also be on the situation of use — how and where the dictionaries are used in South Africa. The luxury of writing dictionaries only to be used in study rooms, libraries and well-equipped classrooms cannot be afforded. As already indicated, dictionary users should also be able to consult the dictionary even in less privileged circumstances.

The compilation of dictionaries in South Africa should in fact be viewed from two angles: improving the dictionaries and improving the users of such dictionaries. PanSALB, AFRILEX and the South African Broadcasting Corporation (SABC) should, for instance, seriously consider incorporating presentations or broadcasts in the education service teaching dictionary use to children, and for that matter, to adults. The National Lexicography Units should also consider making the teaching of dictionary use an item on the agenda of their annual open days. It serves no purpose if dictionaries of good quality are compiled by the National Lexicography Units and also by private individuals but target users are not trained to use them. Issues of text production and text reception are important here.

In the *Collins COBUILD English Language Dictionary*, it is stated that "users expect more and more from their dictionaries, and in particular want to gain confidence in using a word by looking it up in a dictionary". The latter part of this remark underlines the responsibility of the lexicographer to supply enough

productive (encoding) information to the user, and even more importantly, that this information should be on the level of the user.

Attention should also be given to the role of dictionaries in the creation of a dictionary culture. It could be a good idea to start with a small monolingual dictionary for primary school children and in doing so establish a dictionary tradition, then to compile follow-up dictionaries keeping pace with the development of this target group.

All core terms with which dictionary users as laypersons are confronted daily belong in a general dictionary. It helps to establish the authority of dictionaries as instruments of use, for example, by the inclusion of terms from the legal, medical and other professional fields.

In conclusion, aspects such as the following should be highlighted if the mandate of the National Lexicography Units in dictionary compilation is to be met: raising funds to supplement the state grant, commitment to and passion for the work, continued training of staff and use of available expertise, and collaboration with bodies supporting dictionary production.

References

- Laufer, B.** 1992. Corpus-based versus Lexicographer Examples in Comprehension and Production of New Words. Tommola, H. et al. (Eds.). 1992. *EURALEX '92 Proceedings I-II: Papers Submitted to the 5th EURALEX International Congress on Lexicography in Tampere, Finland*: 71-76. Tampere: University of Tampere.
- Lombard, F.J.** 1991. Die aard en aanbieding van die leksikografiese definisie. *Lexikos* 1: 158-182.
- Sinclair, J. (Ed.)**. 1988. *Collins COBUILD English Language Dictionary*. London: HarperCollins.

The Heart of the Matter: Methodological Challenges in Developing a Contemporary Reading Programme for Monolingual Lexicography, from the Perspective of the Dictionary Unit for South African English*

Leela Pienaar, *Dictionary Unit for South African English, Rhodes University, Grahamstown, Republic of South Africa (l.pienaar@ru.ac.za)*

Abstract: This article argues the importance of the reading programme as the pivotal issue in the lexicographic process. It is essentially a practical article which outlines strategies for developing and implementing a reading programme for monolingual lexicography. The arguments are informed by theory, together with an examination of the data-collection procedures followed by the Dictionary Unit for South African English (DSAE) and a survey of current practice in major English dictionary units around the world, namely the Oxford English Dictionary, the Canadian Oxford Dictionary, the Australian National Dictionary Centre and the New Zealand Dictionary Centre. The reading programme for the DSAE is first defined and contextualized within its mission statement. The article then explores the challenges inherent in sampling contemporary written and spoken English in the South African multilingual context. It is intended to inform the DSAE's intake policy, in terms of the following critical issues:

- the definition of South African English,
- the monitoring and selection of print, oral and electronic sources,
- the excerpting of citations and relevant bibliographic information, and
- the recruiting and training of readers.

These interlinked aspects of the reading programme have crucial implications for the quality and authority of the monolingual dictionary on historical principles.

Keywords: READING PROGRAMME, MONOLINGUAL DICTIONARY, HISTORICAL PRINCIPLES, CONTEMPORARY, INTAKE, CITATIONS, STRATEGIES

* This article was presented as a paper at the Tenth International Conference of the African Association for Lexicography, organized by the Sesiu sa Sesotho Lexicography Unit, University of the Free State, Bloemfontein, Republic of South Africa, 13–15 July 2005.

Opsomming: Die kern van die saak: Metodologiese uitdagings by die ontwikkeling van 'n eietydse leesprogram vir eentalige leksikografie, uit die perspektief van die Dictionary Unit for South African English. Hierdie artikel bepleit die belangrikheid van die leesprogram as die kernsaak in die leksikografiese proses. Dit is hoofsaaklik 'n praktiese artikel wat strategieë vir die ontwikkeling en implementering van 'n leesprogram vir eentalige leksikografie skets. Die argumente word ondersteun deur teorie, saam met 'n ondersoek na die dataversamelingsprosedures wat deur die Dictionary Unit for South African English (DSAE) gevolg word en 'n oorsig van die huidige praktyk by belangrike Engelse woordeboekeenhede oor die hele wêreld, naamlik die Oxford English Dictionary, die Canadian Oxford Dictionary, die Australian National Dictionary Centre en die New Zealand Dictionary Centre. Die leesprogram van die DSAE word eers omskryf en geplaas binne sy missiestelling. Die artikel ondersoek dan die uitdagings verbonde aan die toetsing van eietydse geskrewe en gesproke Engels binne die Suid-Afrikaanse veeltalige konteks. Dit is bedoel om die DSAE se opnamebeleid te belig betreffende die volgende kritieke sake:

- die omskrywing van Suid-Afrikaanse Engels,
- die kontrole en keuse van gedrukte, gesproke en elektroniese bronne,
- die ekserpering van sitate en tersaaklike bibliografiese inligting, en
- die werwing en opleiding van lesers.

Hierdie verbandhoudende aspekte van die leesprogram het beslissende implikasies vir die gehalte en gesaghebbendheid van die eentalige woordeboek volgens historiese beginsels.

Slutelwoorde: LEESPROGRAM, EENTALIGE WOORDEBOEK, HISTORIESE BEGINSELS, EIETYDS, OPNAME, SITATE, STRATEGIEË

Introduction

'Most users believe that the lexicographer simply sits down and "writes" a dictionary. This is far from true! The dictionary editors conduct a reading program, excerpting quotations (citations) from many written sources — from books, periodicals, newspapers, letters — and sometimes from spoken sources — tapes, television and radio.' (Kipfer 1984: 1). In terms of Zgusta's theoretical model (1971), the collection of material is the lexicographer's 'first task' in the dictionary compilation process. For a dictionary arranged according to historical principles, like the *Dictionary of South African English* (1996), the collection of material depends on a well-conceived and frequently-monitored reading programme. It is therefore not just the first in a series of tasks, but the pivotal issue in the lexicographic process, with consequences down the entire length of the process.

The early English 'hard-word' dictionaries were the products of a blend of intuition and a selection of material from other dictionaries. Both the intuitive aspect of the practice and the selection of material were highly subjective, usually driven by the lexicographer's personal agenda. Johnson, in *The Plan of a*

Dictionary 1747, stated that his aim was 'to preserve the purity and ascertain the meaning of the English idiom' and he expressed a preference for words found in the works of 'polite writers' (1970: 4). Furthermore, as Crystal (2004: 383) notes: 'The quotations were chosen more for their literary or moral value than for their linguistic clarity.' The use of secondary sources and a reliance on subjective intuition was the accepted methodology of the day.

Dictionary-making today is very different. Svendsen (1993: 40-53) describes the collection of evidence based solely on the lexicographer's linguistic intuition as 'hazardous', and cautions the lexicographer to guard against the influence of personal preferences and prejudices when selecting material on which to base a dictionary. Latter-day wisdom is that the selection of material for a dictionary, especially a monolingual dictionary on historical principles, ought to be from verifiable primary sources. It is by observing this practice that the lexicographer can claim that the created dictionary represents 'authentic' language use.

Definition

The reading programme involves the systematic collection of citations or quotations to enable the lexicographer to construct a picture of the language variety, which will be described in the dictionary. It is a process which is at the heart of all major historical dictionaries. For example, the Oxford English Dictionary (OED) reading programme of 1857 involved the collection of quotations excerpted from a variety of original written texts, ranging from personal letters to scientific treatises, going back as far as Anglo-Saxon literature. The current OED reading programme is continuing this tradition by ensuring that a wide range of modern authors and source types are constantly monitored. Surveying and sampling English in the 21st century for an extensive project such as the OED requires the lexicographer to acknowledge all varieties of World English and to include conventional written texts as well as commercial databases, electronic texts and the Internet.

In the production of the *Dictionary of South African English on Historical Principles* (DSAE Hist.), quotations were amassed over a period of 25 years from sources as diverse as diaries, letters, newspapers, books, periodicals, and even one-liners from labels on medicine bottles. In 1971, the Dictionary Committee of the Institute for the Study of English in Africa at Rhodes University, submitted 50 sample entries for the proposed 'Dictionary of South African English'. The submission was prefaced by the following salutary observation: 'A perfect Dictionary of South African English would be based upon a study of all English texts that might be regarded as South African. Since a study of this kind is possible only "in the dreams of a poet doomed at last to wake a lexicographer" (Johnson 1755), an actual Dictionary must be based upon a sampling of the available material, both written and spoken.' (The Dictionary Committee 1971: 4).

The challenge to the current Dictionary Unit for South African English (DSAE) in drawing up a Contemporary Reading Programme is similar in some respects to the one that confronted Robert Burchfield, the new editor of the OED in 1957. At that time, he established a Reading Programme to target modern-day language and extended the coverage to include more general, popular, and scientific sources (Winchester 2003: 244).

However, the DSAE faces the additional challenges of sampling contemporary written and spoken English in the multilingual South African context, where English is in contact with ten other indigenous languages. The reading programme has been compiled against this dynamic linguistic background and is structured as follows:

- the definition of South African English,
- the monitoring and selection of print, oral and electronic sources,
- the excerpting of citations and relevant bibliographic information, and
- the recruiting and training of readers.

What is South African English?

In the broadest sense, South African English (SAE) might be taken to include all the varieties of English spoken and written in the Republic of South Africa. It includes:

1. the English of people for whom it is a mother tongue or first language, and
2. the English of people for whom it is an additional language.

SAE is not one monolithic variety of English, but an interesting collection and mix of varieties, ranging from the standard form with no local base, to highly localized and specialized varieties. Some of the recognized varieties are:

- Black South African English (De Klerk and Gough 2002),
- South African Indian English (Mesthrie 1996),
- Cape Flats English (Malan 1996),
- Afrikaans English (Watermeyer 1996), and
- Xhosa English (De Klerk 2002).

There are possibly other varieties that may not yet have been formally acknowledged in research literature. They could include:

- regional varieties such as Eastern Cape English, West Coast English, KwaZulu-Natal English and so on, located in the different provinces or geographical areas of South Africa,

- slang associated with specific age groups such as children or adolescents, and
- specialized vocabularies relating to hobbies, trades or professions.

The DSAE's policy regarding the collection and documentation of the language should be inclusive and flexible in keeping with the complexities of the South African linguistic landscape. More specifically it should encompass the range of English from the most conservative usages to the broadest, from speakers and writers for whom it is the sole language of communication, to those for whom it is an additional language in multilingual South Africa. In doing so, it will be upholding the commitment expressed in its mission statement:

Considering the central role of English as a language of communication between South Africans, ... the Unit is committed to the recording of the history and culture of English in South Africa, and the documentation of its changing vocabulary.

Code-mixing and assimilated words

In defining the linguistic boundaries of English in a multilingual country like South Africa, it is realistic to note that code-mixing (*ad hoc* borrowing of lexical items between languages) is a normal occurrence. However, a distinction needs to be made between words used in code-mixing and those which have been assimilated and have stabilized in SAE. The following criteria may be used individually or together, to help the lexicographer discern the 'assimilated' items:

1. Assimilated words occur frequently in several contexts and have generally had a long duration in SAE, e.g. *donga* from Xhosa or Zulu *udonga*, earliest citation 1857 (DSAE Hist.), *naartjie* from Tamil *narattai*, earliest citation 1790 (DSAE Hist.), and *eland* from the Dutch word *eland* meaning 'elk', earliest citation 1786 (DSAE Hist.).
2. Assimilated items follow English grammar rules or orthography, e.g. *braaied* is from Afrikaans *braai*, but past tense formation *-ed* follows English rules, and *winelands* is from Afrikaans *wynlande*, but plural formation follows common English rule of adding *-s*.
3. Assimilated items are not glossed, placed within inverted commas or underlined, e.g. *lobola*, *boerewors*, and *gogga*.
4. Assimilated items undergo a change or extension of meaning between the parent language and English, e.g. *pondok* is from Malay *pondok*: a hut or shed; school and lodgings for students of religion. In modern Malaysia, the *pondok* is a rural, resident, Islamic educational institution conducted in reed huts. In SAE the *pondok* or *pondokkie* refers almost

exclusively to a shack or crudely constructed shanty made of pieces of wood, cardboard, corrugated iron and scrap material. So in South Africa, the original Malay word *pondok* has lost its association with the education of students of religion. In the South African socio-political context, the primary sense of *pondok* encodes details about the fragile materials and the appearance of the structure, together with an associated notion of the economic situation of those who live in pondoks.

Having defined the parameters of South African English and established some guidelines for identifying relevant lexical items, the next task of the editors is to fix the period that the reading programme will cover. The material collected for the DSAE Hist. came from a vast collection of sources dating back to 1589. The present reading programme is targeting contemporary sources after 1996 (when the DSAE Hist. was published) up to the present.

Sources

The field of primary sources for the DSAE contemporary reading programme is divided, according to the medium, into print, oral and electronic sources. In restricting the use of sources to primary material, the editors are aware of the need to expand their holdings of SAE to include the following categories of items:

- words which have only appeared recently,
- new senses of existing words,
- slang and colloquial words and expressions,
- variant spelling forms, and
- a range of varieties of SAE.

Print sources

Newspapers

James Murray acknowledged the importance of newspapers by commenting as follows in the Murray papers of 1882: 'To the philologist & historian of language — newspaper quotations are the most valuable of current instances — they show how the language grows ...' (Muggleston 2000: 14). Newspapers then are an essential part of any contemporary reading programme. They capture and disseminate trends in the language, whether these relate to new coinages such as *vuvuzela*, new senses of existing words such as *laager mentality*, or *rainbow nation* or new spellings of words which have already been recorded such as *takkies* (formerly *tackies*) and *Joburg* (sometimes *Jo'burg*). Crystal (2004:

519) observes: 'On 4 October 1957, the first sputnik was launched; on 5 October, the word *sputnik* was known everywhere, thanks to broadcasting and the press.'

In the South African context, Branford (1976: ix) urged that 'the Dictionary is not intended merely as an historical monument but as a record also of contemporary South African English at its most important points of growth'. That is as true today as ever it was when the idea of collecting evidence towards the DSAE Hist. was first mooted. It is a sentiment that is echoed by lexicographers of major English dictionary centres on both sides of the Atlantic: Nic Shearing (2005), the Reading Programme Manager of the OED: 'If I had to suggest just one thing to start with, it would be reading newspapers.'¹; and Katherine Barber (2005), editor-in-chief of the Canadian National Oxford Dictionary: 'We read the national newspapers of course ... but we also make an effort to acquire small local newspapers from across the country and read them.'² It follows then, that newspapers must be a crucial part of any reading programme that is seeking to collect evidence of contemporary language usage.

The OED consults the *Willings Press Guide* for listings of printed media sources. The South African equivalent is the *Media List*. As it is not possible to read every newspaper, it is necessary to be eclectic, but objective, in order to ensure that coverage of the language is wide and reliable. By using criteria such as target readership, circulation figures and geographical distribution, it is possible to extract a list of newspapers that would contribute to a balanced intake programme.

In the South African context, this means not just reading newspapers from all nine provinces, but also ensuring that these include a range of journalistic styles, from the more formal and conservative types, to the tabloids. The latter are known to be particularly innovative and often feature a mixture of articles in both informal and formal language. It is interesting that the tabloid, the *Daily Voice*, quickly acquired circulation and readership figures to rival that of the more established *Sowetan*. The figures are important as they are linked to the newspaper's potential to influence not only the reception, but also the production of the language.

Until its recent demise, the national daily English newspaper was *This Day*. No other daily national newspaper has taken its place, although several larger regional newspapers, such as Gauteng's *Sowetan* and *The Star* and Kwa-Zulu-Natal's *Daily News*, are distributed in major cities throughout the country. Fortunately the texts of two of the largest newspaper groups in the country are available electronically. They are the Independent Online (IOL), and Media 24. IOL includes the reportage of the *Cape Argus*, the *Cape Times*, the *Daily News*, the *Mercury*, the *Post*, the *Pretoria News*, the *Star*, the *Independent on Saturday*, and the *Sunday Tribune*, and Media 24 includes the *Witness*, the *Daily Sun*, the *City Press* and the *Sunday Sun*.

In addition to the large national newspapers, small local papers, e.g. the *Barrydale Times*, the *Stanger Herald*, and free community newspapers ('knock 'n drops') should not be overlooked. Their coverage of news relating to local affairs and culture often includes rare or regionally-based lexical items.

Magazines and Periodicals

The initial vocabulary for the *Dictionary of South Africanisms* was based on a survey of 34 fields of vocabulary ranging from 'human types' through 'food, dishes and cooking' to 'church and state' (Branford 1976: ix). The current reading programme of the OED has a daunting list of 350 labels to categorize incoming material. For the DSAE, the list of magazines from the *Media List* could be divided into broad interest areas such as Youth and Popular Culture, Women's Magazines, Sport, Leisure and Lifestyle, Men's Interest and Business and Current Affairs. Depending on the constraints of the budget and human resources, one or more magazines from each broad category could be selected for reading and intake. To ensure that the range of magazines within a category is monitored, titles could be rotated to keep the selection balanced.

Purchasing newspapers and magazines is a costly business, but it is possible to obtain free copies from libraries, student unions, regular subscribers and from the publishers themselves. Publishers may be willing to donate recent back issues if they are told that their magazine or newspaper is being consulted as a source for *bona fide* linguistic research and if they are assured that they will be properly acknowledged if any citations are used. They have been found to be particularly eager to cooperate when they are informed of the important part they play in contributing to the documentation of the language history of the country.

Printed books

One advantage of compiling a book list in post-1994 South Africa is that the researcher is not restricted by the censorship laws that existed under the apartheid regime. In the introduction of *Voorloper*, the text which preceded the DSAE Hist., Branford (1976: ix) observed that 'a very substantial body of South African writing' could not be represented in their collection as either the books themselves, or their authors were banned at the time and could therefore not be quoted. In 2005 the quantity and availability of material is both a blessing and a challenge to the lexicographer. Consequently some guidelines are necessary to enable the lexicographer to make a judicious selection. The OED consults publications such as *The Bookseller* and the *Times Literary Supplement* to monitor the appearance of new books on the literary scene. The South African equivalent of those references would be a list compiled from titles supplied by the National English Literary Museum (NELM) and the sources used to produce the S.A. Book Data Award. The titles could then be divided into initial broad categories such as adult fiction, non-fiction, and juvenile fiction. The fiction section might be broadly sub-divided into prose and drama, and the non-fiction might be sub-divided into biographies, letters, essays, criticism and travel, for example.

When one is reading for vocabulary, the literary merit of the books is not at issue: popular best-selling fiction is as worthy of consideration as scholarly

non-fiction. In fact, as with tabloids which enjoy a large readership, so-called 'pulp fiction' is important because of its potential to impact on vocabulary at large.

With regard to poetry, the guideline is to handle it cautiously and with due circumspection. James Murray is reported to have commented in this regard that 'we cannot take the language of poets too seriously' (Mugglestone 2000: 16). From a lexicographic perspective, examples of poetic usage by established literary figures are occasionally included as 'nonce' forms in a dictionary. While they are interesting examples of linguistic creativity, they do not represent common usage. It is true that some of these coinages may acquire currency and may be assimilated into the language. An example of this is Lewis Carroll's blending of *chuckle* and *snort* to produce the portmanteau word *chortle* in the poem 'Jabberwocky' from *Through the Looking Glass* (1872). As for *jabberwocky* itself, the lexicographers originally excluded the word from the first edition of the OED, but they were proved wrong when it was assimilated into general usage, and during a subsequent revision of the OED, it had to be included.

Oral sources

In a world where print sources are readily available, the collection and transcription of examples of oral evidence is a challenge to lexicographers. Many of the oral sources for Branford's *Dictionary of South Africanisms* were contributed by members of the public in response to a competition organized by the Dictionary Unit through the popular *Personality* magazine in 1970. Others, marked simply 'Informant' were what the dictionary team collected through their social networks.

Today, it is easier to access reliable oral sources where electronic corpora of the spoken language exist. Some corpora already available for SAE are:

- the SAE component of the International Corpus of English collected at the Nelson Mandela Metropolitan University,
- a corpus of spoken Xhosa English collected by Prof. V.A. de Klerk at the Department of Linguistics at Rhodes University,
- a corpus of Tswana English being collected by Prof. A.J. van Rooy at the North-West University, and
- corpora which might be part of university research projects.

Another source of spoken language data might be transcripts obtained from radio and television stations. Transcripts of phone-in programmes are particularly valuable in providing unrehearsed, unscripted spoken language, but speakers would need to be identified and traced as there are ethical problems surrounding the use of speech samples without prior permission of the speaker. This is a question that would need to be addressed well ahead of the stage of dictionary compilation.

It is occasionally argued that drama scripts can provide examples of the spoken component of the language. While they offer verifiable print evidence of a word's occurrence in a dramatized speech situation, they cannot be regarded as authentic speech as each character's words have been filtered through one writer's imaginative interpretation.

Electronic sources

Despite the burgeoning wealth of material available on the Internet, the prevailing wisdom (from the OED, for example) is that wherever possible, print sources are preferable, as they allow subsequent researchers to verify findings. However, in the absence of such sources, material should be 'judiciously selected from the open corpus of the web' (John Simpson OED 2005). The general rule in this regard is that the web sources consulted should be stable and durable.

Recruiting and training of readers

The readers are trained paid assistants or volunteers, who read texts and excerpt quotations to assist the lexicographers at various stages of the lexicographic process: sense and part of speech disambiguation, discerning orthography, finding collocational patterns, and selecting citations or example sentences. Most dictionary units, like the DSAE which has a staff of only four editors, cannot rely solely on staff to do the reading and excerpting. It is therefore necessary and in fact essential, in terms of ensuring a broad coverage of the language, to enlist the help of readers.

Quite early Branford (1976: xi) observed that 'the collection of materials for the Oxford English Dictionary Supplement is depending increasingly upon the work of paid readers, and it is clear that the preoccupations and commitments of English-speaking South Africa are such that substantial provision for this kind of help will have to be made in future budgeting for DSAE'. In the 25 years leading up to the publication of the DSAE Hist., the Dictionary Unit conducted a thorough historical reading programme under the editors-in-chief, William Branford and later, Penny Silva. Budget constraints at the time necessitated the use of unpaid volunteer readers but Branford's advice about paid readers is confirmed by the experiences of staff at other dictionary centres. The current editors at the National Dictionary Centres in Australia and Canada have found that using unpaid volunteers is not cost-effective. At both dictionary centres, they agree that the reading programme should not be viewed as a volunteer activity but as an essential part of serious lexicography that has to be budgeted for. Readers also need to be carefully selected and then trained in how to recognize and excerpt relevant material for the dictionary database. Murray's experience with recruiting readers was that the most consistently productive people were not necessarily scholars, but interested lay people. He found that scholars tended to display an interest in unusual or abstruse words

and would overlook familiar words in their search for the extraordinary. For example, Murray reports that he found 50 instances of the word *abusion* and not even five of the more common *abuse* (Winchester 2003: 105).

Excerpting of citations and relevant bibliographic information

The volunteer readers from Britain, North America and the British colonies who responded to Murray's famous 'Appeal to the English-Speaking Public' in 1879, received guidelines about what to excerpt, as well as slips on which to send the quotations they had taken from books. This practice is upheld today. The Australian, New Zealand and Canadian Dictionary Centres all issue their readers with guidelines of what to excerpt and how to excerpt. The overarching rule at the stage of reading and marking is that 'the main objective is the collection of data, not evaluation or editing' (Svensen 1993: 55). Landau (2001: 195-200) offers detailed guidelines about citation gathering. In the case of SAE, the choices can be customized and summarized to include the following categories of words:

- Any word or expression that seem to be particularly South African, e.g. *tik*, *eish*, *papsak*, or *vuuzela*;
- Existing SAE words spelt in unusual ways, e.g. *tekkies* (takkiies), *springbuck* (springbok) or *rooibosch tea* (rooibos tea);
- Existing SAE words that have acquired a new or extended meaning, e.g. *laager mentality* (from *laager*);
- Any word or expression which, although current in World English, is believed to have originated in SAE, e.g. *puff adder*, *apartheid* or *struggle accounting*;
- Any word or expression with a particular SAE sense not found in other varieties of World English, e.g. *township*, *dam*, or *bunny aerial*; and
- Words which, though current in SAE, have fallen out of use in British English, e.g. *bioscope*.

In guiding readers how to excerpt, the following points are relevant:

- Choose the best quotations from a text — the ones which illustrate what the word means, or which capture the word in an unusual context.
- Include sufficient contextual information to enable the lexicographer to draft an entry. More will often be necessary, but less may sometimes suffice. As a general rule, it is better to give too much rather than too little. The context should normally be at least the full sentence in which the highlighted word or phrase occurs. If any words need to be omitted, indicate these by the use of ellipsis '...'.

- Do not take the same word from the same source and with the same meaning more than a specified number of times.
- Excerpts should not be too long or too short but should contain sufficient information to illustrate usage and enable disambiguation of senses.
- Do not use cite words which use 'eye dialect' (sound spellings) such as *wek* for *work* or *situhation* for *situation*.
- Do not cite proper nouns unless they are used generically ('Bhisho has issued a notice to all civil servants...' where *Bhisho* represents the Eastern Cape provincial government department) or are part of a multi-word unit (*Madiba shirt*, where the whole unit represents a new lexical item).
- Do not cite typographical errors as these are not examples of intended usage.
- Avoid figurative language, e.g. metaphors, similes and the kind of creative idiomatic usage found in advertising in particular.

More detailed rules regarding the specific types of words and senses that should be selected, can be formulated according to need. Finally, it is important to use traceable sources, so the excerpters should document everything they can about the source: author, title, page, date, publisher and place of publication.

Conclusion

The electronic age has brought with it both solutions and challenges to lexicographers. An obvious solution is that it is easier and cheaper to update dictionaries if they are in electronic form. In the case of historical dictionaries, ante-dates, inter-dates and post-dates can be easily uploaded, quotations can be corrected during the reverification process and whole entries can be reworked if necessary. But the electronic age has also deluged lexicographers with sources, some more ephemeral than others. One of the challenges to lexicographers is to make judicious choices and to select material that truly represents the language variety being defined in the dictionary. The other challenge is based in Sinclair's notion of the 'empty lexicon'. It is an approach that regards usage as a moving target, and 'learns about vocabulary from the texts, and is constantly being updated. No part of it is absolute or permanent, because the boundary between item and environment is likely to move with new evidence, and meanings may merge or diverge. This is the lexicon of the living language.' (Sinclair 2004: 162).

Endnotes

1. Personal email to the author dated 21 February 2005.
2. Personal email to the author dated 24 February 2005.

References

- Branford, W.** 1976. *Voorloper: An Interim Presentation of Material for a Dictionary of South African English on Historical Principles*. Unpublished report of the Institute for the Study of English in Africa. Grahamstown: Rhodes University.
- Burchfield, R.W.** 1989. *Unlocking the English Language*. London: Faber and Faber.
- Crystal, D.** 2004. *The Stories of English*. London: Allen Lane.
- De Klerk, V.** 2002. Towards a Corpus of Black South African English. *Southern African Linguistics and Applied Language Studies* 20: 25-35.
- De Klerk, V. and D. Gough.** 2002. *Black South African English*. Mesthrie, R. (Ed.). 2002. *Language in South Africa*. Cambridge: Cambridge University Press.
- Johnson, S.** 1970. *The Plan of a Dictionary, 1747*. Facsimile edition. Menston: The Scolar Press.
- Kipfer, B.A.** 1984. *Workbook on Lexicography: A Course for Dictionary Users with a Glossary of English Lexicographical Terms*. Exeter: University of Exeter.
- Landau, S.I.** 2001². *Dictionaries: The Art and Craft of Lexicography*. Cambridge: Cambridge University Press.
- Malan, K.** 1996. Cape Flats English. De Klerk, V. (Ed.). 1996. *Focus on South Africa*: 125-148. Amsterdam/Philadelphia: John Benjamins.
- Mesthrie, R.** 1996. Language Contact, Transmission, Shift: South African Indian English. De Klerk, V. (Ed.). 1996. *Focus on South Africa*: 79-98. Amsterdam/Philadelphia: John Benjamins.
- Mugglestone, L. (Ed.).** 2000. *Lexicography and the OED: Pioneers in the Untrodden Forest*. Oxford: Oxford University Press.
- Silva, P. (Mng. Ed.).** 1996. *A Dictionary of South African English on Historical Principles*. Oxford: Oxford University Press.
- Sinclair, J.** *Trust the Text: Language, Corpus and Discourse*. 2004. London: Routledge.
- Svensén, B.** 1993. *Practical Lexicography: Principles and Methods of Dictionary-Making*. Oxford: Oxford University Press.
- The Dictionary Committee.** 1971. *Towards a Dictionary of South African English on Historical Principles, Fifty Draft Entries*. Unpublished report of the Dictionary Committee, Institute for the Study of English in Africa. Grahamstown: Rhodes University.
- Watermeyer, S.** 1996. Afrikaans English. De Klerk, V. (Ed.). 1996. *Focus on South Africa*: 99-124. Amsterdam/Philadelphia: John Benjamins.
- Winchester, S.** 2003. *The Meaning of Everything: The Story of the Oxford English Dictionary*. Oxford: Oxford University Press.
- Zgusta, L.** 1971. *Manual of Lexicography*. Prague: Academia / The Hague/Paris: Mouton.

De la problématique des articles synopsis dans la compilation des dictionnaires au Gabon*

Gilles Saphou-Bivigat (*13788647@sun.ac.za / saphoubivigat@yahoo.fr*), et
Edgard Maillard Ella (*13751727@sun.ac.za / ellamaillard@yahoo.fr*),
*Département d'afrikaans et hollandais, Université de Stellenbosch,
Stellenbosch, République d'Afrique du Sud*

Abstract: Le caractère multilingue et multiculturel du Gabon résulte de l'environnement linguistique et sociolinguistique, dont il faut tenir compte dans la compilation des dictionnaires. Par rapport aux articles simples les articles synopsis offrent une meilleure approche dans le transfert des données linguistiques et culturelles dans les dictionnaires. Autrement dit, au Gabon l'attention doit porter sur l'utilisation des articles synopsis dans la compilation des dictionnaires, en vue de refléter dans ses ouvrages la diversité linguistique et culturelle de ce pays. Ceci est d'autant plus vrai que l'utilisation des articles synopsis ne dépend aucunement de la typologie du dictionnaire.

Cet article a pour but de suggérer les différentes approches méthodologiques à partir desquelles les lexicographes et compilateurs des dictionnaires peuvent utiliser les articles synopsis pour présenter les données linguistiques et sociolinguistiques. Ces approches méthodologiques prennent en compte l'utilisation de ces articles dans différents types de dictionnaires. Ces développements seront discutés et évalués en fonction de leurs capacités à traiter l'aspect encyclopédique des items lexicaux. L'accent ne sera pas seulement porté sur le traitement des articles synopsis dans la nomenclature, mais aussi sur les post-textes qui peuvent aussi contenir un texte avec une liste de tous les items qui sont les unités de traitement de ces articles dans cette partie centrale du dictionnaire.

En d'autres termes, le point central de cet article est l'utilisation des articles synopsis dans la nomenclature, complétée par les registres alphabétiques contenus dans les post-textes en vue de la compilation des différents types de dictionnaires polyaccessibles en fournissant aux usagers des moyens d'accès rapide pour atteindre les données aussi bien linguistiques que culturelles.

Mots-clés: ARTICLE SYNOPSIS, ITEMS LEXICAUX, ARTICLE SIMPLE, SIGNE LEMME, ASPECTS LINGUISTIQUES, TYPES DE MACROSTRUCTURES, ASPECTS LINGUISTIQUES ET EXTRALINGUISTIQUES, ITEMS LEXICAUX CENTRÉS SUR LA CULTURE

Abstract: The Problematics of Synopsis Articles in the Compilation of Dictionaries in Gabon. Gabon has a multilingual and multicultural character resulting in a

* Cet article a été présenté comme une communication à la Neuvième Conférence Internationale Annuelle de l'Association Africaine pour la Lexicographie (AFRILEX), organisée par le Groupe de Recherche en Langues et Cultures Orales (GRELACO), Université Omar Bongo, Libreville, Gabon, 18-23 juillet 2004.

linguistic and sociolinguistic environment that must be taken into account in the compilation of dictionaries. Compared to simple articles, synopsis articles offer a better approach to the transfer of linguistic and cultural information in dictionaries. In other words, in Gabon, the focus should be on the use of synopsis articles in dictionaries that must be compiled to take into account and convey in these works the linguistic and cultural diversity of this country. This is very important because the utilisation of synopsis articles does not depend on dictionary typology.

The aim of this article is to suggest different methodological approaches for which lexicographers and compilers of dictionaries can use synopsis articles to present linguistic and cultural data. These methodological approaches take into account the use of these articles in different types of dictionaries. These developments will be discussed and evaluated for their ability to present and treat the encyclopedic aspect of lexical items. The emphasis will not only be on the treatment of synopsis articles in the central list, but also on back matter texts that can also contain a text with a list of all the items which are the treatment units of these articles in this central part of the dictionary.

In other words, the focal point of this article is directed at the level of the utilisation of synopsis articles in the central list which will be completed by alphabetical registers in the back matter texts for the compilation of different types of dictionaries which will be polyaccessible in offering to users rapid access means to reach both linguistic and cultural data.

Keywords: SYNOPSIS ARTICLE, TYPES OF MACROSTRUCTURES, SINGLE ARTICLE, LEMMA SIGN, LINGUISTIC AND EXTRALINGUISTIC ASPECTS, TARGET GROUP, DATA CATEGORIES, CULTURE-BOUND LEXICAL ITEMS

1. Introduction

L'un des débats actuels dans les discussions théoriques sur les problèmes lexicographiques met l'accent sur la diversité culturelle et linguistique. Ce débat conduit inéluctablement à l'utilisation des articles synopsis qui garantissent le transfert des données aussi bien linguistiques que culturelles. Dans la recherche métalxicographique actuelle, on distingue les articles simples et les articles synopsis. Les articles simples sont ceux qui présentent le traitement type assigné au signe lemme moyen, c'est-à-dire un traitement linguistique. Quant aux articles synopsis, ils offrent au-delà du traitement linguistique un surcroît de catégorie des données d'une nature encyclopédique permettant d'améliorer le transfert des données.

Par ailleurs, chaque dictionnaire doit être compilé selon les spécificités culturelles et linguistiques de la communauté à laquelle il est destiné. Dans une société multilingue et multiculturelle comme le Gabon, les dictionnaires jouent un rôle social non négligeable dans l'échange et le transfert des connaissances. Comme référence, on citera Wiegand (1997) et Gouws (1999) qui sont favorables à une attention particulière à l'usage des articles synopsis dans les dictionnaires pour une société multilingue et multiculturelle, où ces ouvrages de référence ont en filigrane un rôle de bâtisseur de la nation et de responsabilité sociale. Dans cette lignée d'idées Mavoungou (2002) a déjà mis en relief l'im-

portance de ce type d'articles dans la compilation des dictionnaires incluant les langues gabonaises qui sont des véhicules de cultures très diversifiées.

Ces différentes perspectives attribuées au rôle et au *but véritable* du dictionnaire auront définitivement des implications dans la compilation des dictionnaires au Gabon en incluant des défis intéressants qui font face non seulement aux lexicographes mais aussi aux linguistes, dans cet environnement multilingue et multiculturel que représente le Gabon.

2. Orientation théorique et méthodologique de l'utilisation des articles synopsis

2.1 Caractéristiques des usagers des dictionnaires gabonais

Chaque dictionnaire doit être compilé selon les besoins spécifiques d'un groupe d'usagers bien déterminé. Par conséquent, la compilation de n'importe quel dictionnaire exige auparavant l'identification des usagers indiqués, leurs besoins et habiletés lexicographiques. Par conséquent, la sélection de la macrostructure ainsi que le traitement microstructural des lemmes comme la structure d'adressage doivent être déterminés par le profil de l'utilisateur cible. Compte tenu du rôle et du *but véritable* du dictionnaire dans un cadre de vulgarisation et d'unification des connaissances et d'un environnement multiculturel et multilingue, les besoins spécifiques ainsi que les demandes et habiletés des usagers de provenances diverses doivent être pris en compte dans la planification et la compilation des dictionnaires au Gabon.

De même, les usagers cibles des dictionnaires au Gabon représentent une communauté linguistique et culturelle hétérogène. À part des locuteurs des langues gabonaises il faut aussi prendre en compte les autres communautés étrangères qui voudraient avoir accès à ces langues gabonaises. Les lexicographes ne peuvent pas satisfaire tous ces usagers sur une base égale. Leurs décisions par rapport à la planification et la structure des dictionnaires au Gabon doivent être guidées par le profil sociolinguistique du prototype de l'utilisateur cible. Les lexicographes des dictionnaires gabonais doivent tenir compte des usagers cibles dans leurs divergences et convergences culturelles aussi bien que des points de vue idéologiques. Le dictionnaire ne devrait pas prendre une position impartiale. La sélection macrostructurale aussi bien que le traitement des lemmes doivent faire preuve d'une attitude neutre quant à la culture et quant aux idéologies de certains items lexicaux. L'influence de la culture sur le sens d'un mot peut se refléter sur plusieurs items lexicaux comme *père* par exemple:

Les sens des mots *père* et *mère* varieront dans les langues gabonaises selon que l'on est dans une communauté à tradition matrilinéaire ou dans une communauté à tradition patrilinéaire. C'est ainsi que dans une communauté à tradition matrilinéaire telle que les Vili dans la Nyanga la

mère, nommée par l'unique item lexical «*mame*» [mâ:mə] fera référence à la génitrice mais aussi aux oncles maternels qui jouent le même rôle et ont la même autorité que la mère génitrice. Or dans une communauté à tradition patrilinéaire comme celle des Fang, le père nommé par l'unique item lexical «*tara*» [tárà], fera référence au géniteur, mais il se référera aussi aux oncles paternels qui jouent le même rôle et ont la même autorité que le père géniteur.

2.2 Les articles synopsis dans la macrostructure des dictionnaires gabonais

La compilation des dictionnaires gabonais rendra les langues gabonaises plus accessibles parmi les différentes communautés et jouera un rôle majeur dans la promotion des langues gabonaises parmi les communautés linguistiques étrangères. Ceci a des implications dans l'organisation de la macro- et microstructure et met l'accent sur le besoin de clarté et d'usage facile des structures d'accès. D'un point de vue macrostructural, il est important que la sélection du lemme soit en fonction des critères de sélection avec l'usage des fréquences comme facteur dominant, c'est-à-dire en se basant sur les listes des fréquences, des transcriptions des interviews, de l'usage dans la presse, des programmes de télévision, etc. pour aboutir à la sélection d'items lexicaux constituant le noyau du vocabulaire.

2.3 Les articles synopsis dans la microstructure des dictionnaires gabonais

Le traitement microstructural de chaque lemme a pour but de rendre l'utilisateur apte à obtenir les compétences nécessaires pour l'usage pratique du mot en communication réelle. Ainsi, lorsque la macrostructure offre à l'utilisateur l'accès à une certaine partie du vocabulaire, la microstructure offre l'accès aux caractéristiques linguistiques de ces mots. Dans la situation du Gabon, la microstructure devrait aussi offrir un accès aux caractéristiques culturelles de ces mots. Cet accès devrait être traité d'une manière non ambiguë en prévision du transfert le plus facile possible des données. La simplicité dans la définition est un concept largement respecté en vue de maximiser les chances de l'utilisateur de comprendre les définitions des mots. À cet effet, la notion de grande simplicité devrait être un principe majeur dans l'utilisation des articles synopsis.

3. Orientation théorique et méthodologique pour le traitement de l'accès des usagers aux données

3.1 Les articles synopsis et l'utilisation des structures d'accès

Les dictionnaires ne véhiculent pas seulement les données à travers les signes linguistiques et codes mais aussi par l'usage des conventions ou signes typo-

graphiques. Ces codes ou marqueurs typographiques ont pour but de contribuer au transfert optimal des données. Malheureusement, comme on peut souvent constater, une mauvaise utilisation ou une abondance de ces marqueurs et codes peuvent rendre plus complexes l'accès à la microstructure et peuvent empêcher par cela une bonne interprétation du texte lexicographique. L'usager type des dictionnaires gabonais sera un usager qui ne sera pas encore familier avec l'utilisation des dictionnaires particuliers des langues gabonaises, ainsi qu'avec l'utilisation du dictionnaire en général. Or, cet usager comptera sur ces dictionnaires pour une assistance spontanée sans pour autant être obligé de suivre les renvois d'articles dans les autres composantes du dictionnaire, surtout lorsque ces renvois ne sont pas codés et nécessitent encore d'être décodés pour avoir l'aide souhaitée.

En incluant systématiquement les articles synopsis, les dictionnaires gabonais devraient répondre aux besoins d'un accès instantané de l'information en mettant en place des structures d'accès efficace. Pour cela l'utilisation des structures d'accès doit être faite avec la plus grande simplicité. Cette complémentarité devrait aboutir à diminuer les marqueurs typographiques ainsi que les autres entrées qui rendent l'article trop complexe. Cette combinaison devrait aussi introduire des nouvelles conventions visant à rendre moins ambigu l'accès à chaque catégorie des données contenue dans l'article. Par exemple, la façon dont les indications tonales ont été élaborées dans le dictionnaire fang-français/français-fang, œuvre du révérend Samuel Galley (1919–1959), pourrait être simplement marquée par la représentation phonétique du lemme dans la compilation de nouveaux dictionnaires incluant la langue fang ou n'importe quelle autre langue gabonaise. En effet, Galley a noté, par une lettre entre parenthèse l'indication relative à la tonalité du lemme. Pour le mot *aba*, il se présente comme suit:

ABA (h) n. 4, pl. meba (dialecte atsi). Corps de garde où se tiennent les hommes. Syn.: abeñy (h).

Dans cet article le (h) indique une haute tonalité. Ceci est confus dans la mesure où il n'y a aucune indication quant à la voyelle portant le ton, et représente un accès assez difficile pour les usagers qui doivent s'astreindre à maîtriser cette structure d'accès assez complexe parmi tant d'autres. Le lexicographe ferait mieux de donner l'indication tonale dans la représentation phonétique du lemme ABA dont la première voyelle a le ton haut, et la deuxième une voyelle longue. On pourrait ainsi avoir:

ABA, [ába:] ...

En d'autres termes, dans la compilation des dictionnaires gabonais qui impliqueraient l'utilisation des articles synopsis, les lexicographes devraient limiter les structures d'accès dans la mesure où leur utilisation abondante pourraient aboutir à la confusion chez les utilisateurs.

3.2 La recherche d'une plus grande clarté

Comme cela a été mentionné plus haut l'utilisation des renvois internes d'articles devraient coïncider avec l'objectif d'atteindre une plus grande clarté. Une caractéristique dominante des dictionnaires gabonais qui inclurait les articles synopsis devrait être le clair et explicite accès à travers lequel les données sont transférées. Cela implique donc que l'accès devrait être facilité par des entrées qui offrent un transfert direct des données. L'application de ces principes doit amener à reconsidérer l'utilisation des abréviations comme entrées micro-structurales, c'est-à-dire que non seulement le transfert codé des données grammaticales devrait être reconsidéré, mais devrait être aussi valable pour l'usage dans plusieurs dictionnaires des abréviations des classes des mots. Par exemple, au lieu des entrées:

«afan (*pl me ø-*)», [áfàn] qui signifie *forêt*

le lexicographe fera mieux de donner une présentation plus claire et plus explicite de la forme du pluriel:

«afan (*pl mefan*)».

La plus grande clarté, combinée à une meilleure chance d'accéder aux données, peut être obtenue en abandonnant le système d'indication des données morphologiques, comme les pluriels, les diminutifs et les temps passés en notant seulement le suffixe qui s'y applique, au lieu de l'entrée entière.

Ce genre d'approche rend l'usager apte à obtenir directement et instantanément l'accès à l'information dont il/elle a besoin avec moins de chance de mal interpréter le système d'accès dans le dictionnaire.

3.3 Contrôle de la densité des données dans l'utilisation des articles synopsis

Simplicité et clarté devraient être considérées comme deux points majeurs du succès des transferts des données dans l'utilisation des articles synopsis. Quand bien même ces deux aspects peuvent être véhiculés à travers les structures des renvois internes des articles, les lexicographes des dictionnaires gabonais pourraient aussi bien utiliser d'autres méthodes pour atteindre cet objectif. Cette méthode est d'évaluer et de réguler la densité des données appropriées à chaque dictionnaire. Selon Hausmann (1989: 982, cité par Gouws 1993: 39) la densité des données d'un dictionnaire est la relation quantitative entre la macro- et micro-structure. Une augmentation des données dans la microstructure d'un dictionnaire, sans pour autant en changer la macrostructure impliquerait donc une augmentation des données de cette macrostructure. Ainsi, un dictionnaire pourrait être orienté soit de façon macro- ou microstructurale ou pourrait également maintenir un équilibre des données entre la macro- et microstructure.

Les restrictions d'espace dans le traitement lexicographique peuvent être résolues par le transfert implicite des données. Un exemple de ce transfert implicite des données dans un dictionnaire général bilingue à dimension trilingue peut être la présentation d'un registre d'équivalents dans les posttextes.

Ainsi, dans un dictionnaire général bilingue à dimension trilingue vili/français/anglais, le transfert implicite des données trilingues pourrait être la présentation d'un registre d'équivalents vili/français/anglais dans les posttextes.

3.4 Orientation bidirectionnelle dans l'utilisation des articles synopsis

L'orientation bidirectionnelle dans l'utilisation des articles synopsis interviendra à priori dans la compilation des dictionnaires bilingues. Ainsi, vu la densité des données dans l'utilisation des articles synopsis, les lexicographes des dictionnaires gabonais devraient maintenir un équilibre entre la macro- et micro-structure en adoptant une orientation bidirectionnelle selon la conception de Hausmann et Werner (1991: 2740, cité par Gouws 2001: 118) qui est relative à la notion du dictionnaire biscopal.

Dans l'utilisation des articles synopsis, les lexicographes devraient mettre l'accent sur la délimitation de la microstructure de telle façon que l'accès à l'information ne soit pas obstrué par une grande quantité de données. En prenant les décisions sur la densité des données à inclure dans un dictionnaire spécifique, les lexicographes devraient encore une fois prendre en considération les besoins et les aptitudes du groupe cible. En d'autres termes, la densité des données à inclure dans les articles synopsis dépend du type et du *véritable but* du dictionnaire à compiler, ainsi que des caractéristiques, des besoins et des problèmes du groupe cible, c'est-à-dire l'ensemble des potentiels des usagers pour lesquels un dictionnaire est destiné. Le *véritable but* du dictionnaire est en terme métalexigraphique l'objectif visé par la compilation d'un dictionnaire de transférer, au moyen de données lexicologiques, les données concernant l'ensemble d'items lexicaux qui sont inclus comme unités du traitement pour assurer l'acquisition par le pouvoir linguistique de l'utilisateur cible (cf. Gouws 2001: 103).

4. Pour une approche concise et simplifiée de la macrostructure

Dans un Gabon multilingue et multiculturel l'utilisateur type des dictionnaires gabonais sera premièrement intéressé par l'orthographe et la définition des lemmes ainsi que leur utilisation en situation de communication. Ainsi ces ouvrages de références destinés à vulgariser et à standardiser les langues gabonaises doivent fournir une variété de catégories de données telle que la prononciation, la morphologie et l'étymologie nécessaires. Néanmoins, cela ne signifie pas qu'il faut inclure tous les aspects sémantiques d'un mot mais plutôt de

fournir les données nécessaires en vue de rendre ce mot accessible, compréhensible pour la plus grande variété possible des usagers. En effet, à cause de l'insuffisance des connaissances linguistiques et culturelles liées aux langues gabonaises, le souci premier des usagers des dictionnaires gabonais sera d'avoir à leur portée des dictionnaires d'accès faciles qui leurs permettront d'avoir l'information utile et nécessaire pour la définition et l'usage correct des items lexicaux en langues gabonaises, c'est-à-dire que ces dictionnaires doivent être des outils fonctionnels en ce qu'ils doivent contribuer à parfaire les aptitudes des usagers à communiquer en langues gabonaises.

Toutefois, il est également important de mentionner ici la relation entre le type du dictionnaire à compiler et la façon dont seront utilisés les articles synopsis. En effet, la quantité des données à inclure dans ces articles lorsqu'ils sont destinés à un dictionnaire encyclopédique ne sera pas la même que lorsque ceux-ci sont destinés à un dictionnaire pédagogique scolaire par exemple.

4.1 Les exemples scripturaux

Baunebjerg Hansen (1990: 19, cité par Gouws 1993: 42) considère les exemples scripturaux comme une catégorie lexicographique à part. Comme le soutient Rademeyer (1992, cité par Gouws 1993: 42) dans le même ordre d'idées, les exemples scripturaux peuvent avoir une variété de fonctions dans un dictionnaire. L'utilisation des exemples scripturaux se révélera donc en premier lieu dans le contexte de l'utilisation des articles synopsis qui contiennent entre autres des données culturelles qui doivent être comprises par les usagers. En effet, les exemples scripturaux ont cette fonction pragmatique dans la réception et la production des textes, en éclairant et en distinguant par exemple les différents sens polysémiques d'un lemme, qui permettent une utilisation performante des items lexicaux dans un contexte de communication.

4.2 Les exemples picturaux

Les exemples picturaux, même s'ils sont souvent assez utilisés, sont malheureusement parfois considérés comme redondants et comme des ornements aux articles du dictionnaire. Toutefois, ces exemples ont une grande valeur lexicographique, puisque tout en étant un complément à la définition verbale, ils ajoutent une information ou une explication supplémentaire. Par cela, ils permettent d'améliorer la compréhension de l'utilisateur. L'inclusion des exemples picturaux dans les articles synopsis ne se limitera pas seulement, comme il est de coutume, aux seuls dictionnaires monolingues. Les exemples picturaux devraient être pris en compte dans la compilation de n'importe quel type de dictionnaire incluant les langues gabonaises dès que le besoin se ferait sentir. Leur usage jouerait un rôle encyclopédique non négligeable dans la compréhension des données extralinguistiques incluses dans les articles synopsis. Ainsi,

ils ne seront pas redondants car ils contribueront à un transfert optimal des données nécessaires aux besoins et aux demandes des usagers des dictionnaires de langues gabonaises.

Par exemple, le lemme désignant *mbandja* doit être précédé de l'illustration malgré la définition. En effet, ce lemme pourrait être défini comme une «sorte de hangar» ou «petite maison» qui sont deux concepts qu'un usager européen par exemple aurait du mal à représenter car le *mbandja* représente toute autre chose en réalité qu'un hangar ou une petite maison selon la conception européenne.

5. Conclusion

Les articles synopsis constituent des moyens efficaces lorsqu'il s'agit de véhiculer les données linguistiques et extralinguistiques dans un dictionnaire. Leur traitement doit reposer sur une base théorique et pédagogique du point de vue lexicographique et sociologique, mais aussi sur des principes linguistiques. Dans un Gabon multilingue et multiculturel les dictionnaires de langues gabonaises doivent offrir un accès non seulement à ces différentes langues mais aussi aux différentes cultures véhiculées par ces langues. Ces dictionnaires, en incluant les articles synopsis, pourraient donc non seulement servir à relier les écarts linguistiques et culturels entre les communautés linguistiques du Gabon, mais pourraient également permettre aux locuteurs natifs des langues gabonaises de rendre performante leurs compétences à communiquer dans leurs propres langues.

Références bibliographiques

- Baunebjerg Hansen, G.** 1990. *Artikelstruktur im zweisprachigen Wörterbuch. Überlegungen zur Darbietung von Übersetzungsäquivalenten im Wörterbuchartikel.* Lexicographica. Series Maior 35. Tübingen: Max Niemeyer.
- Galley, S.** 1964. *Dictionnaire fang-français et français-fang, suivi d'une grammaire fang.* Neuchâtel: Henri Messeillier.
- Gouws, R.H.** 1993. Afrikaans Learner's Dictionaries for a Multilingual South Africa. *Lexikos* 3: 29-48.
- Gouws, R.H.** 1999. *Die maatskaplike gerigtheid van die metaleksikografie in 'n meertalige samelewing.* Cours d'ouverture. Stellenbosch: Universiteit van Stellenbosch Drukkery.
- Gouws, R.H.** 2001. Formation lexicographique: approches et thèmes. Emejulu, J.D. (Éd.). 2001. *Éléments de lexicographie gabonaise. Tome I:* 95-134. New York: Jimacs-Hillman.
- Hausmann, F.J.** 1989. Das Definitionswörterbuch. Hausmann, F.J. et al. (Éds.). 1989-1991: 981-988.
- Hausmann, F.J. et R.O. Werner.** 1991. Spezifische Bauteile und Strukturen zweisprachiger Wörterbücher: Eine Übersicht. Hausmann, F.J. et al. (Éds.). 1989-1991: 2729-2769.
- Hausmann, F.J. et al. (Éds).** 1989-1991. *Wörterbücher: Ein internationales Handbuch zur Lexikographie / Dictionaries: An International Encyclopedia of Lexicography / Dictionnaires: Encyclopédie interna-*

tionale de lexicographie. Handbücher zur Sprach- und Kommunikationswissenschaft 5.1-5.3. Berlin/New York: Walter de Gruyter.

Mavoungou, P.A. 2002. Synopsis Articles in the Planning of a Trilingual Dictionary: Yilumbu-French-English. *Lexikos* 12: 181-200.

Rademeyer, L. 1992. *Die funksies van verbale voorbeeldmateriaal in eentalige woordeboeke*. M.A. thèse inédite. Stellenbosch: Universitè de Stellenbosch.

Wiegand, H.E. 1997. Die gesellschaftliche Verantwortung der wissenschaftliche Lexikographie. *Hermes, Journal of Linguistics* 18: 177-202.

HAT veertig jaar — 'n persoonlike oorsig

F.F. Odendal, *Emeritus professor in die Afrikaanse en Nederlandse
Taalkunde, Voormalige Randse Afrikaanse Universiteit, Johannesburg,
Republiek van Suid-Afrika*

Opsomming: In 2005 was dit 40 jaar sedert die verskyning van die eerste uitgawe van HAT (*Verklarende Handwoordeboek van die Afrikaanse Taal*), onder redakteurskap van P.C. Schoonees, C.J. Swanepoel, S.J. du Toit en C.M. Booysen. Die ontwikkeling voor hierdie publikasie en daarna van HAT 2 en 3 onder die alleenredakteurskap van F.F. Odendal en van HAT 4 en 5 onder die gesamentlike redakteurskap van Odendal en R.H. Gouws, word oorsigtelik geskets.

Sleutelwoorde: HAT (VERKLARENDE HANDWOORDEBOEK VAN DIE AFRIKAANSE TAAL), WAT (WOORDEBOEK VAN DIE AFRIKAANSE TAAL)

Abstract: HAT Forty Years — A Personal Survey. In 2005 it has been 40 years since the publication of the first edition of HAT (*Verklarende Handwoordeboek van die Afrikaanse Taal*), then with P.C. Schoonees, C.J. Swanepoel, S.J. du Toit and C.M. Booysen as editors. The development before this publication and subsequently of HAT 2 and 3 under the sole editorship of F.F. Odendal and of HAT 4 and 5 under the combined editorship of Odendal and R.H. Gouws is synoptically sketched.

Keywords: HAT (VERKLARENDE HANDWOORDEBOEK VAN DIE AFRIKAANSE TAAL), WAT (WOORDEBOEK VAN DIE AFRIKAANSE TAAL)

1.1 Die verskyning van die vyfde uitgawe van HAT (*Verklarende Handwoordeboek van die Afrikaanse Taal*) in September 2005, veertig jaar na die publikasie van HAT 1 (1965), en die gelyktydige uittrede van Odendal as lid van die redaksie (waaraan hy verbonde was vanaf 1971), is waarskynlik 'n baie geleë tyd om die hooftrekke van die ontstaan en ontwikkeling van HAT op skrif te stel, ook omdat Odendal die enigste nog lewende redakteur is wat 'n oorsig oor die hele tydperk kan gee, al is daar o.a. weens wisseling van uitgewers, besonderhede waaroor hy nie beskik nie.

1.2 Om die geskiedenis van HAT, veral wat sy ontstaan en die eerste uitgawe te verstaan, is dit nodig om aandag te gee aan die verhouding tussen HAT en die WAT (*Woordeboek van die Afrikaanse Taal*, aanvanklik slegs bekend as *Die Afrikaanse Woordeboek*). Dit is algemeen bekend dat daar voor en gedurende die jare twintig van die vorige eeu ernstige gesprekke deur kundiges gevoer is oor die totstandbrenging van 'n Afrikaanse woordeboek. (Daar is in

een stadium selfs 'n tweetalige woordeboek oorweeg.) Uiteindelik is 'n mate van vastigheid bereik toe 'n ooreenkoms in Maart 1926 aangegaan is tussen die Nasionale Pers en die regering om 'n eentalige verklarende woordeboek van ongeveer die omvang van die destydse Nederlandse *Van Dale* tot stand te bring. Hierdie *Van Dale* was slegs 'n eendelige werk — eintlik 'n handwoordeboek. J.J. Smith, professor in Afrikaans aan die Universiteit van Stellenbosch, sou die redakteur wees, en die werk moes binne drie jaar voltooi word. Die gestelde tydperk was, agterna gesien, heeltemal onrealisties; dit toon egter ook hoe dringend die taak beskou is. Toe die woordeboek nie binne die voorgestelde tyd voltooi is nie, het die Universiteit van Stellenbosch die Nasionale Pers se deel van die kontrak oorgeneem; nou sou die woordeboek binne vyf jaar voltooi moes word (ewe onrealisties as die eerste doelwit).

1.3 Prof. Smith het na ongeveer 20 jaar om gesondheidsredes verlof geneem en nie verder aan die woordeboek gewerk nie. In 1947 is hy deur dr. P.C. Schoonees (skoolhoof op Vryheid) as hoofredakteur opgevolg. Sonder om in besonderhede te tree, moet vermeld word dat die doelwit van die WAT drasties verskuif het. Na 'n Kommissie van Onderzoek aanbevelings gemaak het, is o.m. 'n Raad van Beheer ingestel, wat die WAT-onderneming op 'n vaster basis geplaas het. Die Sekretaris van Onderwys sou as voorsitter dien en die Rektor van die Universiteit van Stellenbosch as administrateur. Belangrik was dat die doelwit van die WAT nou was om 'n omvattende werk te lewer wat die "volledige" Afrikaanse woordeskat sou opteken en verklaar, met as voorbeelde die OED (*The Oxford English Dictionary*) en die WNT (*Woordenboek der Nederlandsche Taal*). Só 'n werk kon bepaald nie binne vyf jaar voltooi word nie.

1.4 Onder Schoonees se deurdrywende leiding het die eerste deel (A–C) in 1951 die lig gesien, die tweede (D–F) in 1955 en deel III (G) in 1957. Dit was nou duidelik dat die Woordeboek, anders as in die eerste 20 jaar, konstante vordering sou maak — en hiervoor verdien Schoonees die nodige lof. Dit was eweneens duidelik dat, ten spyte van 'n behoorlike uitbreiding van die redaksie, die Woordeboek nie binne enkele jare voltooi sou word nie. Daarom het die Beheerraad teen die einde van die jare vyftig, begin van die jare sestig, besluit dat gelyklopend met die omvattende werk ook 'n "handwoordeboek", "standaardwoordeboek", "lessenaarwoordeboek", of watter naam dit ook sou kry, aangepak moes word. Advertensies vir 'n redakteur is geplaas (volgens beplanning sou dit 'n eenpersoonstaak wees), talle aansoeke is ontvang en 'n tweetal kandidate is as aanstelbaar beskou, t.w. dr. M.J. Posthumus en L.R. Heiberg (mondelinge medeling van Schoonees aan Odendal). Die Beheerraad het egter van hierdie besluit afgesien en geen aanstelling is gemaak nie.

Dit is onnodig om hier in verdere besonderhede oor die vordering van die WAT te tree, daar dit HAT weinig geraak het, en volledig uiteengesit is deur Schoonees (s.j.), Snijman (1975²) en Gericke (1991).

2.1 Die gesprek en onderhandelinge rondom 'n handwoordeboek by die WAT het wel 'n baie positiewe resultaat gehad: Schoonees het naamlik die idee opgevat om in sy eie tyd so 'n werk op te stel. Die Beheerraad het sy voorneme goedgekeur, selfs met soveel entoesiasme dat hy toestemming verkry het om die materiaal van die WAT te gebruik. So het 'n noue band ontstaan tussen die WAT en die eerste uitgawe van HAT. Nadat Schoonees oorlede is, het HAT nie verder hierdie band behou nie.

2.2 Schoonees het met sy gebruiklike geesdrif, werkkrag en deursettingsvermoë begin werk aan HAT. Dit het egter gou geblyk dat dit 'n moeilike, selfs onmoontlike taak vir een persoon was om in 'n kort tyd te voltooi. Schoonees het Odendal reeds vroeg genader om op gelyke voet met hom saam te werk, 'n aanbod wat Odendal om diverse redes van die hand gewys het. Hy het egter die naam van C.J. Swanepoel, lid van die redaksie van die WAT, aan Schoonees genoem. Schoonees het die wenk aanvaar en Swanepoel het positief gereageer. Ook met die hulp van Swanepoel het die werk nie vinnig genoeg na Schoonees se oordeel gevorder nie, en is dr. S.J. du Toit, ook lid van die redaksie van die WAT, as derde redakteur bygebring. Nog later het C. Murray Booyesen tot die redaksie van HAT toegetree. Swanepoel was Afrikaans-onderwyser voordat hy by die WAT aangesluit het, Du Toit was reeds afgetree, maar voor sy aanstelling as lid van die redaksie van die WAT, hoofinspekteur van die destydse Oranje-Vrystaat, en Booyesen was voor sy aftrede op Stellenbosch Direkteur van Onderwys in die destydse Natal.

2.3 Die eerste uitgawe van HAT het in 1965 verskyn onder al vier name in 2.2. genoem, alhoewel Booyesen hom reeds om gesondheidsredes aan praktiese deelname onttrek het. Al was hy dit in feite, word Schoonees nie as hoofredakteur aangedui nie. Die uitgewer van HAT 1 was Voortrekkerpers, wat enkele jare later met Afrikaanse Pers Boekhandel saamsmelt om Perskor te vorm, die firma wat HAT 2–HAT 4 uitgegee het (al het Perskor reeds voor die verskyning van HAT 4 opgegaan in Maskew Miller Longman (MML)). HAT 5 is uitgegee deur Pearson Education South Africa (PESA), die afdeling van MML wat opvoedkundige werke hanteer.

3. In die "Woord Vooraf" van HAT 1 word, histories gesien, 'n waardevolle uiteensetting gegee van watter dele van die oorspronklike manuskrip deur watter redakteur bewerk is. Dit gee ook 'n indruk van hoe sekere letters verdeel en onderverdeel is onder die verskillende redakteurs in 'n poging om die projek so spoedig moontlik te voltooi. Gaan 'n mens die verdeling van die manuskrip in terme van gedrukte bladsye na, klop die tantième van die verskillende redakteurs nogal goed met die hoeveelheid wat elkeen gelewer het. Die finale verdeling was: P.C. Schoonees 9,92%, C.J. Swanepoel 4,11%, S.J. du Toit 3,17% en C.M. Booyesen 2,8%.

4.1 Toe Perskor Odendal in 1971 nader met die versoek om die redakteur-skap van HAT oor te neem (die oorspronklike redakteurs was oorlede of, indien nog in die lewe, in Perskor se woorde, "te oud" om die werk voort te sit), het dit as 'n totale verrassing gekom. Nie alleen was sy tyd behoorlik in beslag geneem as professor in Afrikaanse en Nederlandse Taalkunde by die jong Randse Afrikaanse Universiteit, as voorsitter van die Taalkommissie van die S.A. Akademie vir Wetenskap en Kuns, eweneens as voorsitter van die Linguistevereniging van Suider-Afrika en ook as lid van die Komitee vir Taaladvies van die SAUK nie, maar hy het lank geen kontak op enige wyse met HAT gehad nie. Hy het, ten spyte van al sy ander bedrywighede, besluit om die aanbod te aanvaar. Sy eerste oorweging was nie 'n finansiële nie, maar die gedagte dat, as alles voorspoedig met HAT verloop, hy na sy aftrede iets sinvols sou hê om te doen. Dit het dan ook so uitgewerk: hy kon tot sy 79ste jaar met HAT besig wees.

Alhoewel Odendal in hierdie stadium slegs enkele artikels oor die leksikografie laat verskyn het, was hy tog nie volkome onkundig op hierdie gebied nie: hy het drie jaar ondervinding by die Vaktaalburo gehad en tien jaar as mederedakteur en assistenthoofredakteur van die WAT; sy jarelange verbintenis (sedert 1962) met die Taalkommissie het hom ook sydelings met die praktyk van werk-met-woorde in verbinding gehou.

By sy aanvaarding van Perskor se aanbod, het Odendal die firma gevra wat hulle as 'n realistiese tyd vir die voltooiing van die werk beskou. Die antwoord was twee jaar. Odendal het dit onaanvaarbaar gevind want, soos hy dit gestel het, "ek kan nie my naam teken by 'n werk wat ek nie met 'n mate van deeglikheid beskou en desnoods gewysig het nie". Sy skatting van vier jaar is deur Perskor aanvaar; uiteindelik sou dit agt jaar duur. Die ooreenkoms tussen Perskor en Odendal is in 1972 onderteken.

4.2 Odendal het in 1972 'n volle jaar langverlof gehad en kon dus heelwat tyd aan die herbewerking van HAT afstaan. Sy werkwyse was om die galeiproewe van HAT 1 as basis te gebruik en daarop te wysig, aan te vul en te skrap soos nodig.

Een of twee van die oorspronklike redakteurs het sporadies wysigings en byvoegings aangebring. Die wyse waarop hierdie insette egter voorgelê is, was van so 'n aard dat dit 'n oponthoud eerder as 'n hulp was. Odendal het dus vroeg besluit om hierdie opmerkings te negeer. Gevolglik moet hy die volle verantwoordelikheid vir HAT 2 dra.

5.1 'n Nuwe hoofstuk in die geskiedenis van HAT begin dus in 1972, met Odendal as alleenredakteur vir ongeveer 'n kwarteeu.

Die galeie van HAT 1 het 'n uitbreiding van ongeveer 50% ondergaan en daar is gepoog om 'n werklik hersiene uitgawe daar te stel: naas byvoegings is definisies gewysig, argaïese woorde (en betekenisse) geskrap, veral ook woorde en betekenisse wat eerder Nederlands as Afrikaans was, alhoewel in hierdie geval omsigtig te werk gegaan is, veral met die oog op ouer taalgebruikers.

Hier moet die onskatbare bydrae van mev. Estelle Odendal vermeld word, wat dwarsdeur die tyd wat haar man met HAT besig was, op 'n konstante basis nuwe en ontbrekende woorde uit koerante, tydskrifte en boeke opgeteken het. Alhoewel hierdie bydrae telkens in die voorwoord van nuwe uitgawes erken word, is eers in die voorwerk van HAT 5 'n spesiale dankwoord aan haar gerig.

5.2 Een aspek waaraan Odendal by die bewerking van HAT 2 spesiaal aandag gegee het, was die uitbreiding en verbetering van voorbeeldsinne. Dit was, en is, sy oortuiging dat woorde "in aksie" in goed gekose voorbeeldsinne, dus binne sintaktiese verband, hul betekenis(se) duidelik(er) laat blyk as net in 'n (afstandige,) koue, wetenskaplike definisie.

5.3 HAT 2 verskil so ingrypend van HAT 1 dat dit 'n verlies is dat die galeie waarop HAT 2 bewerk is, verlore gegaan het.

5.4 HAT 2 het in 1979 verskyn, onder die name van Odendal plus die oorspronklike vier redakteurs (al het hulle geen bydrae meer gelewer nie). Die bekendstelling het by wyse van 'n geleentheid by die destydse Randse Afrikaanse Universiteit (RAU) geskied, waarby die destydse voorsitter van die Akademie, dr. H.W. Snyman, voorheen voorsitter van die S.A. Mediese Vereniging, 'n leergebonde eksemplaar aan die redakteur oorhandig het. Ook teenwoordig was senior kollegas van die RAU, o.a. die Rektor, prof. J.P. de Lange en ook mnr. C.J. Engelbrecht, bekende taaladviseur by *Die Transvaler* en verwante publikasies, wat lang lyste voorstelle aan Schoonees en later ook Odendal laat toekom het.

5.5 Toe dit bekend word dat Odendal HAT sou herbewerk, het talle kollegas en lede van die publiek (die meeste van lg. onbekendes) voorstelle tot verbetering en vreemde woorde vir opname aan die uitgewers gestuur. Die grootste deel van hierdie voorstelle was waardevol, en dit was 'n voorreg om sommige bydraers in die voorwoord te bedank.

Werk aan HAT 2 was 'n intrigerende oefening, o.a. omdat HAT 1 so ongelyk van gehalte was vanweë die bewerking deur verskillende redakteurs met verskillende benaderings.

5.6 HAT 2 is baie gunstig deur die pers, radio en televisie ontvang. Odendal het op uitnodiging onderhoude gevoer met verskeie koerante en meermale oor die radio; hy het selfs opgetree in 'n populêre televisieprogram wat hoofsaaklik oor die kunste gehandel het.

6. Intussen het Odendal voortgegaan met werk aan HAT met die oog op 'n derde uitgawe. Dit het spoedig geblyk dat HAT die mark in sy bepaalde nis verower het en as gesaghebbend aanvaar is deur die publiek (oud en jonk). Dit het 'n plek gekry in die meeste skole, asook in regeringskantore, by sakelui, e.d.

Van besondere belang was dat dit as outoriteit aanvaar is by die bepaling van die betekenis van 'n woord in hoër en laer howe (soos mondeling aan Odendal berig deur 'n regter van die Hoërhof in Transvaal). Voeg hierby dat koerante en briëfskrywers aan koerante gereeld HAT as gesaghebbende bron aangehaal het as dit oor die betekenis van 'n woord gaan en dit was duidelik dat HAT geleidelik 'n huishoudelike naam geword het.

7.1 Toe Perskor ernstig begin praat oor 'n derde uitgawe, het Odendal prakties voltyds daaraan begin werk. Hy het selfs in 1989, twee jaar voor sy verpligte aftreedatum, die diens van die Universiteit verlaat om HAT te voltooi. Dat HAT 3 eers in 1994 verskyn het, is objektief beskou nie aan die outeur te wyte nie. Die verskyning van HAT 3 is met twee, drie jaar vertraag deur die uitgewer self. Wat ook al die redes daarvoor, die outeur was en is daar nog steeds onbewus van. Een moontlike rede kon wees dat Perskor reeds finansiële probleme ondervind het, wat die uitgawe van 'n lywige, moeilik setbare werk soos HAT problematies gemaak het.

Daarbenewens kon Odendal nie 'n ooreenkoms vir HAT 3 van Perskor kry nie. Daar was geen nuwe ooreenkoms nodig nie, hoogstens 'n aanpassing van tantièmes. Perskor het egter 'n addendum by die ooreenkoms van HAT 2 verkies, maar ook dit het nie gerealiseer nie. Uiteindelik moes Odendal sterk aandrang by die verantwoordelike persoon (nie 'n lid van Perskor se personeel nie) wat die ooreenkoms moes voorberei, om dit te finaliseer.

7.2 Positiewer was die feit dat HAT 3 besonder goed deur die media ontvang is, selfs gunstiger as HAT 2, en Odendal het weer onderhoude met al die media gevoer.

7.3 By die voorbereiding van HAT 3 het Odendal probeer verbeter op al die fasette wat t.o.v. HAT 2 genoem is (vgl. 5.1 en 5.2 hierbo). Die omvang is met 30% vergroot (moontlik gemaak deur die werk in drie i.p.v. twee kolomme te druk). Argaiëse woorde is weer met versigtigheid verwyder. In die besonder is, meer as vroeër, op 'n sistematiese wyse aandag gegee aan etikette, wat eintlik daarop neerkom dat 'n woord sonder etiket, gewone Standaardafrikaans is — in beginsel altans. In die voorwerk word 'n spesiale afdeling afgestaan aan "Gebruiklikste etikette en afkortings". Dat die aanbieding enigsins onsistematies geword het, moet vir die rekening van die uitgewer staan, want dit is nooit aan die redakteur voorgelê nie.

Problematies t.o.v. etikette was veral die "rassistiese" woorde, en in 'n mindere mate wat vandag as "gender"-woorde bekend staan. Daar was sterk druk, ook buite HAT om, om hierdie woorde heeltemal weg te laat om sodoende geen gebruiker van Afrikaans te na te kom nie. Odendal het hom op die standpunt gestel dat dergelike woorde wel opgeneem moet word as 'n volledige beeld van die Afrikaanse woordeskat gegee wil word. Hierdie kategorie woorde moet egter met veel groter omsigtigheid en empatie as in die verlede ge-

hanteer word. Die juiste etikette (en desnoods nog die bewoording van die definisies self, asook selfs 'n sterk afwysing van die gebruik daarvan) behoort die status van die woorde genoegsaam aan te dui.

7.4 Die waarde wat Odendal aan voorbeeldsinne heg, is vroeër reeds onder 5.2 beklemtoon. In HAT 3 het hy verder gegaan. Nie alleen is sy selfgemaakte voorbeelde vermeerder nie, maar hulle is aangevul met 'n groot aantal sitate uit die werke van Afrikaanse skrywers. Hierdie gebruik is algemeen in woordeboeke van groter omvang, maar rondom 1990 was HAT by die skrywer se wete die enigste woordeboek van sy grootte wat uitgebreid uit die literatuur aangehaal het — en dit geld nie net vir Afrikaans nie. Odendal het sy werkwyse in die voorwoord tot HAT 3 "verdedig" deur daarop te wys dat sitate uit skrywers bewys dat gewone én ongewone woorde ook in die skryftaal van erkende skrywers voorkom. Daardeur word verder erkenning gegee aan die belangrike rol wat die hoë gehalte van ons letterkunde in die bestaansreg van Afrikaans speel.

7.5 Soos in die vorige uitgawes is besonder aandag aan uitdrukkings gegee, is hulle uitgelig en volgens 'n vasgestelde wyse telkens aan die einde van 'n lemma gegroepeer.

7.6 HAT 3 was die eerste Afrikaanse woordeboek wat geprobeer het om 'n behoorlike uiteensetting te gee van sy werkwyse en die teoretiese beginsels waarop dit berus. Oor die lys etikette is reeds onder 7.3 uitgewy. Wat ontbreek het, was 'n skematiese voorstelling waaruit die verskillende afdelings van 'n lemma duidelik blyk, iets wat tot groot hulp vir die gebruiker kan dien. Dit is in die 4de en nog beter in die 5de uitgawes reggestel.

8.1 Samewerking tussen Odendal en Perskor t.o.v. HAT 2 en 3 was in die geheel aangenaam. Daar is weinig druk op die redakteur geplaas en voorstelle van sy kant is steeds met groot simpatie bejeën. Minder aangenaam was veral die feit dat die skakelpersoon (eintlik redakteur) by Perskor dikwels vervang is (sonder om Odendal in kennis te stel), wat kommunikasie bemoeilik het; en dat Perskor gereeld toegelaat het dat HAT 2 uit druk was (weer sonder kennisgewing). Ewe lastig was die feit dat Odendal nie altyd op hoogte was met watter firma hy besig was om te werk nie: Perskor, Educum, Kagiso, weer Perskor, weer Educum en so voort. Odendal het eers laat uitvind dat Perskor oorgeneem is deur Maskew Miller Longman (en dat HAT nie in die slag gebly het nie).

8.2 Die bekendstelling van HAT 3 was teleurstellend. Odendal het by Perskor aangedring dat dit tyd was om HAT 3 bekend te stel: die boek was gedruk en gebind met genoegsame eksemplare om die mark te betree, maar daar was geen responsie van Perskor se kant nie. Uiteindelik is hy in kennis gestel dat daar een middag om drie-uur 'n funksie sou wees in die Doornfontein-kantore

van Perskor en dat 'n minimum gaste genooi kon word. Dit was 'n geleentheid waarby slegs sowat twaalf persone teenwoordig was en 'n koppie tee bedien is. Die volgende dag het 'n foto van Odendal, saam met die Besturende Direkteur van Perskor, met 'n onderskrif, in *Beeld* verskyn — en dit was die totale reklame vir HAT 3! Die luukse eksemplaar wat hom by herhaling beloof is, het die redakteur nooit bereik nie.

9. Gedurende die jare neëntig het Odendal Perskor genader met die suggestie dat dit tyd was om 'n tweede, jonger redakteur aan te stel met die oog op 'n vierde en verdere uitgawes. Odendal wou verseker dat HAT in bekwame hande sou wees wanneer hy nie langer kon meewerk nie. Hy het net een naam voorgelê, dié van Rufus H. Gouws (tans professor in die Departement van Afrikaans en Nederlands aan die Universiteit van Stellenbosch). Sy status as teoretiese en praktiese leksikograaf maak hom 'n onbetwiste leier as woordeboekkundige. Sy boek *Leksikografie* het in 1991 die Suider-Afrikaanse Linguistevereniging se Toekenning vir Taalwetenskap ontvang en sy publikasies sluit hoofstukke in boeke en artikels in al die gesaghebbende vaktydskrifte nasionaal en internasionaal in. Hy was in 1986 en 1992 erenavorsingsgenoot aan die Navorsingsentrum van die Universiteit van Exeter en in 1999 gasprofessor aan die Universiteit van Heidelberg. Hy is mederedakteur van verskeie woordeboeke en het kontakte met talle woordeboekinstansies oorsee, o.a. as hoofredaksielid van De Gruyter se internasionale vakwoordeboek in leksikografie en woordeboeknavorsing. Hy is betrokke by meerdere projekte met kollegas in België en Nederland en lid van nasionale en internasionale verenigings. Hy was o.a. voorsitter van AFRILEX (the African Association for Lexicography). Intussen styg sy aansien hier en in die buiteland en gee hy leiding aan 'n verbasende aantal nagraadse studente, veral in die metaleksikografie.

Dit is moeilik om aan 'n geskikter kandidaat as mederedakteur te dink. Gelukkig was sowel Perskor as Gouws bereid om Odendal se voorstel te aanvaar. Ten spyte van die groot verskil in ouderdom (byna 30 jaar) tussen die twee redakteurs, was die samewerking tussen Gouws en Odendal uiters aangenaam.

10. Voordat voortgegaan word met HAT se publikasie deur Maskew, Miller, Longman (MML) moet aandag gegee word aan 'n uitbreiding t.o.v. HAT, nl. die uitgawe in die jare negentig van 'n elektroniese weergawe van HAT 3 op CD-ROM. Hierdie taak is hoofsaaklik aangepak en deurgevoer deur Gouws en die eer van die peil daarvan plus die toegevoegde waarde op verskeie gebiede, moet hom toekom. Dat daar allerlei probleme opgeduik het, was te verwagte; die ernstigste vertraging was te wyte aan die finalisering van 'n kontrak met Next Page (vroeër Folio Views). Gouws het egter deurgedruk en uiteindelik kon ELHAT (soos die elektroniese HAT gou algemeen bekend geword het) op twee agtereenvolgende aande in Oktober 1998 in 'n restaurant in Kaapstad bekend gestel word. Tans word gewerk aan 'n nuwe uitgawe van ELHAT, nou gebaseer op HAT 5.

11.1 By die beplanning van HAT 4 het Gouws en Odendal besluit dat die toegelate tyd 'n indringende hersiening van HAT 3 onmoontlik maak en dat, wat die algemene hersiening betref, die alfabet in twee gelyke dele tussen die outeurs verdeel sou word. Odendal sou die eerste, Gouws die tweede gedeelte vir hul rekening neem. Bepaalde aspekte sou egter deur een outeur vir A-Z bewerk word, bv. die byvoeging van sitate, verfyning van etikette, en kontrole en aanvulling van rekenaarterme. Ten spyte van die tydsbeperking is talle nuwe woorde bygevoeg en argaïsmes geskrap. 'n Nuutjie was wel (hopelik met die nodige omsigtigheid) die opname van 'n aantal Engelse woorde soos *cool*, *sorry*, *show* en *song*. Alhoewel hierdie woorde tot 'n minimum beperk is, het dit, soos te verwagte, kritiek uitgelok (tog minder as wat die outeurs voorsien het). Gelukkig was daar ook ondersteuning vir dié "moedige stap" van die redakteurs. 'n Diagram wat die uiteensetting van artikels illustreer, is as deel van die voorwerk bygevoeg.

11.2 Die samewerking tussen die redakteurs het vlot verloop, daar is flink gevorder met die herbewerking van die nuwe deel en HAT 4 kon in November 2000 bekend gestel word by 'n funksie by Oude Libertas, met mnre. Japie Pienaar as seremoniemeester en M.B. Kumalo as geleentheidspreker.

Weer is 'n uitgawe van HAT goed ontvang, maar met minder fanfare as bv. HAT 3. Alhoewel HAT 4 nog onder die imprimatuur van Perskor verskyn het, was dit reeds volledig onder beheer van MML.

12.1 In 2003 versoek Pearson Education South Africa (PESA), wat nou namens MML die uitgewers van HAT was, ietwat onverwags dat die outeurs 'n vyfde uitgawe van HAT in 2004 moet lewer. Albei redakteurs reageer redelik afwysend hierop: dit het net té min tyd toegelaat. Gouws het met die goeie idee gekom dat 2005 nie net 'n realistieser datum sou wees nie, maar dat dit met 'n veertigste verjaarsdagviering van die verskyning van HAT 1 in 1965 gekombineer kon word. PESA het hierdie voorstel aanvaar.

12.2 Wat betref die bewerking van HAT 5 het die redakteurs besluit om elk weer 'n helfte van die werk vir sy spesifieke verantwoordelikheid te neem, dié keer net die omgekeerde as by HAT 4: Gouws die eerste deel, Odendal die tweede. Soos by HAT 4 het elke redakteur spesiale verantwoordelikheid vir bepaalde aspekte van die werk vir die volledige woordeboek aanvaar: uitbreiding van die etimologieë, 'n verbeterde artikeluitleg, uitbreidings van afkortings en verskuiwing van afkortings as 'n eenheid na die einde van die boek, byvoeging van 'n afdeling geografiese name met hulle afleidings, verdere verfyning van etikette, sorg vir ooreenstemming tussen 'n nuwe *Afrikaanse Woordelys en Spelreëls* en HAT 5, 'n voorteks met volledige gebruikershandleiding, verdere aandag aan geskikte sitate (nie net uit die "bekendste" skrywers nie), en 'n woordkeuse gebaseer op 'n verteenwoordigende, uitgebreide en gebalanseerde elektroniese korpus (waarvoor prof. D.J. Prinsloo van die Universiteit

van Pretoria dank verdien). Verder is nóg "vreemde", maar gebruiklike woorde opgeneem en met omsigtigheid woorde uit ander variëteite as Standaardafrikaans.

12.3 HAT 5 het betyds van die drukker gekom, in 'n besonder aantreklike maroen band met goue letters en met 'n veel aantrekliker en lesersvriendelike bladspieël plus, vir die eerste keer, duimdrukke, en op veel beter papier gedruk as sy voorganger.

12.4 HAT 5 is op 14 September 2005 by die landgoed Spier bekend gestel, met Laetitia Botha van PESA as seremoniemeester en prof. W.T. Claassen van die Universiteit van Stellenbosch as geleentheidspreker. Elk van die redakteurs het 'n luukse eksemplaar in halfleer gebind, ontvang. Odendal sou reeds na die vierde uitgawe uit die redaksie tree, maar het om verskeie redes aangebly tot die voltooiing van HAT 5. Dit was dus ook 'n afskeid vir hom, na 'n verbintenis van meer as 30 jaar. Tot sy verbasing het PESA dit ook 'n huldiging van hom gemaak, met o.a. 'n waarderingswoord deur Gouws en 'n plakiet vir " 'n besondere bydrae tot die leksikografiepraktyk".

13.1 HAT 5 het sekerlik die beste blootstelling van al die uitgawes gehad. Nie alleen het enkele weke voor sy verskyning elke Vrydag 'n woordspeletjie (met 'n prys daaraan verbonde) in *Die Burger* verskyn nie, maar na die bekendstelling is 'n volbladpromosie in kleur in *Die Burger*, *Beeld* en *Die Volksblad* gepubliseer, met foto's van die redaksielede daarby. Hiernaas het Daniel Hugo op Radio sonder Grense (RSG) 'n onderhoud met die redakteurs uitgesaai en het *Rooi Rose* 'n dubbelbladartikel met 'n foto van Odendal geplaas met die titel "'n Man van baie woorde", maar met sterk klem ook op HAT self. Om die kroon op alles te plaas, is 'n geslaagde insetsel op natuurlike wyse op *Sewende Laan* ingewerk, met Oubaas besig met HAT in die boekwinkel.

13.2 Laastens het die uitgewers 'n los "boekie", getiteld *Klein HAT*, as 'n toegif los voor in elke eksemplaar geplaas wat "nuwe woorde, en toevoegings tot die vyfde uitgawe van HAT, woordspeletjies en woordraaisels" bevat.

14. Dit sou tans, ná veertig jaar, nie erg oordrewe wees om te sê: "Wat is 'n huis sonder 'n HAT?" nie.

Bibliografie

- Gericke, W. 1991. *Die Woordeboek van die Afrikaanse Taal — 'n Kulturhistoriese verkenning*. Ongepubliseerde M.A.-skripsie. Stellenbosch: Universiteit van Stellenbosch.
- Schoonees, P.C. s.j. *Die Afrikaanse Woordeboek*. Johannesburg: Voortrekkerpers.
- Snijman, F.J. 1975². *U woorde, u Woordeboek*. Kaapstad: A.B.C. Pers.

H.E. Wiegand Receives an Honorary Doctorate from the University of Stellenbosch

On 20 April 2006, the degree Doctor of Literature, *Honoris Causa*, was conferred on H.E. Wiegand by the University of Stellenbosch. The decision by the Council and the Senate to award him this degree has been taken on the grounds of the following considerations stated in the commendatio read at the graduation ceremony.

Herbert Ernst Wiegand, professor emeritus at the University of Heidelberg in Germany, has made an immense contribution over the years to theoretical lexicography, Germanic studies and linguistics.

In lexicographic research, he has played a leading role in the establishment of a comprehensive and complex theory that offers lexicographers from all language groups a model for the compilation of dictionaries.

As a scientist, Wiegand has made a contribution of unparalleled scope. His impressive list of publications bears testimony to this contribution, particularly his magnum opus, the book *Wörterbuchforschung*, and the most comprehensive contribution to the broader field of linguistics, the authoritative reference series, *Handbücher zur Sprach- und Kommunikationswissenschaft*, which was established by him.

As research manager he initiated the journal *Lexicographica* and the book series *Lexicographica Series Maior*, which contains more than a hundred books for the subject specialist. He compiled the book series *Studien zur neuhochdeutschen Lexikographie*, *Wörterbücher in der Diskussion* and *Studien zur zweisprachigen Lexikographie mit Deutsch*, and is a founding member of the book series *Reihe Germanistische Linguistik*, a series in which more than 260 books have already been published.

In Afrikaans, South African and broader African lexicography, Wiegand's theoretical methods of approach have also resulted in research and dictionary projects. His involvement with Afrikaans lexicography has led him to include Afrikaans as one of the languages in the decalingual *Dictionary of Lexicography and Dictionary Research*.

The Pan South African Language Board established National Lexicography Units for each of the eleven South African official languages. Their training programmes are largely based on Wiegand's model from which adaptations and applications have been made so that each dictionary unit could find guidelines for the compilation of a dictionary directed at the specific needs and reference skills of its target user group.

Wiegand's influence in Africa is also evident from the research of doctoral students from Gabon in the postgraduate programme for lexicography at the University of Stellenbosch. All these students work within the framework of the Wiegand model.

As professor, study supervisor, promoter and academic manager Wiegand has enriched his field of study. He has been promoter of no less than 44 doctor's degree students.

His work has been acknowledged in a variety of ways, for example through a commemorative conference, a Festschrift, the publication of his *Kleine Schriften*, and honorary doctorates from the Aarhus School of Business in Denmark and the University of Sofia in Bulgaria.

At the celebration after the graduation ceremony, Professor Elize Botha, Chancellor of the University of Stellenbosch, while welcoming the guests, referred to the honorary doctorates, especially expressing her appreciation for the attention given to lexicography.

In the absence of Professor Walter Claassen, Vice-Principal: Research, Professor Rufus Gouws proposed a toast to Professor Wiegand to which he responded as follows:

Your appreciative words really mean a lot to me. I am grateful and would like to thank you ever so much. Baie baie dankie! I can assure you that I am proud of this new honorary doctorate of mine. There are three reasons for feeling proud about it.

The first reason is the fact that the University of Stellenbosch is not only regarded as one of the three most distinguished universities within Africa but also has an excellent reputation and standing internationally.

The second reason is based on the definite knowledge that the honorary doctorate from the University of Stellenbosch has been awarded to me as an acknowledgement of my scientific work and not, for example, for being a sponsor or a donor of whatever nature.

The third reason I regard as by far the most important. With this honour bestowed on me today the status of the research area in which I have primarily, although not exclusively, worked during the last two decades, has been enhanced in a way that should not be underestimated.

This gives me the opportunity to make a few brief comments on the role of the University of Stellenbosch in the field of lexicography and metalexicography.

The University of Stellenbosch, represented by Professor Walter Claassen, the Vice-Principal: Research, has acknowledged that lexicography should be promoted in the interest of everyone living in South Africa. Furthermore he has recognized, and this should be noted, that this cannot prevail by merely promoting lexicographic processes as a practice, because this practice needs to be supported by a modern theoretical underpinning. In this regard he is of the same opinion as our colleague, Professor Bernard Lategan, the Director of the Stellenbosch Institute for Advanced Study, whom I would like to thank on this

occasion for the privilege of being a fellow of his institute where I may pursue my research. Both Professor Claassen and Professor Lategan are furthermore in agreement with my esteemed colleague, Rufus Gouws, who has always striven towards this theoretical underpinning, without forgetting that the practice has its own dynamics.

For the past two and more decades the University of Stellenbosch has actively supported the continued endeavours of Professor Gouws to establish contacts on national and international level with the leading exponents in the field of dictionary research. The results are quite evident: The University of Stellenbosch, along with the Bureau of the WAT and its outstandingly edited journal *Lexikos*, in which internationally established as well as emerging African researchers publish, has become an internationally renowned centre for dictionary research. In 1986 at a conference of the European Association for Lexicography, it had been demanded on political grounds that the South African Gouws should leave the conference. This, however, was avoided. Nowadays people travel to Stellenbosch for their research, and Professor Gouws is not only invited to the big international conferences but also to the small elite conferences of Europe. He participates in the most important major international projects in the field of metalexigraphy, for example the *Dictionary of Lexicography and Dictionary Research* and the *Supplementary Volume to the Encyclopedia of Lexicography* in the series *Handbücher zur Sprach- und Kommunikationswissenschaft*, to mention only two projects. This international network, promoted by Professor Claassen, and to the realisation of which I have also directed my efforts, benefits not only the University of Stellenbosch and the Stellenbosch students specifically, but also South African lexicography in general.

That this may also prevail in future, I would like to toast with everyone present today.

Lexique Pove–Français / Français–Pove, Mickala Manfoumbi: Seconde Note de Lecture

Hugues Steve Ndinga-Koumba-Binza, *Stellenbosch University Centre for Language and Speech Technology (SU-CLaST), Stellenbosch, République d’Afrique du Sud (13751719@sun.ac.za)*

Résumé: Cet article fait suite au compte-rendu de lecture du *Lexique Pove–Français / Français–Pove* (Mickala Manfoumbi 2004) fait par Mavoungou (2005). En dehors des réserves sur les questions purement lexicographiques, l’ouvrage de Mickala Manfoumbi suscite des interrogations intéressantes qui dépassent le cadre seul de la métalexicographie. Il s’agit des questions sur le contexte de parution, sur la forme ainsi que la situation de l’ouvrage dans le cadre de la planification stratégique de la lexicographie gabonaise. Ces questions amènent à situer le lexique pove dans le cadre général aussi bien de la linguistique gabonaise que de l’émergente lexicographie gabonaise. Cet article se fixe l’objectif d’apporter un complément au compte-rendu de lecture de Mavoungou (2005) sur ces différentes questions.

Mots-clés: LEXIQUE, POVE, FRANÇAIS, DICTIONNAIRE, LINGUISTIQUE GABONAISE, LEXICOGRAPHIE GABONAISE, LANGUES GABONAISES

Abstract: *Lexique Pove–Français / Français–Pove*, Mickala Manfoumbi: A Second Review. This article is a follow-up to a review of the *Lexique Pove–Français / Français–Pove* (Mickala Manfoumbi 2004) by Mavoungou (2005). Apart from reservations on purely lexicographical issues, the work of Mickala Manfoumbi gives rise to interesting questions which exceed the limited field of metalexigraphy. These relate to matters such as the context of publication, the format structure, as well as the positioning of the publication within the strategic planning of Gabonese lexicography. These questions lead to the placing of the Pove lexicon within the framework of Gabonese linguistics as well as the emerging Gabonese lexicography. This article aims at supplementing the review by Mavoungou (2005) concerning these matters.

Keywords: LEXICON, POVE, FRENCH, DICTIONARY, GABONESE LINGUISTICS, GABONESE LEXICOGRAPHY, GABONESE LANGUAGES

1. Introduction

Cet article fait suite à la note de lecture de Mavoungou (2005) sur le *Lexique Pove–Français / Français–Pove* (LPFFP) de Roger Mickala Manfoumbi (2004). Mavoungou (2005) a tenté de situer le LPFFP dans le cadre des attentes de la métalexicographie gabonaise actuelle qui prône une réflexion théorique et un

travail pédagogique dans le processus de préparation d'un dictionnaire. Nous sommes d'avis avec l'auteur que le LPFFP est d'une "facture particulière" et qu'il "peut rendre de bons services aux locuteurs Pove et à toute personne désireuse d'apprendre le gevove comme langue seconde" (Mavoungou 2005: 84).

En dehors des réserves sur les questions purement lexicographiques — à savoir les prétextes, le public visé, la macrostructure, la microstructure, et la portée et direction — le LPFFP suscite bon nombre d'autres questions qui ne se limitent pas seulement au domaine métalexicographique. Celles-ci soutiennent tout le mérite d'une seconde lecture du LPFFP. Ces questions trouvent toute leur importance quand l'on s'en tient au contexte dans lequel vient de paraître le LPFFP. Après avoir présenté ce contexte, notre communication ici compte relever les questions de forme que suscite l'ouvrage de Mickala Manfoumbi. Nous allons également énoncer quelques perspectives qui pourraient être prises en compte non seulement dans le cadre d'une édition nouvelle du LPFFP, mais aussi dans le cadre des productions dictionnaires futures dans les langues gabonaises. Enfin, nous allons situer le LPFFP dans le cadre de la planification stratégique de la lexicographie gabonaise.

2. Le contexte de parution

Le contexte dans lequel vient de paraître le LPFFP relève différents aspects dont l'aspect sociolinguistique, le double aspect linguistique et lexicographique et l'aspect éditorial. Le contexte de parution du LPFFP prend également en compte la profession et la formation de linguiste de son auteur, c'est-à-dire Roger Mickala Manfoumbi.

2.1 L'aspect sociolinguistique

L'aspect sociolinguistique du contexte de parution du LPFFP relève de la politique linguistique du Gabon avec ses implications sociales et/ou sociologiques. Lesdites implications sont par exemple le statut des langues dites nationales aussi bien dans l'administration et la société gabonaises que dans le système éducatif d'une part, et leur revalorisation d'autre part.

Jusqu'à ce jour, le français reste l'unique langue officielle du Gabon et les langues dites gabonaises n'ont pas sur le plan constitutionnel et juridique le caractère national que leur affuble le commun des Gabonais. Par le statut de la langue française au Gabon, l'on comprend qu'il est de bonne guerre que cette dernière soit d'une part langue-cible et d'autre part langue-source du LPFFP.

De plus, le LPFFP vient à paraître au moment où la politique linguistique du Gabon reste l'ombre d'elle-même malgré les diverses intentions gouvernementales en vue de la promotion des langues (Ndinga-Koumba-Binza 2005a: 60-70, Nzang-Bié 2001). C'est une politique linguistique qui présente un besoin primordial d'une planification linguistique (Kwenzi-Mikala 1991, Ndinga-Koumba-Binza 2005a) quand on considère le contexte de mosaïque de langues

qu'est le territoire gabonais. Une mosaïque qui elle-même fait le paysage linguistique gabonais (Ndinga-Koumba-Binza 2005b et 2005c) avec la prise en compte à la fois des langues locales et des langues étrangères, c'est-à-dire les langues des populations migrantes.

Dans cet ensemble, le pove ne constitue que l'une des variantes de l'unité-langue membre (ou okande-tsogo) tel que Kwenzi-Mikala (1987, 1988 et 1998) ne cesse de le souligner. Dans le groupe okande-tsogo, le pove occupe la deuxième place après le tsogo en ce qui concerne la vitalité, le dynamisme et le nombre de locuteurs¹, laissant loin derrière elle le gepinzi (apindzi), le geviya (evia) et le gehimbaka (simba). Cette place du pove peut remettre en cause ou soutenir la pertinence de la production d'un ouvrage de référence dans ce parler qui est du reste l'un des plus minoritaires de la diversité linguistique du Gabon.

2.2 Avancées scientifiques et état de développement des langues

Le deuxième aspect du contexte de parution du LPFFP est double. Il met ensemble les avancées scientifiques dans les domaines de la linguistique et de la lexicographie gabonaises d'une part, et l'état de développement de ces langues d'héritage culturel ainsi que leur introduction dans le système éducatif d'autre part. La question qui se pose est de savoir si l'état de développement de ces langues reflète effectivement lesdites avancées. Nous n'allons pas donner ici une réponse globale à cette question, puisqu'elle nécessite en effet une étude détaillée malheureusement inappropriée pour le présent volume.

Force est de constater que le LPFFP paraît à une période où la linguistique gabonaise connaît une certaine accélération dans les domaines de la description et de l'étude systématique des langues d'une part, et dans le domaine de la planification du développement de ces langues d'autre part.

Ce dernier domaine concerne les divers projets en cours (Idiata 2003), les divers alphabets et orthographes en proposition pour l'écriture des langues gabonaises (Idiata 2002) ainsi que l'ensemble des études sur les questions de planification linguistique (Kwenzi-Mikala 1990, Dodo-Bounguenza 2002, Ndinga-Koumba-Binza 2005a, 2005b et 2005c), situation linguistique (Mba-Nkoghe 1991, Blanchon 1994, Emejulu et Nzang-Bié 1999), inventaire des langues (Kwenzi-Mikala 1987, 1988 et 1998; Idiata 2002 et 2005), et autres.

Au sujet de l'état de développement, Idiata (2003: 54) est de l'avis suivant: "aussi bien pour la description des langues que pour les travaux de standardisation et de planification, ... le chantier des langues gabonaises est à peine entamé et (qu')il demande encore plus d'énergie et de publications pour arriver à une meilleure connaissance de la situation linguistique du pays." Toutefois, dans son aperçu des travaux réalisés sur les langues gabonaises entre 1975 et 2001, Idiata (2003: 37) note tout de même que "la recherche sur les langues gabonaises paraît assez développée et diversifiée ... on a une idée à peu près claire de la situation de la plupart des langues vernaculaires gabonaises sur les

plans phonétique, phonologique et morphologique". Nonobstant la rareté des travaux en syntaxe, l'on note une floraison des études sociolinguistiques et psycholinguistiques.

Le contexte de parution du LPFFP est donc une période d'efflorescence notoire de la linguistique gabonaise. Les résultats desdits travaux doivent nécessairement relever l'état de développement et le statut social des langues. L'observation est que le LPFFP fait effectivement suite à de nombreux travaux sur la langue pove, mais il y a des réserves quant à son système d'écriture sur lequel nous revenons un peu plus loin.

Parlant des avancées linguistiques, il est sans conteste que le pove est l'une des langues qui connaît de nombreux travaux. Mickala Manfoumbi en est lui-même l'un des spécialistes. Quels sont les travaux sur le pove? L'on peut citer entre autres Van der Veen (1986, 1987 et 1990) et Mickala Manfoumbi (1994). Mickala Manfoumbi en compte aussi près d'une demi-dizaine encore en voie de publication si l'on s'en tient à la note en couverture intérieure du LPFFP.

Comme dans toutes les langues gabonaises, le niveau de développement de ces langues ne reflète pas toujours l'abondance des publications et travaux dans lesdites langues. L'exemple le plus évident est celui de l'écriture. Malgré les multiples propositions, de Raponda-Walker (1932) à la Session de Concertation de 1999 (Idiata 2002), la matérialisation de l'écriture de ces langues se fait toujours attendre tout comme l'adoption officielle de l'une de ces propositions par les instances gouvernementales.

Par ailleurs, il est sans conteste que la lexicographie gabonaise est récente (Ndinga-Koumba-Binza 2005b: 136), quand bien même la production dictionnaire pour les langues gabonaises remonte aux travaux des missionnaires catholiques et protestants et autres administrateurs coloniaux (Nyangone Assam et Mavoungou 2000, Mihindou 2001). Ndinga-Koumba-Binza (2005b: 136) souligne que le plan lexicographique du Gabon — ou planification stratégique, selon Emejulu (2003: 205) — est encore à son stade embryonnaire. Il faut cependant noter que malgré sa récente institution en tant que discipline (Emejulu 2000, 2001a, 2002a, 2003; Mavoungou 2001a), la lexicographie gabonaise connaît un épanouissement riche et rapide quand l'on s'en tient au nombre et au niveau des publications régulières (Emejulu 2001b et 2002b). La recherche actuelle en lexicographie gabonaise s'articule de plus en plus dans un cadre métalexigraphique en ce qui concerne la planification des dictionnaires (Mavoungou 2002a et 2002b, Nzang-Bié 2002, Afane Otsaga 2004), l'amélioration des dictionnaires existants (Nyangone Assam et Mavoungou 2000, Mavoungou 2001b, Ondo-Mebiame 2005, Mabika Mbokou 2006a), la lexicographie pédagogique et enseignement des langues (Ella 2002, Mabika Mbokou 2001 et 2006b, Nyangone Assam 2002 et 2006), la contribution de la lexicographie au développement des langues (Afane Otsaga 2002, Ekwa Ebanega 2002) ainsi que la confection des lexiques spécialisés (Mihindou 2002 et 2006).

La célébration cette année 2006 des troisième, quatrième et cinquième docteurs gabonais en lexicographie dans ce pays dont la population atteint à peine un million et demi, est à suffisance une preuve d'une discipline en plein essor

dans ce pays. Il ne reste plus qu'aux acteurs de cette discipline d'allier l'épanouissement de la discipline sur le plan systématique et épistémologique au développement de la production dictionnaire.

Pour ce qui est de la production dictionnaire spécifiquement, Ndinga-Koumba-Binza (2005b: 138) relève qu'elle fait montre de deux ères distinctes au Gabon: l'ère ancienne et l'ère moderne. La première regroupe les travaux de référence produits par les missionnaires et administrateurs coloniaux depuis 1800 jusque dans les années 1960. La dernière, qui coïncide avec la présentation de la première thèse de doctorat en lexicographie par un gabonais (Mavoungou 2002a), commence en 2002 avec la publication du dictionnaire *geviya* de Van der Veen et Bodinga-bwa-Bodinga (2002). L'on observe que le LPFFP est certes le premier ouvrage de référence dans la langue pove et le deuxième dans l'ère moderne de la production dictionnaire au Gabon, mais il n'est pas le premier dans ce groupe des langues B30 (okande-tsogo). Ce groupe est le seul qui connaît une constante production dans la deuxième ère de la lexicographie gabonaise (cf. les dictionnaires *tsogo* en cours de publications aux Éditions Raponda-Walker). Ce groupe a également bénéficié de nombreux travaux divers sur la langue *tsogo* par Raponda-Walker (1910 et 1996) et sur la langue *geviya* par Bodinga-bwa-Bodinga (1969), par Bodinga-bwa-Bodinga et Van der Veen (1993 et 1995) et par Van der Veen (1991a, 1991b, 1999a, 1999b et 2001). Sur le groupe lui-même de manière générale, il y a aussi les quelques autres travaux de Van der Veen (1988, 1989, 1991c et 2003) et de Jacquot (1983).

Nous sommes de l'avis selon lequel les ouvrages de référence (lexiques, dictionnaires, encyclopédies, etc.) dans une langue doivent être la somme et la finalité des avancées aussi bien linguistiques que métalexigraphiques dans la langue étudiée. Autrement dit, les grammaires, alphabets et autres sont des préalables aux dictionnaires. De même, dans le contexte actuel de la lexicographie moderne, l'étude métalexigraphique s'avère primordiale à la confection d'un ouvrage de référence. Il s'avère également que dans le cadre d'un dictionnaire de langue (monolingue, bilingue, etc.), l'investigation métalexigraphique tient compte des spécificités (grammaire, prononciation, variations dialectales, système d'écriture, etc.) de la langue étudiée.

De notre point de vue, le LPFFP paraît à une période cruciale du développement de la langue pove, même si jusqu'à Jacquot (1983) et Van der Veen (1986 et 1987) le pove "a été rarement retenu comme sujet d'étude" (Mickala Manfoumbi 2004: 19). Bon nombre de travaux sont en cours, mais les quelques travaux publiés peuvent permettre le contenu linguistique d'un dictionnaire. Toutefois, le LPFFP n'a pas fait l'objet d'une réflexion métalexigraphique préalable. Celle-ci aurait fait éviter un certain nombre des réserves relevées par Mavoungou (2005).

2.3 L'aspect éditorial

L'aspect éditorial du contexte de parution du LPFFP se présente premièrement par l'éditeur de l'ouvrage, les Éditions Raponda-Walker, un organe de la Fon-

dation Raponda-Walker pour la Science et la Culture (en abrégé FRW dans le reste de l'article). La FRW, une organisation non gouvernementale, est un acteur majeur de la vie scientifique et culturelle au Gabon (animation, conférences, édition, etc.). Par son département d'édition, ainsi que le souligne Mavoungou (2005: 80), la FRW a publié "de façon extensive dans le domaine de la littérature, de la linguistique, de la lexicographie, etc.". Ce qui lui donne une certaine notoriété non seulement dans le monde gabonais de l'édition, mais aussi dans le système éducatif gabonais (Ndinga-Koumba-Binza 2005a: 74).

Sur le plan spécifique de la lexicographie, la FRW contient une certaine expérience par la publication et/ou la co-édition d'un certain nombre de dictionnaires (Raponda-Walker 1993a, 1995a et 1995b) et des ouvrages à caractère lexicographiques (Raponda-Walker et Sillans 1996, Kwenzi-Mikala 1996, Abaga Ondo 1998, Hubert 1998, Ngou 2003, etc.). Elle a également fait publier chez d'autres éditeurs de nombreux ouvrages de référence et dictionnaires (Raponda-Walker 1993b, 1993c et 1993d).

Le constat est que l'ensemble des dictionnaires édités ou co-édités précédemment sont des rééditions (pour la plupart avec amélioration) des œuvres d'André Raponda-Walker, et sur les parlers dans lesquels ce dernier s'était énormément investi, à savoir le mpongwe et le tsogo. Le LPFFP est le premier dictionnaire propre aux Éditions Raponda-Walker qui sort des langues privilégiées d'André Raponda-Walker (même si le tsogo et le pove sont du même groupe linguistique).

Il faut également noter que le LPFFP est ce que l'on pourrait qualifier de pur produit local. Il est en effet le premier dictionnaire de langue gabonaise publié à Libreville par une maison d'édition locale. La FRW est à l'heure actuelle l'unique éditeur gabonais à publier des dictionnaires.

Le LPFFP se présente ainsi comme l'inauguration d'une ère nouvelle de l'édition dictionnaire non seulement pour la FRW, mais également pour le monde gabonais de l'édition.

2.4 L'auteur: profession linguiste

La présente sous-section vise la présentation de la profession et la formation de l'auteur du LPFFP comme élément du contexte. Elle permet de comprendre les caractéristiques du LPFFP en ce qui concerne aussi bien son contenu que son intitulé.

En dehors de sa formation et profession de linguiste, Mickala Manfoumbi ne diffère pas trop de ses prédécesseurs dans la compilation des dictionnaires des langues gabonaises, à savoir les missionnaires catholiques et protestants et les administrateurs coloniaux. Ils ont en effet en commun la méconnaissance de la métalxicographie et de la lexicographie en tant que discipline scientifique. Ils partagent également le cadre épistémologique, celui de la description et exposition des langues et cultures des peuples du Gabon. C'est ainsi une épistémologie qui navigue entre la linguistique, la lexicographie et l'anthropologie sans trop faire de distinction entre ces disciplines pourtant si différentes.

Par la publication du LPFFP, Mickala Manfoumbi se démarque de son domaine de prédilection, à savoir la description systématique des langues, pour s'initier à la production dictionnaire, ce qui justifie le peu de considération métalexigraphique dans le contenu du LPFFP. S'il est vrai que la pratique des dictionnaires préexiste à la métalexigraphie (Gouws 1989: 15-17, 23-34; 2001: 59; et Mihindou 2001: 7), il est aussi vrai que les pionniers dans la production dictionnaire n'ont pas toujours la formation théorique adéquate. Certains sont des linguistes formés, mais la plupart a souvent été des linguistes autodidactes ou des anthropologues.

Le cas spécifique de Mickala Manfoumbi est qu'en cette ère de floraison de la lexicographie gabonaise, il n'est pas le premier des linguistes du Gabon à s'aventurer dans la discipline lexicographique, mais probablement le seul à avoir publié un dictionnaire de langue. Le chef de file de la lexicographie gabonaise, le Professeur Emejulu, est lui-même d'abord un linguiste et sémioticien. D'autres linguistes gabonais qui produisent en lexicographie sont Nzang-Bié (2002), Ondo-Mebiame (2005) et Ndinga-Koumba-Binza (2005b). Il faut aussi noter que l'ensemble des lexicographes gabonais, titulaires de doctorat en lexicographie, ont la linguistique comme formation de base².

Il est fort probable que la profession et la formation de l'auteur n'ont pas seulement influencé la macrostructure et la microstructure du LPFFP, mais aussi certaines questions de forme.

3. Questions de forme

Les questions de forme que suscite la lecture du LPFFP comprennent, (a) l'identité de lexique donnée à l'ouvrage et non celui de dictionnaire; (b) l'examen de l'alphabet et du système d'orthographe en usage dans le LPFFP; et (c) le système d'écriture emprunté dans l'ouvrage de Mickala Manfoumbi.

3.1 Lexique ou dictionnaire

Le LPFFP est-il un lexique ou un dictionnaire? C'est la question à laquelle la présente sous-section va tenter de répondre en partant des approches définitoires des notions de *lexique* et *dictionnaire*, et en examinant les caractéristiques du LPFFP.

Le terme de lexique est polysémique selon l'approche d'usage. De nombreux auteurs proposent quatre sens (Crystal 1999: 197-198; Matthews 1997: 207-208; et Richards et Schmidt 2002: 307-308). Mais trois définitions seulement peuvent renvoyer au LPFFP, à savoir:

- (a) celui qui marque une synonymie avec le terme de *dictionnaire*, malgré la précision de Richards et Schmidt (2002: 308) selon laquelle le terme serait plus usité pour les langues mortes tels que le latin et le grec;
- (b) l'ensemble de tous les mots et tournures idiomatiques d'une langue; et

- (c) en grammaire générative, la composante de base que forme l'ensemble des mots et phrases ainsi que les informations au sujet de ces mots et phrases.

Sur un tout autre plan, et ce qui peut paraître comme la définition lexicographique du terme (Emejulu 2001b: 230), le lexique est tout simplement l'analyse scientifique du vocabulaire d'une langue donnée. Ce qui le rapproche du premier sens donné par Crystal (1999: 197), à savoir l'ensemble du vocabulaire d'une langue, ses items lexicaux ou lexèmes, spécialement quand ceux-ci sont listés dans le cadre d'un dictionnaire comme l'ensemble des entrées lexicales.

Par ailleurs, le dictionnaire est entendu en lexicographie comme un ouvrage de référence qui liste par ordre alphabétique les mots d'une ou plusieurs langues avec des informations sur leur orthographe, prononciation, catégorie grammaticale, sens, histoire, et usage (Crystal 1999: 197). À ce titre, le dictionnaire se distingue nettement du lexique, du glossaire ou du thésaurus.

Hormis les réserves relevées par Mavoungou (2005), l'ouvrage de Mickala Manfoumbi a tout d'un dictionnaire. Son titre de *Lexique* se comprend par la nuance selon laquelle l'ouvrage ne remplit pas l'ensemble des conditions méta-lexicographiques d'un dictionnaire proprement dit.

Malgré les diverses définitions de la notion de lexique, le premier ouvrage de référence pove n'est pas un simple lexique même si l'auteur modestement l'intitule ainsi. Cependant, il est regrettable que l'auteur n'ait rien dit à ce propos dans les prétextes.

3.2 Les traits encyclopédiques du LPFFP

Quoique partagé entre dictionnaire et lexique, le LPFFP présente quelques traits que l'on peut assimiler aux traits encyclopédiques. En effet, à la question de savoir si le LPFFP est un lexique ou un dictionnaire, il s'ajoute la question de savoir le type de dictionnaire quand on reconnaît le LPFFP comme étant un dictionnaire.

Le LPFFP est plus ou moins un dictionnaire encyclopédique en l'observant de près. Il faut noter que le dictionnaire encyclopédique partage quelques spécificités avec l'encyclopédie, notamment la mise en exergue d'un certain nombre des caractéristiques extralinguistiques du lexique des langues. Nous allons relever ici deux traits encyclopédiques du LPFFP.

La première caractéristique encyclopédique du LPFFP se trouve dans ses prétextes dans lesquels l'auteur présente le peuple pove dans sa situation géographique, ses origines et migrations, son organisation sociale et ses sociétés secrètes. Sur la langue elle-même, l'auteur revient sur l'état des recherches et sa classification ainsi que sur un aperçu du système linguistique (tableaux phonétiques et phonologiques, distribution des allophones et présentation des classificateurs). Dans les prétextes également, l'on note une abondance d'images et cartes géographiques. Toutes les images sont en couleur très nette.

La deuxième caractéristique encyclopédique est dans le texte même du dictionnaire. Elle se manifeste au travers des descriptions des arbres et plantes reprises à Raponda-Walker et Sillans (1996). Un certain nombre des descriptions des plantes comme des animaux et des objets culturels sont illustrés par des images en noir et blanc. Le lecteur peut également trouver dans l'explication de l'équivalence quelques notes historiques et étymologiques de certains items lexicaux. L'on note par exemple:

- | | | | |
|-----|---------------|-------------------------|--|
| (1) | adyé | adyé | Serment (Je jure!)
Ce serment serait-il un emprunt du français « à Dieu »? Nous ne pouvons l'affirmer. Toutefois, le même serment existe dans d'autres langues du Gabon, le punu par exemple. |
| (2) | +bwèzè | 5,6
bweze
mabweze | antichambre du
bwiti (nzimba)
Cet endroit sert à la cérémonie des rites de passage. Il est situé en pleine forêt, loin du village. C'est l'endroit des délibérations, des séances secrètes où l'on conduit les récipiendaires. |

3.3 Alphabet et orthographe

Nous avons mentionné plus haut que les langues gabonaises ont déjà fait l'objet de plusieurs propositions en ce qui concerne l'alphabet et l'orthographe en vue de leur écriture. En outre, la confection d'un dictionnaire nécessite aussi bien un alphabet qu'un système d'orthographe.

Le passage au Gabon des langues parlées aux langues écrites remonte à l'époque coloniale³. De cette époque, il n'est mention d'aucune proposition alphabétique ou orthographique de la langue pove. Le premier alphabet proposé pour l'écriture du pove est celui de Van der Veen (1990) dans le cadre de l'alphabet scientifique des langues gabonaises (ASG).

La proposition de Van der Veen (1990) est un système de 27 symboles dont sept voyelles (**a, e, ε, i, o, ɔ** et **u**) et deux semi-voyelles (**w** et **y**). Les consonnes sont les suivantes: **b, d, dy, γ, k, l, m, mb, n, nd, ng, ny, nz, p, s, t, ts**, et **β**. Van der Veen (1990: 194) propose également d'orthographier les trois tons pertinents du pove, à savoir haut (H), bas (B) et descendant (HB). Cette proposition ne fait aucune mention ni de la longueur vocalique ni de la gémination, mais reconnaît dans le texte d'illustration proposé la marque de la majuscule à l'initiale des phrases et quelques signes de ponctuation (la virgule, les trois points de suspension et le point final). Le rapprochement entre la proposition d'alphabet et le système phonologique montre que Van der Veen (1990) a opté pour un alphabet phonologique pour l'orthographe du pove. Cela paraît être un système simple et fonctionnel assez éloigné des détails phonétiques.

Avec la publication du LPFFP, Mickala Manfoumbi (2004) présente un nouvel alphabet de la langue pove. Celui-ci diffère de celui de Van der Veen (1990) particulièrement au niveau des consonnes. Il en propose 17 au lieu de 18 en supprimant le digraphe **dy**. Il préfère les signes **g** et **v** respectifs au lieu du signe de la constrictive vélaire [**ɣ**] en variation complémentaire avec l'occlusive vélaire [**g**] (Mickala Manfoumbi 2004: 30) et du signe de la constrictive bilabiale [**β**]. Le signe pour [**β**] est /**v**/ dans le tableau phonologique (Mickala Manfoumbi 2004: 29, 31). De même, le digraphe **ny** est remplacé par le signe même du son qu'il représente **ɲ**, i.e. la nasale palatale.

L'alphabet de Mickala Manfoumbi paraît plus phonologique que celui de Van der Veen en ce qui concerne les consonnes. On observe dans les lemmes la notation systématique du ton haut. Le ton bas est noté dans le découpage morphologique, mais disparaît dans les formes phonologiques des mots en troisième colonne. Ce qui permet de comprendre que le système d'orthographe de Mickala Manfoumbi pour l'écriture du pove admet seulement la notation du ton haut. Comme chez Van der Veen (1990), il n'y a aucune trace de longueur vocale (dénuée de toute pertinence phonologique) dans l'orthographe.

Par ailleurs, la ponctuation est quasiment absente. Les lemmes mêmes ne commencent pas par une majuscule. La microstructure manque de phrases et autres textes illustratifs qui nous auraient donné des informations sur le système de ponctuation ou sur la division du mot (écriture disjonctive ou écriture conjonctive) choisie par Mickala Manfoumbi pour l'écriture de la langue pove. Van der Veen (1990) en son temps avait opté pour l'écriture disjonctive comme ci-dessous dans (3).

- (3) Tàbà dí β̀̀lé, β̀̀lé, β̀̀lé ... yú á èbù
Les moutons s'y introduirent pour s'y restaurer.
(Van der Veen 1990: 194)

4. Le LPFFP et la planification stratégique de la lexicographie gabonaise

La situation du LPFFP dans la planification stratégique de la lexicographie gabonaise se présente non seulement par sa contribution à l'essor de cette lexicographie, mais aussi par son impact sur les plans sociolinguistiques et économiques.

Il faut tout de même préciser que le plan stratégique de la lexicographie gabonaise, encore à un stade embryonnaire, est à cinq phases de son développement (Ndinga-Koumba-Binza 2005b: 136), à savoir:

- (a) la formation lexicographique,
- (b) l'adaptation métalexigraphique des dictionnaires,
- (c) la définition de la lexicographie comme carrière,
- (d) la formulation d'un cadre général cohérent de la lexicographie gabonaise, et

- (e) la mise à jour des ouvrages de référence hérités des missionnaires et administratifs coloniaux.

Le LPFFP paraît inadéquat à chacune de ces phases, à moins de suggérer que par cette publication son auteur dénote une volonté d'embrasser une carrière lexicographique.

Nous avons souligné plus haut que le LPFFP appartient à l'ère moderne de la production dictionnaire au sein de la lexicographie gabonaise. C'est déjà une importante contribution que d'enrichir cette ère d'un ouvrage de référence de plus. Sur le plan métalexographique, selon les réserves de Mavoungou (2005), le LPFFP suscite une réflexion dont les conclusions pourraient apporter de nouvelles orientations au sein de la lexicographie gabonaise.

L'impact sociolinguistique du LPFFP est double. C'est d'abord le fait d'apporter un complément lexicographique aux ouvrages existants sur la langue pove. Nous avons évoqué plus haut l'ensemble des travaux publiés sur le pove. De ces travaux, il manquait bel et bien un ouvrage de référence. C'est donc une lacune que le LPFFP vient combler.

Le LPFFP relance également tout le débat sur les défis de la linguistique gabonaise, défis auxquels la lexicographie gabonaise ne peut faire fi. Elle doit d'une manière ou d'une autre y apporter une contribution (cf. Ndinga-Koumba-Binza 2005b: 139-144). Il s'agit notamment de la place des langues locales dans la politique linguistique nationale, de la standardisation de ces langues ainsi que de l'adoption finale d'un système d'orthographe pour l'écriture de ces langues.

Toutefois, sur le plan sociolinguistique, le LPFFP constitue un véritable facteur d'intégration du peuple pove quand à sa valeur nationale et internationale définie dans la préface:

Le peuple pove n'est rien sans cette mère commune qu'est la nation, sans cette mer sans limites qu'est l'humanité ... Puisse le lecteur trouver l'occasion de venir à la rencontre de la culture gabonaise, lovée ici dans le canton Lolo-Wagna. (Nzouba-Ndama 2004: 12).

La place du peuple pove dans la nation gabonaise comme dans le monde pose tout un problème d'impact réel au LPFFP en ce qui concerne non seulement l'intérêt scientifique, mais aussi les avantages économiques. Le problème se pose avec le nombre de locuteurs.

Il est en effet sans conteste que comparé à un nombre réduit d'autres langues, le pove est une langue en voie d'extinction. L'intérêt scientifique d'un travail quelconque dans une telle langue ne peut nullement être la génération des atouts économiques et commerciaux pour un développement national comme le prône la planification stratégique de la lexicographie gabonaise (Emejulu 2000: 65), mais plutôt la sauvegarde des faits de la langue pour des visées historiques et culturelles.

5. Conclusion

Le dictionnaire est un excellent moyen pour la promotion et la sauvegarde des langues. De même, l'introduction des langues gabonaises dans le système éducatif passera également par la mise en place d'outils adéquats tels que les dictionnaires, les grammaires, etc. Ainsi, les langues gabonaises pourraient connaître une importante accélération dans le processus de leur développement si la linguistique et la lexicographie gabonaises montraient un effort orienté vers la production systématique des ouvrages de référence.

Ainsi, la parution du LPFFP est un apport incontestable dans le cadre du processus de développement de la langue pove. Cet article s'est fixé comme objectif d'apporter un complément au compte-rendu de Mavoungou (2005) du LPFFP. Les questions évoquées ici sont principalement sur le contexte de parution, et sur la forme et sur la contribution du LPFFP à la planification stratégique de la lexicographie gabonaise. Bien que le développement de chaque question n'ait pas été exhaustif, le LPFFP tout comme la lexicographie gabonaise pourrait tirer profit des arguments soutenus dans la présente communication.

Le présent article recommande une seconde édition du LPFFP tenant compte aussi bien des réserves de Mavoungou (2005) que des circonspections mentionnées plus haut. Il est également recommandé de le réintituler selon son caractère de dictionnaire.

Notes

1. Observations personnelles basées sur des discussions avec quelques locuteurs pove et tsogo. Toutefois, une étude systématique sur ces questions reste à faire. Nos sincères remerciements à Gino Boussiengui-Boussiengui et à Marcel Ngouama pour les discussions enrichissantes à ce sujet.
2. Titulaire d'une maîtrise en Lettres Modernes (option Littérature Africaine d'expression française), le docteur Blanche Nyangone Assam fait l'exception parmi les lexicographes gabonais.
3. Ndinga-Koumba-Binza (à paraître) a proposé une approche historique des alphabets et de l'écriture depuis les origines au cas spécifique des langues gabonaises. Certaines analyses dans le présent volume viennent de ce texte en cours de publication.

Références

Dictionnaires

- Crystal, David. 1999. *The Penguin Dictionary of Language*. Harmondsworth: Penguin Books.
- Matthews, Peter H. 1997. *The Concise Oxford Dictionary of Linguistics*. Oxford/New York: Oxford University Press.
- Mickala Manfoumbi, Roger. 2004. *Lexique pove-français/français-pove*. Libreville: Éditions Raponda-Walker.

- Raponda-Walker, André.** 1993a. *Dictionnaire étymologique des noms propres gabonais*. Versailles: Classiques Africains.
- Raponda-Walker, André.** 1995a. *Dictionnaire mpongwe-français*. Versailles/Libreville: Classiques Africains/Éditions Raponda-Walker
- Raponda-Walker, André.** 1995b. *Dictionnaire français-mpongwe*. Versailles/Libreville: Classiques Africains/Éditions Raponda-Walker.
- Richards, Jack C. et Richard Schmidt.** 2002³. *Longman Dictionary of Language Teaching and Applied Linguistics*. Londres/New York: Longman.
- Van der Veen, Lolke J. et Sébastien Bodinga-bwa-Bodinga.** 2002. *Gedandedi sa geviya/Dictionnaire geviya-français*. Louvain/Paris/Sterling, Virginia: Peeters.

Autres sources

- Abaga Ondo, Federico.** 1998. *Proverbios fang (bilingue espagnol)*. Libreville: Éditions Raponda-Walker.
- Afane Otsaga, Thierry.** 2002. Le rôle du dictionnaire dans le développement des langues gabonaises. Emejulu, J.D. (Éd.). 2002: 206-229.
- Afane Otsaga, Thierry.** 2004. *The Standard Translation Dictionary as an Instrument in the Standardisation of Fang*. Thèse de doctorat non publiée. Stellenbosch: Université de Stellenbosch.
- Blanchon, Jean A.** 1994. Gabon: Language Situation. Asher, R.E. (Éd.). 1994. *The Encyclopedia of Language and Linguistics. Volume 3*. New York/Oxford: Pergamon Press.
- Bodinga-bwa-Bodinga, Sébastien.** 1969. *Traditions orales de la race eviya*. Paris: T.M.T.
- Bodinga-bwa-Bodinga, Sébastien et Lolke J. van der Veen.** 1993. Plantes utiles des Eviya: Pharmacopée. *Pholia. Revue du Laboratoire de Phonétique et Linguistique Africaine* 8: 27-66.
- Bodinga-bwa-Bodinga, Sébastien et Lolke J. van der Veen.** 1995. *Les proverbes Evia et le monde animal. La communauté traditionnelle evia à travers ses expressions proverbiales (Gabon)*. Paris: Éditions L'Harmattan.
- Dodo-Bounguenza, Eric.** 2002. Les langues du Gabon: le point en 2001. *Le Gabon et l'Unesco, Revue d'Information* 2: 33-39.
- Ekwa Ebanéga, Guy-Modeste.** 2002. Le corpus lexicographique dans le développement des langues gabonaises. Emejulu, J.D. (Éd.). 2002: 188-205.
- Ella, Edgard Maillard.** 2002. Un dictionnaire scolaire de langue de spécialité. Emejulu, J.D. (Éd.). 2002: 305-325.
- Emejulu, James D.** 2000. Lexicography, an economic asset in multilingual Gabon. *Revue Gabonaise des Sciences du Langage/Gabonese Journal of Language Sciences* 1: 51-69.
- Emejulu, James D.** 2001a. Lexicographie multilingue et multisectorielle au Gabon: planification, stratégie et enjeux. Emejulu, J.D. (Éd.). 2001: 38-57.
- Emejulu, James D. (Éd.).** 2001b. *Éléments de lexicographie gabonaise. Tome I*. New York: Jimacs-Hillman Publishers.
- Emejulu, James D.** 2002a. Défis et promesses de la lexicographie intégrale dans les pays en développement. Emejulu, J.D. (Éd.). 2002: 366-381.
- Emejulu, James D. (Éd.).** 2002b. *Éléments de lexicographie gabonaise. Tome II*. New York: Jimacs-Hillman Publishers.
- Emejulu, James D.** 2003. Challenges and Promises of a Comprehensive Lexicography in the Developing World: The Case of Gabon. Botha, W.F. (Éd.). 2003. *'n Man wat beur. Huldigingsbundel*

- vir Dirk van Schalkwyk*: 195-212. Stellenbosch: Bureau du WAT.
- Emejulu, James D. et Yolande Nzang-Bié.** 1999. *Linguistic Perspectives in Gabon*. Communication lue à la rencontre internationale de la Société Internationale de Linguistique, Université de Dakota du Nord, Grand Forks, États-Unis, 20 juillet 1999.
- Gouws, Rufus H.** 1989. *Leksikografie*. Le Cap: Academia.
- Gouws, Rufus H.** 2001. Lexicographic Training: Approaches and Topics/Formation lexicographique: approches et thèmes. Emejulu, J.D. (Éd.). 2001: 58-94.
- Hubert, Jacques.** 1998. *Album du 150^e anniversaire de l'église catholique du Gabon*. Libreville: Éditions Raponda-Walker.
- Idiata, Daniel Franck.** 2002. *Il était une fois les langues gabonaises*. Libreville: Éditions Raponda-Walker.
- Idiata, Daniel Franck.** 2003. *Pourquoi le Gabon doit investir sur ses langues vernaculaires*. Le Cap: The Centre for Advanced Studies of African Society (CASAS).
- Idiata, Daniel Franck.** 2005. *Les langues du Gabon: Données en vue d'une classification fondée sur le critère d'intercompréhension*. Le Cap: The Centre for Advanced Studies of African Society (CASAS).
- Jacquot, André.** 1983. *Les classes nominales dans les langues bantoues des groupes B10, B20, B30 (Gabon-Congo)*. Travaux et Documents de l'ORSTOM 157. Paris: ORSTOM.
- Kwenzi-Mikala, Jérôme T.** 1987. Contribution à l'inventaire des parlers bantu du Gabon. *Pholia. Revue du Laboratoire de Phonétique et Linguistique Africaine* 2: 103-110.
- Kwenzi-Mikala, Jérôme T.** 1988. L'identification des unités-langues bantu gabonaises et leur classification interne. *Muntu. Revue scientifique et culturelle du CICIBA* 8: 54-64.
- Kwenzi-Mikala, Jérôme T.** 1990. Quel avenir pour les langues gabonaises? *Revue Gabonaise des Sciences de l'Homme* 2: 121-124.
- Kwenzi-Mikala, Jérôme T.** 1991. *La politique linguistique des états africains francophones: le cas du Gabon*. Communication lue à la 22^e Conférence Annuelle de Linguistique Africaine (ACAL), Université de Nairobi, Nairobi, Kenya, 15-19 juin 1991.
- Kwenzi-Mikala, Jérôme T.** 1996. *Parémies d'Afrique centrale*. Libreville: Éditions Raponda-Walker.
- Kwenzi-Mikala, Jérôme T.** 1998. Parlers du Gabon: classification du 11.12.97. Raponda-Walker, A. (Éd.). 1998. *Les langues du Gabon*: 127. Libreville: Éditions Raponda-Walker.
- Mabika Mbokou, Ludwine.** 2001. Le rôle du dictionnaire dans le système éducatif. Emejulu, J.D. (Éd.). 2001: 206-222.
- Mabika Mbokou, Ludwine.** 2006a. The *Dictionnaire français-yipounou/yipounou-français* of the Église Évangélique du Sud-Gabon (CMA) 1966. *Lexikos* 16: 104-120.
- Mabika Mbokou, Ludwine.** 2006b. *A Model for the Macro- and Microstructure of a Yipunu-French School Dictionary*. Thèse de doctorat non publiée. Stellenbosch: Université de Stellenbosch.
- Mavoungou, Paul Achille.** 2001a. La mondialisation et la lexicographie trilingue ou plurilingue au Gabon. Emejulu, J.D. (Éd.). 2001: 160-183.
- Mavoungou, Paul Achille.** 2001b. Macro- and Microstructural Issues in Mazuna Lexicography. *Lexikos* 11: 122-138.
- Mavoungou, Paul Achille.** 2002a. *Metalexicographical Criteria for the Compilation of a Trilingual Dictionary: Yilumbu-English-French*. Thèse de doctorat non publiée. Stellenbosch: Université de Stellenbosch.
- Mavoungou, Paul Achille.** 2002b. Vers un dictionnaire du français du Gabon. Emejulu, J.D. (Éd.). 2002: 230-262.

- Mavoungou, Paul Achille.** 2005. Note de lecture sur R. Mickala Manfoumbi: *Lexique Pove–Français/Français–Pove*. *Journal of Education* 4(1): 79-86. Réduit: Mauritius Institute of Education.
- Mba-Nkoghe, Jules.** 1991. Le Gabon linguistique. *Notre Librairie. Littérature gabonaise* 105: 20-23.
- Mickala Manfoumbi, Roger.** 1994. *Essai de grammaire pove: langue bantoue du groupe B30*. Thèse de doctorat. Bruxelles: Université Libre de Bruxelles.
- Mihindou, Guy-Roger.** 2001. Apports des missionnaires à la lexicographie gabonaise: dictionnaires bilingues fang–français/français–fang; français–yipounou/yipounou–français. Emejulu, J.D. (Éd.). 2001: 7-37.
- Mihindou, Guy-Roger.** 2002. Du dictionnaire pour déficients auditifs du Gabon. Emejulu, J.D. (Éd.). 2002: 326-365.
- Mihindou, Guy-Roger.** 2006. *A Theoretical Model for a Yipunu–English–French Explanatory Dictionary of Medical Terms*. Thèse de doctorat non publiée. Stellenbosch: Université de Stellenbosch.
- Ndinga-Koumba-Binza, Hugues Steve.** 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, Hugues Steve.** 2005b. Considering a Lexicographic Plan for Gabon within the Gabonese Language Landscape. *Lexikos* 15: 132-150.
- Ndinga-Koumba-Binza, Hugues Steve.** 2005c. Gabonese Language Landscape: Survey and Perspectives. *South African Journal of African Languages* 25(3).
- Ndinga-Koumba-Binza, Hugues Steve.** À paraître. Alphabet et écriture: approche historique et cas des langues gabonaises. Hubert, J. et P.A. Mavoungou (Éds.). À paraître. *Écriture et standardisation des langues gabonaises*.
- Ngou, Honorine.** 2003. *Pratique de la langue française, pièges et difficultés*. Libreville: Éditions Raponda-Walker.
- Nyangone Assam, Blanche.** 2001. La lexicographie dans l'apprentissage des langues gabonaises. Emejulu, J.D. (Éd.). 2001: 187-205.
- Nyangone Assam, Blanche.** 2002. Le dictionnaire pédagogique scolaire: quelques aspects du traitement lexicographique. Emejulu, J.D. (Éd.). 2002: 292-304.
- Nyangone Assam, Blanche.** 2006. *Dictionaries as Teaching Instruments for Mother Tongue Education: The Case of Fang in Gabon*. Thèse de doctorat non publiée. Stellenbosch: Université de Stellenbosch.
- Nyangone Assam, B. et P.A. Mavoungou.** 2000. Lexicography in Gabon: A Survey. *Lexikos* 10: 252-274.
- Nzang-Bié, Yolande.** 2001. Vers une éducation multilingue au Gabon: première approche. *Revue Gabonaise des Sciences du Langage/Gabonese Journal of Language Sciences* 2: 17-29.
- Nzang-Bié, Yolande.** 2002. Le corpus lexicographique dans les langues à tradition orale: le cas du dialecte fang-mekè. *Lexikos* 12: 111-138.
- Nzouba-Ndama, Guy.** 2004. Préface. Mickala Manfoumbi, Roger. 2004: 11-12.
- Ondo-Mebiame, Pierre.** 2005. De la révision du dictionnaire de Samuel Galley. *Lexikos* 15: 151-163.
- Raponda-Walker, André.** 1910. *Au pays des Isogho: simple récit de voyage*. Messenger du St-Esprit. Réédition. 1993. Libreville: Éditions Raponda-Walker.
- Raponda-Walker, André.** 1932. L'alphabet des idiomes gabonais. *Journal de la Société des Africanistes* 2(2): 139-146. Réédité in Raponda-Walker, André (Éd.). 1998: 7-15.
- Raponda-Walker, André.** 1993b. *3000 proverbes (multilingue)*. Versailles: Classiques Africains.
- Raponda-Walker, André.** 1993c. *1500 proverbes, devises, serments, cris de guerre et devinettes*. Versailles: Classiques Africains.

- Raponda-Walker, André.** 1993d. *Les tribus du Gabon*. Versailles: Classiques Africains.
- Raponda-Walker, André.** 1996. *Éléments de grammaire ghetsoö*. Libreville: Éditions Raponda-Walker.
- Raponda-Walker, André. (Éd.)** 1998. *Les langues du Gabon*. Libreville: Éditions Raponda-Walker.
- Raponda-Walker, André et Roger Sillans.** 1996. *Plantes utiles du Gabon*. Libreville: Sépia/Éditions Raponda-Walker/Centre Culturel Français Saint-Exupéry.
- Van der Veen, Lolke J.** 1986. *Notes en vue d'une description phonologique et morphologique de la langue pouvi (Gabon)*. Mémoire de maîtrise. Lyon: Université Lumière de Lyon (Lyon 2).
- Van der Veen, Lolke J.** 1987. *De l'espace vocalique et des tons en pouvi: notes descriptives*. Mémoire de DEA. Lyon: Université Lumière de Lyon (Lyon 2).
- Van der Veen, Lolke J.** 1988. Caractéristiques principales du groupe B 30 (Gabon). *Pholia. Revue du Laboratoire de Phonétique et Linguistique Africaine* 3: 271-290.
- Van der Veen, Lolke J.** 1989. Doubles réflexes dans quelques langues du groupe Okani (Gabon). *Pholia. Revue du Laboratoire de Phonétique et Linguistique Africaine* 4: 229-243.
- Van der Veen, Lolke J.** 1990. Gevove. *Revue Gabonaise des Sciences de l'Homme* 2: 193-195.
- Van der Veen, Lolke J.** 1991a. *Étude comparée des parlers du groupe Okani, B30 (Gabon)*. Thèse de doctorat. Lyon: Université Lumière de Lyon (Lyon 2).
- Van der Veen, Lolke J.** 1991b. Le système tonal du ge-via (Gabon). *Pholia. Revue du Laboratoire de Phonétique et Linguistique Africaine* 6: 219-257. Republié en 1992 dans *Journal of West African Languages* 22(2): 17-41.
- Van der Veen, Lolke J.** 1991c. Étude dialectométrique et lexicostatistique du groupe B 30 (Gabon). *Pholia. Revue du Laboratoire de Phonétique et Linguistique Africaine* 6: 191-217.
- Van der Veen, Lolke J.** 1999a. *Les Bantous eviya (Gabon-B30): langue et société traditionnelle*. Thèse d'habilitation. Lyon: Université Lumière de Lyon (Lyon 2).
- Van der Veen, Lolke J.** 1999b. La propagation des tons et le statut des indices pronominaux précédant le verbe en geviya (B30). Blanchon, J. et D. Creissels (Éds.). *Issues in Bantu Tonology*: 15-36. *Grammatische Analysen Afrikanischer Sprachen*, Vol. 10. Cologne: Rüdiger Köppe Verlag. Republié dans Muller, C. (Éd.). 2001. *Actes de clitiques et cliticisation*: 443-457. Paris: Honoré Champion.
- Van der Veen, Lolke J.** 2001. The Tone System of Geviya (Bantu B30/Gabon). Communication lue à l'atelier sur la typologie des systèmes prosodiques africains, Université de Bielefeld, Bielefeld, Allemagne, 18-21 mai 2001. Disponible sur <<http://www.spectrum.uni-bielefeld.de/TAPS/proceedings>>.
- Van der Veen, Lolke J.** 2003. The B30 Languages. Nurse, D. et G. Philippson (Éds.). *The Bantu Languages*. Routledge Language Family Series, Vol. 4. Londres/New York: Routledge.

Julie Coleman and Anne McDermott (Editors). *Historical Dictionaries and Historical Dictionary Research: Papers from the International Conference on Historical Lexicography and Lexicology at the University of Leicester, 2002.* 2004, vii + 224 pp. Lexicographica. Series Maior 123. ISBN 3-484-39123-5 (Pb.). Tübingen: Max Niemeyer. Price: €68.

The papers collected in this volume were originally presented at an International Conference on Historical Lexicography and Lexicology organized by Julie Coleman, and meant to fill the gap created by the two International Round Table Conferences in Florence and Leyden in 1971 and 1977 respectively. The purpose of the conference was to bring together scholars and academics working in the field of historical lexicography, whether researching the history of historical dictionaries or whether compiling dictionaries with a historical point of departure. Accordingly the collection, competently edited and introduced by Julie Coleman and Anne McDermott, is arranged in two sections, reflecting the distinction between those researchers tracing the historical development of dictionaries and those practitioners faced with the problems and challenges of the compilation of historical dictionaries. The first part dealing with the history of dictionaries consists of twelve papers (two-thirds of the book), while the second part dealing with the practice of historical dictionaries consists of six papers (one-third of the book).

Although the twelve papers in the first section, with the exception of four, all discuss English dictionaries, the characteristics of these dictionaries revealed in the discussions have enough generalizing value also to be applicable and comparable to similar types in languages other than English.

These papers are headed by John Considine's contribution: "Du Cange: Lexicography and the Medieval Heritage". After starting off with a rehearsal of the life of Du Cange and the publication histories of his dictionaries, Considine discusses the claim that Du Cange's work with its Latinist and Byzantinist bias was essentially a patriotic, Francocentric enterprise. Considine shows how Du Cange's account of medieval Latin, the *Glossarium latinitatis*, illuminated not only the history of French institutions, but also the origins of the French language itself. Furthermore, his research on Byzantine Greek, which culminated in the *Glossarium graecitatis*, forms part of a network of historical scholarship and polemic whose tendency, and even design, was to bring the Byzantine heritage to France. Considine subsequently advances the idea that the case of Du Cange provides support for the argument that historical lexicography may often or even always lie in the reconstruction of a lost cultural heritage or the rebuilding of a lost cultural homeland, which can only be remembered and mapped as well as possible in a dictionary. This argument makes the paper of Considine a fitting introduction to the following eleven articles on the history of dictionaries.

The next two papers, that of Reiko Takeda and that of Ian Lancashire, are thematically connected, both dealing with manuscript dictionaries, Takeda with

a single manuscript and Lancashire with a whole collection of manuscripts. In "Cambridge, Trinity College Library MS 0.5.4: A Fifteenth-century Pedagogical Dictionary?", Takeda sketches the background features and significance of the manuscript and especially the complete dictionary included in it. The manuscript is a comprehensive manual, compiled and produced specifically for the teaching and learning of Latin. In the dictionary, however, some English definitions appear alongside the Latin, but some entries contain English descriptions only, indicating a departure from the practice of glossing Latin texts. This shows a shift towards the use of English in teaching Latin. In "Lexicography in the Modern English Period: The Manuscript Record", Lancashire gives an overview of the manuscript record between 1475 and 1700 in the web-based corpus of several hundred lexical texts, The Lexicons of Early Modern English (LEME). There are among others ten types of manuscript collections that need attention, bilingual dictionaries, glossaries and topical vocabularies, prose treatises and grammars, books annotated by contemporaries, reference works on proper and place names and monolingual glossaries, the wide range already indicating the reasons for the realization and compilation of these manuscripts. These manuscripts from the early modern period exist for various reasons, for instance, they are late survivals of medieval glossaries, antiquarian research papers, personal study aids, and also works for private consumption. Lexicons are valuable not only for the recovery of the meanings of words, but also as instruments of religious, political and social power. The founding of the Elizabethan Society of Antiquaries marked the mid-1580s as a watershed in English speakers' lexical self-awareness.

Olga Karpova's paper "Author's Lexicography with Special Reference to Shakespeare's Dictionaries" states that English author lexicography represents the largest groups of reference works. As all these concordances differ in the manner in which the material is presented, the number and variety of information categories and the type of notational devices, she uses Shakespeare's dictionaries as the most comprehensive group to trace the peculiarities of their development. She concludes by showing that this group of dictionaries has three main features: corpus, citation and label, each of which can be complete or differential. Such a classification makes it possible not only to trace the main trends in the history of authors' concordances, but also to typify other kinds of writers' dictionaries, taking into consideration their own information categories. Additional advantages of these concordances are that they could be used as reliable sources for authors' quotations in the compilation of other types of reference works and that their limited number of information categories makes the complete inclusion of all words from an author's works possible.

In "An Analysis of a Seventeenth Century Conceptual Dictionary with an Alphabetical List of Entries and a Network Definition Structure: John Wilkens' and William Lloyd's *An Alphabetical Dictionary* (1668)", Natascia Leonardi describes the basic dependence of the dictionary as a by-product of the more comprehensive plan of the Tables of the *Universal Philosophy*, which form the

second of the four parts of Wilkens' *Essay towards a Real Character and a Philosophical Language* (1668). By focusing on the concept 'science' and the related word 'knowledge', she shows that the words in the *Alphabetical Dictionary* are defined by reference to the concepts contained in the Tables of the *Universal Philosophy* and in this way positioned within the hierarchical structure of the classificatory system of the *Essay*. While the hierarchical structure gives information about the elements contained in it, the wider relational network provided by the Tables specifies how an item relates to and differs from the other elements included in the same conceptual semantic field. Because the main concern of the *Alphabetical Dictionary* is not so much the form of natural language as the communicative effectiveness of language, the work displays the semantic value of lexical units rather than the formal properties of the lexemes.

The next three papers by Rowena Fowler, Joan Beal and Julie Coleman discuss individual dictionaries of a specific kind published in the eighteenth and nineteenth centuries. In "Text and Meaning in Richardson's Dictionary", Fowler discusses Charles Richardson's *New Dictionary of the English Language* (1836–1837) which, without any definition field, relies almost entirely on quotations to establish and demonstrate meaning. These quotations, arranged chronologically within four broad historical periods, illustrate all the derivatives of a headword, making no distinction between primary and transferred meanings. This allows the user to rethink the value and application of quotations as the source, establishment and illustration of meaning, and the selection, ordering and contextualization of quotations in a dictionary. Beal's paper "An Autodidact's Lexicon: Thomas Spence's *Grand Repository of the English Language* (1775)", deals with a pronouncing dictionary in which each entry is respelt in a phonetic script of the compiler's own devising, meant for those lacking the possibilities or means for normal schooling to learn to read more easily. Beal's intention is to consider the sources of Spence's words. She argues that Spence, unlike other compilers of pronouncing dictionaries who leaned heavily on Samuel Johnson's *Dictionary of the English Language*, did not simply attach his respellings to Johnson's definitions, but shows independence both in the words chosen and in some of the definitions of words also found in Johnson. The most likely explanation for the mixture of words in the *Grand Repository* is that Spence used a short dictionary as basis for his work, one that either depended on Johnson, or, more likely, one that shared sources with Johnson. In "The Third Edition of Grose's *Classical Dictionary of the Vulgar Tongue: Bookseller's Hackwork or Posthumous Masterpiece?*", Coleman considers the evidence of Eric Partridge's claim in the foreword to the 1931 reprint of the third edition from 1796 of Francis Grose's dictionary that, because no additional editor's name is given, this edition, published after Grose's death in 1791, should be seen as the best representation of Grose's own work. After a careful examination of Grose's working method in the first two editions of his dictionary, Coleman concludes that, although there is circumstantial evidence to support Partridge's claim that some of the new material may well be from Grose's

notes, other additions and the revision of existing entries were made by his editor. The evidence is strong to believe that the publishers Hooper and Wigstead hoped that potential purchasers would, as Partridge did, accept the third corrected and enlarged edition as Grose's own authoritative work.

In his contribution "Roget's Thesaurus, Deconstructed", Werner Hüllen sets out to solve the problem of defining the dictionary type to which Roget's work belongs. Hüllen starts with two assertions, namely that Roget's *Thesaurus* is (a) a cumulative dictionary of synonyms and (b) a topical (or onomasiological) dictionary, into which he integrates two other assertions: (c) both dictionary types mentioned in (a) and (b) had their own history and popularity before Roget, and (d) the new and seminal characteristic of Roget's *Thesaurus* was the amalgamation of characteristics resulting from assertions (a) and (b). Firstly Hüllen discusses assertions (a) and (c), tracing the history of cumulative synonym dictionaries and showing that Roget's dictionary was not unique, but that similar dictionaries, although not topically arranged, were published earlier and simultaneously with his. Secondly Hüllen treats assertions (b) and (c) by discussing the characteristics of topical, in comparison with alphabetical, dictionaries and by tracing the history of topical dictionaries. The article ends with Hüllen's discussion of assertion (d), showing how Roget combined two traditions: the notion of an endless network of synonymy resulting from words expressing ideas and a comprehensive ordering of the universe by ideas. Roget's achievement was the finding of a meaningful order of arrangement, different from the meaningless alphabet, for gaining access to the synonymous vocabulary. This, according to Hüllen, was obtained from the general scientific outlook of the so-called utilitarians of the second half of the nineteenth century of which Roget had an ample knowledge.

Of a completely different nature than the preceding papers, is Thora van Male's contribution "From Incipit to Iconophor" in which she introduces iconophoric dictionary ornaments through a chronological series of examples from French dictionaries dating from the twelfth to the nineteenth centuries. Tracing the origins of iconophors back to medieval manuscripts in which the scribes used various signposting devices, among others the words INCIPIT and EXPLICIT (to indicate the beginning and end of a text), she shows how the old tradition of illumination had been adapted to and adopted by enough dictionaries to create new conventions and trends. By using two tables listing the different features of these iconophoric dictionary ornaments, which create an important lexicographical paratext Van Male eventually lays out the basic research parameters, both in form and in content, initially to be addressed in studying and analyzing these ornaments with a view to revealing certain conventions and trends in their use.

The last two contributions in this section, that of Maria-Pilar Perea and that of Gregory James, deal with two extraordinary foreign-language dictionaries. Perea starts her paper, "The History of a Multi-dialectal Catalan Dictionary: The *Diccionari Català-Valencià-Balear*", with a description of the two tenden-

cies, which dominated Catalan lexicography in the nineteenth century: regional fragmentation, which resulted in autonomous dictionaries for the different dialects and the preponderance of Spanish, which found its way into dictionaries through translation. This was the situation in which Antoni M. Alcover found himself when, inspired by the ideology of the *Renaixença*, the Catalan Renaissance, he started collecting material for his planned dictionary, which was continued after his death by Francesc de B. Moll who compiled and edited the material in ten volumes. After highlighting some aspects of the history, objectives and methodology of the project, Perea ends with an indication of the future of the dictionary: its updating and computerization. In his paper "Culture and the Dictionary: Evidence from the First European Lexicographical Work in China", James discusses Michele Ruggieri and Matteo Ricci's Portuguese–Chinese glossary to which Pasquale D'Elia gave the title *Dizionario portoghese–cinese*. After identifying the source of the headwords as Jerónimo Cardoso's Portuguese–Latin dictionary, he discusses the headword selection. Furthermore he deals with the romanizations and transcriptions, as well as mistakes, which show that the glossary was not only a reference source at semantic level, but also, importantly for illuminating the way Ruggieri and Ricci learnt to read and write Chinese, at the grapho-phonetic level. The Portuguese headwords are a reflection that Portuguese was the lingua franca among the Jesuits, recruited from all over Europe, both in Goa and Macau where they practised their mission.

The range of subjects, which the six papers in the second section treat are reasonably restricted, two dealing with polysemy, one with polysemy and synonymy, one with etymology, one with citations and one with ordering. These subjects are further limited, because the papers discuss only specific dictionaries, and, as far as two of the papers are concerned, specific examples: in the one case loanwords and in the other Shakespeare's informal English. In spite of this, however, valuable insights into certain lexicographical aspects and implications can be gained from the discussions of these subjects.

The first two articles in this section, those by Antonette diPaolo Healey and Robert E. Lewis, are connected through the subject matter they treat. In "Polysemy and the *Dictionary of Old English*", Healey deals with the ordering of senses in the *Dictionary of Old English*, stating that the evidence of the citations themselves are used as primary guide to sense-division. In this process, two conflicting principles often come into play: should polysemy be reduced by making generalizations or increased by making distinctions? The former presents a sparser description of a word, but perhaps leads to a clearer treatment, the latter results in richer semantic detail and perhaps a truer, more representative structure. However, as the dates of most of the texts are uncertain, the ordering of senses in the *Dictionary of Old English* has to take place according to a logical pattern. This aspect of logical ordering is elaborated further by Lewis in "Aspects of Polysemy in the *Middle English Dictionary*" when he differentiates between the logical approach of the *Middle English Dictionary* and that of the

Oxford English Dictionary, the first being a relational, the second a developmental logical approach. Lewis is of the opinion that in English historical lexicography, the logical has come to be equated primarily with the developmental, rather than the relational. This is revealed by the fact that the *Middle English Dictionary* follows the *Oxford English Dictionary* in giving precedence to most of the traditional concepts of ordering such as, for example, general precedes specific, concrete and general precedes abstract, metaphorical and specific, etc.

Eric Stanley's paper, "Polysemy and Synonymy and How these Concepts Were Understood from the Eighteenth Century Onwards in Treatises, and Applied in Dictionaries of English", is related to the previous two articles. As its title indicates, Stanley discusses polysemy and synonymy within the context of considerations of these terms and how lexicographers interpreted them. In fact, in the conclusion to the paper he says that polysemy and synonymy, although subjects within linguistic theory, have arisen from lexicographical practice: polysemy determines dictionary entries, while polysemy is used in dictionary definitions. He starts off by stressing that words often called synonyms are usually not wholly equivalent. He traces how far some earlier compilers of synonym dictionaries such as Piozzi, Crabb and Wilkens, were aware of this. This leads him to show how synonymy, considered in relation to two or several languages, demonstrates the indeterminacy and impossibility of exact translation. He then illustrates his conclusions by discussing the semantic set of words expressing nothing, and furthermore the pronouns of the second person, which additionally exemplify delimitations of custom and register. Cursorily also referring to Locke who, even when propounding the use of concrete language, could himself not avoid employing figurative language, Stanley considers it a generally sound lexicographical principle that, in accordance with European tradition, concrete usage should precede abstract, figurative and transferred applications of a word in a dictionary. Taking his cue from his exposition of the figurative use of organs such as the heart, Stanley proceeds to discuss outdated theories about the origin of language and supposed superabundant synonymy in "original" languages, propagated by Herder and Müller. Although highly polysemous words may have widely divergent senses, according to Stanley, the single etymology will unify them into one entry. He ends his paper by showing how homonymic confusability may not only lead to misunderstanding, but may also be exploited in paranomasia.

In the article "Culinary Exchanges: An Investigation of the Etymologies of some Loanwords in the Third Edition of the *Oxford English Dictionary*", Tania Styles aims at giving an overview of the key features of the etymological work on OED3, the issues it raises and the results it can yield. As illustration she uses the presentation of the etymological component of one class of lexical item, loanwords, mostly from French, and more specifically culinary terms. For each loanword, she stresses, it is attempted to establish that the immediate etymon offered for it was actually available for borrowing, in the right form, by the first recording date of the English word.

Marijke Mooijaart states in her paper "Citations in the *Woordenboek der Nederlandsche Taal*" that about two-thirds of the completed Dutch dictionary consists of citations. As this citation corpus is valuable for different research purposes, it is necessary to evaluate the composition of this collection. She discusses it under three headings: the quality of the citations themselves (length, reliability and correctness), the accessibility and usability of the electronic citations and the non-linguistic information, and the quality of the citations as a corpus (representativeness and diversity). She comes to the conclusion that the WNT citation collection meets the needs of empirical linguistic research.

In the last paper of this section, "Ordering a Historical Dictionary: The Example of Shakespeare's Informal English", N.F. Blake questions whether the alphabetical arrangement of entries is the best way for historical dictionaries. By considering five different cases from Shakespeare's informal English, he illustrates that ordering entries according to those elements, which are semantically the most meaningful results in bringing together words from the same semantic field or with the same morphological composition. Such an arrangement reveals new shades of meaning in words and allows an easier comparison of mutually related words. He therefore advocates putting words in their wider structural context by grouping them according to their morphological endings or listing compounds by their second rather than their first elements.

The collection is rounded off by a selective index, supplementing the Table of Contents.

Apart from the information contained in them, these articles can serve as examples to anybody who wishes to do research on the history and methods of historical dictionaries.

References

- Accademia della Crusca.** 1973. *Tavola rotondo sui grandi lessici storici, Firenze, 3-5 maggio 1971*. Florence: Accademia della Crusca.
- Pijnenburg, W. and F. de Tollenaere (Eds.).** 1980. *Proceedings of the Second International Round Table Conference on Historical Lexicography*. Dordrecht/Cinnaminson: Foris Publications.

J.C.M.D. du Plessis
Bureau of the Woordeboek van die Afrikaanse Taal
Stellenbosch
Republic of South Africa
(wat@sun.ac.za)

Thomas Herbst and Kerstin Popp (Editors). *The Perfect Learners' Dictionary (?)*. 1999, XII + 320 pp. ISBN 3-484-30995-4. Lexicographica. Series Maior 95. Tübingen: Max Niemeyer. Price: €88.

This collection of articles originates from a 1997 symposium in which participated some of the most eminent scholars researching lexicography and linguistics, that, since the 1970s, have widely developed to the extent of becoming a thriving market for publishers. Looking back at the evolution of EFL dictionaries for learners and ahead to the use of corpus linguistics for the compilation of a new generation of dictionaries, this volume is of particular interest to lexicographers, although some of its references may now appear slightly outdated.

Straight from the introduction, Herbst, who has published widely in the field of lexicography, indicates that the concept of a perfect dictionary is, of course, not to be taken for granted, as there is no such thing. This volume questions what the aims and targets for lexicographers and dictionary compilers should be, discussing the merits and demerits of dictionaries and how they can be improved. The main focus of this collection, connecting the articles, is the assessment of four major dictionaries published in 1995, which have since been re-edited: the fifth edition of the *Oxford Advanced Learner's Dictionary*, the third edition of the *Longman Dictionary of Contemporary English*, the second edition of the *Collins COBUILD English Dictionary for Advanced Learners* and the *Cambridge International Dictionary of English*. It may be argued that focusing on four dictionaries is restrictive, but judging by the size of the dictionaries reviewed in detail, the analysis of the contents of a bigger corpus of dictionaries would be quite a complex undertaking.

The initial paper by Cowie looks at the pioneering works produced in English lexicography in Japan and East Asia in the 1920s which left their mark and initiated the gradual emergence of EFL dictionaries, establishing the historical background for the collection as a whole. There are very few extensive and detailed investigations of this kind, comparing particular aspects of English learners' dictionaries as regards phonological, syntactic, morphological or lexical issues, using a qualitative approach as well as quantitative analyses, such as Aarts's article on the syntax of verbs or Klotz's study of adjectives and substantives in the four dictionaries investigated. Leech and Nesi's article which concludes the volume deals with the potential of the first generation of electronic dictionaries and other interactive tools or applications available to language learners and sets out perspectives for the future.

It would have been particularly useful had greater emphasis been laid not only on the contents and structures of the dictionaries, but also on their users. Although much of the material and many of the references in this book would certainly need to be updated (the *Collins COBUILD English Dictionary for Advanced Learners* is now being published in its sixth revised edition), it is particularly comprehensive and as a reference book, this research is mostly destined for lexicographers and dictionary compilers. Both teachers and learners

are also targeted. A great deal of research still needs to be done in the development of electronic dictionaries, for which there is a growing demand, and on the suitability of such dictionaries. Lexicographers' dreams (see De Schryver 2003) for the 'perfect' dictionary of tomorrow should benefit greatly from a contribution of this type to English lexicography. It is hoped that similar work would be conducted in other foreign languages, as studies are scarce in French, Italian, Spanish, German or Russian learners' dictionaries. Besides, there is a growing demand for bilingual dictionaries such as a wide range of newly published Japanese–English dictionaries, which might be worth comparing. As the market of EFL dictionaries is continuously expanding, language learners are in constant need of guidance, and such research is always in need of revision, although it provides invaluable help for the compilation of the next generation of dictionaries.

Reference

- De Schryver, G.-M. 2003. Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography* 16(2): 143-199.

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

Thomas Herbst, Gunter Lorenz, Brigitta Mittmann and Martin Schnell (Editors). *Lexikografie, ihre Basis- und Nachbarwissenschaften. (Englische) Wörterbücher zwischen ‚common sense‘ und angewandter Theorie.* 2004, VI + 207 pp. ISBN 3-484-39118-9. Lexicographica. Series Maior 118. Tübingen: Max Niemeyer. Price: €72.

The main aim of this volume, a collection of contributions by different authors, is to illustrate that dictionary making draws on other disciplines, and that lexicographical practice, in its turn, can feed neighbouring disciplines. The authors assert that linguistics is the basic discipline underlying lexicography, but the practice of dictionary making can, vice versa, expose problems of language analysis and especially crucial features of language, which should be taken into consideration in language theory. Often, lexicographical practice is characterized by devising ad hoc solutions to such problems that might positively influence the stringent and consistent approach in theoretical studies. In this sense, the relation between lexicography and linguistics should be seen as a mutual one, with possibilities of cross-fertilization in both ways. The authors want to show which trends in theoretical linguistics could be, or have already been, reflected in the compilation of dictionaries. They also want to demonstrate that many branches of linguistics, from semantics, via phraseology to syntax, can be enriched by lexicographical description and problems of lexicographical description. According to the introduction of the volume, the subtitle, "between 'common sense' and applied theory", especially characterizes the nature of lexicography practised in a meaningful way on the one hand, and on the other aptly refers to lexicographers themselves in their daily task.

This volume consists of contributions made at a colloquium in 2002 at the University of Augsburg, Germany, on the occasion of the sixtieth birthday of Prof. Dr Dieter Götz. Prof. Götz himself, according to the introduction to the volume, always succeeded in combining insights from and developments in linguistic theory with the practice of lexicography — with the necessary 'common sense'. Therefore, the theme of this volume was considered suitable for honouring him on this important occasion, and for acknowledging his influence on his students and colleagues.

The scant reference to Hausmann's distinction between metalexicography and other types of lexicography, and the claim that metalexicographical dealing with dictionaries should primarily be seen as a branch of applied linguistics, does not really do justice to Wiegand's (1998) detailed and scholarly exposition of the nature of lexicography, metalexicography and their relationship with other linguistic disciplines. This volume, however, makes easy reading — especially in comparison with other publications in the Lexicographica Series Maior. The issues dealt with will be of interest to all students in lexicography, to practising lexicographers, as well as to linguists.

Although all the contributions are written in German, except the one by

Jehle which is in English, the German is not as technical as is usual in German publications, and the intricate constructions so characteristic of German academic writing are absent here. Since the contributions focus on, among others, examples from English dictionaries (such as OALD6, LDOCE3, COBUILD3 and CIDE), this volume will be useful to English-speaking readers, even though their reading abilities of academic German might not be on a very high level.

The volume contains contributions on several aspects of lexicography, including a historical look at the role of the dictionary in the teaching of a foreign language (Schröder), and the place and features of dictionaries within the new media landscape (Schäffler). Within the scope of semantics, it looks at the application possibilities for the "scenes-and-frames" theory (Bublitz and Bednarek) and prototype semantics (Jehle) as well as the effect of different cultural approaches to linguistic description on dictionaries from these cultures (Rothe), and the possibilities for the description of meanings (Wellmann). Other contributions focus on stylistic issues (Kamm), the corpus-based analysis of *time* (Lenk) and prefabricated units in spoken language (Mittmann). Aspects of bilingual lexicography and contrastive linguistics are investigated with a view to the description of idioms (Lorenz) and special-field languages (Schnell). Descriptive problems encountered in English and their treatment in dictionaries are dealt with when looking at formation compounds (Burgschmidt) and at valency (Herbst).

In his contribution "Zur Rolle des Wörterbuchs in der Fremdsprachendidaktik der frühen Neuzeit", Konrad Schröder gives an overview of the multilingual situation in Europe from 1500–1800, where aristocrats as well as the middle classes had to speak many languages in their daily lives. Thousands of reference works and learning materials resulted during this period, such as conversation books, grammars, lexicons, and pronunciation guides. About 20% of these publications focus on several languages at the same time, some including as many as eleven languages. Lexicons and vocabularies constitute roundabout 40% of the corpus under discussion, followed by grammars and conversation books. The borders between the genres are, however, vague. Up to the present, mainly works dealing with the lexicon have been investigated; nevertheless, conversation books and grammars will also have useful information and insights for linguists.

Schröder discusses the factors which motivated the creation of lexicography during the period 1500–1800. These were, on the one hand, of a political nature, in order to promote the particular language in an international forum, and, on the other hand, of a practical kind in order to facilitate communication during trade and business encounters. The publications functioned as normative guidelines, especially to enhance the receptive competence of their users — understanding and reading — in a comparative approach. Idioms, phrases, and discourses were included, to help the user to read and understand literary texts in the foreign language. Syntactic information was obtained in the stride.

Hildegard Schäffler's contribution "Wörterbücher und Medienwandel aus bibliothekarischer Sicht" deals with modern media from the point of view of the librarian, and the role of dictionaries in the library. Since so many publications, including dictionaries, have become available in electronic format, she claims there has been a shift in use. One of the implications is that users have to be trained for successful consultation of these electronic publications. Nevertheless, Schäffler concludes that the electronic versions are much more user-friendly than the printed versions.

She continues with a discussion of the content and range of the OED Online (2001–), based on that which was already available as CD-ROM in 1992, and as printed version in 1989 (second edition), but also containing several additions from the years 1993–1997, as well as updates made every three months. The main advantage of this online version is that it enables regular updates from which the user largely benefits, obviating the consultation of the older, stagnant printed version. The original entries are still present, and are linked to the revised versions, which are provided with the dates on which they were added. This possibility of hyper-linking naturally has advantages over the linear approach of printed versions.

The online entry contains all the information of the printed version, but a date chart, which presents the sources in chronological order is added. Techniques of presentation not possible in a linear publication are used, such as hyper-linking between entries, and cross-references to etymology or sources. In addition, certain information can be faded out, better to highlight others. Each entry also has a so-called "entry map", which displays the structure of all the information for that particular entry. There is also an alphabetical list of entries on the left-hand side of the screen, by means of which the entries directly preceding and following the entry can be accessed.

The possibilities for retrieval of information in the online version are different from those in the printed version. Apart from a simple search, an advanced search, combining search questions by applying logic operators can be performed. Schäffler illustrates this with the example of searching for citations from Jane Austen's *Mansfield Park*.

She mentions two possibilities which have, up to now, not been included in the electronic version of the OED Online. These are hyper-linking to other Oxford dictionaries or linking to a corpus or even publications of competing publishers. (Incidentally, according to a footnote on page 25, COBUILD CD-ROM (2001) does have extracts from the Bank of English corpus.) Schäffler also feels that the OED Online could do more to visually present different levels of information with the technique of fading out or in. For example, encyclopedic information or word-field analyses could be presented in such a way that they can be faded in when needed, and afterwards faded out again.

The rest of her contribution deals with the licensing of online products, the types of licenses and the conditions for licenses.

In their contribution "Nur in begrenzten Rahmen: *Frames* im Wörterbuch", Wolfram Bublitz and Monika Bednarek focus on the use of frames in dictionaries. The concept of frames, which was developed by Minsky, Fillmore and others, is undoubtedly useful in semantic-pragmatic and text-analytical language description, even though it may be one of the concepts in cognitive linguistics which perhaps exhibits the highest degree of fuzziness. The authors deal with the question whether frames can be used to convey potentially context-dependent meanings in dictionary entries, in accordance with the belief in corpus linguistics that meaning cannot be separated from use.

After a critical discussion of the earlier contributions by Wierzbicka and Fillmore within the field of semantics, Bublitz and Bednarek look at the possibilities for using frames in an online dictionary or databank, by means of which a user may consult the databank in a basic or advanced search. Because frames can be linked, it becomes possible for a word (in this case, the word *blunt* used as illustration by them) to appear not only as an alphabetically ordered lemma, but also as a component of one or more frames which can be accessed in different ways. An electronic dictionary conceptualized in this way will differ drastically from a conventional, printed dictionary. Whereas the usefulness of frame-semantics is doubtful for traditional lexicography, it can certainly enhance electronic dictionaries.

In his article "Prototype Semantics and Learners' Dictionaries of English", Günter Jehle compares the semantic descriptions of the verbs *scurry*, *scuttle* and *hurtle* in OALD6, LDOCE3, COBUILD3, and CIDE. His analysis shows that OALD6, LDOCE3 and CIDE use the so-called Aristotelian definition or feature semantics approach. COBUILD3, however, employs complete sentences "which highlight the syntactic as well as the semantic valencies of the lexical item to be defined" (Herbst et al. 2004: 58). According to Jehle, COBUILD3 combines feature semantics and prototype theory, and therefore helps the foreign user to "come to terms with peripheral category members" (Herbst et al. 2004: 58), because it gives excellent guidance about the semantic restrictions applying to verbs in certain contexts.

Jehle proposes that dictionaries should follow a prototypical approach, because this corresponds to the way native learners learn their mother tongue. They link the visual, extralinguistic impressions in reality to the corresponding phonological form of the linguistic sign. Therefore, he continues, the dictionary of the future, which implies that it will be electronic or online should also provide illustrations such as video clips and animations to help foreign language learners to link the linguistic information to the extra-linguistic phenomenon at hand. Each lemma should have a short entry, but should be accompanied by a longer, more detailed entry containing information on the grammar, semantics, semantic prosody, collocation, colligation, as well as examples and video clips or animations.

Ulrike Rothe has done previous work on the influence of cultural aspects and linguistic trends within specific cultures on the content and nature of dictionaries (Rothe 2001). In her contribution to this volume, "Das einsprachige Wörterbuch als Produkt von ‚Kultur‘: Lexikographische Definitionen und Artikelbaupläne im Licht semantischer Theorien", she investigates the same issues, namely, whether lexicographical definitions and the structures of dictionary articles are influenced by specific semantic theories in the countries where the dictionaries originate. In other words, she poses the question whether the methods used in lexicography are determined by cultural factors. If this is the case, she argues, then dictionaries from one specific country should be conceptualized in the same way, showing common features, and it should be possible to distinguish them from those compiled in other countries.

In French linguistics, structural semantics were very influential during the 1960s. In the Anglo-American world, however, prototype semantics was popular in the 1980s. In view of this, Rothe investigates lexicographical definitions in several English and French dictionaries, only to come to the conclusion that the French dictionaries indeed draw on structural semantics, and the English ones on prototype semantics. She warns, however, against the claim that there is a direct link between the linguistic model of meaning and the techniques of writing lexicographical definitions.

In addition, the structures of French dictionary articles differ from English ones, showing French articles to have more hierarchical levels of description than English ones (which usually have only two levels). This can, once again, be attributed to the viewpoints of structural semantics and prototype semantics respectively.

The next contribution, that by Hans Wellmann, "Der Definitionswortschatz des einsprachigen Wörterbuchs", investigates the vocabulary used in definitions of monolingual dictionaries, in this case German dictionaries for advanced foreign learners. Wellmann asserts that learners' dictionaries should be simple, clear, understandable and applicable. Users should be able to find the looked-for information quickly. Cross-references often cause them to give up their search.

There are several interesting differences between the vocabulary used to describe lemmas in German and English learners' dictionaries. These can be attributed to structural differences between the two languages. For example, stylistic labels differ: some words which are regarded as "colloquial" in German, such as *Ding*, *Objekt*, *brechen*, and *tun*, are part of standard speech in English, namely, *thing*, *object*, *break*, and *do*. In selecting vocabulary to define lemmas, however, it is necessary to use "ordinary language", which might include the basic words normally learnt by foreign learners right at the beginning of their language studies, and which can also be easily understood. Internationalisms (e.g. "Telefon" instead of "Fernsprecher") should also be included, because it increases the chances that users will understand them.

Another distinguishing characteristic of German vocabulary is the tendency to form compounds. These may cause problems when used in descriptions, especially with changes in meaning occurring when the meanings of the separate constituents of the compound are no longer transparent.

In her article "Stilistische Angaben in einsprachigen Lernerwörterbüchern des Englischen: Fortschritte, Probleme, Grenzen", Heike Kamm investigates stylistic items in several monolingual English learners' dictionaries, and compares the labelling systems used in the OALD5, CIDE, COBUILD2 and LDOCE3.1. Especially with regard to special-field labels, and labels indicating regional variety, differences occur among the dictionaries. Kamm makes out a case for including the stylistic information in the definition of the meaning, because she believes that users may not always take the time to consult the introduction of the dictionary, where the labelling system is explained. In addition, assumptions should not be made about any previous theoretical-linguistic knowledge of users.

In Uta Lenk's contribution, "Korpuslinguistik und Lexikographie am Beispiel eines hochfrequenten Lexems", corpus linguistics and lexicography are scrutinized on the basis of the frequent lexeme *time*. Certain so-called "stabilized expressions" containing the lexeme *time* are analyzed, such as *from time to time*, *at the same time*, and *take your time*, and their function as semantic prosodies. Lenk proves that stabilized expressions have not been studied in detail, even in the era of corpus linguistics.

In her contribution "Pragmatik und Wörterbücher: *prefabs* und gesprochene Sprache", Brigitta Mittmann discusses so-called "prefabs (prefabricated expressions)" such as *You're welcome* in spoken language, and their treatment in dictionaries. Once again, the examples she analyzes are from English dictionaries such as OALD6, LDOCE3, COBUILD3, and CIDE. According to Mittmann, one interesting aspect of spoken language, namely intonation, is not reflected in any of the four dictionaries she investigated. For example, falling or rising tones may have an effect on the meaning of spoken phrases, such as *There you are*: in the case where it means "So, I've found you!", a falling tone will be heard on the component *there*, in the case where it means "This is for you", a fall-rise will be heard, and in the case where it means "That supports or proves what I've said", a falling tone on *are* will be heard. The question remains whether dictionaries, which ought to focus on the needs of language learners, are the most suitable means to convey such detailed descriptions of spoken language. She refers to Moon (1998), who proposes a completely new concept for a dictionary of spoken language, with a completely different structure and special consideration for phraseological-collocational and pragmatic aspects, as well as new models for the description of different viewpoints. Summers (1999), according to Mittmann, also argues that spoken language has in the past few years been part of the examinations for foreign learners, which implies that there should be dictionaries dealing with typical features of spoken language.

Mittmann suggests that the spoken demographic part of the British National Corpus or the Longman Spoken American Corpus could serve as basis for such a type of dictionary. Of specific use would be information on the frequency of prefab clusters, for, at the moment, LDOCE3 and COBUILD2 only give information on the frequency of single words. Supplying such information would be very difficult, however, because expressions such as *That's that* could stand on its own, but could also form part of *That's that man over there*. But, providing this type of information could also be useful for research on phraseology and in psycholinguistics.

Indications of word stress, descriptions of the situational frame within which expressions can be used, and the question whether the suggested dictionary should be ordered alphabetically or onomasiologically are some of the issues discussed by Mittmann.

In his contribution "*Gilding the Lily? Überlegungen zur zweisprachigen Idiom-Lexikographie*", Gunter Lorenz investigates the treatment of idioms in bilingual lexicography. From a lexicographical point of view, idioms are difficult to lemmatize. For example, it is not clear under which headwords the individual components of the idiom *to cut a long story short* should be placed in the alphabetical order of the dictionary. Variants such as *to make a long story short* complicate matters even more. It has to be decided whether the target group will be native speakers or foreign learners, whether a general dictionary or a special dictionary will be compiled, and whether it will be a monolingual or bilingual dictionary. Of course, when idioms are considered from a contrastive perspective, there are many "idiomatic voids", where one language has an idiom which is not present in another language. There are also so-called "false friends", "near-misses" and "almost-congruences", which complicate the bilingual treatment of idioms.

Martin Schnell's contribution, "*Zum Problem der Äquivalenz in zweisprachigen juristischen Fachwörterbüchern Englisch-Deutsch/Deutsch-Englisch*", which deals with bilingual (i.e. German-English) special-field dictionaries of law, connects with Lorenz's argument, because he too illustrates this situation of incongruence. For example, the German *Notar* and the English *notary* have different meanings within the judicial systems of the two countries. Therefore, lexicographers have to make use of options to describe partial equivalence relations. They could even, in cases where equivalents do not exist, work with so-called "pseudo-terms", which are basically neologisms. Taken for granted is the fact that lexicographers have to provide glosses in which more detailed information can be given. In electronic publishing, this will be increasingly possible, because limitations of space will in future be less problematic.

The last two contributions by Ernst Burgschmidt and Thomas Herbst deal with the way word-formation theory and valency theory respectively are reflected in lexicography. In his article "*Theorie der Wortbildung und ihr Reflex im Wörterbuch*", Burgschmidt illustrates his discussion by means of typical

types of word formation in English, compared to that in German. Herbst's article, "Valenzlexikografie und Valenztheorie — Grenzen der Beschreibbarkeit eines sprachlichen Phänomens", analyzes several examples from dictionaries such as the *Valency Dictionary of English*, which was still forthcoming at the time of publication of this volume.

In spite of certain terminological problems — e.g. Rothe and Wellmann both use the term *definition* in a sense rejected by Wiegand, who feels that it should only be employed with reference to scientific terms — this volume will be informative to several user groups, and it is therefore strongly recommended.

Dictionaries

- CIDE = Procter, Paul (Ed.). 1995. *Cambridge International Dictionary of English*. Cambridge: Cambridge University Press.
- COBUILD2 = Sinclair, John (Ed.). 1995². *Collins COBUILD English Dictionary*. London: HarperCollins.
- COBUILD3 = Sinclair, John (Ed.). 2001³. *Collins COBUILD English Dictionary for Advanced Learners*. Glasgow: HarperCollins.
- COBUILD CD-ROM (2001) = *Collins COBUILD. Resources from the Collins COBUILD Range: English Dictionary for Advanced Learners, English Grammar, English Usage, Thesaurus, 5-Million-Word Wordbank from the Bank of English*. London: HarperCollins.
- LDOCE3 = Gadsby, Adam et al. (Eds.). 2001³. *Longman Dictionary of Contemporary English*. Harlow: Pearson Education.
- LDOCE3.1 = Summers, D. (Ed.). 1995^{3.1}. *Longman Dictionary of Contemporary English*. Harlow: Longman.
- OALD5 = Crowther, J. (Ed.). 1995⁵. *Oxford Advanced Learner's Dictionary of Current English*. Oxford: Oxford University Press.
- OALD6 = Wehmeier, Sally (Ed.). 2000⁶. *Oxford Advanced Learner's Dictionary of Current English*. Oxford: Oxford University Press.
- OED Online (2001–) = *Oxford English Dictionary Online*. 2001. Oxford: Oxford University Press. <<http://oed.com>>.
- Valency Dictionary of English* = Herbst, Thomas et al. (Eds.). 2004. *Valency Dictionary of English: A Corpus-based Analysis of the Complementation Patterns of English Verbs, Nouns, and Adjectives*. Topics in English Linguistics. Berlin/New York: Mouton de Gruyter.

Corpora

- British National Corpus <<http://info.ox.ac.uk/bnc/>>
- Longman Spoken American Corpus <<http://www.longman-elt.com/dictionaries/>>

References

- Moon, Rosamund. 1998. On Using Spoken Data in Corpus Lexicography. Fontenelle, Thierry, Phillipe Hiligsmann, Archibald Michiels, André Moulin and Siegfried Theissen (Eds.). 1998.

EURALEX'98 PROCEEDINGS. Papers submitted to the Eighth EURALEX International Congress on Lexicography in Liège, Belgium: 347-355. Liège: English and Dutch Departments, University of Liège.

- Rothe, Ulrike.** 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.
- Summers, Della.** 1999. Coverage of Spoken English in relation to Learners' Dictionaries, especially the *Longman Dictionary of Contemporary English*. Herbst, Thomas and Kerstin Popp (Eds.). 1999. *The Perfect Learners' Dictionary (?)*: 257-264. Lexicographica. Series Maior 95. Tübingen: Max Niemeyer.
- Wiegand, Herbert Ernst.** 1998. *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie. 1. Teilband.* Berlin/New York: De Gruyter.

Maria Smit
Department of Music
University of Stellenbosch
Stellenbosch
Republic of South Africa
(msmit@sun.ac.za)

Werner Hüllen. *English Dictionaries 800–1700: The Topical Tradition.* 1999, 525 pp. ISBN 0-19-823796-0. Oxford: Clarendon.

Werner Hüllen. *A History of Roget's Thesaurus: Origins, Development, and Design.* 2004, 410 pp. ISBN 0-19-925472-9. Oxford: Oxford University Press, 2004.

Under section **559. Word**, my edition of Roget's *Thesaurus* includes a sub-paragraph:

dictionary, lexicon, wordbook, word-stock, word list, glossary, vocabulary; gradus, thesaurus, wordhoard; compilation, concordance, index. (Roget 1984)

The semi-colons between *vocabulary* and *gradus* and between *wordhoard* and *compilation* are crucial. They indicate a relationship looser than synonymy. They signify that *dictionary* and *thesaurus* are not interchangeable.

Monographs attempting to treat the history of English dictionaries have to restrict their field in some way. They might do this by choosing to treat monolingual lexicography as if it developed in isolation from bilingual lexicography, for example, or by ignoring or underplaying the influence of lexicographical developments in the treatment of other languages. Specialist dictionaries are usually mentioned only in passing, and anything that departs too far from the linear history of dictionary development is set aside as aberrant and therefore irrelevant. It is rare that such histories treat thesauruses as anything other than a minor sideline, confirming the general feeling expressed in Roget's classification, that a thesaurus is not a dictionary. In his populist account of lexicographers and their dictionaries, Green makes only one reference to Roget's *Thesaurus*, and that serves to emphasize its marginality:

Paradoxically, it might be noted, the latest editions of Roget's *Thesaurus* — the supreme exemplar, outside an actual encyclopedia, of the thematic method — are being recast as alphabetical word-books. (Green 1996: 23)

Landau, unusually, does give a few pages to Roget, but does not distinguish between two types of work that preceded it: synonym dictionaries and synonym books (Landau 1984: 24). It is this gap that Hüllen seeks to fill, and these ambiguities that he seeks to clarify.

As Elizabeth Whately was to note in 1851, English is particularly rich in synonyms because its Germanic base has been supplemented by borrowings from Latin, French, and many other languages (Hüllen 2004: 269). This wealth of vocabulary has long created problems both for native speakers of English and for non-native learners. Their sales demonstrate that synonym dictionaries and thesauruses have long been seen as possible solutions. Across these two monographs, Hüllen traces the various traditions that culminated in Roget's *Thesaurus*. They are complementary volumes, in that they treat different aspects of the same tradition, though they cover little of the same ground. Both empha-

size the importance of treating their subjects as products of their intellectual and cultural contexts.

English Dictionaries begins by discussing the history of lexicography. Hüllen remarks that:

There never was *lexicography* without word-lists and/or dictionaries, but there were for a long time (and still are) word-lists and/or dictionaries without lexicography. Thus, lexicography has its *raison d'être* in a language-bound technique which preceded it in time and which may even today exist in its own right. (Hüllen 1999: 3)

His purpose, across these two books, is to explore the developments that led to the types of dictionaries and word-lists that contemporary users take for granted. Dictionaries do not just provide an account of the cultural activity known as language, they are also part of that general culture and, more specifically, of a dictionary culture. The study of the development of lexicography is, Hüllen argues, part of the general history of ideas, and he presents four main sources of information for modern lexicography: linguistics, research into the needs of dictionary users, the history and theory of lexicographical practice, and encyclopaedic knowledge. In topical dictionaries, this culturally determined encyclopaedic knowledge determines the very structure of the word-list.

Hüllen describes a 'forgotten' early twentieth-century discussion of the difference between dictionaries organized 'according to the alphabet' and those arranged 'according to nature' (Hüllen 1999: 16-17), and concedes that:

Alphabetical dictionaries have conquered the market. This fact will not change in the near future. One consequence of this is that they have also conquered the attention of scholars. Yet it seems high time that non-alphabetical dictionaries should at least regain their history. (Hüllen 1999: 21)

It is clear that onomasiological ordering of words formed an important part of language learning before alphabetical lists became widely available. Semantically linked groups of words were memorized as part of the process of mastering a foreign language. While scholarship was still conducted largely in Latin, memorization of such word-lists would have been an important stage in acquiring any body of knowledge.

Hüllen begins his account of the history of the onomasiological tradition with word-lists in Egyptian, Chinese, and Sanskrit, and also covers Graeco-Coptic and Arabic-Syriac lists. Rather than arguing that these form part of the same tradition as later English lists, he suggests that they reveal instead universal features of language and of the uses to which humanity puts language. Although in some cases it is possible to demonstrate links both between these works and also with the later onomasiological lists, Hüllen prefers to consider his material as an *autonomous tradition*: a tradition that comes about 'simply because the nature of the phenomenon makes it natural to behave in this way' (Hüllen 1999: 36). This is a sensible decision, not least because it protects

against the inclination to base assertions of influence on the grounds of similarity of content or structure that might have arisen by chance. It leaves the way open to consider these works as texts belonging to a genre rather than, necessarily, to an unbroken tradition.

The English onomasiological tradition, Hüllen argues, can be traced back to classical roots, where it belongs to the field of *ars memorativa*, by which ideas and words are stored in the memory to be retrieved in a given order, by organizing them in a spatial arrangement. Memory was, and is, central to language learning, but contemporary learners have printed books and note-paper to supplement their powers of recall; before the Renaissance only the teacher (if anyone) would have possessed a book, and pupils would have been able to make only temporary notes on tablets of wax or slate. Language had to pass from the memory of the teacher into the memory of the pupils more or less directly.

The English lexicographical tradition begins with glosses providing Old English translations between the lines of Latin texts. Such glosses 'allow a glimpse into the medieval classroom' (Hüllen 1999: 55) in that they would clearly have served as aids to both teaching and learning. These interlinear glosses were collected into glossaries: lists of Latin terms defined by their English glosses. If the ordering principle is the sequence of words in the text, these glossaries still serve a text-dependent purpose, but adopting an alphabetical or topical approach renders them valuable to a wider audience. Even within a text-dependent glossary, there will sometimes be topical nests of vocabulary, where terms belonging to the same semantic area cluster because they are found together in the source text. The glossary appended to Ælfric's grammar receives particular attention, not only because it is the first of its kind, but also because it aims 'to cover the whole word with words' (Hüllen 1999: 65) and because of its clear and obvious debts to Isidore. The organizational principles underlying the fifteenth-century *Mayer Nominale* are also described in some detail.

Given that many early, topically arranged word-lists were designed for use in language teaching, it is fitting that Hüllen should also discuss contemporary colloquies and dialogues used in the same setting. He pays particular attention to Caxton's *Dialogues in French and English*, which is actually a monologue with role-playing dialogues inserted. There were, of course, many similar works providing dialogues in various combinations of continental European languages, some covering up to eight languages. The order in which these languages were added as the tradition developed is an interesting indication of their contemporary importance. These works were clearly intended as aids to commercial and political interaction, though many of the dialogues are domestic in nature.

The onomasiological word-lists also share features with treatises on terminology, and Hüllen looks in detail at those dealing with four specializations: husbandry, rhetoric, geometry, and seafaring. These works sometimes merely

translate Latin materials into the vernacular, but sometimes involve original writing. Of the 133 such works printed between 1480 and 1640, fifty-two are arranged topically (Hüllen 1999: 141). They are, of course, closer to encyclopaedias than more general onomasiological dictionaries.

John Withals' *Short Dictionary for Young Beginners* (1553) is designed to help learners of Latin, but does so by presenting its bilingual entries with the English term first. Its many editions, gradually growing in size, demonstrate how popular the work was, and Hüllen contends that its contribution to the tradition, earlier than is normally accepted, is the idea of learning words as names for things, rather than merely as words: of the idea that one learns a language by learning reality.

James Howell's *Lexicon Tetraglotton, An English–French–Italian–Spanish Dictionary* (1660) included an onomasiological section, entitled *A Particular Vocabulary, or Nomenclature*. The appearance of only one edition suggests that this was less successful than Withals' work, which may indicate that its slightly greater emphasis on English than the other languages included was not what the market was looking for at this time.

Topical dictionaries were not merely designed as aids in learning foreign languages or perfecting one's mastery of English. They were also motivated by the philosophy of universal language: the idea that there was (or could be) a language used by all, which would be based on the underlying structures believed to be common to all languages. Proponents of universal language argued from theological but also from practical standpoints. For John Wilkins:

[The] world was identical to the sum of all physical and mental objects ... Lexemes were names for these. Thus, the universal lexicon of a universal language was nothing other than a list of words arranged 'philosophically'. (Hüllen 1999: 246)

Wilkins was not the first to arrange language 'philosophically', but he was the most important in the seventeenth century, and probably because of the complexity of the task he set himself, still does not receive the attention he deserves. Roget certainly used the *Essay* as the model for his *Thesaurus*, and Hüllen discusses a number of less important works that were also influenced by it.

Hüllen states that 'the historical development of foreign-language teaching and learning can hardly be adequately treated in historiography according to nations' (1999: 361), but he treats 'the European scene' of the fifteenth century after considering English publications of the seventeenth. There is a mass of information in this volume, and it must have been hard to decide how to structure it, but for this reader it felt like a step backwards, especially since some of the main points had necessarily already been made in setting the scene for the English tradition. However, this all provides a useful introduction to Hüllen's analysis of the work of Comenius, and since most histories of English dictionaries do tend to define their field 'according to nations', it was certainly useful material for me. Twenty-five editions of Comenius's *Janua linguarum in-*

cluded English and were published in England, and it appeared in various editions, covering various languages, throughout Europe. Comenius was motivated by his philosophical and pedagogical ideals: not only would his students acquire an understanding of reality through language, but their minds would be structured in harmony with it, which would have a positive influence on the future structure of society. This, in turn, would strengthen the harmony between the individual and reality:

The harmony between all the members of this chain is grounded in the essential analogy of all the domains of the world, which for Comenius was a fact guaranteed by God. (Hüllen 1999: 373)

An interesting feature of Hüllen's treatment of topical dictionaries are his remarks on how they present 'the world of women'. The *Mayer Nominale*, for example, treats menstruation among the diseases of women. Withals tends to provide terms in their masculine form, so the provision of feminine forms indicates activities in which women would have been involved. It is perhaps understandable, in a text for schoolboys, that Withals treats men's clothing in great detail, from underwear outwards, but concentrates on outer clothing for women. Apparently aimed at a more adult audience, Howell's organization of terms for female clothing gives 'the image ... of a woman undressing in order to attend to her toilet' (Hüllen 1999: 242-243).

Despite the long tradition of topical dictionaries treated in *English Dictionaries*, Hüllen begins his *History of Roget's Thesaurus* by saying that 'when it appeared (Roget 1852), it was the first of its kind' (Hüllen 2004: 1). This monograph builds on *English Dictionaries* to demonstrate that although Roget was by no means the first to attempt to arrange the lexis of English by meaning, he did so in new ways and for new reasons. He was influenced by earlier thinkers (rather more than by earlier linguists), but his innovations are what set him apart. What made Roget's *Thesaurus* unique, Hüllen argues, is that it brought together two separate traditions: the synonym dictionary and the topical dictionary.

Hüllen argues that Roget's work is a product both of naïve 'word-centred language awareness' (Hüllen 2004: 30) and of philosophical traditions long since forgotten. We can only understand the structure of and motivation behind the *Thesaurus* by examining these two strands. This is why Hüllen explores classical treatments of synonymy and considers that way that synonyms are used in literature, including *Beowulf* and *Love's Labour's Lost* to demonstrate 'that synonymy lies at the heart of many profound philosophical deliberations and language-bound projects of the human mind and that it is one of the most powerful resources for creative language use in literature and in ordinary discourse' (Hüllen 2004: 75). Almost any literary texts could have been chosen to make the same point, but I am not convinced that it was necessary to make it at such length. It was harder to see the relevance of his overview of recent develop-

ments in semantics until later in the book, and it would have been useful to have had more explicit explanations for some of the inclusions.

Hüllen's account of early English hard-word dictionaries demonstrates that synonymy was an important part of their method: terms borrowed from Latin, Greek and French are explained, where possible, by providing more accessible synonyms rather than through definitions. Thus the conventional division between semasiological and onomasiological dictionaries is shown to be less clear-cut than it might seem to be.

The account of John Locke's *Essay Concerning Human Understanding* (1689) compares Locke's work with those of present-day semanticists to see whether they deal with similar problems and find related solutions. Locke's ideas mark the turning point between referential or speculative lexicography and mental lexicography:

i.e. where the assumption that vocabulary follows reality in the way in which a name follows a thing turns into the assumption that vocabulary identifies and gives order to reality because it provides the inalienable signs for the expression of ideas. (Hüllen 2004: 170)

Hüllen goes on to demonstrate the influence of Lockean ideas on the thinking behind Johnson's dictionary, expressed in the *Plan* and the preface, and also within individual entries. In the end, though, Johnson had a practical task to complete, and appears to have 'accepted the notion that, if no perfect definition was possible, an imperfect one would do' (Hüllen 2004: 186). Johnson used many such imperfect definitions from Isaac Watts' Lockean-influenced writings to illustrate terms included in the dictionary. Moreover, like the hard-word lexicographers, Johnson often included lists of synonyms along with his definitions.

Abbé Gabriel Girard's *La justesse de la langue française* was published in 1718, and 'almost created a new text genre' (Hüllen 2004: 206), part way between the synonym dictionary and the treatise on synonymy. Girard worked on the assumption that there were no semantically identical terms in the French language, which enabled speakers to differentiate between closely related objects or ideas with great precision, but that semantically similar lexemes could be used to avoid repetition in a text. His influence is seen in the books on synonymy and synonym dictionaries published in many European countries. As time passed, these works moved away from the discussion of selected examples found in Girard's work towards increasingly comprehensive treatment of entire lexicons. Among these European works, which Hüllen lists but does not discuss in detail, he detects a group-identity which, even after the publication of the *Thesaurus*, excludes Roget.

Girard's imitators in Britain were John Trusler, Hester Lynch Piozzi, and William Taylor. Trusler's oddly punctuated *The Difference, Between Words, esteemed Synonymous* (1766) made clear its debt to Girard, and in some cases merely translated articles from its French source. Interestingly, unlike the topi-

cal dictionaries discussed in *English Dictionaries*, this work steered clear of the physical world and concentrated instead on mental and emotional phenomena in conversation. Piozzi's *British Synonymy* (1794) emphasized that synonymy — selecting the appropriate word or phrase in polite conversation — is a study more appropriate to women than to men (who should instead concern themselves with grammar and logic). She mentioned in particular the help that her work would offer to foreigners learning English. Although she probably knew Trusler's book, and certainly knew Girard's, her far more discursive work appears to be an independent compilation. Taylor's *English Synonyms Discriminated* (1813) is a more self-consciously erudite work, but not necessarily any the better for that. It is heavily dependent on Girard. These English synonymists do not adapt Girard's nationalistic sentiments about French to their discussion of the English language.

Hüllen provides a detailed account of English synonym dictionaries preceding Roget's *Thesaurus*, and identifies James Barclay's *Complete and Universal English Dictionary* (1774) as the first. Barclay wrote:

The Synonymous part of our Dictionary we modestly assert to be entirely new; the use of which, both in speaking and writing, must at first view appear to every intelligent Reader; and we have endeavoured to execute it without running into whimsical notions, or fantastic, affected niceties. (quoted in Hüllen 2004: 237)

Although Barclay's dictionary arranged its entries alphabetically, the content of those entries is clearly influenced by the earlier synonym books. Like many of the topical dictionaries discussed in Hüllen (1999), some of the early synonym dictionaries were designed for didactic purposes, including Ellin Devis' *Miscellaneous Lessons, Designed for the Use of Young Ladies* (1782).

William Perry's *The Synonymous, Etymological and Pronouncing English Dictionary* (1805) is based on Johnson's dictionary with the quotations taken out. Its important innovation is the idea of a *radical synonym*. These are highlighted by capitalization, which indicates that the word's various meanings are listed in its own alphabetically placed entry. What this demonstrates is that Perry's book is on the threshold of seeing synonymy not as a question for isolated exposition of tricky examples of usage, but as a fundamental part of word meaning. George Crabb's *English Synonyms Explained* (1816) 'is different from Perry's ... because it combines the expositions of meanings, modelled on the early books by Trusler, Piozzi, and Taylor, with the perfection of a dictionary' (Hüllen 2004: 254). His 'scientific' principle was to include all the synonyms in the language in alphabetical order, moving within entries from the more general to the most specific. Crabb also included definitions and information about etymology and word-formation. John Platts' *Dictionary of English Synonymes* (1825) was a short version of Crabb's work, designed for use in schools. One of the ways in which he compressed Crabb's material was to conflate articles on synonymous terms.

Hüllen's account of the continued production of synonym dictionaries throughout the nineteenth century demonstrates that some of the features familiar to us from Roget's *Thesaurus* were already well established in this genre, particularly some of the typographical features used to show relationships between terms.

Having discussed the synonym dictionary in some detail, Hüllen turns to the other tradition that fed into Roget's *Thesaurus*: the topical dictionary. 'Only a few other works in the onomasiological tradition were published between Wilkins and Roget ... and none of them would be so closely related to Roget's *Thesaurus* as Wilkins's "Tables" ' (Hüllen 2004: 285). Unfortunately, in order to avoid going over the same ground as in Hüllen (1999), this discussion of Wilkins's work is rather briefer than might be helpful, especially given its complexity.

Peter Mark Roget was a doctor rather than a linguist by training, with interests in public health and physiology. He was, therefore, influenced by taxonomical classifications of the natural world in the construction of the *Thesaurus* as well as by publications in linguistics, although Hüllen carefully asserts that he does 'not assume that similarities with Roget's work are due to his knowledge of such ideas' (Hüllen 2004: 324). As a scientist, Roget made some minor discoveries of his own, but his main concern was to make scientific knowledge available and useful to the general population. The *Thesaurus* is similarly motivated. He appears to have been working on 'a compilation' of words for almost fifty years, but concentrated on it after his retirement. It was, of course, to outshine and outlast all of his earlier scientific work. Early comments, belied by the evidence of sales, were generally rather negative.

Hüllen identifies three perspectives from which Roget wished his book to be seen: that it moves from the idea to the word, that it is intended to be of practical value (including, possibly, in the development of a 'strictly Philosophical Language' which could be used for communication between nations), and that the creativity of the human mind is revealed in the varieties of word meaning. What Roget does not do is to discuss synonymy in his introduction: he is writing for native speakers and assumes that such a discussion is unnecessary. In the structure of Roget's classification, Hüllen finds 'a world-view typical of post-Cartesian philosophy and the post-Newtonian sciences' (Hüllen 2004: 334). He characterizes the difference between Wilkins and Roget thus:

John Wilkins presents a world of experience ordered according to the abstract ideas of traditional philosophy. Peter Mark Roget presents a world of human ideas. Whereas Wilkins, as a lexicographer, faced the world and attached names to its items, Roget, as a lexicographer, entered the human mind and unpacked what he found there. (Hüllen 2004: 334)

Roget's work could be seen as a forerunner of contemporary cognitive semantics, in which the structures of language are seen as expressions of the structures of the human mind. It also presents semantic fields a hundred years before the concept was introduced to linguistics.

Although Hüllen emphasizes the practical intentions behind Roget's work, he does not appear to address Landau's observation that:

However ingenious the hierarchy of concepts, it was pure fantasy to suppose that any conceptual arrangement of the vocabulary of English was natural to most native speakers. It cannot be doubted that few readers could have found the meaning sought without recourse to the index, which has been expanded greatly through the work's many editions. (Landau 1984: 107)

It is a pity in a monograph on Roget's *Thesaurus*, that Hüllen did not spend more time discussing its reception and use. Given that the structure of the classification is of little practical value to the thesaurus-user, who has first to look in the index in order to find the appropriate place in the classification, why did Roget's work eclipse earlier synonym dictionaries in which synonyms could be located with half the labour?

Taken together, these two volumes provide an impressive account of the English onomasiological and synonym dictionary traditions and their background. Hüllen writes with great authority across a wide range of texts, and scholars in the field will be grateful for his contribution. However, *English Dictionaries* is printed in a small font with tightly spaced lines. For this reason, I found it rather difficult to read. It uses an odd form of the Anglo-Saxon letter þ (e.g. p. 65), which was not immediately recognizable.

The division of material between the two books must have presented a number of difficult decisions. *English Dictionaries* is densely argued and packed full of information. It clearly would not have been practical to cover Roget's *Thesaurus* in this work as well. 'There was no space to draw parallels between the topical tradition and that of synonymy, and the philosophical changes after 1700, mainly instigated by the works of John Locke, could not be dealt with adequately in the onomasiological context, either' (Hüllen 2004: viii). Having discussed the earlier topical dictionaries in the first of these two monographs, it was unnecessary to cover them again in *A History of Roget's Thesaurus*. However, that left the question of what to cover, and it was not always immediately apparent why Hüllen had chosen to include some sections. As a result, the second of these two monographs seemed somewhat disjointed at times. Perhaps because this work will be of interest across a number of academic fields, I sometimes felt that the background information was more extensive than I needed. Readers in other fields will no doubt have the same feeling about completely different parts of the text. *A History of Roget's Thesaurus* ended rather abruptly. I wanted conclusions or a summary to help me process everything that I had learnt.

References

- Green, J. 1996. *Chasing the Sun. Dictionary-Makers and the Dictionaries They Made*. London: Jonathan Cape.

Landau, S.I. 1984. *Dictionaries. The Art and Craft of Lexicography*. Reprinted 1989. Cambridge: Cambridge University Press.

Roget, P.M. 1984. *Roget's Thesaurus*. Revised, updated and abridged by Susan Lloyd. Hamersmith: Penguin.

Julie M. Coleman
Department of English
University of Leicester
Leicester, United Kingdom
(jmc21@leicester.ac.uk)

Gabriele Stein. *Better Words: Evaluating EFL Dictionaries.* 2002, x + 246 pp. ISBN 0 85989 719 2 (Hb.). Exeter: University of Exeter Press. Price: £39.50.

Gabriele Stein, a Professor of English Linguistics at the University of Heidelberg, is one of the most authoritative academics in the study of English as a foreign language, and has published a wide range of research in the field of English lexicography. This book which is a collection of reprinted articles and previously unprinted papers, forms a cornerstone in EFL lexicography. It is a useful reference book for both EFL teachers and students alike, as well as for lexicographers.

Stein shows how lexicography has evolved over a 25-year period, and how certain information (word definitions, grammar information, idiomaticity, collocations, phrasal verbs, registers and pictorial illustrations amongst others which dictionaries tend to incorporate) has been included to suit the needs and expectations of its users, whatever their skills and background. The section on the way dictionaries have made use of illustrations more or less successfully to help language learners grasp the meaning and the cultural component which words are often unable to express, often verging on stereotyping and caricaturing, is particularly fascinating.

Reviewing the major developments that have taken place in lexicography over the last decades, she focuses on the suitability and efficiency of both monolingual general-purpose and monolingual EFL dictionaries, as well as bilingual EFL dictionaries in a learning and teaching environment. Armed with a dictionary, lay learners of English, for lack of expertise in the handling of dictionaries and grammars and with little knowledge of the often abstruse terminology they use, do not always find the linguistic information they are looking for or a suitable tool adapted to their needs and skills. This book intends to give teachers and learners of English guidelines on how to find their way through the dictionaries on hand, to make the most of them and to select the most adequate ones. It not only contains research into the compiling of dictionaries and the growing intense competition among various publishers in quest of an ideal corpus-based dictionary with compact and clearer definitions, but it also furnishes valuable help and suggestions on what has worked and what has not, thus preparing the way for a future dictionary in the making, which would be more user-friendly and transparent to its users.

As Stein admits, there is still a great deal of work to be done, not only in the compiling of definitions, but also in the explicit reference to grammatical behaviour, the use of authentic data to establish a core vocabulary which would be selected as sub-entries and to illustrate meanings, and the use of illustrations to ease the understanding and retention process. Reference to pragmatic use of language is something on which dictionaries should lay more emphasis. Also, the treatment of archaisms is something that would need further investigation.

The language used in the book is not too academic and makes it accessible to a wide audience. Nevertheless, it may be regrettable to see papers already published elsewhere collected in this book without a proper introduction or conclusion, which would have enabled the author to synthesize the main issues addressed and the innovative solutions provided. Can there still be innovation in the compilation of EFL dictionaries? Certainly Stein's research enables us to reflect on new possibilities which would improve the content and layout of dictionaries, enhance our knowledge of dictionaries and enable lexicographers to develop new strategies. As such, the collection is a landmark in the field of EFL lexicography and its reading can be but highly recommended.

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

Ursula Wingate. *The Effectiveness of Different Learner Dictionaries: An Investigation into the Use of Dictionaries for Reading Comprehension by Intermediate Learners of German.* 2002, X + 301 pp. ISBN 3-484-39112-X. Lexicographica. Series Major 112. Tübingen: Max Niemeyer. Price: €88.

Unlike the long tradition of EFL dictionaries, German learners' dictionaries are quite recent and were first marketed in the 1990s. Ursula Wingate's book is quite unique, since studies in German learners' dictionaries are very few, as is empirical research conducted in dictionary use.

Ursula Wingate is responsible for the development of students' study skills at King's College, London. Her expertise in applied linguistics and in learning strategies has been a particular asset in this research. This book particularly focuses on language learners' needs and investigates the linguistic suitability and effectiveness of dictionaries by direct observation of a target group of intermediate learners. This investigation entails studying a group of Hong Kong Chinese learners of German, assessing the efficiency of monolingual dictionaries and bilingual foreign-language German–English dictionaries in their vocabulary acquisition and in their reading comprehension. Statistics show that, on the whole, language learners of German prefer the use of bilingual German–English dictionaries to bilingual Chinese–English and monolingual dictionaries. Because of the cultural similarity between English and German, Hong Kong Chinese learners generally opt for the German–English version. Only 1% of the sample investigated happens to favour a monolingual dictionary.

This book, with its emphasis on language learners' needs, is particularly aimed at dictionary compilers and publishers, as well as teachers, who will find help in determining the most useful dictionary for their students at all stages of their acquisition process. Despite the general belief that a monolingual dictionary is too demanding at such a level, quantitative studies showed that monolingual dictionaries are more effective than Hong Kong Chinese learners' preferred bilingual German–English dictionaries and should be recommended.

In a further stage of this study, the effectiveness of dictionary definitions in monolingual dictionaries was assessed. The think-aloud studies conducted on a sample of language learners shed light on major strategy mistakes in dictionary consultation. Such a comprehensive study highlights the importance of training dictionary users in dictionary consultation, and the little awareness of teachers that reference skills could clearly enhance language acquisition. The habit of using bilingual dictionaries often accounts for unsuccessful performance when shifting to monolingual dictionaries. The think-aloud studies also provide insight into particular weaknesses in the presentation of words or definitions found in dictionaries.

As suggested by Wingate, it would in future be advisable to assess the suitability of both monolingual and bilingual dictionaries by a group of advanced learners. The findings of such research brings to light lexicographers'

and teachers' lack of awareness of users' needs and point out users' inefficient ways or lack of strategies in using dictionaries, as well as how these dictionaries sometimes fail in providing effective information to language learners. Wingate's recommendations and strategies are a major contribution to an effective dictionary for language learners, not only to German learners specifically, but also to language learners generally.

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

Rufus H. Gouws, Ulrich Heid, Wolfgang Schweickard and Herbert Ernst Wiegand (Editors). *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Special Focus on Computational Lexicography.* Berlin/New York: Mouton de Gruyter.*

The trilingual encyclopedia *Dictionaries. An International Encyclopedia of Lexicography* (edited by Franz Josef Hausmann, Oskar Reichmann, Herbert Ernst Wiegand and Ladislav Zgusta) was published between 1989–1991 in three volumes, consisting of a total of 335 articles divided into 38 chapters and printed on 3 355 pages including the index. The envisaged supplementary volume which will be published in English will consist of 22 chapters, containing 136 articles. This will not exceed 1 360 pages.

The supplementary volume pursues the objectives formulated in the preface of the first volume (HSK 5.1) which are complemented by others, following from the practice-internal and theoretical developments of the last 15 years. With the supplementary volume it is consequently endeavoured to achieve the following objectives:

- Expansion and extension of the presentation of lexicography — especially the printed and electronic linguistic lexicography — of all language circles of the world
- Continuation of the presentation of the status and function of lexicographic reference works up to the present within the cultural systems of societies, with special reference to the new digital forms of lexicography
- Continuation of the already presented history of lexicography by means of selected examples of larger cultures up to the present
- Expansion, deepening and modification of presented excerpts of a general theory of lexicography in its four sections, i.e. research into the use of dictionaries, research of dictionary criticism, and historical and systematic dictionary research, with the aim that metalexicography can document its academic status as scientific discipline with an independent formation of a theory, that has been established during the last decade
- Complementation of the description of all phases of the lexicographic process, especially by means of the presentation of the new methods based on electronic corpora, by means of the discussion of the new pos-

* The editors welcome suggestions for additions to these articles. Please send suggestions before 31 December 2006 to the following address: herbert.ernst.wiegand@gs.uni-heidelberg.de. The editors will give the necessary consideration to suggestions received.

sibilities of computer-assistance as well as the consideration of the linking possibilities of different types of lexicographic processes

- Presentation of lexicographic training and lexicographic institutions
- Presentation of new metalexigraphic methods, and, as focus of the volume: An extensive and detailed presentation of computational lexicography with regard to its historical development as well as all the important research fields, research perspectives, methods and tools of the last two decades.

The conception and grouping, with which the outlined objectives of the supplementary volume will be achieved, are briefly sketched in the following paragraphs. The Arabic numbers refer to the classification given at the end.

The supplementary volume has four chapters directed at new lexicographic concepts and the further development of new beginnings.

In Chapter I, **Dictionaries in Modern Society: Current Status and Perspectives**, the first article gives the present and future prospects, problems and development possibilities of printed dictionaries in their different typological forms. The collaboration in the lexicographic processes of printed dictionaries with corpus linguistic approaches is emphasized. In the second article a corresponding presentation is given of computational lexicography and its influence on linguistic lexicography and other fields of research.

In the last decade, textual structures of printed dictionaries have often been in the focus of lexicographic research. In Chapter II, **New Developments in Lexicographic Theory I: Textual Structures**, the eight articles (nr. 3-10) attempt to present this research as exhaustively as possible in the following manner: A review article briefly indicating all the types of textual structures is followed by a review article on the following selected types of structures: macrostructures, access structures, microstructures and mediostructures as well as textual architectures. The next article treats addressing and addressing structures and in the tenth and final article of this chapter a presentation is given of the different types of dictionary articles.

In Chapter III, **New Developments in Lexicographic Theory II: Dictionary Types and Functions**, new dictionary types and new developments in selected dictionary types are presented in ten articles (nr. 11-20). Consideration is also given to the lexicography of language for general purposes and languages for special purposes. Dictionaries of association, new types of cognitive dictionaries, language contact dictionaries, bridge dictionaries and recent developments in English, French as well as German learner dictionaries are treated. There are also review articles of new linguistic dictionaries as well as all the existing special purpose dictionaries of lexicography and those in progress. Article 20 concludes the chapter with a treatment of dictionary functions.

Chapter IV, **New Developments in Lexicographic Theory III: Selected Dictionary Subjects**, consists of eight articles (nr. 21-28). From the numerous themes, directed especially at aspects of dictionary subject matter, treated in recent metalexigraphic research, relevant and innovative aspects have been selected in such a way that consideration is given to a relative wide spectrum of problems. Culture-bound lexical items in monolingual and bilingual lexicography, language policy and lexicography, sensitive items, and meaning in prototype theory and in frame theory are treated in separate articles. Furthermore the concept of semiotaxis is presented as well as new developments in the lexicographic treatment of collocations and lexicographic examples.

The last chapter dealing with new developments in the theory of lexicography is Chapter V, **New Developments in Lexicographic Theory IV: Research in Dictionary Production and Use**, consisting of eight articles (nr. 29-36). The chapter starts with a presentation of the structure of lexicographic processes and their phases, contains a presentation of dictionary management, an article on the theory of dictionary use and review articles on empirical usage research and on methods developed and employed in recent research on dictionary use. Furthermore there are articles on the concept of simultaneous feedback, research on usage in the domain of electronic dictionaries and ways and means of establishing a dictionary culture.

The sixth chapter deals with organizational questions and is entitled: **New Developments in Lexicographic and Metalexigraphic Organization**. The five articles (nr. 37-41) present review articles on the training of lexicographers and training possibilities, the establishment of lexicographic units, the spectrum of existing lexicographic associations, principles of the evaluation of dictionaries and dictionary projects and aids playing a role in metalexigraphic research.

Chapter VI is followed by six chapters dealing with recent developments in the lexicography of single languages and the lexicography of single languages that could not have been dealt with in HSK 5.1-5.3.

Chapters VII to XII are dedicated to the lexicography of individual languages. The mutual aim of the descriptions are to document the development of lexicographic research since the start of the nineties of the 20th century. The point of departure of these supplementary articles is the treatment of the lexicography of individual languages presented in HSK 5.1-5.3. With this background the latest developments with regard to methodology and form, from both a synchronic and a diachronic perspective, are discussed and commented on. The following focal areas have been identified for the supplementary volume 5.4:

Chapter VII, **New Developments in the Lexicography of Individual Languages since 1990 I: The Ancient Languages of the Near East and the Classical Languages**, treats, by means of examples, the lexicography of the ancient Eastern languages as well as Greek and Latin. Chapter VIII, **New Developments in the Lexicography of Individual Languages since 1990 II: The Ro-**

mance Languages, presents the lexicography of the Romance languages (with the focus on the Iberoromance languages, French, Italian and Romanian). Chapter IX, **New Developments in the Lexicography of Individual Languages since 1990 III: The Germanic Languages**, is dedicated to the complete field of Germanic languages (Scandinavian languages, English and American English, German, Dutch and Afrikaans). Chapter X, **New Developments in the Lexicography of Individual Languages since 1990 IV: The Slavic Languages**, treats the lexicography of selected Slavic languages, divided into the domains of Eastern, Southern and Western Slavic languages.

Chapter XI, **Lexicography of Selected Asian Languages**, illustrates by means of the lexicography of Chinese, Japanese and Korean a selection of the total domain of the Asian languages. Chapter XII, **Lexicography of Selected African Languages**, contains eight articles (nr. 64-71) focusing on the lexicography of certain language families (the Nguni and Sotho languages), individual languages (Shona, Chiluba, Fang and Swahili) and the languages of certain geographical regions (Central Africa and Western Africa). The articles in this chapter complement the limited presentation of the lexicography of the African languages presented in HSK 5.3 and recognize the rapid development of lexicography on the African continent.

After the treatment of the lexicography of selected single languages follow ten chapters on computational lexicography.

The section on computational lexicography starts with Chapter XIII, **The History of Computational Lexicography**, giving an overview of the computational lexicographic work of the 1950s and 1960s and, in a second article, an overview of the time before 1970–1980.

The subsequent two chapters deal with typologies of electronic dictionaries, one for interactive dictionaries and one for dictionaries for natural language processing (NLP).

Chapter XIV, **Typology of Electronic Dictionaries I: Electronic Dictionaries for Human Use**, consists of nine articles (nr. 74-82). The first three cover design criteria for interactive electronic dictionaries, the 'added value' that the user may get from the computational medium, as well as principles of user interface design for electronic dictionaries for human use. The subsequent articles are devoted to electronic versions of printed dictionaries, Internet dictionaries, electronic learner's dictionaries and context-sensitive dictionaries, i.e. those that guide the reader of an online text to the relevant (e.g. idiomatic) reading of a word. Finally, the overview is completed by articles on large-scale documentary electronic dictionaries, electronic encyclopedias and products that combine dictionary look-up with corpus search.

Chapter XV, **Typology of Electronic Dictionaries II: Electronic Dictionaries for Machine Use**, is still typological in nature, but focuses on NLP applications. Individual articles describe lexical needs, representation techniques, descriptive approaches and major realizations in each of the following subdomains of NLP: speech recognition and speech synthesis, text processing and

spelling or grammar checkers, information retrieval, information extraction and data mining, question answering, dialogue systems, natural language understanding systems, text generation systems, machine translation and computer-assisted language learning. The last article discusses generic large-scale electronic dictionaries intended for multiple applications.

Chapter XVI focuses on lexical presentation formalisms: **Models for the Representation of Dictionaries: The Form Aspect**. In seven articles (nr. 93-99), it deals with the major representation formalisms for electronic dictionary resources: XML (and SGML), databases, feature structure-based formalisms, formalisms of knowledge representation as used in artificial intelligence, and hypertext. One article discusses dedicated formalisms for lexical representation and one describes objectives and topics of international standardization in the field of representation formalisms for electronic dictionaries.

Chapter XVII, **Models for the Representation of Linguistic Data in Electronic Dictionaries: The Content Aspect**, is subdivided into twelve articles (nr. 100-111). The first seven articles are each devoted to one 'level of linguistic description', its needs, data types, coding and representation schemes. These articles cover pronunciation, inflectional and word formation morphology, syntax, semantics and pragmatics, as well as dictionaries which explicitly attempt describing and representing the interaction between syntax and semantics. The next article treats idiomaticity in electronic dictionaries, followed by one on multimodal dictionaries. Two articles are devoted to international standardization work for the linguistic description of lexical material in monolingual as well as in bilingual and multilingual electronic resources. The last article of this chapter describes typologies of metadata for electronic dictionaries.

Chapter XVIII, **Models for Non-alphabetical Computational Dictionaries**, is entirely devoted to WordNet and its developments and applications, to resources with world knowledge (such as CYC) and to both linguistic ontologies and ontologies constructed in research towards the Semantic Web. Besides reporting facts about the current state-of-the-art, these four articles (nr. 112-115) are also intended to allow for a comparison between the different approaches.

So far, all chapters of this part of the handbook concern the electronic dictionary as a product: its types, its contents, and its form. Chapters XIX, XX and XXI are devoted, however, to dictionary making and to all its computational aspects. As most dictionary making is nowadays supported to some degree by computational tools, these chapters obviously concern both electronic and paper dictionaries, for men and machines. One chapter is devoted to corpus design for lexicography, one to tools for data acquisition and one to support for dictionary writing.

Chapter XIX, **Computer-based Dictionary Making I: Acquisition of Lexical Data from Corpora — Corpus Design**, encompasses four methodological articles, two on corpus design criteria for lexicography (monolingual and bilingual), one on existing large corpora for lexicography (as e.g. many of the na-

tional corpus projects), and a last one on possibilities and limitations of the use of the World Wide Web for lexicography. Chapter XX, **Computer-based Dictionary Making II: Acquisition of Lexical Data from Corpora and Machine-readable Dictionaries — Tools and Procedures**, deals with tools for lexicographers. Six of its nine articles (nr. 120-125) cover the major aspects of corpus-based data provision for dictionary making: corpus pre-processing, the corpus-based design of a macrostructure, concordancing, the extraction from corpora of data on syntactic and collocational properties of words, and tools for the use of parallel and comparable corpora. The three other articles (nr. 126-128) concern the use of electronic versions of dictionaries as data sources: retrodigitization, the analysis of electronic versions of print dictionaries and the reversal of bilingual dictionaries $A \rightarrow B$ to make them usable as a starting point for a dictionary of the language pair $B \rightarrow A$. Finally, Chapter XXI, **Computer-based Dictionary Making III: Computational Support for the Writing of Dictionary Articles**, gives a state-of-the-art overview of tools for dictionary writing and for consistency control, of workbenches for corpus-based lexicography and of tools supporting lexicographers in dictionary updating by means of a comparison between data from an existing dictionary and the results of corpus analysis.

Computational work in terminology is the topic of Chapter XXII: **Computational Terminography**. As terminology and terminography are extensively dealt with in the HSK handbook on specialized language, only truly computational aspects of terminography are dealt with here: the representation of terminological data in terminological databases and in formats for the exchange of terminological data, and approaches to and challenges in the extraction of terminological data from texts. Finally, the chapter includes an article on existing large-scale termbanks.

I. Dictionaries in Modern Society: Current Status and Perspectives

1. Printed dictionaries in the electronic era
2. The impact of computational lexicography

II. New Developments in Lexicographic Theory I: Textual Structures

3. Textual structures in printed dictionaries: An overview
4. Macrostructures in printed dictionaries: An overview
5. Access structures in printed dictionaries: An overview
6. Microstructures in printed dictionaries: An overview
7. Mediostructural components in printed dictionaries: An overview
8. Textual architectures in printed dictionaries: An overview
9. Addressing and addressing structures in printed dictionaries: An overview
10. Types of dictionary articles in printed dictionaries: An overview

III. New Developments in Lexicographic Theory II: Dictionary Types and Functions

11. Dictionary of associations
12. New types of cognitive dictionaries
13. Dictionary of language contact

14. Bridge dictionaries
15. New developments in learners' dictionaries I: English and French
16. New developments in learners' dictionaries II: German
17. New developments in lexicography for special purposes I: An overview of linguistic dictionaries
18. New developments in lexicography for special purposes II: An overview of dictionaries of lexicography
19. New types of historical dictionaries
20. Lexicographic functions
- IV. New Developments in Lexicographic Theory III: Selected Dictionary Subjects**
21. The treatment of culture-bound items in dictionaries
22. Language policy and lexicography
23. The treatment of sensitive items in dictionaries
24. The treatment of meaning in dictionaries and prototype theory
25. The treatment of meaning in dictionaries and frame theory
26. The concept of semiotaxis
27. New tendencies in the treatment of collocations
28. New developments in the selection of examples
- V. New Developments in Lexicographic Theory IV: Research in Dictionary Production and Use**
29. The lexicographic process
30. Theory of dictionary management
31. Theory of procedures of dictionary use
32. Empirical research in dictionary use since 1990
33. Methods in the research of dictionary use
34. The concept of simultaneous feedback
35. User research in the field of electronic dictionaries: Methods, first results
36. Ways and means to establish a dictionary culture
- VI. New Developments in Lexicographic and Metalexicographic Organization**
37. Lexicographic training: An overview
38. Establishing lexicography units
39. Lexicographic associations
40. Evaluation of dictionaries
41. Aids in metalexicographic research
- VII. New Developments in the Lexicography of Individual Languages since 1990 I: The Ancient Languages of the Near East and the Classical Languages**
42. The ancient languages of the Near East
43. The ancient languages of Greek and Latin
- VIII. New Developments in the Lexicography of Individual Languages since 1990 II: The Romance Languages**
44. Iberoromance I: Historical and etymological lexicography
45. Iberoromance II: Synchronic lexicography
46. French I: Historical and etymological lexicography
47. French II: Synchronic lexicography

- 48. Italian I: Historical and etymological lexicography
- 49. Italian II: Synchronic lexicography
- 50. Romanian
- IX. New Developments in the Lexicography of Individual Languages since 1990**
 - III: The Germanic Languages**
 - 51. Scandinavian languages
 - 52. English and American I: Historical and etymological lexicography
 - 53. English and American II: Synchronic lexicography
 - 54. German I: Historical and etymological lexicography
 - 55. German II: Synchronic lexicography
 - 56. Dutch
 - 57. Afrikaans
- X. New Developments in the Lexicography of Individual Languages since 1990**
 - IV: The Slavic Languages**
 - 58. Eastern Slavic languages
 - 59. Southern Slavic languages
 - 60. Western Slavic languages
- XI. Lexicography of Selected Asian Languages**
 - 61. Chinese
 - 62. Japanese
 - 63. Korean
- XII. Lexicography of Selected African Languages**
 - 64. Lexicography of the Nguni languages
 - 65. Lexicography of the Sotho languages
 - 66. Lexicography of Shona
 - 67. Lexicography of the languages of Central Africa
 - 68. Lexicography of Chiluba
 - 69. Lexicography of Fang
 - 70. Lexicography of Swahili
 - 71. Lexicography of the languages of Western Africa
- XIII. The History of Computational Lexicography**
 - 72. The early history of computational lexicography: The 1950s and 1960s
 - 73. Computational lexicography and corpus linguistics until ca 1970–1980
- XIV. Typology of Electronic Dictionaries I: Electronic Dictionaries for Human Use**
 - 74. Design criteria and 'added value' of electronic dictionaries for human users
 - 75. Principles of user interfaces for electronic dictionaries for human users
 - 76. Electronic versions of printed dictionaries
 - 77. Electronic pedagogical dictionaries
 - 78. Context-sensitive electronic dictionaries
 - 79. Monolingual and bilingual electronic dictionaries on the Internet
 - 80. Large-scale documentary dictionaries on the Internet
 - 81. Electronic encyclopedias
 - 82. Combined products: Dictionary and corpus

XV. Typology of Electronic Dictionaries II: Electronic Dictionaries for Machine Use

83. Electronic dictionaries for speech recognition, for its applications and for speech synthesis
84. Electronic dictionaries for text processing and correction systems
85. Electronic dictionaries for information retrieval, information extraction and text mining
86. Electronic dictionaries for question answering and natural language database access
87. Electronic dictionaries for computer-assisted language learning
88. Electronic dictionaries for natural language understanding
89. Electronic dictionaries for spoken and multimodal dialogue systems
90. Electronic dictionaries for applications of natural language generation
91. Electronic dictionaries for machine translation and machine-aided translation
92. Generic electronic dictionaries for multiple uses in NLP

XVI. Models for the Representation of Dictionaries: The Form Aspect

93. Representing human and machine dictionaries in markup languages (SGML, XML)
94. Representing computational dictionaries in relational databases
95. Representing computational dictionaries in feature structure-based representation formalisms and typed feature logic
96. Representing computational dictionaries in AI-oriented knowledge representation formalisms: Dictionaries within expert systems
97. Representing computational dictionaries in hypertextual form
98. Representing computational dictionaries in dedicated formalisms: DATR, Word-Manager
99. Standardization of the formal representation of lexical information for NLP

XVII. Models for the Representation of Linguistic Data in Electronic Dictionaries: The Content Aspect

100. Coding pronunciation in computational dictionaries
101. Coding inflectional morphology in computational dictionaries
102. Coding word formation morphology in computational dictionaries
103. Coding syntactic properties of words in computational dictionaries
104. Coding semantic properties of words in computational dictionaries
105. Computational dictionaries relating semantic and syntax: the FrameNet approach
106. Coding pragmatic properties of words in computational dictionaries
107. Coding idiomatic expressions in computational dictionaries
108. Coding of multimodal dictionaries
109. Standardization of the coding of monolingual linguistic information in computational dictionaries
110. Standardization of the coding of bilingual and multilingual linguistic information in computational dictionaries
111. Typologization of metadata of electronic dictionaries

XVIII. Models for Non-alphabetical Computational Dictionaries

- 112. Representing world knowledge: CYC
- 113. WordNet: Principles, developments and applications
- 114. Ontologies for natural language processing
- 115. Ontologies for the Semantic Web: Principles, formal considerations, resource building, tools and applications

XIX. Computer-based Dictionary Making I: Acquisition of Lexical Data from Corpora — Corpus Design

- 116. Criteria for the design of corpora for lexicography I: Monolingual dictionaries
- 117. Criteria for the design of corpora for lexicography II: Bilingual dictionaries
- 118. Large corpora designed for lexicographic work
- 119. The World Wide Web as a resource for lexicography

XX. Computer-based Dictionary Making II: Acquisition of Lexical Data from Corpora and Machine-readable Dictionaries — Tools and Procedures

- 120. Corpus pre-processing for lexicography
- 121. Tools to support the design of a macrostructure (nomenclature)
- 122. Concordance tools for lexicography
- 123. Tools and procedures for the acquisition of morphological and syntactic information from corpora
- 124. Tools for the acquisition of significant lexical combinatorics
- 125. Tools for the lexicographic use of parallel and comparable corpora
- 126. Analyzing machine-readable dictionaries — Reformatting and reinterpretation
- 127. Reversal of bilingual dictionaries
- 128. Retrodigitization: From paper dictionaries to electronic dictionaries

XXI. Computer-based Dictionary Making III: Computational Support for the Writing of Dictionary Articles

- 129. Tools for dictionary writing and consistency control
- 130. Workbenches for corpus-based lexicography
- 131. Tools for dictionary updating: Comparing dictionary and corpus data

XXII. Computational Terminography

- 132. Models for the representation of terminological data on the computer I: Terminological databases
- 133. Models for the representation of terminological data on the computer II: Terminological exchange formats
- 134. Large terminological databases
- 135. Acquisition of terminological data from text I: Approaches
- 136. Acquisition of terminological data from text II: Term variation, concept relations, etc.

Stefan J. Schierholz and Herbert Ernst Wiegand (Editors). *Wörterbücher zur Sprach- und Kommunikationswissenschaft (WSK)* [Dictionaries of Linguistics and Communication Science]

The WSK series consists of alphabetically organized specialized dictionaries for the subfields of the overall conceptual domain ("linguistics and communication science"). At present 23 volumes have been planned, while additional volumes are envisaged. As can be seen from the following list of dictionaries, the structure of the WSK volumes chiefly follows traditional dispositions. The names of the editors for the volumes whose collaboration is certain are given next to the titles.

1. *Grammar* (2 subvolumes: 1.1 *Morphology*; 1.2 *Syntax*) (Christa Dürscheid and Stefan J. Schierholz)
2. *Word Formation* (Peter O. Müller and Susan Olsen)
3. *Historical Linguistics* (Mechthild Habermann and Markus Hundt)
4. *Phonetics and Phonology* (Tracy Alan Hall and Bernd Pompino-Marschall)
5. *Writing* (Martin Neef and Rüdiger Weingarten)
6. *Text Linguistics and Stylistics* (Jannis Androutsopoulos and Stephan Habscheid)
7. *Language Technology and Computational Linguistics* (Dafydd Gibbon and Ulrich Schmitz)
8. *Dialectology* (Heiko Girth and Franz Patocka)
9. *Quantitative and Formal Linguistics* (Peter Grzybek and Reinhard Köhler)
10. *Semantics and Pragmatics* (Ralf Klabunde and Beatrice Primus)
11. *Linguistic Theory and Methodology* (Johannes Kabatek and Bernd Kortmann)
12. *Lexicology and Phraseology* (Christiane Fellbaum and Ingo Warnke)
13. *Media Sciences* (Irmela Schneider and Peter M. Spangenberg)
14. *Cognitive Grammar* (Susanne Niemeier and Doris Schönefeld)
15. *Sociolinguistics* (Stephan Elspaß and Ekkehard Felder)
16. *Philosophy of Language* (Christoph Demmerling and Pirmin Stekeler-Weithofer)
17. *Language Teaching: Native and Foreign Language* (Jörg Kilian and Jutta Rymarczyk)
18. *Language Typology/General Linguistics* (Johannes Helmbrecht and N.N.)

19. *Semiotics* (Guido Ipsen and N.N.)
20. *Conversation Linguistics*
21. *Lexicography and Dictionary Research*
22. *Psycho- and Neurolinguistics* (Christina Kauschke and Prisca Stenneken)
23. *Language and Variety Names*

The intended users of the WSK can be divided into three groups:

- (a) Students in all philological or linguistic disciplines, both in Germany and abroad, who need to read professional literature in German and English. These are considered semi-experts.
- (b) Academic instructors, in Germany and abroad, who are the teachers of the students in group (a). These are considered experts in some fields and semi-experts in others.
- (c) University graduates who have taken their degrees in a philological and/or linguistic discipline and whose professions draw on their field(s) of study. These primarily include secondary school teachers but also smaller groups such as specialized translators and terminologists considered experts in some fields and semi-experts in others.

All WSK volumes will be multifunctional, with the various dictionary functions weighted as follows:

- (a) Primary functions of the dictionary include:
 - aid in the comprehension of texts ("comprehension function"), and
 - provision of specialized information on a given field ("information function").
- (b) The secondary function of the dictionaries is:
 - aid in translation ("translation function").

These dictionary functions cover at least 120 clearly distinguishable types of usage situations for a particular dictionary.

The WSK volumes contain a front matter section, a word-list and a back matter section.

The most important part of the front matter is the systematic introduction to the special field covered by the entries in the word-list. This introduction is divided into paragraphs and stretches across three or four printed sheets. At the end of each paragraph there are references to all synopsis articles belonging together; it is possible to refer to paragraphs of the systematic introduction in dictionary articles, creating bidirectional references.

The word-list consists of single articles, synopsis articles and cross-reference articles. Single articles convey the specialized knowledge pertaining to a specific term. Synopsis articles contain a synopsis of the specialized knowledge concerning semantically related terms, which in turn are covered in single articles. The structure of single and synopsis articles is identical, despite quantitative differences: The German lemma is followed by its definition, then by the English equivalent and its definition. Each article thus conveys semantic knowledge in German and English at its very beginning. What follows is the position for the further explanation which is written in German and which supplements the terminological knowledge with specialized encyclopedic knowledge. Two optional article positions follow, namely the position for synonyms and the position for antonyms, then succeeded by the obligatory position for cross-references to the systematic introduction, other articles or other WSK volumes. Each article ends with an obligatory position for literature.

The most important text of the back matter is the index for English–German equivalents which enables the user to access dictionary articles when reading or translating an English text. The WSK volumes constitute the dictionary type "specialist study and reference dictionary" and take the status of English as a scientific and research language into consideration. Hence all bear the following subtitle: *Study and Reference Dictionary. With a Systematic Introduction and English Translations*. The specific properties of the WSK volumes help to ensure quick look-ups as well as systematic dictionary-intrinsic searches.

The first volumes of the WSK are due to be published during 2008–2009. Over 40 000 terms of the overall conceptual domain "linguistics and communication science" will be covered in all currently planned volumes. After the release of the first volumes, these will continually be converted into one online version.

Further information on the WSK is available at <www.wsk.uni-erlangen.de> and in the following publications:

- Schierholz, Stefan J. and Herbert Ernst Wiegand.** 2005. Die Wörterbücher zur Sprach- und Kommunikationswissenschaft (WSK). Eine neue Konzeption der linguistischen Fachlexikographie und ihre computergestützte Praxis. *Lexicographica* 20: 164-264.
- Wiegand, Herbert Ernst.** 2004. *Die Wörterbücher zur Sprach- und Kommunikationswissenschaft (WSK) und ihre Benutzungsmöglichkeiten im Fach Deutsch als Fremdsprache*. Paper presented at the 3. Internationales Kolloquium zur Wörterbuchforschung, held in Varna, Bulgaria, October 2004. [Appears 2006]
- Wiegand, Herbert Ernst.** 2006. Das Lern- und Konsultationswörterbuch. Ein neuer Fachwörterbuchttyp am Beispiel der *Wörterbücher zur Sprach- und Kommunikationswissenschaft (WSK)*. *Lexikos* 16: 205-221.

Publikasieaankondigings / Publication Announcements

John Bennet and Nthuseng Tsoeu (Editors). *Multilingual Illustrated Dictionary: English, IsiZulu, Sesotho, IsiXhosa, Setswana, Afrikaans, Sepedi*. 2006, 373 pp. ISBN-10: 07021-6712-6 / ISBN-13: 978-07021-6712-6. Cape Town: Pharos/Juta Gariep.

Julie Coleman and Anne McDermott (Editors). *Historical Dictionaries and Historical Dictionary Research: Papers from the International Conference on Historical Lexicography and Lexicology at the University of Leicester, 2002*. 2004, vii + 224 pp. *Lexicographica*. Series Maior 123. ISBN 3-484-39123-5 (Pb.). Tübingen: Max Niemeyer. Price: €68. (Review in this issue.)

R.H. Gouws and D.J. Prinsloo. *Principles and Practice of South African Lexicography*. 2005, i + 210 pp. ISBN 1-919980-88-1 (Pb.). Stellenbosch: SUN Press. Price: R200.

Thomas Herbst and Kerstin Popp (Editors). *The Perfect Learners' Dictionary (?)*. 1999, XII + 320 pp. ISBN 3-484-30995-4. *Lexicographica*. Series Maior 95. Tübingen: Max Niemeyer. Price: €88. (Review in this issue.)

Thomas Herbst, Gunter Lorenz, Brigitta Mittmann und Martin Schnell (Herausgeber). *Lexikografie, ihre Basis- und Nachbarwissenschaften: (Englische) Wörterbücher zwischen "common sense" und angewandter Theorie*. 2004, VI + 207 pp. ISBN 3-484-39118-9. *Lexicographica*. Series Maior 118. Tübingen: Max Niemeyer. Price: €72. (Review in this issue.)

Henry Hitchens. *Dr Johnson's Dictionary: The Extraordinary Story of the Book that Defined the World*. 2005, 278 pp. ISBN 0-71-956631-2. London: John Murray.

Werner Hüllen. *English Dictionaries 800–1700: The Topical Tradition*. 1999, xvii + 525 pp. ISBN 0-19-823796-0 (Hb.). Oxford: Oxford University Press. Price: £110. (Review in this issue.)

Werner Hüllen. *A History of Roget's Thesaurus: Origins, Development, and Design*. 2004, ix + 410 pp. ISBN 0-19-925472-9 (Hb.). Oxford: Oxford University Press. Price: £75. (Review in this issue.)

Christa Kilian-Hatz. *Khwe Dictionary. With a Supplement on Khwe Place-names of West Caprivi* by Matthias Brenzinger. 2003, 431 pp. ISBN 10 3-89645-083-2. *Namibian African Studies 7*. Cologne: Rüdiger Köppe Verlag. Price: €52.80.

- Roger Mickala Manfoumbi. *Lexique pove-français/français-pove*. 2004, 761pp. ISBN 2-912776-43-0. Libreville: Éditions Raponda-Walker. (Article de revue dans ce numéro.)
- B.M. Mini (Director and Editor-in-Chief), S.L. Tshabe (Editor), F.M. Shoba and P.N. van der Westhuizen (Assistant Editors), and T.A. Ndungane, A.M. Pienaar and T.X. Mfana (Former Editors). *The Greater Dictionary of IsiXhosa. Volume 2: K to P*. 2003, xlviii + 1126 pp. ISBN 0 9499 7485 4. Alice: IsiXhosa National Lexicography Unit, University of Fort Hare. Price: R390.
- Lynda Mugglestone. *Lost for Words: The Hidden History of the Oxford English Dictionary*. 2005, xxi + 273 pp. ISBN 0 30 010699 8. New Haven (Conn.): Yale University Press.
- Gabriele Stein. *Better Words: Evaluating EFL Dictionaries*. 2002, x + 246 pp. ISBN 0 85989 719 2 (Hb.). Exeter: University of Exeter Press. Price: £39.50. (Review in this issue.)
- Simon Winchester. *The Meaning of Everything: The Story of the Oxford English Dictionary*. 2003, xxv + 260 pp. ISBN 0 19 860702 4. Oxford: Oxford University Press.
- Ursula Wingate. *The Effectiveness of Different Learner Dictionaries: An Investigation into the Use of Dictionaries for Reading Comprehension by Intermediate Learners of German*. 2002, X + 301 pp. ISBN 3-484-39112-X. Lexicographica. Series Maior 112. Tübingen: Max Niemeyer. Price: €88. (Review in this issue.)

VOORSKRIFTE AAN SKRYWERS

(Tree asseblief met die Buro van die WAT in verbinding (wat@sun.ac.za) vir 'n uitvoeriger weergawe van hierdie instruksies of besoek ons webblad: <http://www.sun.ac.za/wat>)

A. REDAKSIONELE BELEID

1. Aard en inhoud van artikels

Artikels kan handel oor die suiwer leksikografie of oor implikasies wat aanverwante terreine, bv. linguïstiek, algemene taalwetenskap, rekenaarwetenskap en bestuurskunde vir die leksikografie het.

Bydraes kan onder enigeen van die volgende rubrieke geklassifiseer word:

(1) **Artikels:** Grondige oorspronklike wetenskaplike navorsing wat gedoen en die resultate wat verkry is, of bestaande navorsingsresultate en ander feite wat op 'n oorspronklike wyse oorsigtelik, interpreterend, vergelykend of krities evalueerend aangebied word.

(2) **Resensieartikels:** Navorsingsartikels wat in die vorm van 'n kritiese resensie van een of meer gepubliseerde wetenskaplike bronne aangebied word.

Bydraes in kategorieë (1) en (2) word aan streng anonieme keuring deur onafhanklike akademiese vakgenote onderwerp ten einde die internasionale navorsingsgehalte daarvan te verseker.

(3) **Resensies:** 'n Ontleding en kritiese evaluering van gepubliseerde wetenskaplike bronne en produkte, soos boeke en rekenaarprogramme.

(4) **Projekte:** Besprekings van leksikografiese projekte.

(5) **Leksikonotas:** Enige artikel wat praktykgerigte inligting, voorstelle, probleme, vrae, kommentaar en oplossings betreffende die leksikografie bevat.

(6) **Leksikovaria:** Enigeen van 'n groot verskeidenheid artikels, aankondigings en nuusvystellings van leksikografiese verenigings wat veral vir die praktiserende leksikograaf van waarde sal wees.

(7) **Verslae:** Verslae van konferensies en werksessies.

Bydraes in kategorieë (3)-(7) moet almal aan die eise van akademiese geskrifte voldoen en word met die oog hierop deur die redaksie gekeur.

2. Wetenskaplike standaard en keuringsprosedure

Lexikos is deur die Departement van Onderwys van die Suid-Afrikaanse Regering as 'n gesubsidieerde d.w.s. inkomstegenererende navorsingstydskrif goedgekeur.

Artikels sal op grond van die volgende aspekte beoordeel word: taal en styl; saaklikheid en verstaanbaarheid; probleemstelling, beredenering en gevolgtrekking; verwysing na die belangrikste en jongste literatuur; wesenlike bydrae tot die spesifieke vakgebied.

3. Taal van bydraes

Afrikaans, Duits, Engels, Frans of Nederlands.

4. Kopiereg

Nóg die Buro van die WAT nóg die African Association for Lexicography (AFRILEX) aanvaar enige aanspreeklikheid vir eise wat uit meewerkende skrywers se gebruik van materiaal uit ander bronne mag spruit.

Outeursreg op alle materiaal wat in *Lexikos* gepu-

bliseer is, berus by die Beheerraad 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 rekenaar-skyf (360 KB tot 1.44 MB) voorgelê word. 'n Rekenaardrukstuk van die artikel moet die skyf vergesel. Elke artikel moet voorsien wees van 'n Engelse opsomming van tussen 150 en 250 woorde, sowel as tussen 10 en 30 Engelse sleutelwoorde.

2. Grafika

Een stel duidelike oorspronklike illustrasies, tabelle, grafieke, diagramme, of kwaliteitsafdrukke daarvan, moet voorgelê word. Die plasing van grafika binne die teks moet duidelik aangedui word.

3. Bibliografiese gegewens en verwysings binne die teks

Kyk na onlangse nommers van *Lexikos* vir meer inligting.

INSTRUCTIONS TO AUTHORS

(For a more detailed version of these instructions, please contact the Bureau of the WAT (wat@sun.ac.za) or refer to our website: <<http://www.sun.ac.za/wat>>)

A. EDITORIAL POLICY

1. Type and content of articles

Articles may deal with pure lexicography or with the implications that related fields such as linguistics, general linguistics, computer science and management have for lexicography.

Contributions may be classified in any one of the following categories:

(1) **Articles:** Fundamentally original scientific research that has been done and the results that have been obtained, or reflecting existing research results and other facts in an original, synoptic, interpretative, comparative or critically evaluative manner.

(2) **Review articles:** Research articles presented in the form of a critical review of one or more published scientific sources.

Contributions in categories (1) and (2) are subjected to strict anonymous evaluation by independent academic peers in order to ensure the international research quality thereof.

(3) **Reviews:** An analysis and critical evaluation of published scientific sources and products, such as books and computer software.

(4) **Projects:** Discussions of lexicographical projects.

(5) **Lexiconotes:** Any article containing practice-oriented information, suggestions, problems, questions, commentary and solutions regarding lexicography.

(6) **Lexicovaria:** Any of a large variety of articles, announcements and press releases by lexicographic societies which are of particular value to the practising lexicographer.

(7) **Reports:** Reports on conferences and workshops.

Contributions in categories (3)-(7) must all meet the requirements of academic writing and are evaluated by the editors with this in mind.

2. Academic standard and evaluation procedure

The Department of Education of the South African Government has approved *Lexikos* as a subsidized, i.e. income-generating research journal.

Articles will be evaluated on the following aspects: language and style; conciseness and comprehensibility; problem formulation, reasoning and conclusion; references to the most important and most recent literature; substantial contribution to the specific discipline.

3. Language of contributions

Afrikaans, Dutch, English, French or German.

4. Copyright

Neither the Bureau of the WAT nor the African Association for Lexicography (AFRILEX) accepts any responsibility for claims which may arise from contributing authors' use of material from other sources.

Copyright of all material published in *Lexikos* will be vested in the Board of Control of the Woordeboek van die Afrikaanse Taal. Authors are free however to use their material elsewhere provided that *Lexikos* (AFRILEX Series) is acknowledged as the original publication source.

5. Originality

Only original contributions will be considered for publication. Authors bear full responsibility for the originality and factual content of their contributions.

6. Free offprints and copies

Authors will receive five free offprints of each of their articles or review articles published, as well as one complimentary copy of the issue containing such article(s). Authors of purely evaluative reviews and of contributions to the categories Lexiconotes, Lexicovaria, Projects, and Reports receive five free offprints of their contributions. In the case of the latter four categories, the editors may, however, depending on the nature and scope of the contributions, decide to grant the author a copy of the issue concerned.

7. Invitation and editorial address

All interested authors are invited to submit contributions for publication in *Lexikos* to:

The Editor: LEXIKOS
Bureau of the WAT
P.O. Box 245
7599 STELLENBOSCH
Republic of South Africa

B. PREPARATION OF MANUSCRIPTS

Manuscripts of articles must meet the following editorial requirements:

1. Length and format

Contributions should not exceed more than 20 type-written A4 pages with double spacing and ample margins (about 2,5 cms). Manuscript should preferably be in electronic form on a (360 KB to 1.44 MB) floppy disk as either ASCII text, fully-formatted Microsoft Word (DOS or Windows) or WordPerfect (DOS or Windows) files. A computer printout of the article should accompany the disk. Each article must be accompanied by an English abstract of 150 to 250 words, and between 10 and 30 English keywords.

2. Graphics

One set of clear original drawings, tables, graphs, diagrams or quality prints thereof must be submitted. The locations of graphics must be clearly indicated in the text.

3. Bibliographical details and references in the text

Examine recent issues of *Lexikos* for details.

HINWEISE UND RICHTLINIEN FÜR AUTOREN

(Nehmen Sie bitte mit dem Büro des WAT Kontakt auf (wat@sun.ac.za) für eine ausführlichere Wiedergabe dieser Hinweise oder besuchen Sie unsere Webseite: <<http://www.sun.ac.za/wat>>)

A. REDAKTIONELLE ZIELSETZUNGEN

1. Art und Inhalt der Artikel

Es können Artikel aufgenommen werden, die sich mit Themen der Lexikographie befassen oder mit Zusammenhängen, die zwischen der Lexikographie und benachbarten Fachgebieten wie z.B. Linguistik, allgemeiner Sprachwissenschaft, Lexikologie, Computerwissenschaft und Management bestehen.

Die Beiträge sollten einer der folgenden Kategorien entsprechen:

(1) **Artikel**, die grundlegend über neue Forschungsansätze und deren Ergebnisse berichten, oder die bestehende Forschungsergebnisse und andere Informationen 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 Aufsichtsrat unseres Büros übertragen. Es steht Autoren jedoch frei, ihren Beitrag anderweitig zu verwenden, vorausgesetzt, *Lexikos* (AFRILEX-Serie) wird als Originalquelle genannt.

5. Originalität

Nur Originalbeiträge werden begutachtet. Autoren tragen die volle Verantwortung für die Originalität und den sachlichen Inhalt ihrer Beiträge.

6. Sonderdrucke und Freixemplare

Autoren erhalten fünf Sonderdrucke ihrer veröffentlichten Artikel oder Rezensionsartikel gratis sowie ein Freixemplar der betreffenden Ausgabe. Rezensenten und Autoren von Beiträgen zu den Kategorien Notizen zum Lexikon, Lexikovaria, Projekte und Berichte erhalten fünf Sonderdrucke ihrer Beiträge gratis. Die Redaktion kann sich jedoch, abhängig von der Art und dem Umfang der Beiträge der letztgenannten vier Kategorien, vorbehalten, dem Autor ein Freixemplar der Ausgabe zu überlassen.

7. Einladung und redaktionelle Adresse

Alle Autoren, die interessiert sind, Beiträge für *Lexikos* zu liefern, sind herzlich willkommen. Sie werden gebeten, ihre Artikel an die folgende Adresse zu schicken:

Der Redakteur: LEXIKOS
Buro van die WAT
Postfach 245
7599 STELLENBOSCH
Republik Südafrika

B. VORBEREITUNG DES MANUSKRIPTS

Ein Artikelmanuskript muß den folgenden redaktionellen Anforderungen entsprechen:

1. Umfang und Format

Beiträge sollen nicht länger als 20 getippte A4-Seiten in zweizeiligem Abstand und mit Randabständen von ca. 2,5 cm sein. Das Manuskript sollte möglichst als elektronischer Text auf einer (360 KB bis 1.44 MB) Diskette vorgelegt werden, entweder im ASCII-Format, oder in formatiertem Microsoft Word (DOS oder Windows) bzw. WordPerfect (DOS oder Windows). Ein Ausdruck des vollständig formatierten Artikels soll mit der Diskette eingereicht werden. Jedem Artikel ist eine Zusammenfassung im Umfang von 150-250 Wörtern beizufügen. Ferner sollen etwa 10-30 inhaltskennzeichnende Stichwörter zu jedem Artikel angegeben werden.

2. Abbildungen

Ein reproduktionsfähiger Satz der originalen Abbildungen, Illustrationen, Tabellen, Graphiken und Diagramme oder Qualitätsabdrucke muß vorgelegt werden. Der Text selber sollte klare Hinweise auf die Position der Abbildungen enthalten.

3. Bibliographische Einzelheiten und Hinweise im Text

Zu Einzelheiten des bibliographischen Systems sind neuere Ausgaben von *Lexikos* einzusehen.

INSTRUCTIONS AUX AUTEURS

(Pour une version plus détaillée de ces instructions, contacter le Bureau du WAT (wat@sun.ac.za)
ou consulter notre website: <<http://www.sun.ac.za/wat>>)

A. POLITIQUE ÉDITORIALE

1. Caractéristiques et contenu des articles

Les articles seront consacrés à la lexicographie pure, ou aux rapports entre la lexicographie et les disciplines voisines telles que la linguistique, la linguistique générale, l'informatique et le management.

Les contributions pourront appartenir à l'une des catégories suivantes:

- (1) **Articles:** Recherches scientifiques originales, avec leurs résultats; ou présentations originales, synoptiques, interprétatives, comparatives, évaluatives et critiques des résultats de recherches en cours;
- (2) **Articles bilans:** Articles de recherche présentés sous forme de bilan critique de travaux scientifiques déjà publiés.

Les contributions appartenant aux catégories (1) et (2) seront soumises de manière anonyme à des experts spécialistes indépendants afin d'en assurer la qualité scientifique au niveau international.

- (3) **Recensions:** Analyses et évaluations critiques de travaux de recherche et de productions scientifiques, telles que livres ou logiciels;
- (4) **Projets:** Présentations de projets lexicographiques;
- (5) **'Lexiconotes':** Textes contenant des informations pratiques, ou des suggestions, des problèmes, des questions, des commentaires et des solutions concernant des activités lexicographiques;
- (6) **'Lexicovaria':** Articles, annonces, communiqués de presse émanant de centres de lexicographie et qui revêtent un intérêt particulier pour les lexicographes;
- (7) **Rapports:** Rapports sur des colloques et ateliers.

Les contributions dans les catégories (3) à (7) devront répondre aux exigences de qualité des publications scientifiques et seront évaluées dans cette optique.

2. Critères et procédures d'évaluation

La revue *Lexikos* est reconnue et subventionnée par le Ministère de l'Éducation du gouvernement Sud-Africain comme revue devant générer des revenus.

Les articles seront évalués selon les critères suivants: langue et style, concision et clarté, formulation de la problématique, raisonnement et conclusion, référence aux travaux les plus importants et les plus récents, contribution substantielle à la discipline.

3. Langue des contributions

Afrikaans, allemand, anglais, français ou néerlandais.

4. Copyright

Le Bureau du WAT ou l'*African Association for Lexicography* (AFRILEX) décline toute responsabilité en cas de réclamations motivées par l'utilisation d'autres sources par les auteurs.

Les droits d'auteurs des documents publiés dans *Lexikos* appartiennent au Board of Control du *Woordeboek van die Afrikaanse Taal* (WAT). Cependant, les au-

teurs 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*.