

Lexikos 11

Lexikos 11

Redakteur

Editor

J.C.M.D. du Plessis

Resensieredakteur

Review Editor

T. Harteveld



African Association for Lexicography

AFRILEX-REEKS 11:2001

AFRILEX SERIES 11:2001



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 © 2001 deur die uitgewer
Alle regte streng voorbehou
Eerste uitgawe 2001

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 59 Goodwood

ISBN 0 9584120 9 X

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

Adviesraad / Advisory Board

Prof. H. Béjoint (Frankryk/France)
Dr. 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)
Dr. L.S. Vikør (Noorweë/Norway)
Prof. H.E. Wiegand (Duitsland/Germany)
Prof. L. Zgusta (VSA/USA)

Redaksiekomitee / Editorial Committee

Prof. A. Carstens (RSA)
Prof. W.A.M. Carstens (RSA)
Mnr. E. Chabata (Zimbabwe)
Dr. A.R. Chuwa (Tanzanië/Tanzania)
Prof. C.J. Conradie (RSA)
Prof. L.G. de Stadler (RSA)
Dr. M.M.R. Dube (RSA)
Prof. J.D. Emejulu (Gaboen/Gabon)
Dr. A.E. Feinauer (RSA)
Prof. E.F. Kotzé (RSA)
Me. K.P. Kavanagh (RSA)
Dr. M. Lisimba (Gaboen/Gabon)
Mnr. K.J. Mashamaite (RSA)
Dr. J.S. Mdee (Tanzanië/Tanzania)
Mnr. T.X. Mfafa (RSA)
Mnr. M.H. Mpungose (RSA)
Prof. F.A. Ponelis (RSA)
Prof. D.J. Prinsloo (RSA)
Prof. R. Sitaram (RSA)
Prof. P.H. Swanepoel (RSA)
Dr. J. Tsonope (Botswana)
Prof. M.W. Visser (RSA)

Inhoud / Contents

Voorwoord	x
Foreword	xii
<i>J.C.M.D. du Plessis</i>	
'n Woord van AFRILEX	xiv
A Few Words from AFRILEX	xv
<i>Danie Prinsloo</i>	
Redaksionele doelstellings	xvi
Editorial Objectives	xvii
Redaktionelle Ziele	xviii
Navorsingsartikels / Research Articles	
The Lexicographic Treatment of Days in Sepedi, or When Mother-Tongue Intuition Fails	1
<i>Gilles-Maurice de Schryver and B. Lepota</i>	
Orature and Morpholexical Deconstruction as Lexicographic Archeological Sites: Some Implications for Dictionaries of African Languages	38
<i>Francis Matambirofa</i>	
Beskouende artikels / Contemplative Articles	
The Role of the Introductory Matter in Bilingual Dictionaries of English and Arabic	60
<i>Hashan Al-Ajmi</i>	
Lexicography versus Terminography	71
<i>Mariëtta Alberts</i>	
Beiträge zu einer kognitiv ausgerichteten Lexikographie	85
<i>Udo L. Figge</i>	

The Use of an Improved Access Structure in Dictionaries <i>Rufus H. Gouws</i>	101
The Compilation of Bilingual Dictionaries between African Languages in South Africa: The Case of Northern Sotho and Tshivenda <i>Kwena J. Mashamaite</i>	112
Macro- and Microstructural Issues in Mazuna Lexicography <i>P.A. Mavoungou</i>	122
The Compilation of Electronic Dictionaries for the African Languages <i>D.J. Prinsloo</i>	139
Dictionary Quality and Dictionary Design: A Methodology for Improving the Functional Quality of Dictionaries <i>Piet Swanepoel</i>	160
Supplying Syntactic Information in a Quadrilingual Explanatory Dictionary of Chemistry (English, Afrikaans, isiZulu, Sepedi): A Preliminary Examination <i>Elsabé Taljard and Rachéle Gauton</i>	191
Projekte / Projects	
From Business Corpus to Business Lexicon <i>Li Lan and Grahame T. Bilbow</i>	209
Working on an Historical Dictionary: The Swedish Academy Dictionary Project <i>Per Stille and Bo-A. Wendt</i>	222
Leksikonotas / Lexiconotes	
Foreign Words as a Problem in Standardisation/Lexicography: English and Afrikaans Loan-words in isiXhosa <i>Anja Drame</i>	231
Problems of Equivalence in Shona-English Bilingual Dictionaries <i>Nomalanga Mpofu</i>	242

"Nonmorphological Derivations" and the Four Main English Learner's Dictionaries <i>Geart van der Meer</i>	252
Lexikovaria / Lexicovaria	
The Hausa Lexicographic Tradition <i>Roxana Ma Newman and Paul Newman</i>	263
Resensieartikels / Review Articles	
<i>Tweetalige Frasewoordeboek / Bilingual Phrase Dictionary: 'n Fokus op funksies en enkele probleemareas</i> <i>Anna Nel Otto</i>	287
Herbert Wiegand with a "Metalexicographical Panga" in the Jungle: An Unlocking of <i>Wörterbuchforschung</i> <i>Maria Smit</i>	296
Resensies / Reviews	
Dorothea Mantzel and Bernd Schulz. <i>Francolin Illustrated School Dictionary for Southern Africa</i> <i>M. Benjamin</i>	311
Harry Demeersseman, Piet Vermeer, Roy Pheiffer, Fred Pheiffer, Anneli Terre Blanche en Sief Veltkamp-Visser. <i>Kramers woordenboek Zuid-Afrikaans-Nederlands Nederlands-Zuid-Afrikaans</i> <i>Hans den Besten</i>	314
K.H. van Dalen-Oskam, K.A.C. Depuyt, W.J.J. Pijnenburg and T.H. Schoonheim. <i>Dictionaries of Medieval Germanic Languages</i> <i>J.C.M.D. du Plessis</i>	322
<i>Pharos Woordeboeke/Dictionaries 5 in 1</i> <i>Phillip Louw</i>	324
<i>Pharos Speller: Speltoetsen en Woordafbreker vir Afrikaans</i> <i>Phillip Louw</i>	328

F. Heyvaert, A. Moerdijk, M. Mooijaart, M. Smits en R. Tempelaars. <i>Het grootste woordenboek ter wereld: Een kijkje achter de kolommen van het Woordenboek der Nederlandsche Taal (WNT)</i> Gerhard van Wyk	330
A. Moerdijk en R. Tempelaars. <i>Van A tot Z en verder ...: Lezingen bij de voltooiing van het WNT</i> Gerhard van Wyk	338
Publikasieaankondigings / Publication Announcements	342
Voorskrifte aan Skrywers	345
Instructions to Authors	346
Richtlinien für Autoren	347

Voorwoord

Lexikos het altyd 'n tweeledige redaksionele beleid gehad: om artikels te publiseer nie net oor teoretiese kwessies nie, maar ook oor praktiese aspekte van die leksikografie. Dit is belangrik vir die teoretiserende metaleksikograaf om deeglik bewus te wees van die probleme wat die praktiserende leksikograaf teëkom, en dit is ewe belangrik vir die praktiserende leksikograaf om kennis te neem van die insigte wat die teoretiserende metaleksikograaf bied. Hierdie twee, die teoretiserende metaleksikograaf en die praktiserende leksikograaf, kan maklik van mekaar vervreem raak. Dit is daarom belangrik dat hierdie twee voortdurend bewus is van en voeling hou met mekaar om teorie en praktyk te versoen, en andersom. Die bedoeling met die artikels wat *Lexikos* in elke uitgawe insluit, is om hierdie versoening te bewerkstellig.

Maar die metaleksikograaf en die leksikograaf in Afrika moet ook rekening hou met die vernuwende denke en praktyke in die internasionale leksikografiese wêreld. Dit is belangrik vir die metaleksikografie en leksikografie in Afrika om voort te beweeg volgens die internasionale rigtings deur die jongste navorsing en tegnologie te gebruik en aan te pas vir die ontsluiting van die ryk taalverskeidenheid van Afrika. *Lexikos* publiseer dus ook artikels wat die nuutste leksikografiese denke en praktyke in die res van die wêreld weerspieël.

Redaksiekomitee

Lexikos en AFRILEX word nie net internasionaal bekend nie, maar maak ook kontak met die res van Afrika. Dit is daarom aangenaam om prof. J.D. Emejulu van Gaboen as nuwe lid van die Redaksiekomitee te verwelkom. *Lexikos* is jammer om afskeid te neem van een van sy Suid-Afrikaanse lede prof. J.A. du Plessis, maar net so bly om prof. M.W. Visser in sy plek te verwelkom.

Dankbetuiging

Daar is twee groepe mense sonder wie die publikasie van *Lexikos* onmoontlik sal wees: die bydraers van artikels en die keurders van hierdie artikels. Die redakteur wil graag altwee bedank: die bydraers vir die manier waarop hulle bereid is om die kritiek en voorstelle van die keurders te akkommodeer, en die keurders vir hul bereidwilligheid om artikels te beoordeel, dikwels op baie kort kennisgewing. Deur dié waardevolle wisselwerking tussen hierdie twee groepe, kan *Lexikos* artikels van 'n aansienlike gehalte aanbied.

Vanjaar het die span verantwoordelik vir die produksie van *Lexikos* bestaan uit mes. Tanja Harteveld en Riette Ruthven. Ek wil hulle graag altwee bedank vir hul entoesiasme en geduld by die afhandeling nie net van die dikwels moeilike setwerk nie, maar ook van die uitgebreide korrespondensie wat met die produksie van *Lexikos* gepaardgaan. Ek wil ook graag me. Harteveld bedank vir haar bydrae as resensieredakteur van *Lexikos*.

J.C.M.D. du Plessis

Buro van die Woordeboek van die Afrikaanse Taal

Foreword

Lexikos has always had a dual editorial practice: to publish articles not only on theoretical issues, but also on practical aspects of lexicography. It is important for the theorizing metalexigrapher to be well aware of problems the practising lexicographer encounters, and it is equally important for the practising lexicographer to take note of the insights the theorizing metalexigrapher offers. These two, the theorizing metalexigrapher and the practising lexicographer, can easily become estranged from each other. It is therefore important that these two constantly remain aware of and keep in touch with each other to reconcile theory and practice, and *vice versa*. The intention with the articles *Lexikos* includes in each issue is to bring about this reconciliation.

But the metalexigrapher and the lexicographer in Africa must also take account of innovative thoughts and practices in the international lexicographical world. It is important for metalexigraphy and lexicography in Africa to advance according to international trends, employing and adapting the latest research and technology for the unlocking of the rich language diversity of Africa. Therefore *Lexikos* also publishes articles reflecting the metalexigraphical and lexicographical thoughts and practices in the rest of the world.

Editorial Board

Lexikos and AFRILEX are not only becoming known internationally, but are also establishing contact with the rest of Africa. It is therefore pleasant to welcome Prof. J.D. Emejulu from Gabon as new member of the Editorial Board. *Lexikos* is sorry to bid farewell to one of its South African members Prof. J.A. du Plessis, but is equally happy to welcome Prof. M.W. Visser in his stead.

Acknowledgments

There are two groups of people without whom the publication of *Lexikos* would be impossible: the contributors of articles and the referees of these articles. The editor wishes to thank both: the contributors for the way they were willing to accommodate the criticism and suggestions of the referees, and the referees for their willingness to judge articles, often at very short notice. With this valuable interaction between these two groups *Lexikos* can offer articles of a considerable standard.

This year the team responsible for the production of *Lexikos* consisted of Mss Tanja Harteveld and Riette Ruthven. I want to thank them both for their enthusiasm and patience in handling not only the often difficult typesetting, but also the extensive correspondence which accompanies the production of *Lexikos*. I also want to thank Ms Harteveld for her contribution as review editor of *Lexikos*.

J.C.M.D. du Plessis

Bureau of the Woordeboek van die Afrikaanse Taal

'n Woord van AFRILEX

Met die 11de uitgawe van *Lexikos* en die 6de internasionale konferensie van die African Association for Lexicography (AFRILEX) agter die rug, kan mens met reg trots wees op die suksesverhaal van AFRILEX en *Lexikos*. *Lexikos* verwerf toenemend nasionaal en internasionaal bekendheid as 'n leksikografietydskrif van besonder hoë gehalte. AFRILEX tel eweneens onder die toonaangewende leksikografieverenigings in die wêreld en 'n rekordaantal nuwe lede het vanjaar aangesluit. AFRILEX bied 'n tuiste aan almal wat belangstel in een of ander aspek van die leksikografie en speel 'n belangrike ondersteunende rol in die totstandbrenging van woordeboekeenhede vir die Afrikatale deur die Pan Suid-Afrikaanse Taalraad. Benewens die jaarlikse internasionale konferensie, fasiliteer AFRILEX 'n aantal seminare, werksinkels en tutoriale en kommunikeer met lede deur sy nuusbrief en verskeie omsendbriewe.

Die bydrae tot die leksikografieproses van twee AFRILEX-lede vanjaar verdien besondere vermelding. Prof. Rufus Gouws het 'n leidende rol gespeel in verskeie PANSAT-geïnisieerde leksikografieopleidingsgeleenthede en as medeaanbieder van die tweedagseminaar oor die samestelling van vakwoordeboeke. Mnr. Gilles-Maurice de Schryver het baanbrekerswerk verrig met die samestelling van elektroniese korpusse vir die Afrikatale en deur die *Onoma* rekenaarprogram aanpasbaar te maak vir die samestelling van Afrikataalwoordeboeke.

Dit moet aanvaar word dat woordeboekprojekte in Suid-Afrika toenemend finansiële druk sal ervaar. So, byvoorbeeld, kon slegs 'n klein persentasie van die minimumbegroting wat nodig is vir die totstandbrenging van woordeboekeenhede vir die Afrikatale tot dusver bewillig word. In baie gevalle is woordeboekeenhede die feitlik onbegonne taak opgelê om met hoofsaaklik deeltydse personeel te funksioneer. Dié toedrag van sake plaas 'n groot morele verpligting op gevestigde woordeboekeenhede, AFRILEX en moedertaalsprekers om met raad en daad die leksikografiese proses te ondersteun.

Die Raad van AFRILEX glo dat lidmaatskap van AFRILEX (slegs R80 per jaar insluitende 'n gratis kopie van *Lexikos*) uitstekende waarde vir geld is. Ons spreek ook 'n woord van dank aan die Buro van die WAT vir die publikasie van *Lexikos* en in besonder aan dr. J.C.M.D. du Plessis vir die uitnemende werk wat hy as redakteur verrig.

Danie Prinsloo

Voorsitter: AFRILEX

AFRILEX tuisblad <http://www.up.ac.za/academic/libarts/afrilang/homelex.html>

A Few Words from AFRILEX

With the 11th edition of *Lexikos* and the 6th international conference of the African Association for Lexicography (AFRILEX) now behind us, we can be proud of the success story of AFRILEX and *Lexikos*. *Lexikos* is becoming increasingly well known, both nationally and internationally, as a quality lexicographic journal. Similarly, AFRILEX now counts as one of the leading lexicographic associations in the world. A record number of new members joined this year. AFRILEX provides a home for everyone interested in some aspect or other of lexicography and plays an important supporting role in the establishment of dictionary units for the African languages through the Pan South African Language Board (PANSALB). In addition to the annual international conference, AFRILEX facilitates a number of seminars, workshops and tutorials, and communicates with its members through its newsletter and circulars.

Two AFRILEX members' contributions to the lexicography process this year deserve particular mention. Prof. Rufus Gouws played a leading role in several PANSALB-initiated lexicographic training opportunities, and was the co-presenter of a two-day seminar on the compilation of dictionaries for special purposes. Mr Gilles-Maurice de Schryver executed groundbreaking work both in the compilation of electronic corpuses for the African languages and through adapting the *Onoma* computer program for the compilation of African language dictionaries.

We should accept that dictionary projects in South Africa will, increasingly, face financial pressure. For example, only a small percentage of the minimum budget necessary for the establishment of dictionary units for the African Languages has been allocated. In many cases, dictionary units are faced with the virtually impossible task of functioning with mainly part-time staff. This state of affairs places a moral obligation on established dictionary units, AFRILEX and mother-tongue speakers to support the lexicographical process both with advice and action.

The Board of AFRILEX believes that membership fees of AFRILEX (only R80 annually, including a free copy of *Lexikos*) represent extremely good value for money. We would like to thank not only the Bureau of the WAT for the publication of *Lexikos*, but also Dr J.C.M.D. du Plessis in particular for the excellent work that he, as editor, does.

Danie Prinsloo

Chairperson: AFRILEX

AFRILEX home page <http://www.up.ac.za/academic/libarts/afri-lang/homelex.html>

Redaksionele doelstellings

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

Die strewe van die AFRILEX-reeks is:

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

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

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

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

Editorial Objectives

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

The objectives of the AFRILEX Series are:

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

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

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

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

Redaktionelle Ziele

Lexikos ist eine Zeitschrift für Fachleute der Lexikographie, die in der AFRILEX-Serie erscheint. "AFRILEX" ist ein Akronym für "Lexikographie in und für Afrika". Von der sechsten Ausgabe dient *Lexikos* als amtliches Mundstück des *African Association for Lexicography* (AFRILEX), u.a. weil das Büro des WAT gerade das ausgesprochene Ziel mit der Ausgabe der AFRILEX-Serie hatte, die Gründung solches 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 zum lexikographischen Gespräch 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.

The Lexicographic Treatment of Days in Sepedi, or When Mother-Tongue Intuition Fails*

Gilles-Maurice de Schryver, *Research Assistant of the Fund for Scientific Research – Flanders (Belgium) and Department of African Languages, University of Pretoria (schryver@postino.up.ac.za), and*
B. Lepota, *Unit for Language Skills Development, University of Pretoria (lepota@postino.up.ac.za), Pretoria, Republic of South Africa*

Abstract: Since May 2001 the Sepedi National Lexicography Unit officially started the actual dictionary-writing of a pioneering Explanatory Sepedi Dictionary (PyaSsaL). The compilation is undertaken within the theoretical framework of Simultaneous Feedback, is fully corpus-based, and follows an onomasiological approach to the Sepedi lexicon. In this article the various compilation aspects are examined and illustrated by means of one onomasiological sub-field, namely the days of the week. It is shown how a balanced combination of mother-tongue intuition, data from existing (bilingual) dictionaries, fieldwork results, corpus queries, and grammarians' conjectures can and should lead to a sound lexicographic treatment.

Keywords: SEPEDI NATIONAL LEXICOGRAPHY UNIT (NLU), SIMULTANEOUS FEEDBACK (SF), PRETORIA SEPEDI CORPUS (PSC), ONOMASIOLOGICAL APPROACH, DAYS OF THE WEEK, MOTHER-TONGUE INTUITION, FIELDWORK, QUESTIONNAIRES, INTERVIEWS, GRAMMARIANS' CONJECTURES, NOUN CLASSES, REPETITIVE INSERTED TEXTS

Kakaretšo: Tshakatsheko ya matšatši polelong ya Sesotho sa Leboa go ya ka dinyakwa tša go ngwala Pukuntšu, goba ge dikakanyo tša mmoledi wa polelo ye di sa atlege. Go thoma ka Mei 2001 Yuniti ya Bosetšhaba ya Pukuntšu ya Sesotho sa Leboa e thomile semmušo go ngwala Pukuntšutlhaloši ya Sesotho sa Leboa (PyaSsaL). Thulaganyo ya yona e laolwa ke teori ya *Simultaneous Feedback*, e theilwe godimo ga sešegontšu (khophase), gape e latela tsela ya go hlaloša manšu a Sesotho sa Leboa go ya ka direrwa tše di fapanego. Mo taodišwaneng ye go tsinkelwa mekgwa ya go fapanafapana ya thulaganyo gape mekgwa yeo e šupetšwa ka karolwana e tee ya direrwa tša PyaSsaL, e lego matšatši a beke. Go bontšhwa ka fao tekanelo ya kopanyo ya dikakanyo tša mmoledi wa Sesotho sa Leboa, tshedimošo go tšwa go dipukuntšu (tša malemepedi) tšeo di šetšego di le gona, dipelo tša nyakišišontle, diphatišišo ka gare ga sešegontšu, le dikakanyo tša borapopapolelo di ka kgonago ebile di swanetše go fihliša tshaka-tshekong ya kgonthe go ya ka dinyakwa tša go ngwala pukuntšu.

Mantšu a bohlokwa: YUNITI YA BOSETŠHABA YA PUKUNTŠU YA SESOTHO SA LEOBA (YBP), SIMULTANEOUS FEEDBACK (SF), KHOPHASE YA SESOTHO SA LEOBA YA

* An earlier version of this article was presented at the Sixth International Conference of the African Association for Lexicography, organised by the Sepedi National Lexicography Unit, University of the North, Pietersburg, 2-4 July 2001.

TSHWANE (KST), DIRERWA, MATŠATŠI A BEKE, KAKANYO YA MMOLEDI WA SESOTHO SA LEBOA, NYAKIŠIŠONTLE, MANANEOPOTŠIŠO, DIPOLEDIŠANO, DIKAKANYO TŠA BORAPOPAPOLELO, MAGORO A MAINA, KHII YA GO TSENYWA KA GO BOELETŠWA

1. South Africa's 'golden opportunity'

In a recent publication Gouws (2000: 114) refers to the process of establishing a new lexicographic dispensation in South Africa as a 'golden opportunity' and a unique occasion in international terms. This golden opportunity is materialising at this very moment, with the Pan South African Language Board (PANSALB) having established National Lexicography Units (NLUs) for each of the official South African languages in 2000, and the actual compilation of dictionaries already being undertaken by some of them. In May 2001 the Board of the Sepedi NLU appointed two full-time mother-tongue lexicographers who promptly started their activities. In addition, the Board appointed one Ph.D. student as part-time lexicographer, and accepted the offer from another one to act as facilitator. During the first three months, the Head Office joined the Branch Office at the University of Pretoria (UP), where the lexicographic team was supplemented by two part-time corpus builders from UP's Department of African Languages.¹

2. The theoretical framework and computational support

An extensive discussion of the methods underlying the current compilation procedures and all the facets of the computational support will be described elsewhere. In short, however, we can point out that the compilation is undertaken within the theoretical framework of Fuzzy Simultaneous Feedback (cf. e.g. De Schryver and Prinsloo 2001), which can be considered as the electronic continuation of the concept of Simultaneous Feedback (SF) (cf. e.g. De Schryver and Prinsloo 2000, 2000a). Briefly, SF can be understood as entailing a dictionary-making method in terms of which the release of several small-scale Parallel Dictionaries triggers off feedback that is instantly channelled back into the compilation process of a Main Dictionary.

It is well-known that "[t]he line function of a unit, as stated by PANSALB, should eventually be the compilation of a comprehensive monolingual explanatory dictionary" (Gouws 2000: 111), and the Sepedi NLU is giving heed to this. From the start, the facilitator trained the team members in the writing of explanatory definitions, sketched the main features of the dictionary to be compiled and the structure of the articles, and approached the querying of the electronic corpus hands-on. Within two weeks a first little test dictionary was produced and circulated among mother-tongue speakers. Feedback was retrieved, fed back into the project, and the compilation adapted accordingly. Within two months the First Parallel Dictionary was printed, and the cycle repeated. This procedure will be part and parcel of the entire compilation process.

As far as the computational support is concerned, we can briefly observe that the data are entered in the Onoma Lexical Workbench, a software package developed by Lexilogik in Sweden.² At all times the facilitator has been — and will be for some time into the future — in close contact with the software developers in order to adapt Onoma to the requirements of the Unit. SQL³, the database underlying Onoma, has been stored on the facilitator's computer (the server side), whilst the computers of the three mother-tongue lexicographers (the client side) have been linked to the server through a network. Another crucial facet of the computational support is the electronic corpus. In this respect the Sepedi NLU is rather fortunate, as it can make free use of the Pretoria Sepedi Corpus (PSC), a corpus that was painstakingly assembled during the past decade by D.J. Prinsloo and G.-M. de Schryver. Currently, PSC stands at 5.8 million running words.⁴ At the moment, PSC has not been integrated into Onoma. Rather, PSC is analysed with WordSmith Tools, a versatile corpus query software package developed by Mike Scott in the UK.⁵

3. An onomasiological approach to dictionary compilation

With PSC at hand (or better: 'on screen'), the compilation of the *Pukuntšutlhaloši ya Sesotho sa Leboa* (PyaSsaL) 'Explanatory Sepedi Dictionary', is fully corpus-based. For every compilation aspect — from the selection of the lemma signs up to the writing of the dictionary articles themselves — the corpus is queried. In De Schryver and Prinsloo (2000c) the different steps one needs to follow in order to compile a corpus-based macrostructure have been reviewed. One starts by extracting a word-frequency list from the corpus, this list is then transformed into a lemmatised frequency list, after which the latter is turned into a lemma-sign list. However, taking the lemma-sign list and working through the alphabet from A to Z, rarely results in a sound end product. Indeed, such an approach is more often than not marred by inconsistencies and poor definitions. Instead, the compilation of PyaSsaL follows an onomasiological approach to the Sepedi lexicon. Different semantic fields are chosen by each lexicographer, and each then tries to cover all the basic/frequent items from that field (where the selection is based on PSC). One lexicographer will then read all the definitions from a certain field to the others to see whether or not the others can pinpoint the correct lemma sign. If not, the definition must be adapted or rewritten. Such an approach has already proven to have many advantages, foremost among them the fact that the compilers are forced to differentiate every item from every other one and to make sure circular definitions are avoided.

The latter point can be illustrated with an example taken from the *Collins COBUILD English Dictionary* (COBUILD2, Sinclair 1995²) — one of four reference works many lexicographers consider to be among the best learners' dictionaries available for English. The first definition (and in most cases the only one users read) for *minute* in COBUILD2 is shown in (1).

- (1) A **minute** is one of the sixty parts that an hour is divided into. People often say 'a **minute**' or '**minutes**' when they mean a short length of time.

As can be seen, the definition of *minute* is based on *hour*. A user who does not know the meaning of *hour* will have to consult that item in order to understand the meaning of *minute*, only to find (2) as the first definition.

- (2) An **hour** is a period of sixty minutes.

Here, the definition of *hour* is based on *minutes*. Such circular definitions are unacceptable. By forcing the mother-tongue compilers to work within fields, such circularity is combated. Even though the compilation follows an onomasiological approach, the editors will have the choice to present the data stored in the database in a semasiological way (meaning that the lemma signs are listed in an alphabetical sequence) or in an onomasiological way (thus as a dictionary with a thematic character). Yet the onomasiological-compilation approach also enables the lexicographers to transcend the paper dimension. As a matter of fact, every sense of each lemma sign of a particular onomasiological field is labelled with the same 'classifier'. These data are entered in a hidden slot in Onoma — 'hidden' in that it is not shown in the printed version and can be hidden in the electronic version of PyaSsaL. Ultimately every lemma sign will contain several classifiers, and the idea is to facilitate searches in the electronic version where users go from concept to word rather than from word to concept. A user will be enabled to input some classifiers, keywords, style labels, Boolean operators, etc. after which the software will run through the multi-indexed data to suggest the item(s) the user is seeking (cf. also Geeraerts 2000).

So far, several dozen of onomasiological fields, together with their classifiers, have been treated: COLOURS, DISEASES, TIME, CROPS, VEHICLES, KITCHENWARE, CLOTHES, etc. etc. The field TIME proved to be a particularly hard one, and was classified into numerous sub-fields. DAYS is just one of these many sub-fields, and, taken at face value, one could assume that it can be treated in just a few hours. However, this is not so, and this sub-field will be taken as an example to illustrate how the different compilation aspects can and must be combined in order to arrive at a sound lexicographic treatment.

4. When mother-tongue intuition fails

Just as in English, the days are nouns in Sepedi. This immediately implies that they belong to a certain gender, i.e. a fixed singular + plural class — crucial information as all the syntactical concords follow suit. Not all, but the great majority of nouns have both a singular and a plural form. See for instance (3) for some examples (with the gender indicated after the equal-signs).

- (3) *moeta* 'vessel' / *meeta* 'vessels' = 3/4
letšhollo 'diarrhoea' = 5/0
lepai 'cotton blanket' / *mapai* 'cotton blankets' = 5/6
maloba 'the day before yesterday' = 0/6
sehla 'season' / *dihla* 'seasons' = 7/8
hlogo 'head' / *dihlogo* 'heads' = 9/10

Most mother-tongue speakers know the form (where applicable) of the singular and the plural of a particular noun. However, the fact that *not all* speakers have this intuitive knowledge, and certainly not for rare or borrowed words, forces the compilers to guide the future users of PyaSsaL. It was therefore decided to include gender information for every noun, with the full treatment by default at the singular (and the reverse only where the plural is more frequent). The plural contains, apart from the comment on form, no more than a cross-reference to the singular (or where the plural is more frequent, a cross-reference from the singular to the plural). The gender information is given with the generally accepted numbering system illustrated in (3), where the class number of the treated word is printed in boldface. In addition, a repetitive inserted text at the bottom of the page (in the paper version) or a pop-up window (in the electronic version) briefly summarises the meaning and the various concords. (For a more extensive argumentation of the followed procedure when lemmatising nouns, see Prinsloo and De Schryver (1999), and for more details on the repetitive inserted text, see De Schryver and Prinsloo (2000: 200-203).)

With this procedure, the treatment of the sub-field DAYS seems simple: (a) write down all the items belonging to the closed set DAYS, including the variants, (b) query the corpus to retrieve frequency information and example sentences, and (c) treat the items accordingly, with cross-references from the lesser-used to the more-frequent ones. A sample of the first attempt at this procedure (here for Thursday) — retrieved from an early draft — is shown in (4).

- (4) **Labone** *leina* 5/6, 5/10, 5/2a
 Ke letšatši la bone la beke, le thoma ka morago ga Laboraro gomme la fela ge go thoma Labohlano: O nteleditše mogala ka Labone
Malabone?? *leina* 5/6 BONA **Labone**
Dilabone *leina* 5/10 BONA **Labone**
Bolabone?? *leina* 5/2a BONA **Labone**

What at this point worried the mother-tongue lexicographer who had entered this into the database, was that the plurals of most days (here *Malabone*, *Dilabone* and *boLabone*) had ended up in up to *three different classes* — a rather surprising result. After consulting with the other members of the team, the compilers realised that their intuition let them down, for they could not agree on the genders of the days. A reanalysis of PSC revealed that the corpus seemed to contain only a few occurrences of the plurals of any day. At this stage, the

Popular	✓	-	-	✓	✓	✓	✓	✓	-	✓	✓	-	✓
E-S	-	-	-	-	-	-	-	-	-	-	-	-	-
Gerber	✓	-	✓	-	✓	✓	✓	✓	-	✓	✓	✓	✓
A-S	-	-	-	-	-	-	-	-	-	-	-	-	-
LEGEND	S: Sepedi / E: English / A: Afrikaans Z & M: Ziervogel and Mokgokong 1975 / New E: Kriel 1976 ⁴ / Klein: Ziervogel and Mokgokong 1988 ⁴ / T & O: Departmental Northern Sotho Language Board 1988 ⁴ / V Wyk: Kriel et al. 1989 ⁴ / Prinsloo: Prinsloo et al. 1996, 1997 / Popular: Kriel et al. 1997 ⁴ / Gerber: Gerber 2000												

From (5) one can see that all dictionaries taken together offer 13 alternatives for the seven days of the week. For every dictionary, the top line shows the inclusion (✓) or exclusion (-) of the singular, and the bottom line whether or not the plural form is suggested (through the indication of class or gender information). Since we are primarily concerned with the formation of the plurals, we will first look into this aspect. Ziervogel and Mokgokong suggest, both in their comprehensive (1975) and their pocket (1988⁴) dictionary, that the plural of *Sontaga* 'Sunday' is *Disontaga*. Van Wyk, in his revision of the third edition of Kriel's *Pukuntšu* (Kriel 1983³), added class information to all the singular days (Kriel et al. 1989⁴). If a (non-proficient?) user follows the cryptic rules in the front matter of Van Wyk's revision, this user might derive the plurals listed in (6).

- (6) Class information and derived plurals in Van Wyk's *Pukuntšu* (Kriel et al. 1989⁴)

Day	Singular	Class information	Derived plural
Monday	Mošupulogô	n. dev. class 3	Mešupulogô
	Mošopologô	n. dev. class 3	Mešopologô
Tuesday	Labobêdi	poss. conc. class 5 + n.	(Mabobêdi / Malabobêdi / ?)
Wednesday	Laboraro	poss. conc. class 5 + n.	(Maboraro / Malaboraro / ?)
Thursday	Labonê	poss. conc. class 5 + n.	(Mabonê / Malabonê / ?)
Friday	Labohlano	poss. conc. class 5 + n.	(Mabohlano / Malabohlano / ?)
Saturday	Mokibêlô	n. dev. class 1a	bôMokibêlô
Sunday	Sôntaga	n. class 1a	bôSôntaga
	Lamorêna	poss. conc. class 5 + n.	(Mamorêna / Malamorêna / ?)
	Lamodimo	poss. conc. class 5 + n.	(Mamodimo / Malamodimo / ?)

The lexicographers were rather doubtful at this point, and the next step, field-work, was the consequence. However, it is necessary further to look at the lexicographic treatment of the days of the week in the above-mentioned dictionaries.

When describing the onomasiological approach to dictionary compilation in par. 3, we pointed out two crucial aspects leading to better dictionary articles when compiling within this framework: (a) the combat against circularity, cf. extracts (1) and (2), and (b) the avoidance of inconsistencies. Even a cursory

glance at any of the sets of days shown in Appendix A confirms the need for the latter. No set is consistent throughout. On a first level, one notes the inconsistencies in punctuation and layout, the random inclusion or omission of parts of speech, and the haphazard presence or absence of loanword labels. On a second level, and more problematic, one observes the variation in orthography between the different sections of these dictionaries (e.g. *La Morêna* in one direction, *Lamorêna* in the other). Furthermore, numerous discrepancies can be found among the various alternatives given in the different sections (e.g. *Mošupulogo* and *Mošopologo* in one direction, yet *Mošopologo* and *Mantaga* in the other). Luckily, present-day computational support helps to avoid many of these problems. There is however a third level that is wholly unacceptable when dealing with a closed set such as the days of the week. The supposedly most complete dictionary available for Sepedi, the *Comprehensive Northern Sotho Dictionary* (Ziervogel and Mokgokong 1975), treats only five days of the week, totally neglecting the existence of Monday and Saturday. To make things even worse, these two days belong to the three most-frequently used days (cf. (5) for frequencies in the 5.8 million PSC). Even the little pocket derivation of this comprehensive dictionary does better, as it contains those two days in the direction Afrikaans/English to Sepedi (1988⁴). Finally, but this cannot be seen from Appendix A, some items (e.g. *Sôntaga* in Kriel et al. 1989⁴) are even in the wrong alphabetical place in the dictionary, making it truly hard for users to find any day at all.⁶ Sometimes the user is confronted with a case of 'impossible to find' (or dead reference), when the item in question has not been included in the dictionary. The treatment of Sunday in the *Comprehensive Northern Sotho Dictionary* is shown in (7).

(7) SÓN'TAGA, (se-)/di- (*Sôntaga*) (< Afr.), cf. LÁMODÍMO, Sondag // Sunday

Yet, when trying to follow up the cross-reference, the user will find that *Lamodimo* is not lemmatised.⁷ Again, proper alphabetical order and sound cross-references are aspects where basic software can and should assist the modern lexicographer.

A tenth dictionary that is available to the team is *Basic English–N.Sotho* (Hartshorne et al. 1984). This dictionary is essentially a monolingual English learners' dictionary, with an appendix containing translation equivalents in Sepedi. A numeric-alphabetic reference system links these two sections. As such, and in dictionary-typological terms, this reference work is a bilingualised (or semi-bilingual) dictionary (cf. e.g. Laufer and Melamed 1994). In the front matter to Hartshorne et al.'s dictionary (1984: introduction), one reads:

Basic English has been compiled for *the pupil* who is studying English as a second or foreign language. [...] In the first place a comprehensive range of textbooks in English and English literary works, frequently used by pupils, was subjected to a *computer analysis* in order to identify the vocabulary being used at this level. [...] This initial list was then checked against existing authoritative

international basic word lists, in order to make sure that all those words which are frequently used in English were included.

Nevertheless, although much effort seemingly went into the compilation of this dictionary, one will look in vain for days of the week. The closest one comes to them is in the article for the lemma sign *week**. This article is repeated verbatim in Appendix B, together with the corresponding data from that dictionary's appendix. One cannot but fail to see that four of the seven days (namely Monday, Friday, Saturday and Sunday) are used to 'define' and 'illustrate' *week**. This brings another section from the front matter (Hartshorne et al. 1984: introduction) to mind:

The final stage was to include [...] certain words which were needed in this dictionary's definitions and illustrative sentences.

It is therefore surprising that the names of the week were omitted, especially since COBUILD2 assigns four black diamonds to each day of the week, meaning that the names of the days belong to the top 1 900 most-frequent words of the English language. The only other major English learners' dictionary including frequency information is the *Longman Dictionary of Contemporary English, Third Edition* (LDOCE3, Summers 1995³). Yet, although the top-3 000 items in both spoken and written language have been marked in LDOCE3, none of the days of the week was given a frequency marker. The reason might lie in the fact that, in LDOCE3, insufficiently word-like items, including "numbers, closed sets such as nationalities and currencies, non-standard forms, and variants" (Kilgarriff 1997: 142), were not given a frequency annotation. However, this procedure of not marking, for instance, closed sets, jeopardizes the basic aim whereby frequency data are entered because "[a] central fact about a word is how common it is. The information is particularly valuable for language learners, as it immediately indicates how important it is to learn a word" (Kilgarriff 1997: 135).

6. Fieldwork: questionnaires and interviews

So far we have seen that both the mother-tongue lexicographers and the existing (bilingual) dictionaries could not provide the necessary data for a sound treatment of (the plurals of) the days. It was therefore decided to carry out some fieldwork, mainly aimed at retrieving mother-tongue speakers' suggestions for the plurals of the days. To do so, numerous techniques exist, the one having more flaws than the other. From the various alternatives, the utilisation of an informal questionnaire was chosen. Precautions were taken on two complementary levels in order to be able to 'trust' the results. On the one hand, feedback was retrieved among mother-tongue speakers in the Northern Province, Mpumalanga and Gauteng, and this in two distinct ways. 80% of the questionnaires were printed and distributed, whereby the speakers were asked

to write down their suggestions on the spot. For the remaining 20%, the paper version of the questionnaire was simulated through face-to-face interviews, whereby the interviewers tried to *trigger* spontaneous use of the plurals of the days. These interviews were recorded unknowingly. Following each interview, the interviewees were informed and their input only kept (and later transcribed) if they agreed. On the other hand, two versions of the questionnaire/interview were set up. In version 1 the participants were asked to say something about themselves first, after which they were presented with some singulars for which they had to provide plurals. Then followed a few general questions about their background. In version 2 however, the personal questions were moved to the end, and one made-up day (**Lamathomo* for Monday) was included in the list, together with more (low-frequency) day alternatives. Also, the order of the days was slightly permuted. With these various procedures, the very integrity of the methodology was tested. The two paper versions of the questionnaires can be found in Appendixes C and D.

In total, exactly 100 opinions were collected. In the preliminary analysis, the written input was kept separate from the oral input, and within these two groups, the answers for versions 1 and 2 were differentiated. Yet, after a careful study of the four types of feedback, it was noticed that there were no statistically relevant differences between them. The four types will therefore be discussed together below. That there were no such differences was surprising. On the one hand, one could have expected that participants react differently when filling in a questionnaire compared to natural spoken use. On the other hand, it is generally accepted that the retrieval of personal questions at the start of a questionnaire/interview skews the results. Fortunately, not too many participants tried to force a plural onto **Lamathomo* — indicating that they were not answering a battery of questions 'on automatic'.

The analysis of the questionnaires/interviews could best be started with Saturday, since this is the day with the least variation. The results of the fieldwork are shown in (8)(a), together with the corresponding frequency in PSC for each suggestion.

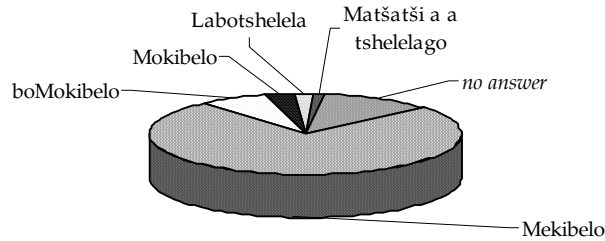
(8)(a) Fieldwork results: Saturday

Singular	Freq.	Plural	Freq.	Option 1	Option 2	Option 3
Mokibelo	219	Mekibelo	10	75	7	0
		boMokibelo	2	7	5	0
		Mokibelo	219*	3	0	0
		Labotshelela	0	2	0	0
		Matšatši a a tshelelago	0	1	0	0
		Dimekibelo	0	0	1	2
		<i>no answer</i>		12	87	98

The participants could suggest as many possibilities as they wanted, yet no one offered more than three. Since 75% of the mother-tongue speakers opted for *Mekibelo* as the first and another 7% as the second possibility, it is clear from

the fieldwork that this noun belongs to gender 3/4, and therefore classes *mo-* for the singular and *me-* for the plural. Option 1 is shown graphically in (8)(b).

(8)(b) Graphic presentation: Saturday



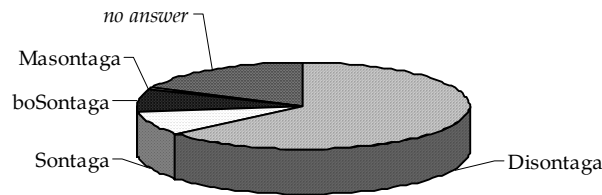
As regards Sunday, one can see from (5) that the most-frequently used form is a loanword (*Sontaga*), and not the term prescribed by the language board (*Lamorena*, see Appendix A). Likewise, as far as the suggested plurals are concerned, there is also more agreement on the former than on the latter, cf. (9)(a).

(9)(a) Fieldwork results: Sunday

Singular	Freq.	Plural	Freq.	Option 1	Option 2	Option 3
Sontaga	163	Disontaga	12	64	4	0
		Sontaga	163*	9	0	0
		boSontaga	1	8	6	1
		Masontaga	0	1	3	1
		Sontageng	1*	0	1	0
		boDisontaga	0	0	0	1
		no answer		18	86	97
Lamorena	63	Malamorena	1	24	3	0
		boLamorena	0	14	8	0
		Amorena	0	13	0	2
		Dilamorena	0	7	5	0
		Lamorena	63*	6	0	0
		Malaborena	0	3	0	0
		A Lamorena	0	1	0	0
		Ga Morena	206*	0	1	0
		Ba Morena	71*	0	0	1
		Lamerena	0	0	0	1
		no answer		32	83	96
Lamodimo	1	Lamodimo	1*	14	0	0
		Malamodimo	0	8	2	0
		Amodimo	0	4	0	0
		boLamodimo	0	2	2	0
		Alamodimo	0	2	0	0
		Melamodimo	0	2	0	0
		no answer		68	96	100

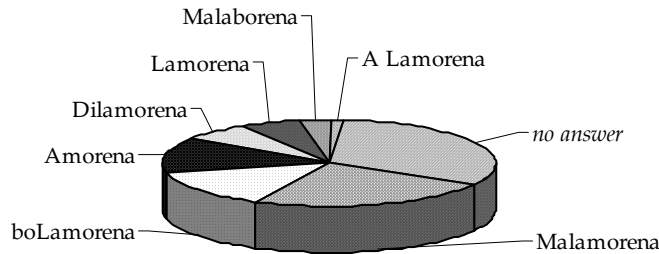
From the data in (9)(a), one must conclude that the fieldwork shows that *Sontaga* follows the pattern of most loanwords, i.e. it ends up in gender 9/10, taking *di+* as the plural prefix. Option 1 for the plurals of *Sontaga* is shown graphically in (9)(b).

(9)(b) Graphic presentation: Sunday (i)



The other alternative for Sunday which should also be treated in a dictionary according to the fieldwork is *Lamorena*, which takes the prefix of class 6 *ma+* for the formation of the plural, i.e. *Malamorena*. Option 1 for the plurals of *Lamorena* is shown graphically in (9)(c).

(9)(c) Graphic presentation: Sunday (ii)



Whilst Saturday is the most-frequently and Sunday the second most-frequently used day in Sepedi, Monday is the third. The summary of the fieldwork for Monday is shown in (10)(a).

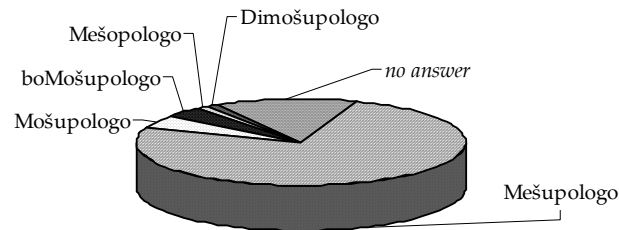
(10)(a) Fieldwork results: Monday

Singular	Freq.	Plural	Freq.	Option 1	Option 2	Option 3
Mošupologo	144	Mešupologo	0	75	3	0
		Mošupologo	144*	5	0	0
		boMošupologo	0	4	6	0
		Mešopologo	0	1	2	0
		Dimošupologo	0	1	1	0

		boMešupologo	0	0	1	0
		Mešupologong	0	0	1	0
		boMedišupologo	0	0	0	1
		<i>no answer</i>		14	86	99
Mantaga	0	Dimantaga	0	25	6	1
		Memantaga	0	14	2	1
		Mebantaga	0	12	2	0
		boMantaga	0	6	12	1
		Mantaga	0	6	0	0
		Mantageng	0	1	0	0
		<i>no answer</i>		36	78	97
*Lamathomo	0	Lamathomo	0	14	0	0
		Amathomo	0	6	2	0
		Malamathomo	0	6	0	0
		boLamathomo	0	2	2	0
		Dilamathomo	0	2	2	0
		Memathomo	0	0	0	2
		<i>no answer</i>		70	94	98

As can be seen from (10)(a), it is reassuring to notice that up to 70% of the participants did not even try to suggest a plural for the made-up day *Lamathomo, whilst another 14% simply repeated the singular. As this item was merely inserted to test the validity of the followed methodology, we should of course not discuss it in the context of the search for 'plurals of days'. As far as the methodology itself is concerned, we must conclude that it is valid given a fair margin of error. Furthermore, just as for the loanword *Sontaga* the fieldwork shows that the loanword *Mantaga* ends up in gender 9/10, yet with much less persuasiveness. Given the fact that this item has a zero frequency in PSC, we will not discuss it any further. Conversely, the mother-tongue speakers are rather unanimous when it comes to the plural of *Mošupologo*. Option 1 for the plurals of *Mošupologo* is shown graphically in (10)(b).

(10)(b) Graphic presentation: Monday



Just as for *Mokibelo*, the fieldwork shows that *Mošupologo* belongs to gender 3/4.

As far as the other days of the week are concerned, the analysis of the fieldwork shows a clear pattern, except for yet another loanword, *Foreitaga*, 'Friday'. Since PSC has again not a single occurrence of this alternative, we will therefore not discuss it below. The fieldwork data for the days Tuesday to Friday have been summarised in (11) to (14).

(11) Fieldwork results: Tuesday

Singular	<i>Freq.</i>	Plural	<i>Freq.</i>	Option 1	Option 2	Option 3
Labobedi	61	Malabobedi	0	31	6	1
		boLabobedi	0	14	9	0
		Dilabobedi	0	13	6	1
		Labobedi	61*	6	0	0
		Labobedi a mabedi	0	1	0	0
		Labobeding	0	1	0	0
		Mabobedi	0	1	0	0
		Matšatši a mabedi	96*	1	0	0
		Melabobedi	0	0	0	2
		Abobedi	0	0	0	1
		boMedilabobedi	0	0	0	1
		<i>no answer</i>		32	79	94

(12) Fieldwork results: Wednesday

Singular	<i>Freq.</i>	Plural	<i>Freq.</i>	Option 1	Option 2	Option 3
Laboraro	68	Malaboraro	0	29	6	1
		boLaboraro	0	16	10	0
		Dilaboraro	0	12	6	0
		Laboraro	68*	8	0	0
		Laboraro a mararo	0	1	0	0
		Maboraro	0	1	0	0
		Matšatši a mararo	126*	1	0	0
		Melaboraro	0	0	0	2
		Aboraro	0	0	0	1
		boMedilaboraro	0	0	0	1
		<i>no answer</i>		32	78	95

(13) Fieldwork results: Thursday

Singular	<i>Freq.</i>	Plural	<i>Freq.</i>	Option 1	Option 2	Option 3
Labone	59	Malabone	0	27	4	1
		boLabone	0	17	10	0
		Dilabone	0	14	5	1
		Labone	59*	8	0	0
		Amane	45*	1	0	0

Labone a mabedi	0	1	0	0
Labone ye mebedi	0	1	0	0
Melabone	0	0	0	2
Abone	3*	0	0	1
boMedilabone	0	0	0	1
<i>no answer</i>		31	81	94

(14) Fieldwork results: Friday

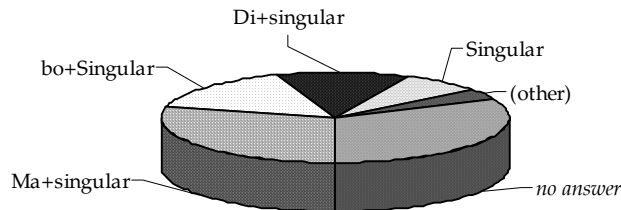
Singular	Freq.	Plural	Freq.	Option 1	Option 2	Option 3	
Labohlano	127	Malabohlano	0	27	5	1	
		boLabohlano	0	17	9	0	
		Dilabohlano	0	11	4	2	
		Labohlano	127*	8	0	0	
		Labohlano a mabedi	0	2	0	0	
		Amahlano	0	1	0	0	
		Labohlano ye mebedi	0	1	0	0	
		Matšatši a mahlano	13*	0	1	0	
		Melabohlano	0	0	0	2	
		Abohlano	0	0	0	1	
		boMedilabohlano	0	0	0	1	
		<i>no answer</i>			33	81	93
		Foreitaga	0	Diforeitaga	0	40	4
Foreitaga	0			8	0	0	
boFreitaga	0			4	4	0	
Maforeitaga	0			2	0	0	
<i>no answer</i>					46	92	98

Even though we are dealing with up to 10 or 11 possible plurals for one form of the singular — and hence, even though *single* mother-tongue intuition definitely fails —, the *average* mother-tongue intuition indicates that the preferred strategy is to prefix these days with *ma+*, followed by the strategies to prefix with *bo+* and *di+*. Actually, the sequence and the percentages for these four days are so similar that one can safely make an average of the strategies for the days in the range Tuesday to Friday. These averages are shown in (15).

(15)(a) Fieldwork results: Tuesday to Friday

Plural strategy	Tue.	Wed.	Thu.	Fri.	Average
Ma+singular	31	29	27	27	28.5
bo+Singular	14	16	17	17	16
Di+singular	13	12	14	11	12.5
Singular	6	8	8	8	7.5
(other)	4	3	3	4	3.5
<i>no answer</i>	32	32	31	33	32

(15)(b) Graphic presentation: Tuesday to Friday



7. The value of the corpus

In par. 4 we reviewed the lexicographers' failure to pinpoint the plurals of the days of the week; in par. 5 we showed clearly that all existing (bilingual) dictionaries of Sepedi poorly treat the singulars of the days, and are even more vague (and partly wrong) when it comes to the plurals; and in par. 6 we saw how balanced fieldwork can bring a solution. Yet, there is a fourth and even a fifth level that must be taken into account. Level four consists of results from corpus queries, and level five of facts derived from the existing scientific literature. The former will be discussed in this paragraph, the latter in the next paragraph.

Calzolari (1996: 4) adequately summarises why data culled from corpora differ from the results obtained by means of other types of information retrieval:

Carefully constructed, large written and spoken corpora are essential sources of linguistic knowledge if we hope to provide extensive and adequate descriptions of the concrete use of the language in real text. These types of descriptions certainly remain impossible if we only rely on introspection and native speaker intuition [...].

As noted in par. 2, PSC currently stands at 5.8 million running words. In building PSC, sections were sampled from several hundreds of written sources and a corpus of 5.8 million words roughly corresponds to the equivalent of 300 books. In building PSC, utmost care was taken to structure the corpus in such a way as to keep it stable. (For more information on how to build African-language corpora, see De Schryver and Prinsloo (2000b), and for an extensive discussion of the notion of corpus stability, see Prinsloo and De Schryver (forthcoming).)

Compared to intuition, informant elicitation and grammatical conjectures, the corpus has the big advantage in that it shows *real* language use. Any corpus query shows the 'attested and authentic usage *average*' of several hundreds of mother-tongue speakers. In (5) we have listed the PSC frequencies for the

various singular-day alternatives. For a corpus-based dictionary like PyaSsaL, this immediately implies that the loanwords *Mantaga* and *Foreitaga* will not be included in the dictionary, nor the low-frequency variant *Mošopologo* for Monday or the alternative *Lamodimo* for Sunday. The fieldwork satisfactorily supports this, as the mother-tongue speakers were very unsure when it came to the plurals of these items, or they simply disregarded them. Conversely, the situation with the loanword *Sontaga* is totally different. Indeed, as the second most-frequently used day of the week, it must not only be included, but must also be given a comprehensive lexicographic treatment.

Together with the fieldwork results presented in (8) through to (15), the PSC frequencies were also indicated. These frequencies are telling indeed, yet one must also carefully consider the context in order to see whether the mother-tongue suggestions are truly plurals of the days of the week. Wherever there is a severe problem with this, a star follows the frequency. As a matter of fact, the corpus shows that *Mokibelo*, *Sontaga*, *Lamorena*, *Lamodimo*, *Mošupologo*, *Labobedi*, *Laboraro*, *Labone* and *Labohlano* are all always referring to a singular. Furthermore, other suggestions have nothing to do with a plural, such as *Ga Morena* 'at the place of God', *Ba Morena* '(people) of God', *Amane* 'involves' and *Abone* 'when he/she saw'. And finally, PSC shows that still other suggestions for plurals are simply counts of days, such as *Matšatši a mabedi* 'two days', *Matšatši a mararo* 'three days' and *Matšatši a mahlano* 'five days'.

In a multicultural and multilingual environment like South Africa, it is crucial to make sure whether or not words from the (former) dominant languages are not perhaps used instead of indigenous ones. In the entire PSC, in a search for all possible singulars and plurals in English and Afrikaans, only two English days occur. These are shown in (16).

- (16) *Ke Friday* today, *ge re fetša fela mo o name o tsebe. Go a iwa.*
 'It is *Friday* today, once we finish here, you must know, we go.'
Karabo yona ya re, "Ka Sontaga — next Sunday, today a week — at 7 a.m."
 'The answer was: "On Sunday — next *Sunday*, today a week — at 7 a.m."'

We can therefore safely assume that, at least in the written language, the Sepedi forms are used. One would have to see whether or not an oral corpus component would alter this finding, especially since informal observation indicates that code switches are much more frequent in spoken than in written language. Frequency markers derived from the corpus should therefore include a label differentiating between written and spoken frequencies — as is done in LDOCE3.

A corpus can also be queried with the use of wildcards, and this reveals one single instance of a plural that was *not* discovered through the fieldwork. This is shown in (17).

- (17) *Gona fao ka matšatši a boSatertaga, le boSontaga go fetoga Sodoma ye nnyane, ka ge botagwe le tšohle tše mpe di direga.*

'There, days such as Saturdays and Sundays are turned into a small Sodom because drunkenness and other bad things occur.'

Single occurrences are however not enough to base conclusions on.

Finally, the loanword *Sabatha* is used in the Bible and just one other book. Here the problem is that this item has no distribution across a variety of different sources to warrant inclusion in PyaSsaL. Compare in this regard Knowles (1983: 188) who claims that "a word must occur evenly in a large number of the stratified sub-samples rather than excessively often in a small number of them, given that these two very different cases could show identical 'total-corpus' frequencies".

8. The grammarians' point of view compared to the results thus far

Unfortunately, as far as the days of the week are concerned, a systematic trawl through existing textbooks, journals and monographs did not reveal anything substantial. To make up for this, we consulted with two Sepedi grammarians, Dr E. Taljard and Prof. L.J. Louwrens. Both scholars were presented with the results of the study presented above, and their comments will now be interwoven with an overview of those results.

Firstly, Louwrens (*personal communication*, 15 June 2001) argues that the accepted way to form plurals in the field of TIME is by affixing *bo+* at the beginning of a word that already has a class prefix. To illustrate this point, he puts forward the examples listed in (18) as 'acceptable'.

(18)	<i>Kresemose</i>	>	<i>boKresemose</i>
	'Christmas day'		'Christmas days'
	<i>Go bunwa mabele ka Julae.</i>	>	<i>Go bunwa mabele ka boJulae.</i>
	'Maize is harvested in July.'		'Maize is harvested around July.'
	<i>Ba goroga ka iri ya lesome.</i>	>	<i>Ba goroga ka iri ya bolesome.</i>
	'They arrive at the hour of ten.'		'They arrive at around the hour of ten.'
	<i>Mošupologo</i>	>	<i>boMošupologo</i>
	'Monday'		'Mondays'
	<i>Ke letšatši la Labohlano ...</i>	>	<i>Ke matšatši a boLabohlano ...</i>
	'It is (the day of) Friday ...'		'It is (the days of) Fridays ...'
	<i>Ke letšatši la Mokibelo ...</i>	>	<i>Ke matšatši a boMokibelo ...</i>
	'It is (the day of) Saturday ...'		'It is (the days of) Saturdays ...'

This would imply that every single day of the week can take the prefix *bo+* to form some kind of plural. The fieldwork supports this.

Secondly, Louwrens emphasises that the plural prefix *di+* is productively used for the formation of plurals of loanwords, and gives the examples shown in (19).

(19)	<i>tafola</i> 'table'	>	<i>ditafola</i> 'tables'
	<i>khomphuta</i> 'computer'	>	<i>dikhomphuta</i> 'computers'

This would substantiate the finding that the plural of the loanword *Sontaga* is *Disontaga*.

Thirdly, Louwrens has his doubts as to the formation of plurals by prefixing *ma+* to the singulars of days that are non-loanwords.

Taljard (*email*, 18 June 2001) agrees that *bo+* can be prefixed to all days of the week, yet she remarks:

There is a distinct semantic difference between *boSontaga* (Bosontaga?) and *Disontaga*, but this will only become clear when these forms are tested within context. The prefix *bo+* is often used to indicate associative plurality. Usually, ordinary plurals express distributive plurality, thus *monna* 'one man', but *banna* 'many men'. When one uses *bo+* it can indicate the same kind of plurality, but it can also indicate associative plurality. Thus *botate* can mean either 'fathers' in the sense of one, two, three, four fathers (e.g. These kids all have different fathers), but it can also mean 'father and company' i.e. other people who are associated with him on the basis of certain shared semantic features (e.g. father and his brothers, father and the other members of the *kgoro*, etc.). The prefix *bo+* very often expresses associative plurality when used together with adverbs, which is a function frequently fulfilled by the days of the week. [...] (I have used *Sontaga* as an example, but it will also be valid for the other days of the week.)

Further, Taljard also agrees with Louwrens when it comes to the formation of the plurals of loanwords, i.e. through the prefixing of *di+*.

Finally, from a strictly grammatical point of view, she suggests the analogy shown in (20).

- (20) **L**Etšatši **LA** bobedi > **MA**tšatši **A** bobedi
 'The second day' > 'The second days'

Therefore, a "logical" conclusion could be as follows:

Labobedi	>	A bobedi	>	Abobedi
'Tuesday'				'Tuesdays'
Laboraro	>	A boraro	>	Aboraro
'Wednesday'				'Wednesdays'
Labone	>	A bone	>	Abone
'Thursday'				'Thursdays'
Labohlano	>	A bohlano	>	Abohlano
'Friday'				'Fridays'

The fieldwork brought up each of these options, yet each only once and in each case only as the third option. In addition, in PSC only *Abone* occurs, yet with the meaning 'when he/she saw', which should have been spelt as two words, namely *A bone*. Both the fieldwork and PSC therefore do not support Taljard's grammatical speculation. What is clear from (20) however, is the way in which the days Tuesday to Friday were formed — as the second, third, fourth and

fifth day, where the word *letšatši* 'day' was dropped and the remainder written conjunctively. In this context, the invented **Lamathomo* (*< (*Letšatši*) *la mathomo* 'The first (day)') for Monday (cf. par. 6) was not so eccentric. Furthermore, it is also clear from (20) that the singulars of the days Tuesday to Friday belong to class 5. This is confirmed in PSC, as all concords for these days are class 5 concords. This same method of looking at concords can also be used to pinpoint the classes of the plurals. For instance, the fieldwork suggests that the plural of *Lamorena* 'Sunday' (< (*Letšatši*) *la morena* 'the Lord's (day)') is *Malamorena*, thus gender 5/6. The concords in PSC confirm this too, as can be seen in (21).

- (21) *Ke lemoga lebaka leo bjale ka gore e šetše e le Malamorena a mararo a go hlomagana, o sa thiše kerekeng.*
'I am aware of that, now that it has been three consecutive Sundays that you have not absented yourself from church.'

The fieldwork and the corpus therefore indicate that Tuesday to Friday (*Labobedi*, *Laboraro*, *Labone* and *Labohlano*), and Sunday (ii) (*Lamorena*) belong to gender 5/6. The fieldwork and the corpus are also unanimous when it comes to assigning gender 3/4 to Monday (*Mošupologo*) and Saturday (*Mokibelo*), and gender 9/10 to Sunday (i) (*Sontaga*). All this information will have to be indicated in PyaSsaL.

The only type of plural that needs further study at this point is *bo*+Singular. Van Wyk, in his revision of Kriel's *Pukuntšu*, suggests as plurals for *Mokibelo* and *Sontaga*, *boMokibelo* and *boSontaga* respectively. This possibility is confirmed in PSC, with both associative and distributive meanings.⁸ Furthermore, the fieldwork indicates that all days of the week can take *bo*+ as prefix, and also Louwrens and Taljard suggest this possibility. The question arises however, whether the possibility of the plural *bo*+Singular should be indicated in PyaSsaL. In an enlightening article, Van Wyk (1987: 34) claims that "the morpheme *bo*- [...] can be used as a pluralizer and a nominalizer with an almost unlimited range of nouns, other parts of speech, phrases, and sentences". With this knowledge, it is obviously not a good idea to tell the dictionary user at every day of the week (and at almost every noun, for that matter) that one can add *bo*+ to the singular to form some kind of plural. As far as the days of the week are concerned, Van Wyk (1987: 37) himself gives the example shown in (22).

- (22) *ba bantši ba hwile ka bo-labone*
'many died on or around Thursday'

It is now appropriate to bring all the data together, i.e. (a) the information one can find in existing (bilingual) dictionaries, (b) the results from the fieldwork, (c) the occurrences in the corpus, and (d) the grammarians' input. As explained in the previous paragraph, cross-comparing corpus frequencies for the singulars of days with the data from the fieldwork, compels us to keep only one pos-

sibility per day, except for the variant for Monday and the alternative for Sunday. Also, we will only focus on the truly frequent suggestions.

(23) Cross-comparing the various data sources for the plurals of the days of the week

Days	Singulars	PSC 5.8M	Plurals	Z&M 1975	V Wyk 1989	Field. 2001	PSC 5.8M	Louwr. 2001	Taljad 2001
Mon.	Mošupologo	✓	Me- bo+	- -	- -	✓ ✓	- -	✓ ✓	✓ ✓
	Mošupulogo	✓	Me- bo+	- -	✓ -	- -	- -	✓ ✓	✓ ✓
Tue.	Labobedi	✓	Ma+ bo+	- -	(✓) -	✓ ✓	- -	? ✓	- ✓
Wed.	Laboraro	✓	Ma+ bo+	- -	(✓) -	✓ ✓	- -	? ✓	- ✓
Thu.	Labone	✓	Ma+ bo+	- -	(✓) -	✓ ✓	- -	? ✓	- ✓
Fri.	Labohlano	✓	Ma+ bo+	- -	(✓) -	✓ ✓	- -	? ✓	- ✓
Sat.	Mokibelo	✓	Me- bo+	- -	- ✓	✓ ✓	✓ ✓	✓ ✓	✓ ✓
Sun.	Sontaga	✓	Di+ bo+	✓ -	- ✓	✓ ✓	✓ ✓	✓ ✓	✓ ✓
	Lamorena	✓	Ma+ bo+	- -	(✓) -	✓ ✓	✓ -	? ✓	- ✓

With all this information it is now possible to treat the closed set 'days of the week' in such a way that the average of all the approaches is reflected. This will be done in the next section.

9. Towards a sound lexicographic treatment of days in Sepedi

A full treatment of all the days listed in (23) can be found in Appendix E. Firstly, compared to major learners' dictionaries for English such as COBUILD2 and LDOCE3, one will notice that circular definitions within this closed set were combated. Secondly, as PyaSsaL is fully corpus-based, all example sentences were culled directly from PSC, and this with minimal editing. Where attested in PSC, both singular and plural example sentences were included. Thirdly, moving to the macrostructure, one notices that the frequent items have been marked much as is done in LDOCE3, with **N** (*mo polelong ye e ngwadilwego*) 'W (in the written language)' introducing frequencies in written Sepedi. A ① indicates that the lemma sign belongs to the top 1 000 items, a ② that it belongs to the top 2 000, and a ③ that it belongs to the top 3 000. As an illustration, the data for Saturday and Sunday are shown in (24) and (25) respectively, together with an approximate English translation.

(24) Towards a sound lexicographic treatment of Saturday

N② **Mokibelo** [Mokibêlô] *leina 3/4*
Ke letšatši la bošupela la beke, le
thoma ka morago ga Labohlano
gomme la fela ge go thoma Sontaga: *Ka Mokibelo re tsogela pitšong; O e apare fela ka Mekibelo le Disontaga, le gona ge go na le phathi; Ka boMokibelo ba inwela malwa le bašomi ba bangwe ka fao lokheišeneng*

Mekibelo [Mekibêlô] *leina 3/4*
BONA **Mokibelo**

W② **Saturday** /'sætədeɪ || -ər-/ *noun, singular*
The sixth day of the week, starting after Friday and ending when Sunday begins: *On Saturday we go to the meeting; You must wear it only on Saturdays and Sundays, and only when there is a party; On Saturdays they drink beers together with other workers in the township*
Saturdays /'sætədeɪz || -ər-/ *noun, plural*
SEE **Saturday**

(25) Towards a sound lexicographic treatment of Sunday

N③ **Sontaga** [Sôntaga] *leina 9/10*
Ke letšatši la bošupa la beke, le
thoma ka morago ga Mokibelo
gomme la fela ge go thoma Mošupologo {LEHLALOŠETŠAGOTEE
Lamorena}: *O be a fela a etla ga gešo gagolo ka Sontaga ka morago ga kereke; Baruti ba theogela ka Disontaga*

Disontaga [Disôntaga] *leina 9/10*
BONA **Sontaga**

Lamorena [Lamorêna] *leina 5/6*
Ke letšatši la bošupa la beke, le
thoma ka morago ga Mokibelo
gomme la fela ge go thoma Mošupologo {LEHLALOŠETŠAGOTEE
Sontaga}: *Ge e le Lamorena ga re reke felo; Ke lemoga lebaka leo bjale ka gore e šetše e le Malamorena a mararo a go hlomagana, o sa thiše kerekeng*

Malamorena [Malamorêna] *leina 5/6*
BONA **Lamorena**

W③ **Sunday** /'sʌndeɪ/ *noun, singular*
The seventh day of the week, starting after Saturday and ending when Monday begins {SYNONYM **the Lord's Day**}: *He used to come to my place mostly on Sunday after church; Priests work on Sundays*

Sundays /'sʌndeɪz/ *noun, plural*
SEE **Sunday**

the Lord's Day /'ðə lɔ:ds || lɔ:ds deɪ/ *noun, singular*
The seventh day of the week, starting after Saturday and ending when Monday begins {SYNONYM **Sunday**}: *On the Lord's Day we have nowhere to buy; I am aware of that, now that it has been three consecutive Lord's Days that you have not absented yourself from church*

the Lord's Days /'ðə lɔ:ds || lɔ:ds deɪz/ *noun, plural*
SEE **the Lord's Day**

In the paper version of PyaSsaL, the bottom of every left-hand side or even page displays the even repetitive inserted text (ERIT) shown in (26), whilst the bottom of every right-hand side or odd page displays the odd repetitive inserted text (ORIT) shown in (27).⁹

(26) Even repetitive inserted text (ERIT) in the paper version of PyaSsaL

Lešala	1 †	1 ††	2 †	2 ††	1	2	1a	2a	3	4	5	6	7	8	9	10	14	15	16	17	18
Hlogo	nna	rena	wena	lena	mo-	ba-	-	bo-	mo-	me-	le-	ma-	se-	di-	N-	diN-	bo-	go-	fa-	go-	mo-
† (†)	ke	re	o	le	o	ba	o	ba	o	e	le	a	se	di	e	di	bo	go	go	go	go
⊗ (⊗)	N-	re	go	le	mo	ba	mo	ba	o	e	le	a	se	di	e	di	bo	go	go	go	go

(27) Odd repetitive inserted text (ORIT) in the paper version of PyaSsaL

1 †	Botee bja mmoledi	1 ††	Bontši bja mmoledi	Ⓛ, Ⓜ le Ⓝ	mantšu a 1 000, 2 000 le 3 000 a mathomo
2 †	Botee bja mmoledišwa	2 ††	Bontši bja mmoledišwa	N	mo polelong ye e ngwadi- lwego
† (†)	Lekgokasediri	⊗ (⊗)	Lekgokasedirwa	B	mo polelong ye e bolelwago
Dinomoro tše di tšwelelago ka morago ga 'leina' di bontšha gore leina leo le hlalošwago ke la legoro lefe.					

The gender information that accompanies every noun is linked with ERIT and ORIT. With ERIT and ORIT, dictionary users are — at a single glance — not only provided with data on how the plural can be derived from the singular (or *vice versa*), but also on how to go about concordial relationships and concordial references. It should further be obvious, although the data have been grouped in (24), (25) and Appendix E, that singulars and plurals are scattered in a semasiological dictionary. The data for Sunday(s), for instance, will be found under the alphabetical categories **D**, **L**, **M** and **S**. In the electronic version of PyaSsaL, ERIT and ORIT are replaced with grammatical pop-up windows, whilst the actual 'placement' of the data is of less concern to the electronic-dictionary users.

10. Retrieving feedback on the followed lemmatisation approach

In line with the theoretical framework of Simultaneous Feedback (SF, cf. par. 2), feedback was retrieved on the lemmatisation approach illustrated in Appendix E, and (26) and (27). PyaSsaL 1.0, i.e. *PyaSsaL's First Parallel Dictionary* (De Schryver 2001), contained an earlier version of the data grouped in Appendix E. During a special session at the Sixth International Conference of the African Association for Lexicography, PyaSsaL 1.0 was presented and distributed among some 50 conference attendees. A questionnaire had also been prepared in Sepedi and English, and the aim was to retrieve a first impression of PyaSsaL 1.0 from the mixed audience of both seasoned lexicographers and lexi-

cographers-to-be, and from mother-tongue speakers, second- and third-language speakers, as well as from those for whom Sepedi is unknown. Of the 15 questions, only one (question 6) explicitly dealt with plurals of the days. This question is shown in (28).

- (28) *Go ya ka pukuntšu ye Labobedi ka bontši ke eng mo polelong ya Sesotho sa Leboa?*
 'According to this dictionary, what is the plural of **Labobedi** (Tuesday) in Sepedi?'

It can be expected that the brief presentation of PyaSsaL 1.0, together with the information (both in Sepedi and English) in the front matter of this dictionary, would be sufficient a basis for arriving at the correct answer. The analysis of the answers to (28) is shown in (29).

- (29) Retrieval of feedback on the followed lemmatisation approach (here for *Tuesdays*)

Suggested plurals of Labobedi	Mother-tongue speakers	2nd/3rd-language speakers	Foreign-language speakers
Malabobedi	29 %	50 %	28 %
Bolabobedi	–	–	6 %
Di- Labobedi	–	10 %	–
Mabobedi	–	–	22 %
Amabedi	–	–	6 %
Kalabobedi	–	–	6 %
<i>wrong explanation</i>	57 %	20 %	13 %
<i>no answer</i>	14 %	10 %	19 %
	100 %	100 %	100 %

The analysis presented in (29) is in line with previous rounds of feedback retrieval in South Africa (cf. De Schryver and Prinsloo 2000: 205-208). Firstly, it is not surprising that foreign-language speakers struggle with the conventions in a monolingual Sepedi dictionary. Only 28% of them pinpointed the correct plural using the gender information together with ERIT and ORIT. Secondly, up to half of the second- and third-language speakers were able to decode the conventions. Here one must bear in mind that many attendees in this group are actually lecturers of Sepedi (at university level), and are thus quite familiar with the use of dictionaries. Lastly, the mother-tongue speakers performed the worst of all. From par. 4 we know however that mother-tongue intuition fails in this context. The inability to answer correctly, especially considering the fact that the respondents had only just received PyaSsaL 1.0, can be attributed to a presumed lack of dictionary culture. From a metalexicographic perspective, this has been pointed out e.g. by Gouws (1999: 7, 11), while Atkins (1998: 3) has observed: "The speakers of African languages have not in their formative years had

access to dictionaries of the richness and complexity of those currently available for European languages. They have not had the chance to internalize the structure and objectives of a good dictionary, monolingual, bilingual or trilingual."

In this context it is interesting to compare the answers to question 9 in the same questionnaire. This question is shown in (30).

- (30) *Ke ka lebaka lang ge dinomoro tše dingwe di swiswadišwe mola go tše dingwe go se bjalo, mohlala **dinku leina 9/10** ?*
'Why are some numbers in bold and others not, for example **dinku leina 9/10** ?'

Here only 43% of the mother-tongue speakers answered correctly, while up to 60% of the second- and third-language speakers, and 56% of the foreign-language speakers knew the answer. The presumed lack of dictionary culture among mother-tongue speakers of Sepedi might again explain this finding.

Imbedded in SF is the fact that potential users continuously guide the compilers during the entire compilation process. The unabated retrieval of feedback can therefore be considered as the main strength of the methodology. According to Atkins and Varantola (1997: 1), "[t]here are two direct routes to more efficient dictionary use: the first is to radically improve the dictionary: the second is to radically improve the users". If one looks at the percentages in (29), one sees that if one would want to make PyaSsaL more accessible to foreign-language speakers, one should envisage a more explicit guidance when it comes to the formation of the plurals. Instead of just *leina 5/6* under **Labobedi**, one could — in order to improve the dictionary — for instance consider *leina 5/6 (Ø/Ma+)*, hereby explicitly telling the user that the plural is formed by *adding* Ma+ to the singular (hence **Malabobedi**, and not **Mabobedi**). Nonetheless, PyaSsaL is a dictionary aimed at mother-tongue speakers, and for them the percentages in (29) clearly indicate that one should walk the second route suggested by Atkins and Varantola, i.e. 'radically improve the users'. Therefore, while the Second Parallel Dictionary is being compiled, important efforts are directed towards the explicit and systematic teaching of dictionary skills in the classroom, as e.g. suggested by Chi (1998: 566). Formulated differently, while the Main Dictionary is being compiled, the future users are simultaneously being trained in using it, as early feedback (here only exemplified for the plural of one day of the week) indicated that there is a great need for this.

11. In conclusion

In this article we have examined the various compilation aspects to be confronted by the lexicographers active within the Sepedi NLU. We have seen that a sound treatment of the lexicon might require more than introspection and the tools present in a standard dictionary-compilation office, i.e. existing (bilingual) dictionaries, large electronic corpora, and grammar books. Indeed, for

many semantic (sub-)fields, additional fieldwork might well turn out to provide the most reliable data. It is the interplay between all these different types of data that ultimately enables the mother-tongue lexicographers to compile sound (i.e. reliable and truly representative) dictionary articles.

We have also shown clearly that an onomasiological approach to the lexicon combats circularity between dictionary definitions, and enables more consistency. Working within the framework of Simultaneous Feedback has the added advantage that the work can be brought out to the future target users, and amended if need be while the Main Dictionary is still under compilation.

Finally, this article has further indicated how practical lexicographers can be brought to engage in fundamental scientific research, if they are truly willing to provide the future users with the state of the art of their own language. As far as the days of the week are concerned, the original research reported on here has pinpointed the genders of these days for the very first time.

Endnotes

1. The two full-time lexicographers are M.P. Mogodi and M.C. Mphahlele, the part-time lexicographer is B. Lepota, the two corpus builders are S. Nong and B.P. Sathekge, and the facilitator is G.-M. de Schryver.
2. For more information on Onoma, the home page of *Lexilogik* can be consulted at: <http://www.lexilogik.se>
3. Structured Query Language.
4. Actually, corpora for all South African languages have been built at UP's Department of African Languages. The sizes of these are in constant evolution. For the latest developments, the home page of *ELC for ALL* (Electronic Corpora for African Language and Linguistics) can be visited at: <http://www.up.ac.za/academic/libarts/afrilang/elcforall.htm>
5. For more information on *WordSmith Tools*, the home page of Mike Scott can be consulted at: <http://www.lexically.net> (or its mirror site: <http://www.liv.ac.uk/~ms2928>).
6. A thorough study has indicated that the *Pukuntšu* as revised by Van Wyk (Kriel et al. 1989⁴) contains an average of two errors in alphabetical order per page.
7. The *Comprehensive Northern Sotho Dictionary* being a stem-based dictionary, this is particularly unsatisfactory. Indeed, the user first tries to find *Lamodimo*. Upon realising that this item has not been included as such, the user will try to find *Lamodimo* under *-modimo*, then under *-dimo*, and finally under *-mo* — all to no avail.
8. In the corpus, just as in grammar books, the morpheme *bo+* is written in small letters and prefixed to whatever form, whether that form starts with a capital or not. Hence, for instance *boSontaga* and not *Bosontaga*.
9.

Lešala	=	'Pronoun'
Hlogo	=	'Prefix'
1 †	=	First person singular
1 ††	=	First person plural
2 †	=	Second person singular
2 ††	=	Second person plural
† (†)	=	Subject concord
‡ (‡)	=	Object concord

①, ② le ③ = first 1 000, 2 000 and 3 000 items

N = in the written language

B = in the spoken language

Dinomoro tše di tšwelelago ka morago ga 'leina' di bontšha gore leina leo le hlalošwago ke la legoro lefe.

'The numbers appearing after the part of speech "noun" indicate the gender of the word that is being treated.'

References

- Atkins, B.T. Sue.** 1998. *Some Discussion Points Arising from Afrilex-Salex '98*. Unpublished course evaluation document. Pretoria: University of Pretoria.
- Atkins, B.T. Sue and Krista Varantola.** 1997. Monitoring Dictionary Use. *International Journal of Lexicography* 10(1): 1-45.
- Calzolari, Nicoletta.** 1996. Lexicon and Corpus: A Multi-faceted Interaction. Gellerstam, Martin, Jerker Järborg, Sven-Göran Malmgren, Kerstin Norén, Lena Rogström and Catarina R. Pappmehl (Eds.). 1996. *Euralex '96 Proceedings I-II, Papers submitted to the Seventh EURALEX International Congress on Lexicography in Göteborg, Sweden*: 3-16. Gothenburg: Department of Swedish, Göteborg University.
- Chi, Man L.A.** 1998. Teaching Dictionary Skills in the Classroom. Fontenelle, Thierry, Philippe Hilgsmann, Archibald Michiels, André Moulin and Siegfried Theissen (Eds.). 1998. *ACTES EURALEX'98 PROCEEDINGS, Communications soumises à EURALEX '98 (Huitième Congrès International de Lexicographie) à Liège, Belgique / Papers submitted to the Eighth EURALEX International Congress on Lexicography in Liège, Belgium*: 565-577. Liège: English and Dutch Departments, University of Liège.
- Departmental Northern Sotho Language Board.** 1988^a. *Northern Sotho Terminology and Orthography*. No. 4. Pretoria: Government Printer.
- De Schryver, Gilles-Maurice** (Ed.). 2001. *Pukuntšutlhaloši ya Sesotho sa Leboa 1.0 (PyaSsaL's First Parallel Dictionary)*. Pretoria: (SF)² Press.
- De Schryver, Gilles-Maurice and D.J. Prinsloo.** 2000. Dictionary-Making Process with "Simultaneous Feedback" from the Target Users to the Compilers. Heid, Ulrich, et al. (Eds.). 2000: 197-209.
- De Schryver, Gilles-Maurice and D.J. Prinsloo.** 2000a. The Concept of "Simultaneous Feedback": Towards a New Methodology for Compiling Dictionaries. *Lexikos* 10: 1-31.
- De Schryver, Gilles-Maurice and D.J. Prinsloo.** 2000b. The Compilation of Electronic Corpora, with Special Reference to the African Languages. *Southern African Linguistics and Applied Language Studies* 18(1-4): 89-106.
- De Schryver, Gilles-Maurice and D.J. Prinsloo.** 2000c. Electronic Corpora as a Basis for the Compilation of African-Language Dictionaries, Part 1: The Macrostructure. *South African Journal of African Languages* 20(4): 291-309.
- De Schryver, Gilles-Maurice and D.J. Prinsloo.** 2001. Fuzzy SF: Towards the Ultimate Customised Dictionary. Lee, Sangsup (Ed.). 2001. *ASIALEX 2001 PROCEEDINGS, Asian Bilingualism and the Dictionary*: 141-146. Seoul: Center for Linguistic Informatics Development, Yonsei University.

- Geeraerts, Dirk.** 2000. Adding Electronic Value. The Electronic Version of the *Grote Van Dale*. Heid, Ulrich, et al. (Eds.). 2000: 75-84.
- Gerber, Hendri H.** 2000. *Woordeboek Afrikaans – Noord-Sotho / Pukuntšu Seburu – Sesotho sa Leboa*. Eldoraigane: Arbeidsprestasie BK.
- Gouws, Rufus H.** 1999. *Die maatskaplike gerigtheid van die metaleksikografie in 'n meertalige samelewing*. Inaugural lecture. Stellenbosch: Universiteit van Stellenbosch Drukkery.
- Gouws, Rufus H.** 2000. Toward the Formulation of a Metalexigraphical Founded Model for National Lexicography Units in South Africa. Wiegand, Herbert E. (Ed.). 2000. *Wörterbücher in der Diskussion IV. Vorträge aus dem Heidelberger Lexikographischen Kolloquium*: 109-133. (Lexicographica Series Maior 100.) Tübingen: Max Niemeyer.
- Hartshorne, Kenneth B., J.H.A. Swart and Edgar Posselt.** 1984. *Dictionary of Basic English – N.Sotho across the Curriculum*. Johannesburg: Educum Publishers.
- Heid, Ulrich, Stefan Evert, Egbert Lehmann and Christian Rohrer** (Eds.). 2000. *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000, Stuttgart, Germany, August 8th-12th, 2000*. Stuttgart: Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.
- Kilgarriff, Adam.** 1997. Putting Frequencies in the Dictionary. *International Journal of Lexicography* 10(2): 135-155.
- Knowles, Frank.** 1983. Towards the Machine Dictionary, "Mechanical" Dictionaries. Hartmann, Reinhard R.K. (Ed.). 1983. *Lexicography: Principles and Practice*: 181-193. London: Academic Press.
- Kriel, Theunis J.** 1976⁴. *The New English – Northern Sotho Dictionary, English – Northern Sotho / Northern Sotho – English*. Johannesburg: Educum Publishers.
- Kriel, Theunis J.** 1983³. *Pukuntšu woordeboek, Noord-Sotho – Afrikaans / Afrikaans – Noord-Sotho*. Pretoria: J.L. van Schaik.
- Kriel, Theunis J., Egidius B. van Wyk and S.A. Makopo.** 1989⁴. *Pukuntšu woordeboek, Noord-Sotho – Afrikaans / Afrikaans – Noord-Sotho*. Pretoria: J.L. van Schaik.
- Kriel, Theunis J., D.J. Prinsloo and Bethuel P. Sathekge.** 1997⁴. *Popular Northern Sotho Dictionary, Northern Sotho – English / English – Northern Sotho*. Cape Town: Pharos.
- Laufer, Batia and Linor Melamed.** 1994. Monolingual, Bilingual and "Bilingualised" Dictionaries: Which are More Effective, for What and for Whom? Martin, Willy, Willem Meijs, Margreet Moerland, Elsemiek ten Pas, Piet van Sterkenburg and Piek Vossen (Eds.). 1994. *Euralex 1994 Proceedings, Papers submitted to the 6th EURALEX International Congress on Lexicography in Amsterdam, The Netherlands*: 565-576. Amsterdam: Vrije Universiteit.
- Prinsloo, D.J. and Bethuel P. Sathekge.** 1996. *New Sepedi Dictionary, English – Sepedi (Northern Sotho) / Sepedi (Northern Sotho) – English*. Pietermaritzburg: Shuter and Shooter.
- Prinsloo, D.J., Bethuel P. Sathekge and Lizeth Kapp.** 1997. *Nuwe Sepedi Woordeboek, Afrikaans – Sepedi (Noord Sotho) / Sepedi (Noord Sotho) – Afrikaans*. Pietermaritzburg: Shuter and Shooter.
- Prinsloo, D.J. and Gilles-Maurice de Schryver.** 1999. The Lemmatization of Nouns in African Languages with Special Reference to Sepedi and Cilubà. *South African Journal of African Languages* 19(4): 258-275.
- Prinsloo, D.J. and Gilles-Maurice de Schryver.** Forthcoming. Monitoring the Stability of a Growing Organic Corpus, with Special Reference to Sepedi and Xitsonga. *Dictionaries: Journal of the Dictionary Society of North America* 22.
- Sinclair, John M.** (Ed.). 1995². *Collins COBUILD English Dictionary*. London: HarperCollins.

- Summers, Della.** (Director). 1995³. *Longman Dictionary of Contemporary English, Third Edition*. Harlow: Longman Dictionaries.
- Van Wyk, Egidius B.** 1987. Proclitic *bo* of Northern Sotho. *South African Journal of African Languages* 7(1): 34-42.
- Ziervogel, Dirk and Pothinus C.M. Mokgokong.** 1975. *Pukuntšu ye kgolo ya Sesotho sa Leboa, Sesotho sa Leboa – Seburu/Seisimane / Groot Noord-Sotho-woordeboek, Noord-Sotho – Afrikaans/Engels / Comprehensive Northern Sotho Dictionary, Northern Sotho – Afrikaans/English*. Pretoria: J.L. van Schaik.
- Ziervogel, Dirk and Pothinus C.M. Mokgokong.** 1988⁴. *Klein Noord-Sotho woordeboek, N.-Sotho – Afrikaans – English/Afrikaans – N.-Sotho/English – N.-Sotho*. Pretoria: J.L. van Schaik.

Appendix A: Days in existing (bilingual) Sepedi dictionaries

Z & M: Ziervogel and Mokgokong 1975

LÁBÓBEDÍ (< tšatši la bobêdi) (Labobêdi)

Dinsdag // Tuesday; (< lentšu la bobedi) (labobêdi) altstem // alto (voice)

LÁBÓRÁRO (< letšatši la boraro) (Laboraro)

Woensdag // Wednesday; (< lentšu la boraro) tenoor(stem) // tenor (voice)

LÁBÓNE (< tšatši la bone) (Labonê) Don-

derdag // Thursday; (< lentšu la bone) (labonê) bas(stem) // bass (voice)

LÁBÓHLÁNO (< tšatši la bohllano) Vrydag

// Friday

SÓN'TAGA, (se-)/di- (Sôntaga) (< Afr.), cf.

LÁMODÍMO, Sondag // Sunday

LÁMORENA (Lamorêna) (< letšatši la

Morêna) Sondag // Sunday

New E: Kriel 1976⁴

Mošupo'logo, n., Monday.

'Labo'bedi, n., Tuesday.

'Labo'raro, n., Wednesday.

'La'bone, n., Thursday.

'Labo'hllano, n., Friday.

Mokibelo, Mo'ki'bêlô, n., Saturday.

'Lamo'dimo, n., Sunday.

Monday, Mošupologo, Mantaga (A.).

Tuesday, n., Labobedi.

Wednesday, Laboraro.

Thursday, Labone.

Friday, Labohlano.

Saturday, n., Mokibelo.

Sunday, Sontaga, Lamodimo.

Klein: Ziervogel and Mokgokong 1988⁴

Lábóbedí (Labobêdi) Dinsdag/Tuesday

Lábóráro Woensdag/Wednesday

Lábóne (Labonê) Donderdag/Thursday

Lábóhláno Vrydag/Friday

Sóntaga (se- kl./cl.) Sondag/Sunday

La'morena (Lamorêna) (< letšatši la Morêna)

Sondag/Sunday

Máandag Mošupologô

Dínsdag Labobêdi

Wóensdag Laboraro

Dónderdag Labone

Vrýdag Labohlano

Sáterdag Mokibêlô

Són dag Sôntaga

Monday Mošupologô

Tuesday Labobêdi

Wednesday Laboraro

Thursday Labonê

Friday Labohlano

Saturday Mokibêlô

Sunday Sôntaga

T & O: Departmental Northern Sotho Language Board 1988⁴

Days of the week / Dae van die week:

Moš. (Mošupologo)

Lbb. (Labobedi)

Lbr. (Laboraro)

Lbn. (Labone)

Lbh. (Labohlano)

Mok. (Mokibelo)

Lam. (Lamorena)

Matšatši a beke:

Moš (Mošupologo)

Lbb (Labobedi)

Lbr (Laboraro)

Lbn (Labone)

Lbh (Labohlano)

Mok (Mokibelo)

Lam (Lamorena)

V Wyk: Kriel et al. 1989⁴

Mošupulogô, snw. dev. kl 3, LLLLL: Maandag.

Mošopologô, snw. dev. kl 3, LLLLL: Maandag

Labobêdi, bsk. kl 5 + snw. HHLH: Dinsdag.

Laboraro, bsk. kl 5 + snw. HLHL: Woensdag.
Labonê, bsk. kl 5 + snw. HHL: Donderdag.
Labohlano, bsk. kl 5 + snw. HLHL: Vrydag.
Mokibêlô, snw. dev. kl 1a, LHHL: Saterdag.
Sôntaga, snw. kl 1a, HLLL: Sondag.
Lamorêna, bsk. kl 5 + snw. HHLL: Sondag,
dag van die Here.
Lamodimo, bsk. kl 5 + snw. HLHL: Sondag,
dag van die Here.

Maandag, Mošupologô, Mantaga.
Dinsdag, Labobêdi.
Woensdag, La boraro.
Donderdag, Labonê.
Vrydag, La bohllano, Forêitaga.
Saterdag, Mokibêlô.
Sondag, La Morêna, Sôntaga.

Prinsloo: Prinsloo et al. 1996

Mošupologô Monday
Labobêdi Tuesday
Laboraro Wednesday
Labonê Thursday
Labohlano Friday
Mokibêlô Saturday
Sôntaga Sunday
Lamorêna Sunday

Monday Mošupologô
Tuesday Labobêdi
Wednesday Laboraro
Thursday Labonê
Friday Labohlano
Saturday Mokibêlô
Sunday Sôntaga, La Morêna

Prinsloo: Prinsloo et al. 1997

Mošupologô Maandag
Labobêdi Dinsdag
Laboraro Woensdag
Labonê Donderdag

Labohlano Vrydag
Mokibêlô Saterdag
Sôntaga Sondag
Lamorêna Sondag

Maandag Mošupologô
Dinsdag Labobêdi
Woensdag Laboraro
Donderdag Labonê
Vrydag Labohlano
Saterdag Mokibêlô
Sondag Sôntaga, La Morêna

Popular: Kriel et al. 1997⁴

Mošupologô Monday
Mošupologô Monday
Labobêdi Tuesday
Laboraro Wednesday
Labonê Thursday
Labohlano Friday
Mokibêlô Saturday
Sôntaga Sunday
Lamorêna Sunday
Lamodimo Sunday

Monday Mošupologô, Mantaga
Tuesday Labobêdi
Wednesday Laboraro
Thursday Labonê
Friday Labohlano
Saturday Mokibêlô
Sunday Sôntaga, Lamodimo

Gerber: Gerber 2000

Maandag Mošupologô, Mošupologô
Dinsdag Labobêdi
Woensdag Laboraro
Donderdag Labonê
Vrydag Labohlano
Saterdag Mokibêlô
Sondag Lamorêna, Lamodimo, Sôntaga

Appendix B: Week* in *Basic English – N.Sotho* (Hartshorne et al. 1984)

W46 **week** (n) the seven consecutive days which begin with Monday. *There are roughly four weeks in each month and 52 weeks in a year.* 1 **weekday** (n) one of the days from Monday to Saturday. 2 **week-end** (n) the end of the week, usually from Saturday to Sunday. 3 **long week-end** *When a holiday falls on a Monday a long week-end is usually from Friday afternoon until Monday evening.* 4 **weekly** (adj) once a week or every week. **Weekly** newspapers are usually published on Sundays but some are published on other days of the week.

W46 beke 1 letšatši la beke 2 mphelabeke 3 mphelabeketelele 4 ka beke

Appendix C: Questionnaire to retrieve the form of day plurals, version 1

MATŠATŠI A BEKE A SESOTHO SA LEBOA

Thaetlele (Mna/Mtšana/Moh./Ngaka/Prof./ bj.bj.): _____

Sefane: _____

Leina: _____

Bong (Monna/Mosadi): _____

Mengwaga (Mohlala: 20): _____

HLOKOMELA: PAMPIRI YE E NA LE DIPOTŠIŠO TŠE HLANO FELA. O KGOPELWA GO ARABA DIPOTŠIŠO TŠE KA MOKA.

1. Ngwala bontši bja mantšu a a latelago ka mo tlase. Ge go kgonega, efa dikarabo tša go feta e tee.

Mohlala: Kgomo > (a) Dikgomo (b) Magomo

BOTEE	BONTŠI (a)	BONTŠI (b)	BONTŠI (c)
Lamorena			
Sontaga			
Mošupologo			
Mantaga			
Labobedi			
Laboraro			
Labone			
Labohlano			
Mokibelo			

2. Ke lengwalo lefe la thuto le o nago nalo? (Swaya ka X)

Praemari	
Matriki	
Kholetšheng	
Tekhnikon	
Yunibesithi	

3. O šoma eng? (Hlaloša ka boripana)

4. O dula kae? (Swaya ka X)

Polaseng	
Nagalegaeng	
Lokheišeneng	
Toropong	

5. Ke polelo goba dipolelo dife tše o di šomišago ka gae, mošomong, le ge o na le bagwera? (Swaya ka X)

Polelo	Gae	Mošomong	Bagwera
Sesotho sa Leboa			
Setswana			
Sesotho sa Moshoeshoe			
Setebele			
Sezulu			
Seswatsi			
Sethosa			
Setšhakane			
Sevenda			
Seisemane			
Seafrikanse			

RE LEBOGA TEMA YA GAGO MOŠOMONG WO!!!

Appendix D: Questionnaire to retrieve the form of day plurals, version 2

MATŠATŠI A BEKE A SESOTHO SA LEOA

HLOKOMELA: PAMPIRI YE E NA LE DIPOTŠIŠO TŠE HLANO FELA. O KGOPELWA GO ARABA DIPOTŠIŠO TŠE KA MOKA.

1. Ngwala bontši bja mantšu a o tsebago a kgona go tšwelela ka bontšing. Mo go kgonegago, efa dikarabo tša go feta e tee.

Mohlala: Kgomo > (a) Dikgomo (b) Magomo

BOTEE	BONTŠI (a)	BONTŠI (b)	BONTŠI (c)
Mošupologo			
Mantaga			
Lamathomo			
Labobedi			
Laboraro			
Labone			
Labohlano			
Foreitaga			
Mokibelo			
Lamorena			
Sontaga			
Lamodimo			

2. Ke lengwalo lefe la thuto le o nago nalo? (Swaya ka X)

Praemari	
Matriki	
Kholetšheng	
Tekhnikon	
Yunibesithi	

3. O šoma eng? (Hlaloša ka boripana)

4. O dula kae? (Swaya ka X)

Polaseng	
Nagalegaeng	
Lokheišeneng	
Toropong	

5. Ke polelo goba dipolelo dife tše o di šomišago ka gae, mošomong, le ge o na le bagwera? (Swaya ka X)

Polelo	Gae	Mošomong	Bagwera
Sesotho sa Leboa			
Setswana			
Sesotho sa Moshoeshoe			
Setebele			
Sezulu			
Seswatsi			

Sethosa			
Setsonga			
Sevenda			
Seisemane			
Seafrikanse			

Re kgopela gore o re botše ka bowena ka boripana gomme tshedimošo ye e ka se fiwe motho yo mongwe.

Thaetlele (Mna/Mtšna/Mdi./Ngaka/Prof./ bj.bj.): _____

Sefane: _____

Leina: _____

Bong (Monna/Mosadi): _____

Mengwaga (Mohlala: 20): _____

RE LEBOGA TEMA YA GAGO MOŠOMONG WO!!!

Appendix E: A sound lexicographic treatment of the days of the week based on the research presented in this article

Monday

N **Mošupologo** [Mošupologô] *leina 3/4*
 Ke letšatši la mathomo la beke, le thoma ka morago ga Sontaga gomme la fela ge go thoma Labobedi: *Ka Mošupologo ge re fihla sekolong morutiši a re biiša ka o tee ka o tee*
Mešupologo [Mešupologô] *leina 3/4*
 BONA **Mošupologo**
Mošupulogo [Mošupulogô] *leina 3/4*
 BONA **Mošupologo**
Mešupulogo [Mešupulogô] *leina 3/4*
 BONA **Mošupulogo**

Tuesday

Labobedi [Labobêdi] *leina 5/6*
 Ke letšatši la bobedi la beke, le thoma ka morago ga Mošupologo gomme la fela ge go thoma Laboraro: *E be e le Labobedi ge sefatanaga sa Ariel se ikgaša gare ga mpa ya Tshwane*
Malabobedi [Malabobêdi] *leina 5/6*
 BONA **Labobedi**

Wednesday

Laboraro *leina 5/6*
 Ke letšatši la boraro la beke, le thoma ka morago ga Labobedi gomme la fela ge go thoma Labone: *Ka Laboraro mo gare ga beke ye nngwe ba bolela gore ba tla tsoga ba eya Tshwane*
Malaboraro *leina 5/6*
 BONA **Laboraro**

Thursday

Labone [Labonê] *leina 5/6*
 Ke letšatši la bone la beke, le thoma ka morago ga Laboraro gomme la fela ge go thoma Labohlano: *O nteleditše mogala ka Labone*
Malabone [Malabonê] *leina 5/6*
 BONA **Labone**

Friday

Labohlano *leina 5/6*
 Ke letšatši la bohloano la beke, le thoma ka morago ga Labone gomme la fela ge go thoma Mokibelo: *O tla swara taxi ge a seno tšwa sekolo ka Labohlano*
Malabohlano *leina 5/6*
 BONA **Labohlano**

Saturday

N **Mokibelo** [Mokibêlô] *leina 3/4*
 Ke letšatši la boselela la beke, le thoma ka morago ga Labohlano gomme la fela ge go thoma Sontaga: *Ka Mokibelo re tsogela pitšong; O e apare fela ka Mekibelo le Disontaga, le gona ge go na le phathi; Ka boMokibelo ba inwela malwa le bašomi ba bangwe ka fao lokheišeneng*
Mekibelo [Mekibêlô] *leina 3/4*
 BONA **Mokibelo**

Sunday

N **Sontaga** [Sôntaga] *leina 9/10*
 Ke letšatši la bošupa la beke, le thoma ka morago ga Mokibelo gomme la fela ge go thoma Mošupologo {LEHLALOŠETŠAGOTEE **Lamorena**}: *O be a fela a etla ga gešo gagolo ka Sontaga ka morago ga kereke; Baruti ba theogela ka Disontaga*
Disontaga [Disôntaga] *leina 9/10*
 BONA **Sontaga**
Lamorena [Lamorêna] *leina 5/6*
 Ke letšatši la bošupa la beke, le thoma ka morago ga Mokibelo gomme la fela ge go thoma Mošupologo {LEHLALOŠETŠAGOTEE **Sontaga**}: *Ge e le Lamorena ga re reke felo; Ke lemoga lebaka leo bjale ka gore e šetše e le Malamorena a mararo a go hlomagana, o sa thiše kerekeng*
Malamorena [Malamorêna] *leina 5/6*
 BONA **Lamorena**

Orature and Morpholexical Deconstruction as Lexicographic Archaeological Sites: Some Implications for Dictionaries of African Languages¹

Francis Matambirofa, *Department of African Languages and Literature,
University of Zimbabwe, Harare, Zimbabwe (francism@compcentre.uz.ac.zw)*

Abstract: This article takes a multidisciplinary approach to African lexicographic practice. It has as its primary premise the assumption that without words there can be no dictionaries to compile and discuss. Owing to this fact, the article focuses on a specific strategy for collecting words which belong in a special category of their own, namely archaic or obsolete words. Such words are important because of the need to mark them as such in any general purpose dictionary. Most, if not all dictionaries of African languages seem not to have this category, giving the misleading impression that there are no such words in African languages. Apart from digging up archaic or what we have also referred to as artefact words, the article also argues that the words have a substantial and intrinsic etymological value. Thus they can be used in specialised etymological dictionaries of African languages or even in standard general dictionaries. The multidisciplinary aspect resides in the methodology proposed for the recovery of archaic words. It is considered necessary that disciplines such as oral literature, oral history, historical linguistics and to a limited extent theoretical linguistics and computational linguistics, and, symbolically, archaeology and lexicography itself, be brought to bear on the subject of inquiry. The article is also an attempt at working out a method which African lexicographers can employ as an instrument to dig up artefact words for etymological and other such purposes. Hopefully the method can be refined further.

Keywords: ORATURE, MORPHOLEXICAL, DECONSTRUCTION, RECONSTRUCTION, ARCHAEOLOGICAL SITES, LEXICOGRAPHY, AFRICAN LANGUAGES, DICTIONARIES, DIACHRONIC, SYNCHRONIC, HISTORICAL LINGUISTICS, PROVERBS, RIDDLES, IDIOMS, ORAL TRADITIONS, MNEMONIC, ETYMOLOGY, ARCHAIC, SHONA, DEVERBATIVE NOUNS, VERB ROOT, SEMANTIC, ARTEFACT WORDS

Opsomming: **Mondelinge literatuur en morfoleksikale dekonstruksie as leksikografiese argeologiese terreine: 'n Aantal implikasies vir woordeboeke van Afrikatale.** Hierdie artikel het 'n multidissiplinêre benadering tot die leksikografiese praktyk in Afrikatale. Dit het as primêre uitgangspunt die veronderstelling dat dit sonder woorde onmoontlik is om woordeboeke op te stel en te bespreek. Op grond van hierdie feit fokus die

artikel op 'n spesifieke strategie vir die versameling van woorde wat tot 'n kategorie van hul eie behoort, naamlik argaïese of verouderde woorde. Sulke woorde is belangrik vanweë die noodsaaklikheid om hulle as sulks te merk in enige woordeboek vir algemene gebruik. Die meeste, indien nie alle woordeboeke van Afrikatale nie skyn nie hierdie kategorie te hê nie, wat die misleidende indruk wek dat daar nie sulke woorde in Afrikatale is nie. Behalwe om argaïese, of waarna ons ook as artefakwoorde verwys het, op te diep, word daar ook in hierdie artikel geredeneer dat die woorde 'n wesenlike en werklike etimologiese waarde het. Daarom kan hulle in gespesialiseerde etimologiese woordeboeke van Afrikatale of selfs in algemene standaardwoordeboeke gegee word. Die multidisiplinêre aspek lê in die voorgestelde metodologie vir die terugvind van argaïese woorde. Dit word as noodsaaklik beskou dat vakgebiede soos mondelinge literatuur, mondelinge geskiedenis, historiese linguïstiek, in 'n beperkte mate teoretiese linguïstiek en rekenaaringuïstiek, en, simbolies, argeologie en leksikografie self, aangewend word by die onderwerp van ondersoek. Hierdie artikel is ook 'n poging om 'n metode uit te werk wat leksikograwe van Afrikatale as 'n instrument kan gebruik om artefakwoorde op te diep vir etimologiese en ander doeleindes. Hopelik kan die metode verder verfyn word.

Sleutelwoorde: MONDELINGE LITERATUUR, MORFOLEKSIKAAL, DEKONSTRUKSIE, ARGEOLOGIESE TERREINE, LEKSIKOGRAFIE, AFRIKATALE, WOORDEBOEKE, DIACHRONIES, SINCHRONIES, HISTORIESE LINGUISTIEK, SPREEKWOORDE, RAAISELS, IDIOME, MONDELINGE TRADISIES, MNEMONIESE WOORDE, ETIMOLOGIE, ARGAIËS, SJONA, DEVERBATIEWE SUBSTANTIEWE, WERKWOORDELIKE STAM, SEMANTIES, ARTEFAKWOORDE

Introduction

This article deals with a sphere, the supporting evidence of which, chiefly coming from oral literature, is seldom the concern of linguistics. The broader sphere to which it belongs, is historical linguistics. The primary concern of this article is to develop an apparatus with which to extract and recover aspects of the diachronic state of Shona from its synchronic dimension for the benefit of African lexicography in general. Such a desire is motivated by the fact that Shona, in comparison with other Bantu languages, has a relatively short period of literacy from which to study its diachronic state. One could, broadly speaking, accord written Shona a little less than a century. While much effort to reduce the language into a written code was evinced by missionaries shortly before, but mainly after colonisation in 1890², this period is too short in terms of the scope of interest of this article.

Owing to this relatively short period of the written form of the language, we propose to use some disciplines of oral literature (henceforth orature) as lexicographic archaeological sites from which we will, in spite of the attendant limitations, glimpse into the past with the aim of extracting archaic words for lexicographic purposes. It is important to mark specified entries as archaic in Shona lexicography as well as in that of other African languages because this aspect of lexicography is erroneously being ignored. The only possible reason

for not paying attention to this category is to falsely assume that African languages are not changing and are therefore not casting off old words and usages which are worthy of the lexicographer's notice. But if they are, which is in fact the case, then African lexicographic practice must equally conform with and/or reflect standard international lexicographic practice where archaic words, just like other categories of entries, are indeed marked according to their status in different types of dictionaries.

Having noted the above, we hope to illustrate the point of the article from the major oral genres which are proverbs, riddles and idioms. The term archaeology, here used metaphorically, is borrowed from the branch of knowledge which deals with the cultural remains of a people. Using the term in its denotative sense, Finegan (1989: 255) points out:

Archaeological records enable researchers to make educated guesses about where our ancestors come from and where they migrated to, as well as how they lived and died.

The quest to glimpse into the past state of Shona, among other things, already suggests the lack of written records of the language, equally stretching back in the past as say English to, for instance, the times of Chaucer and Shakespeare. Benefiting from the long period of literacy of the English language, Finegan (1989: 252) is able to tell us:

About a thousand years ago, the English verb *starve* (Old English *steorfan*) meant simply "die" (by any cause); today *starve* refers principally to deprivation and death by hunger (or, by metaphorical extension, deprive of affection.)

Different writers (Fromkin and Rodman 1993, Finegan 1989, O'Grady et al. 1989, Aitchison 1995, etc.) all agree that all languages change imperceptibly over long periods of time. Referring directly to the importance of the written records of a language for historical research, Fromkin and Rodman (1993: 320), for instance, stress: "Many of the changes are revealed when languages have written records." Needless to say, Shona, alongside its sister languages, does not, as mentioned earlier, possess such written records from which such changes can be revealed.

On the Use of Orature or Oral Traditions

The study of the African past with *inter alia* the corroboration of oral records has become somewhat of an academic tradition in its own right. Many disciplines in the humanities take recourse to it in varying degrees. Included in these are history, anthropology, literature, cultural studies and to some extent archaeology. To these should be added linguistics and especially historical linguistics.

Before we look at how orature and oral traditions can benefit historical linguistic studies, it is necessary to define these terms. Orature can be described as a body of literary knowledge present in living traditions, the defining features of which comprise, among others, proverbs, tales, riddles, aphorisms, dictums, idioms etc. The term "oral tradition" is largely used in the domain of history. Vansina as quoted by Beach (1994: 251) defines it as "verbal messages which are reported statements from the past beyond the present generation ... spoken, sung or called out". Arguing from a purely historical point of view, Beach (1994: 244) informs us:

Oral traditions can be used to good effect to recover the pre-colonial past, providing that one does not expect too much.

A definite thread therefore links orature from the aspect of literary research and oral traditions to the study of history. The difference, if worth anything, is that oral traditions may only be viewed as specialised abstractions from the gamut of orature that is largely concerned with a people's history. In other words, oral traditions are a subgenre of orature and they therefore logically take on the broad defining aspects of orature itself.

Here we have established the characteristics of orature and it is therefore necessary to indicate its relevance to what we want to achieve. It should already be obvious that if indeed oral traditions can be put to "good effect to recover the pre-colonial past" then surely, apart from the recovery of historical and literary material, we can, *inter alia*, recover from it lexical materials for diachronic linguistic analysis. This is so because, after all, is it not the verbal and therefore linguistic base that conveys the literary material which is orature? In recovering the past from oral traditions or orature heed must however be taken of Beach's earlier cautionary comments where he says, "providing that one does not expect too much".

What to Recover and How to Recover it?

In this section, we want to give a brief outline of the following aspects on which this article will focus, viz.:

- (i) what we particularly aim to recover;
- (ii) the subgenres of orature targeted for the recovery procedure;
- (iii) the methodology of morpholexical deconstructive recovery;
- (iv) the methodology of reconstructive recovery; and
- (v) a cross-linguistic comparative analysis of lexical items.

From a linguistic point of view, we hope that lexical items especially can best be recovered *vis-à-vis* syntactic or phonological structures. We argue that the easiest to recover are unattached or freestanding lexemes as opposed to mor-

pholexically attached ones. For example, freestanding lexical items are **mwana** (child), **n'ombe** (cattle), **imba** (hut), **tsime** (well), etc. A host of lexical items are often connected in morpholexical word-formation processes such as derivation and inflection. In agentive nominal lexical items such as **musvuti** (smoker), **murimi** (father), **muroti** (dreamer), **muroyi** (witch) and **muurayi** (killer), we can recover related lexical items belonging to the grammatical category of verbs. A recovery demonstration will be given on two of these when we embark later in the discussion on its methodological aspect.

Fromkin and Rodman (1993: 322) point out that "no part of the grammar remains the same over the course of history". This indicates that an historical linguist can therefore look into the diachronic state of the language by means of a fairly long agenda dealing with phonology, syntax, morphology, semantics and pragmatics among others. The limiting factor can only be whether or not a particular area of interest is practically recoverable. Not all areas of linguistic interest can be studied with the same degree of success. Take for instance phonology, it is doubtful to what extent the "historical" tones on Shona historical names such as Munhumutapa, Changamire, Gatsi Rusere and many others can accurately be captured. There may indeed be no conjecture on the part of the present speakers that the tone they assign to different segments may after all be historically incorrect because there is no clue as to how the said names were pronounced in ancient times.

Moreover, it is not every aspect or subgenre of orature that passes for a productive, let alone plausible "linguistic archaeological site". With this in mind, our major aim is to unearth aspects of the language that specifically pertain to the lexical domain. It should therefore follow that we will identify subgenres of orature that have an inherent "lexical fossilisation" and "lexical preserving" tendency. Some such subgenres of orature are proverbs, riddles, idioms, incantations, songs, children's games and others. We shall, for our purposes in this article, however, largely explore proverbs. For illustrative purposes proverbs alone can, in our opinion, adequately stand in the stead of riddles and idioms because the latter two differ from the former mainly in terms of the proverb's extended length relative to that of riddles and idioms. It should, however, be henceforth assumed that where mention is made of proverbs, it shall implicitly be inclusive of riddles and idioms as well, unless the contrary is stated.

The respective "lexical or linguistic archaeological sites" that have been designated above, have the capacity to yield archaic words. We do not, however, work with the assumption of the *exclusive middle* that a word can only be either archaic or current. Rather, we take the pragmatic position that the lexicon can be put on a word-currency scale or continuum ranging from obsolete to current. Archaic words, on the one hand, would reside on the lower reaches of such a continuum, while words in current usage, on the other hand, would be found on its upper reaches. We shall explain, under methodology, the criteria determining the extraction of a particular word from say a proverb or a rid-

dle as a potential archaic word. We shall explicate methods through which we arrive at such an assertion. It is, at this juncture, necessary to answer the question why proverbs, riddles and idioms have been chosen as "lexical archaeological sites".

On the Choice of Proverbs, Riddles and Idioms

Although proverbs, riddles and idioms are categorically different in their form, they are, however, conceptually complementary in so far as they are all repositories and therefore carriers of ancient wisdom and lore. It is because of this that we shall, as indicated earlier, only concentrate on characterising the proverb, the characterisation of which, by extrapolation, shall be assumed to refer equally to riddles and idioms, at least in terms of their oral transmission and linguistic fixedness.

Bukenya (1994: 36-37) correctly observes that the term proverb has been defined "in various ways by different scholars". Some of those he lists are Kipury, Finnegan, Nandwa and Bukenya as well as Sunkuli and Miruka. Bukenya resists giving another definition but instead synchronises these various definitions to come up with what may be called the characteristic features of this genre. We shall abstract from these the characteristics of the proverb that best suit the argument of this article. The prominent features of proverbs show them as:

- short, terse or brief, and
- relatively invariable/fixed.

Bukenya (1994: 40) further points out that proverbs are "small poems". Poems, by their very nature, have an inherent mnemonic feature which supports our delineation of them as recorders of the archaic words of a language.

Pongweni (1989: 4) agrees with the view of the linguistic fixedness of the proverb when he makes the assertion regarding the "frozen character of the syntax of the proverb". He also quotes Hasan-Rokem who remarks that the proverb "bears many marks of a tight linguistic entity". Hasan-Rokem aptly captures the nature of the proverb as follows:

Even more than the idiom, the proverb is a fixed word-order not subject to inflectional changes according to linguistic context.

Lastly, the historicity of the proverb cannot be in doubt as most writers on the subject have emphasised the "time-immemorial" aspect of its existence. This comes out quite unmistakably through the characterisation of proverbs as "messages coded by tradition and transmitted from generation to generation" (Pongweni 1989: 5). Bukenya (1994: 12) describes the historicity of myths in a

manner that also applies to other genres of orature, proverbs included, when he writes:

Myth, to illustrate our argument, behaves like time: it has a tendency to be chronologically transcendental and can therefore defy the boundaries of historical time.

Proverbs, riddles and idioms as well as other subgenres of orature can be studied from many dimensions according to the research interests of the student or scholar. In respect of the above-mentioned characteristics and also from the viewpoint of linguistic research, the proverb or the riddle is most suited for what we hope to achieve because each of them relies for its transgenerative transmission on: (i) its relative linguistic fixity over long passages of time; (ii) its shortness which in turn renders it (iii) highly mnemonic due to its poetic form; (iv) its "chronologically transcendental tendency" owing much to its social regulatory function as well as (v) the respect accorded to griots, sages and all people of sound thinking. This by itself can act as a worthy incentive to carry it on in the people's collective memory over numerous generations.

The case we are making in respect of orature as a window into the past is, however, quite modest. We will indeed not claim, and neither can anyone else, to be able to use proverbs, idioms and other aspects of orature for reconstructing ancient Shona or any other Bantu language, dating say a thousand years back. It would be expecting too much from orature, especially if empirically better-placed disciplines such as archaeology, as Bourdillon (1987: 3) asserts, "can yield ... very limited [information] in the absence of literary deposits". Beach (1994: 21) provides some demarcation back in time regarding the reliability of Shona traditions as historical sources. In this regard, he (Beach 1994: 273) makes the following comments:

I can sum up by claiming that Shona oral traditions give us a reasonable basis for a history of the Zimbabwean plateau, but one going back only to about 1700 and often not as far.

There is therefore a limit to which one can successfully recover the past via orature. This should not, however, be misconstrued to be an underestimation of the relative wealth of information, especially in the absence of any written records, that one can still unearth, as will be demonstrated shortly, from this source.

On the Method of Recovery³

In an article of this kind, the ways along which we arrive at certain conclusions must be shown. This is necessary largely for three reasons. The first is that the elucidation of our methods enables independent judgment of the suitability of

the outlined method. The second and corollary reason is that, should our proposed methods be judged correct, they can then be applied, not only to Shona, but also to other Bantu languages. The third and final reason is that if the proposed procedure is not sound, perhaps different and more effective ways can be found to achieve the same goal of extracting diachronic lexemes for the benefit of African lexicographic practice.

We adopt, as a broad methodological principle, what Katamba refers to as *discovery methods*. Katamba (1993: 23) makes some incisive remarks with regard to these methods when he notes:

The main principle used in the analysis of words is the principle of contrast. We contrast forms that differ (i) in phonological shape due to sounds used and (ii) in meaning broadly defined to cover both lexical meaning and grammatical function.

As indicated earlier, there are basically two targets from which we hope to obtain the desired lexical items. The first of these are free-standing lexemes as opposed to those we will disentangle from constructions of a higher order, such as verb phrase complexes and larger nominal phrases. Examples of such free-standing words have already been given from synchronic Shona. We will now proceed to explain each of these lexical targets in the following sections.

Our task is to justify the selection of certain words at the exclusion of others. Potential words are first selected from any relevant source, such as a book of Shona proverbs, riddles or idioms on the basis of being unfamiliar to the language archaeologist or lexicographer. Consider for instance the following proverbs from Hamutyinei and Plangger (1974: 394, 436 and 437 respectively):

- (1) (a) **Manga** chena inoparira parere nhema
(A white pumpkin betrays where the black one lies)
- (b) **Sonzi** rakatuma nzou
(An ant sends an elephant [on an errand])
- (c) Muromo **chiponza** unopotsera
(A mouth is like a knobkerrie which you can throw)

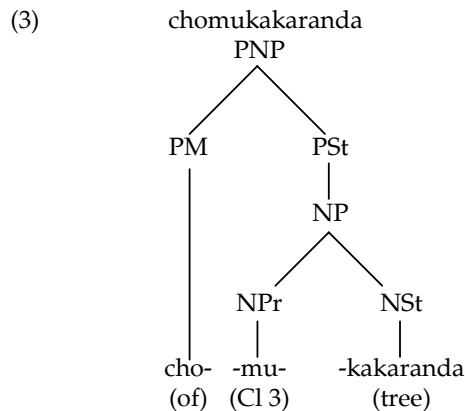
In (1) above, the words in bold type, **manga**, **sonzi** and **chiponza**, are unfamiliar to the writer of this article. After having identified such words, we propose, as the next step, that the existing dictionaries of Shona⁴ are checked. If the word is present, this might be an indication that the word is not necessarily archaic or this might be a suggestion that the word is becoming obsolete. It is also necessary to check the word with other speakers whose dialects may be different from that of the researcher's. The word can also be checked against the ALLEX Shona database and corpus.⁵ Failure to find the word in the corpus technically qualifies it for the designation "archaic". If, on the other hand, the word is in the corpus but showing a low frequency, i.e. after running a concordance of the word, then that word also qualifies as one becoming obsolete. If, however, the

particular word only occurs in the context of that proverb or riddle, then it can be considered archaic and can therefore be marked as such in the dictionary.

Apart from free-standing vocabulary items, we also find words that are part of larger constructions occurring in proverbs, riddles, idioms and other subgenres of orature. The following riddles contain such cases of potentially archaic words (see Pongweni and Chiwome 1995: 7, 46, 47 respectively):

- (2) (a) **Zifuridzo** mukombe une buri
(The rumbling instrument (which is like) a ladle with a hole)
- (b) Chinguri ndakazvarwa handisati ndamboona **bwabwa rechibvimo**
(Since I was born, I have never seen the token of betrothal of my mother)
- (c) Ndagumburwa nechitsiga **chomukakaranda**
(I knocked my foot against a mukakaranda tree stump)

In (2)(a) **zifuridzo** is a noun belonging to class 21 as marked by the **zi-** prefix. From the main form of such a noun we can abstract **furidzo** as a potential archaic word. In (2)(b) and (c) potentially archaic words are encapsulated in possessive phrases. The possessive phrases **bwabwa⁶ rechibvimo** and **chomukakaranda** contain within them lexemes on which our interest is focused. In (3) below, we, by a binary deconstruction methodology, disentangle the noun of interest from the higher possessive phrase complex down to its respective morphemic components.



Through the above process, we have been able to isolate the target lexeme **mukakaranda**, a type of tree, from the PNP within which it was bound. This process is a function of the language type to which Shona belongs. From the five language types which Katamba lists⁷, it is evident that Shona is an agglutinating or agglutinative language. Katamba (1993: 57) describes an agglutinative language as a "kind of language [where] there tends to be a more or less

one-to-one matching of morphemes with morphs". He gives Turkish as a typical agglutinative language and in passing mentions Luganda as well. Being a Bantu language like Shona, Luganda belongs, according to Katamba's classification, to the agglutinative type, which, by inference, marks Shona as being agglutinative too. It is important to note that the kind of binary morpholexical deconstruction as demonstrated in (3) above is occasioned by the language type to which Shona belongs.

We also argue in this article that it is possible, in reverse order, to recover from the synchronic state of the language, aspects that cast some meaningful light on its diachronic state, i.e. in so far as the lexicon is concerned. Note that here, unlike what we posited earlier, we do not particularise orature as a "lexical archaeological site". In this particular argument, we posit that deverbatives can be a rich source of verbs that may no longer be in use although they may possibly have been, at one stage in the history of the language, commonly used.

It should be noted that our technique of morpholexical deconstruction corresponds with Dale's observation regarding the obvious derivational origins of deverbative nouns in Shona. Dale (1981: viii) clearly spells out how the Shona view the nature of deverbative derivational lexemes when he says in the introduction to his bilingual dictionary *Duramazwi: Shona-English Dictionary*:

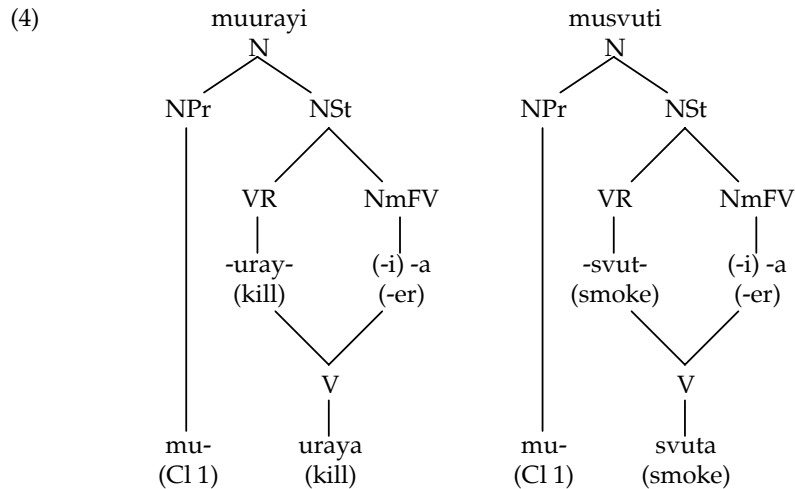
Deverbative nouns (e.g. *mubiki* from *-bika*) normally do not have the verb of origin suggested. This is because to the Shona speaker the relationship is very obvious.

We consider such lexical rederivations with the same belief as that of English speakers in the historical existence of words such as *sweveninges* (dreams), *lesynges* (lies), *cinn* (chin), *cirice* (church), *cuman* (come) as recorded from Chaucer's and Shakespeare's literary works (cf. O'Grady et al. 1989: 189 and 191) regardless of the paucity of evidence from the present state of English that such words in fact ever existed.

In the preceding we followed Dale's lead in respect of disentangling deverbative nouns to their minimal morphemic constituents. After having done so, we then "clinically" extracted and thereby recovered the seminal or nucleic verb root. Consider the verbal nouns **muurayi** (killer) and **musvuti** (smoker) in (4) below.

The branching diagram is a morpholexical deconstruction and disentanglement of a basic lexical form which is the underived (or seminal) root, being in this case **uraya** and **svuta** respectively and the agentive nominal forms derived thereof, which are **muurayi** and **musvuti**. The underived forms are what Katamba (1993: 92) refers to as "the kernel of the word". Note that in the derivation process the prefixal morpheme, **mu-** in either case, and the final vowel **-i**, are both viewed as the nominalising agent triggers (hence they are glossed as **-er**). There is no doubt, however, that **mu-** is relatively more significant in its nominalisation input to the verb, at least in so far as it assigns the class of the "new" word *vis-à-vis* the vowel, which is introduced phonologically

into the word-formation process.



It would not be necessary to draw elaborate branching diagrams for the range of deverbatives that fall within this category. What has been done in (4) methodologically suffices for the point we are making. In (5) below, we show a random list of lexical items situated within such a deverbative derivational morpholexical paradigm.

(5) Deverbative Noun	Gloss	Verb Root	Original Verb	Gloss
musika	stirring stick	-sik-	sika	stir
chibhako	moulding stick	-bhak-	bhaka	mould sadza
chigaro	chair	-gar-	gara	sit
murimi	farmer	-rim-	rima	farm/plough
musevenzi	worker	-sevenz-	sevenza	work
mufemberi	prophet	-fember-	fembera	prophecy/guess
chisvusvuro	breakfast	-svusvur-	svusvura	eat breakfast

The above nouns fit into the following verb-to-noun derivational schema:

(6) Cl Morpheme + VR + Nominal FV ⇒ Deverbative Noun

All the deverbative nouns in (5) are obvious synchronic examples that are part of the competence module of current speakers of Shona. The morpholexical deconstruction process that is worked out in (4) rederives, in reverse order, a verb from a noun. As has already been explained, the kernel for **murayi** (killer) is the verb root **uray-** (kill). We are of the opinion that this process can be

applied to "deverbative" nouns that have lost their corresponding verb roots serving as a technique through which historical/archaic verbs can be defossilised. In a way, we basically reverse the process in (6) and this yields (7) whose input schema is provided below.

(7) - Cl Morpheme + VR - Nominal FV \Rightarrow Diachronic Verb

It is evident from this input schema that our concern above is largely with the suspected diachronic or what we may call the diachronic "seminal" verb that has been fossilised⁸ within the synchronic deverbative noun. We may use a less controversial illustration such as the deverbative noun **mupakato** (handle) to illustrate this. A morpholexical deconstruction of this noun is done within the paradigm of (7), and it is represented as (8) below.

(8) - Cl Morpheme + VR - Nominal FV \Rightarrow Diachronic Verb
 - mu + pakat - o \Rightarrow pakata

Note that this particular "rederived" verb is not necessarily diachronic as such. It is only employed for illustrative purposes. Speakers of Shona may agree with the writer of this article that the verb **pakata** (hold) is almost exclusively associated with a handle. Instead, in other contexts, speakers of Shona would use the verb **bata** (touch or hold) which is clearly close to **pakata**, both in terms of form and attested Bantu sound shifts. The verb **pakata** belongs almost solely within the context of the proverb **Kunzi pakata, hakuzi kunzi ridza** (To be asked to hold [a whip] is not the same as being asked to crack it). The occurrence of **pakata** in this proverb and seldom in other contexts, gives credence to our argument which singles out proverbs and other genres of orature as repositories of archaic words.

The efficacy of our methodology can be illustrated with lexical examples that are within the competence of synchronic native speakers of the language. Here we propose to analyse words intermediate in terms of archaic status, e.g. words that are within the bounds of "incontrovertible" etymological recoverability such as those given in (9) below:

(9) (a) mupakato (handle)
 (b) nhovo (shield)
 (c) museve (arrow)
 (d) pfumo (spear)

First it must be shown that all these lexical items are deverbative nouns since this is the *sine qua non* for their subjection to the morpholexical deconstruction procedure as outlined in (7). The next step, therefore, is to run the deverbative nouns through this procedure in order to arrive at their verb kernels. This is done in (10) below:

(10)	- CI Morpheme	+ VR	- Nominal FV	⇒	Diachronic Verb
(a)	- mu	+ pakat	- o	⇒	pakata
(b)	- ∅	+ tav	- o	⇒	tava
(c)	- mu	+ sev	- e	⇒	seva
(d)	- ∅	+ pfum	- o	⇒	pfuma

The "recovered" verb **pakata** in (10)(a) has already been discussed. What is interesting to note, however, is that apart from **mupakato** (handle), the nouns **bakatwa** (knife), **mubatwa/nhapwa** (slave), **ubatwa** (slavery), **mubati** (hard worker) and **mbato** (tongs) are etymologically all derivatives of the verb **pakata**⁹.

All the rederived verbs in (10) which we give here in their infinitive mode as **kutava** (to shield/protect), **kuseva** (to dip into soup/relish), and **kupfuma** (to sew/pierce) are still current. The point of the argument is that, by a logical extension, this methodology can successfully be applied to deverbatives whose corresponding nucleic verbs are no longer in current use. However, with regard to the other rederived verbs in (10), we argue that **nhovo** (shield) is derived from the verb **tava** (to shield) and that the deverbative noun is phonologically different from the kernel verb owing to morphophonemic changes that are quite consistent and widespread with class 9 and other classes, such as 10 and 5¹⁰ which display similar morphophonological behaviour. We argue that nouns such as **teve** (protective charm), **nhovo** (shield), **tavira** (protect a home or person with charms against evil spirits) and **nhovo**¹¹ (protector) may have been derived semantically and etymologically from the kernel verb **tava**. To this end, we argue that the nouns **nhova** (fontanelle) and **dovo** (animal hide) are figuratively, though not literally, related to the verb **tava**.

Verbs such as **seva** and **pfuma** may not be so difficult to explain owing to the current sense of the verb **seva** which refers to dipping say a morsel of **sadza**¹² into soup or relish in the process of eating. Etymologically nouns such as **musuva** (morsel), **usavi** (relish), **museve** (arrow), **musevi** (?deeper — as in soup/relish), **museve** (sun's rays), **museve** (fast runner)¹³ as well as **seva** (to follow closely behind), it is argued here, are very closely related. There is a defining featural sense of ± enter/or touch with (or be touched by) the tip or front end of some instrument. The notation ± refers to the object as having a property of entering such as **musuva** and **museve**, or showing a property of being entered such as **usavi**, for instance.

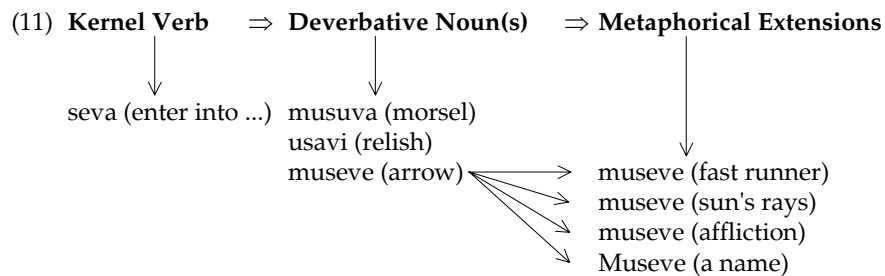
The other verb that has been rederived using the morpholexical deconstruction procedure as outlined in the schema in (7), is **pfuma**. We have already argued that **pfuma** as a verb means to pierce, i.e. in the infinitive mode. The noun **pfumo** from which we have derived **pfuma** could therefore be translated into English as a piercer, although the correct rendition of it is a spear. The sense of pierce is etymologically blended and extended both literally (i.e. denotatively) and abstractly (i.e. idiomatically) into the following words:

pfuma (sew), **pfumo** (spear), **pfumo** (army), **pfumo** (serious predicament) and **Mapfumo** (name). It is necessary to comment on the surfacing of different though related shades of meaning ascribed to a single diachronic kernel verb.

Semantically and phonologically the correlation between what we have referred to as kernel verbs and their derivatives do not have a mathematical one-to-one correspondence. In other words, the rederived verbs **tava**, **pfuma** and **seva** and their corresponding derivatives such as **nhova**, **nhovo**, **pfumo**, **pfuma**, **museve** and **musuva** do not have a readily discernible semantic and to some extent phonological correspondence. Traugott (quoted by Heine et al. 1991: 14) explains this phenomenon of semantic fading resulting from linguistic change when he makes the following observation:

Meanings [with time] tend to come to refer less to objective situations and more to subjective ones (including speaker point of view), less to the described situation and more to the discourse situation.

We argue that meaning changes from "objective situations" to "subjective ones" are evident in a continuum of semantic extensions which proceed from the verb to derived nouns and from the deverbative nouns to metaphors. In (11) below we illustrate this phenomenon using the rederived verb **seva**.



The diagram in (11) is largely self-explanatory as it is an illustration of semantic shifts from the verb **seva** to its derivative forms. **Seva** can be marked as the point of "objective situations" within the semantic shift continuum. The deverbative noun **museve** assumes the role of a semantic bridge across which metaphorical and therefore "subjective" meanings pass. In this process the metaphorical word retains its form. This does not only apply to **museve** in (11) but also to other nouns such as **nhovo** and **pfumo**.

It has already been said that language change is a multiple process. Among these changes are phonetic processes as well. Such processes are responsible for the phonological differences that sometimes occur between earlier and later forms of the words, apart from the differences that are cognates of the same phonological form. This is evident even in languages with a long literacy history such as, for instance, English. The following English

historical words and their corresponding synchronic forms already referred to are a case in point:

- (12) *sweveninges* ⇒ dreams
lesynges ⇒ lies
cinn ⇒ chin
cirice ⇒ church

Unquestionably it is difficult to decipher the phonological correspondences between the diachronic and synchronic forms of the words given in (12). It may in fact take somebody with a knowledge of English etymology rather than any ordinary speaker of the language to unravel the correspondences. This similarly applies to Shona. Any semblance of a relationship between **tava** and its derivatives such as **nhova**, **teve**, **tavira**, **nhovo**, and the names **Matavire** and **Manhovo** may not readily be seen. This may be analogous to the relationship that binds genetically related languages, for instance Bantu languages. To say that XiTsonga, Zulu, KiSwahili, Sotho and Shona are all Bantu languages, does not mean that speakers of these languages understand one another. Similarly, the relatedness of elements internal to a language may not be evident at first sight, due to what Heine et al. (1991: 15) describe as "adaptation, erosion, fusion, and loss" with respect to phonetic change.

The Cross-linguistic Comparative Aspect¹⁴

In a procedure such as the one we are advocating, one of the greatest frustrations is a lack of adequate feedback. Unearthing possible obsolete words and verbs in a language with no written records can easily be ridiculed as a worthless undertaking since there will be no records against which to check the authenticity of the findings. Trust in the methodological stringency corroborated (when one is fortunate) with cross-linguistic comparative evidence is sometimes the only proof on which to rely. Consider the examples in (13) below:

(13)

DN	Gloss	-CIM	+VR	- Nm FV	⇒ DV	Gloss
chibako	tobacco pouch	- chi	+ bak	- o	⇒ baka/paka	put into
musika	stirring stick	- mu	+ sik	- a	⇒ sika	spin/stir
muromo	mouth	- mu	+ rom	- o	⇒ roma	bite
mugoti	cooking stick	- mu	+ got	- i	⇒ gota	?cook
mugwaku	ladle	- mu	+ gwak	- u	⇒ gwaka	?to ladle
bakwa	stake of wood	- ∅	+ bak(w)	- a	⇒ baka/paka	stake/put into ¹⁵
mbiko	hiding place	- ∅	+ bik	- o	⇒ bika	hide
muvi	bad person	- mu	+ v	- i	⇒ vi	bad ¹⁶
muuya	good person	- mu	+ uy	- a	⇒ uya	good ¹⁷
shomodzi	messenger	- ∅	+ shomodz	- i	⇒ shomodza	announce ¹⁸

Most, if not all the verb artefacts, are decidedly unfamiliar to current speakers of Shona¹⁹. The recovered items can be a source of archaic words as well as a source of etymological derivatives of current words — information which is important in the dictionary-making process. The following pairs of verbs, for instance, make an interesting comparison:

(14) roma	←	ruma (bite)
bika	←	viga (hide)
baka	←	paka (stake/put into)
shomodza	←	shevedza (call), shambadza (announce/sell)

In each case the "diachronic verb" is given first, followed by the current verb which is separated from the former by the notation ← which means "has changed to the following synchronic form".

Both the adjectives **-vi** and **uya** have more or less been lexicalised and fossilised within the grammatical noun category as used in (13). This further applies to **-vi** in **chivi** (sin) and **Chivi** (name of both a chieftainship and district of Masvingo province)²⁰, and to **uya** in **muuya** (good person). It is interesting to note that the same adjectives have also been fossilised in the idiom **mavi nemauya** (good and bad [words]). These words are almost obsolete outside the few usages that have been mentioned. It is on account of this that we technically qualify such words as archaic. Adjectives, as Fortune (1980) observes, are a long series of stems in that they are semantically compatible with a wide range of the nominal affixes of the noun classes²¹. The nominal lexicalisation and concomitant fossilisation of these adjectives are proof enough that they may be survivors from some past linguistic epoch.

As stated above, one way of getting feedback would be to check one's inventory of lexical artefacts with related languages. The need for such a procedure arises from the rationale that related languages might in some instances still be using "parent" or prototypical forms of words that may, however, have become synchronically obsolete in others, and vice versa. This is for instance the case with Shona and Nguni as well as Sotho-Tswana languages with regard to locatives. Meinhof gives Ur-Bantu locative prefixes respectively as **pa-**, **ku-** and **mu-**. These locative prefixes or their cognates are by and large no longer productive in Nguni and the Sotho-Tswana languages. They have, at most, become lexicalised in locative nouns such as **phezu** (at/on the top of), **phansi** (down/underneath) and **fatshe** (down), **hodimo** (above), and **morao** (at the back)²². However, the same locative prefixes are highly productive in Shona as is evidenced by the existence of fully fledged noun classes with prefixes: 16 **pa-**, 17 **ku-** and 18 **mu-**. These locative prefixes, unlike in Nguni, are active and productive to a point where literally any other word, save for its intrinsic semantic constraints *vis-à-vis* a given locative morpheme, can be locativised through the usual prefixing morpholexical process.

In the table under (13) there are some artefact verbs whose possible dia-

chronic existence can be proved from the suggested cross-linguistic perspective. The adjective **-vi**²³ (bad/evil) can be taken as an illustrative case. We have already said that **-vi** is a Shona adjective stem that has been rendered redundant and only survives lexicalised in fossil form in a few words as given above. On the contrary, this adjective stem, rendered as **-bi** in Nguni orthography, is highly productive in languages such as Ndebele and Zulu among others. The adjective phrases in (15) below are ample witness to this fact.

(15) Ndebele/Zulu Phrase ²⁴	Gloss
umfazi omubi	ugly/evil woman
umoya omubi	bad/evil heart
indlu embi	bad/ugly house
umganu omubi	ugly plate
isitulo esibi	bad/ugly stool
ilizwe elibi	bad country

It is clear from what has been shown above that cross-linguistic comparisons can sometimes be taken recourse to in order to validate findings in a related language. It is interesting to note that in the case of the above adjective **-bi**, Seto and Seto (1997: 476) argue that it is the same form as that which occurs in Papuan Austronesian languages of Southeast Asia whose forms of the adjective are given in (16):

(16) Kewa	wae
Malas	pi
Pila	fiwi
Moere	fia
Mawak	fi:wi
Hinihon	fi:wi
Zulu	-bi

The argument of Seto and Seto (1997: 195) to justify the above comparison, which they in fact also apply to other different languages of the world, is as follows:

Chronologically significant lexical affinities [of different languages of the world] indicate the migratory derivation of the words for village life from a common ancestral homeland in tropical Mainland Southeast Asia.

The major obstacle to such a procedure, however, is the relative paucity of artefact words that are cross-linguistically attestable through a methodological paradigm such as the one performed in (15) and (16) above. If, however, Seto and Seto's claim is anything to go by, then one can perhaps look for cross-linguistic comparative evidence from much further afield.

Conclusion

The foregoing discussion has been an attempt to harness methodologies from some different though related disciplines in the humanities for the advancement and enrichment of lexicography. Some of the disciplines which have been found useful for our purposes are historical linguistics, oral history and traditions as well as oral literature. All these have been identified as rich and specialised areas in which African lexicographers can excavate for archaic and etymologically significant words to be included in both general and specialised dictionaries.

We have in the foregoing discussion dwelt on archaic words and how to gather them since they are part of the raw material needed for the dictionary-making process. We have delineated orature as a symbolically rich linguistic archaeological site. This is so because, as we have demonstrated, orature acts as a preservative of archaic word forms, due not only to its syntactic fixedness, but also due to its mnemonic proclivity and important cultural significance. We have, however, specifically singled out proverbs, riddles and idioms as some of the subgenres of orature that easily lend themselves to the procedure under discussion. Archaic words emanating from these sources can be marked in the same way as those in any dictionary. It is instructive to note that in the lexicographic traditions in contemporary African languages in Southern Africa, both in terms of the proliferating monolingual and the traditional bilingual dictionaries, lexicographers are quick to mark entries as loan-, taboo or offensive words, but very few dictionaries mark archaic or obsolete entries²⁵. Shall we, therefore, on the basis of this observation, arrive at the fallacious conclusion that African languages, unlike all other known languages, have no archaic or obsolete words?

Verbs specifically have been our main source of exemplification owing to their inherent capacity for prolific deverbative morpholexical noun formation processes. While the synchronic deverbative noun may, by historical fortuity, have lasted up to the present day, this study has, by implication, argued that the diachronic kernel or nucleic verb itself could over many generations have become obsolete to a point where it only remains preserved as a grammaticalised morpholexical fossil. A method of morpholexical deconstruction and disentanglement has been proposed. The efficacy of this method in extracting and recovering especially artefact verbs has been demonstrated. A similar method can be developed for other word forms depending on their idiosyncratic inflectional and derivational behaviour. In this regard, a method of dealing with possessive noun phrases has been suggested in (3). Here we have explored the methodology of verbal disambiguation by morpholexical deconstruction.

It has also been shown in the discussion that by employing the morpholexical deconstruction process, we can possibly recover an artefact verb or word to which other synchronic words can etymologically be traced back. This has been shown to be possible in (10) and (11). There are, however, cases in

which certain artefact words may be well beyond the scope of etymological recoverability. It has been suggested that a cross-linguistic comparative study may be undertaken as a way of ascertaining whether etymologically irrecoverable artefacts may be found in a related language or other languages. This was illustrated in (15) and (16). We have justified such an undertaking by observing that while certain diachronic word forms may have altogether vanished from the synchronic state of a target language, they may probably be found in the synchronic form of a sister language or, as Seto and Seto seem to suggest, any other language. Examples giving credence to such an exercise have been adduced by comparing Shona and Nguni as well as the Sotho-Tswana group of languages. Evidence for the cross-linguistic use of the adjective stem **-bi** has been seen to occur in some Papuan Austronesian languages. To this end, it has been demonstrated that, while the locative prefixes are no longer as active in the Nguni and Sotho-Tswana group of languages, they are nevertheless still very active and highly productive in the Shona language. The reverse was found to be true in respect of the adjective stem **-bi**, which is still highly active in Nguni while it has fallen dormant in Shona. Artefact word forms that are supported by the facts from related or other languages for that matter, may in all probability be authentic and should therefore not be dismissed as a figment of the historical linguist's or etymologist's imagination.

Notes

1. I wish to acknowledge the valuable contributions and comments, some of which I have incorporated, that were made when I initially presented a version of this article as a paper in the Linguistics Department of the University of Zimbabwe in 1999. I single out the following for special mention: Mr Liveson Tatira, Dr Carolyn Harford, Prof. Norris Dembetembe, Mr Jairo Kangira, Mrs Hakani Moyo and Dr Ann Jefferies.
Apart from the contributions received from those named above, I want to mention the fact that a version of this article was also read as a paper during the Fourth International Conference on Lexicography (5-6 July 1999) that was organised by AFRILEX and hosted by the University of Pretoria in the Republic of South Africa. I note with pleasure the remarks of the keynote and opening address delivered by Prof. A.C. Nkabinde where he, among other issues, challenged Bantu lexicographers to explore the importance of historical linguistics in lexicographic work. I take the present effort as one of the many dimensions through which historical linguistics can fruitfully be brought to bear upon African lexicography. I acknowledge with thanks the useful questions, remarks and encouragement that I received from among the audience during my presentation. Apart from that, I extend my sincere thanks to two anonymous referees who made useful suggestions in terms of how to improve the general import of the article, some of which have been gladly incorporated into its final version.
2. Rhodesia (the present-day Zimbabwe) was colonised by the British through Cecil John Rhodes's British South Africa Company.
3. In this section, the following abbreviations will be used: PNP: Possessive Noun Phrase; NPr: Noun Prefix; NmFV: Nominalising Final Vowel; NST: Noun Stem; VR: Verb Radical; N:

Nominal; V: Verb; Cl: Class Affix followed by Gender Number; PM: Possessive Marker; NP: Noun Phrase; PSt: Possessive Stem; DN: Deverbative Noun; DW: Derived Word; – Cl Morpheme: Delete Morpheme.

4. Existing Shona dictionaries include the following: Chimhundu et al. (1996), Dale (1981), Hanan (1981) and Barnes (1932).
5. ALLEX is the acronym for African Languages Lexical Project, which is a dictionary-making undertaking in partnership with collaborators from the universities of Oslo in Norway and Gothenburg in Sweden. The first monolingual Shona dictionary, *Duramazwi ReChiShona*, was compiled by the ALLEX Project. The ALLEX Shona corpus has at present over two million running words.
6. Note that **bwabwa**, though not our main focus for explanation by virtue of it being a stand-alone noun, is a word no longer in common use, if used at all.
7. Katamba gives the following as the five known language types: (i) analytic, (ii) agglutinating, (iii) inflecting, (iv) incorporating, and (v) infixing languages.
8. As we discuss the aspect of fossilisation, it is instructive to note that Heine et al. (1991: 15) discuss grammaticalisation as an aspect of language change. Here they state that "this process [of change] affects all levels of language structure". They distinguish between functional, morphosyntactic, and phonetic processes. In their discussion of the morphosyntactic which interests us, they outline the following as morphosyntactic processes of language change, i.e. permutation, compounding, cliticization, **affixation** and **fossilisation** (the emphasis is mine).
9. To this list we can also add **mbati**, for one caught red-handed, as in the idiom **Mbavha imbati** (A thief is caught in the act). This **mbati**, a deverbative as Shona speakers may agree, is exclusively used in the context of this idiom.
10. Fortune (1980) argues that diachronically class 5 had an overt syllabic prefix **ri-** but synchronically it is no longer overtly expressed except in a few cases. He further argues that the disappearance of the **ri-** morphemic prefix triggered off the wide-spread morphophonemic changes associated with nouns that fall in this class such as the systematic nominal stem onset consonant change $g < k$, $b < p$, $f < bv$ and $j < ch$ found between the singular and the plural forms respectively. The morphophonemic changes in classes 9 and 10 are associated with the influence of nasalisation which is quite common in these classes.
11. Here **nhovo** (protector) is used as a metaphor, especially within the context of the Christian faith where God is often metaphorically referred to as such. **Nhova** (fontanelle) and **dovo** (animal hide) may possibly be figuratively related to the verb **tava** (shield). We would like to leave it to the reader to arrive at their own judgement on this matter.
12. **Sadza** is the staple food of much of Southern Africa. It is prepared from corn meal and made fairly thick so that it can be shaped it into a ball. Its name varies according to the particular language: it is called **nsima** in ChiChewa, **isithwala** in Ndebele and **stywepap** in Afrikaans.
13. **Museve**, as in sun's rays and fast runner, are taken from Chimhundu et al. (1996: 299).
14. In this section, the following additional abbreviations will be used: DN: Deverbative Noun; Cl M: Class Marker/Prefix; DV: Diachronic Verb/Rederived Verb.
15. Note that **bakwa** is implicationally related to, if not the same as, **chibako**.
16. **-vi** is decidedly not a verb root, but an adjective stem.
17. **Uya** is the opposite of **-vi**.

18. **Shomodza**, we infer, can mean to announce or to deliver a message.
19. Most speakers of Shona do not agree with this claim. Some, however, begin to see some resemblances only after being given certain explanations.
20. Masvingo, one of the major provinces of Zimbabwe, lies within the south and south western zone of the country.
21. Officially Shona has 21 noun classes but technically they are actually 20. Class 20 was originally assigned to some Kalanga nouns. This provision became redundant after the recognition of Kalanga as a separate language from Shona, therefore leaving Shona with 20 and not 21 noun classes.
22. The locatives **phezu** and **phansi** are taken from Taylor (1994) and the locatives **fatshe**, **hodi-mo** and **morao** from Moeketsi (1994).
23. Note that the difference between **-vi** and **-bi** is only orthographic while they are phonologically and semantically identical. Shona orthographises the phoneme through the spelling **-vi** while the other languages which were examined, use the form **-bi**.
24. I would like to thank Ms. Karren Dube (of the Department of African Languages and Literature, University of Zimbabwe) who made some valuable spelling corrections to the words in the table.
25. I have verified with Mr. Samukele Hadebe (of the Department of African Languages and Literature, University of Zimbabwe) that both Zulu and Ndebele dictionaries do not mark archaic words. In an informal discussion at the Fourth International AFRILEX Conference, Prof. Nkabinde also confirmed that Zulu dictionaries, in spite of their long tradition, have up to the present not marked entries as being archaic. The same practice prevails in Shona lexicography. I have, therefore, arrived at the conclusion that the picture may in all probability be the same for other African languages in Southern Africa.

References

- Aitchison, J. 1995. *Linguistics: An Introduction*. London: Hodder and Stoughton.
- Barnes, B.H. 1932. *Vocabulary of the Dialects of Mashonaland*. London: The Sheldon Press.
- Beach, D.N. 1980. *The Shona and Zimbabwe 900-1850: An Outline of Shona History*. Gweru: Mambo Press.
- Beach, D.N. 1994. *A Zimbabwean Past: Shona Dynastic Histories and Oral Traditions*. Gweru: Mambo Press.
- Bourdillon, M.F.C. 1987. *The Shona Peoples*. Gweru: Mambo Press.
- Bukenya, A. (Ed.). 1994. *Understanding Oral Literature*. Nairobi: Nairobi University Press.
- Chimhundu, H., N.E. Mberi, R.B. Batidzirai, M.B. Chitauro, E. Mangoya, A. Masasire and F. Matambirofa. 1996. *Duramazwi ReChiShona*. Harare: College Press.
- Dale, D. 1981. *Duramazwi: A Basic Shona-English Dictionary*. Gweru: Mambo Press.
- Finegan, E. 1989. *Language: Its Structure and Use*. Orlando: Harcourt Brace and Co.
- Finnegan, R. 1970. *Oral Literature in Africa*. Oxford: Oxford University Press.
- Fortune, G. 1980. *Shona Grammatical Constructions, Part 1*. Harare: Mercury Press.
- Fromkin, V. and R. Rodman. 1993. *An Introduction to Language*. Fort Worth: Harcourt Brace Jovanovich College Publishers.

- Hamutyinei, M.A. and A.B. Plangger.** 1974. *Tsumo-Shumo: Shona Proverbial Lore and Wisdom*. Gweru: Mambo Press.
- Hannan, M.** 1981. *Standard Shona Dictionary*. Harare: College Press.
- Heine, B., U. Claudi and F. Hunnemeyer.** 1991. *Grammaticalization: A Conceptual Framework*. Chicago: University of Chicago Press.
- Katamba, F.** 1993. *Morphology*. Houndmills: The Macmillan Press Ltd.
- Louwrens, L.J.** 1992. The Conceptualisation of Spatial Relationships as Expressed by Locative Structures in Northern Sotho. *South African Journal of African Languages* 12(3): 107-111.
- Moeketsi, R.** 1994. *The Versatile Nature of the Southern Sotho Demonstrative*. Unpublished paper presented at a Workshop on Conceptions of Space in Language, Cognition and Social Interaction in Southern Africa held at the University of Botswana, January 1994.
- O'Grady, W., M. Dobrovolsky and M. Aronoff.** 1989. *Contemporary Linguistics: An Introduction*. New York: St. Martin Press.
- Pongweni, A.J.C.** 1989. *Figurative Language in Shona Discourse*. Gweru: Mambo Press.
- Pongweni, A. and E. Chiwome.** 1995. *Zvirahwe Zvakare Nezvitsva: Traditional and Modern Shona Riddles*. Eiffel Flats: Juta Zimbabwe.
- Seto, G. and F. Seto.** 1997. *Ancient Tropical Mainland Southeast Asia: Linguistic and Cultural History, Vol. 2*. Mandaluyong: Africa-Asia Educational and Cultural Publications.
- Spencer, A.** 1991. *Morphological Theory: An Introduction to Word Structure in Generative Grammar*. Oxford: Basil Blackwell Ltd.
- Taylor, J.R.** 1994. *The Syntax and Semantics of the Zulu Locative*. Unpublished paper presented at a Workshop on Conceptions of Space in Language, Cognition and Social Interaction in Southern Africa held at the University of Botswana, January 1994.

The Role of the Introductory Matter in Bilingual Dictionaries of English and Arabic

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

Abstract: Monolingual and bilingual dictionaries for learners of English as foreign language differ from each other in many aspects. Among the differences are the types of information a dictionary provides outside the A–Z word list, especially in its introductory matter. Yet, dictionary introductions have been referred to rather peripherally in the literature on dictionary users and uses. Within the context of lexicographic practice, this study aims to outline the differences between EFL learner's dictionaries and bilingual English–Arabic dictionaries, and to determine to what extent a given set of popular dictionaries provide varied and sufficient information in their introductions. Three categories of dictionaries will be the subject of comparison: EFL dictionaries, European bilingual dictionaries, and bilingual dictionaries in the Arab world.

Keywords: BILINGUAL DICTIONARY, INTRODUCTORY MATTER, ENGLISH–ARABIC, EFL LEARNER'S DICTIONARY, ORIENTALIST DICTIONARY

Opsomming: Die rol van die inleidende gedeelte in tweetalige woordeboeke van Engels en Arabies. Eentalige en tweetalige woordeboeke vir aanleerders van Engels as vreemde taal verskil in baie opsigte van mekaar. Een van die verskille is die tipes inligting wat 'n woordeboek buite die A–Z-woordelys verskaf, veral in die inleidende gedeelte. Tog is daar slegs in die verbygaan na woordeboekinleidings in die literatuur oor woordeboekgebruikers en -gebruik verwys. Hierdie studie het dit ten doel om, binne die konteks van die leksikografiese praktyk, die verskille uit te stippel tussen EVT-aanleerderswoordeboeke en tweetalige Engels–Arabiese woordeboeke, en te bepaal in watter mate 'n gegewe groep gewilde woordeboeke veelsoortige en genoegsame inligting in hul inleidings verskaf. Drie kategorieë woordeboeke sal die onderwerp van hierdie vergelyking vorm: EVT-woordeboeke, Europese tweetalige woordeboeke en tweetalige woordeboeke binne die Arabiese wêreld.

Sleutelwoorde: TWEETALIGE WOORDEBOEK, INLEIDENDE GEDEELTE, ENGELS–ARABIES, EVT-AANLEERDERSWOORDEBOEK, ORIËNTALISTEWORDEBOEK

1. Introduction

A dictionary introduction is widely regarded in lexicographic circles as an essential source of information in all types of dictionaries. It should be noted, however, that a terminological disagreement exists in this regard; some diction-

aries use the term "introduction" while others begin their texts with the title "preface". Still, many lexicographers refer to the pages where the dictionary's microstructure and macrostructure are described as the "introductory matter". Here, one has to differentiate between an introduction in a book and one in a dictionary. An introduction or preface in a book aims to provide its readers with a general overview and prepare them for a series of related ideas in numerically ordered chapters. On the other hand, because a dictionary is a collection of different articles sharing nothing in common except the structure and organization of information, a dictionary introduction performs a different function. It limits itself to explaining this structure and addressing issues such as the symbols, abbreviations and cross-reference system used in the dictionary.

From a user perspective, one can determine the effectiveness of an introduction according to the extent to which it addresses the dictionary user's needs in relation to the information retrieval process. It may be of greater importance when a user is new to dictionary look-up operations but it loses its significance gradually as he/she becomes more familiar with the dictionary's conventions, mostly through trial and error. Indeed, it has been found that only a small percentage of dictionary users refer to their dictionary introductions for help. Béjoint (1981), for example, found that French advanced learners using EFL dictionaries rarely utilized the information contained in their dictionary guides.

It should also be noted that lexicographic studies have so far dealt with dictionary introductions in a peripheral fashion, the emphasis being traditionally placed on the components of the entry: morphology of words, meanings, pronunciation and syntactic structures. Yet a bilingual dictionary is, after all, a practical tool used in an intellectual activity such as reading, writing, or translating from or into a foreign language. Accordingly, there is an obvious need for lexicographers to pay attention to those aspects related to the user's capability of accessing the required dictionary information. This can be achieved by a lexicographic re-examination of the role of the introduction to ensure that it provides simplified and sufficient information that may lead to profitable use of the dictionary.

Despite the rare reference to dictionary introductions in lexicographic literature, one can classify lexicographers' opinions in this connection into three approaches. Some believe that it is the duty of the dictionary maker to provide the users with all the information they need at the point of entry, because these users rarely refer to their dictionary introductions (Gimson 1973, Stein 1984, Wiegand 1984, Berkov 1990). Another group of lexicographers indicate that many language learners come to the foreign-language learning setting without any background in the use of mother-tongue dictionaries and that there is a need to encourage users to read dictionary introductions and to train them in how to find the required information through a division of work between the dictionary and foreign-language textbooks (Rossner 1985, Crystal 1986, Cowie

1987, Snell-Hornby 1987). A third view is expressed by Hartmann (1986, 1987) who says that training in dictionary use is primarily the teacher's responsibility not the dictionary's, because, although dictionary guides set out a number of possible uses of a dictionary, these are based not on an analysis of the reference needs of the learner, but on the various information categories that the dictionary is traditionally expected to contain, i.e. meaning, spelling, grammar, etymology and pronunciation.

The problem that arises when one tries to inculcate users with reference skills, is that dictionaries, especially those designed for foreign learners of English, are now developing more rapidly than before, thus widening the gap that already exists between the complexity of information and the users' often rudimentary reference skills (Cowie 1987).

An introduction usually addresses certain categories of dictionary information such as the dictionary macro- and microstructure. Yet introductions vary in the way they treat this information. Some concentrate on certain aspects of the entry and ignore others, whereas other introductions may not specify the intended user category, probably because of factors relating to distribution. Also, compilers do not follow standardized and clear criteria in writing an introduction or dividing it into sections. Therefore, many monolingual and bilingual dictionaries contain very long introductions which are not easy to read and their excessive length and use of specialized linguistic terms would represent a burden on the user who tries to find a specific point of guidance.

According to Stein (1989), an introduction performs different functions. She indicates that it publicizes the dictionary: it may include claims of superiority over similar dictionaries, and at the same time it provides invaluable information on the content and structure of the dictionary. She also points out that the amount of information provided in the introduction about the lexicographic team, the accuracy of description and transparency, gives an indication of their expertise and attitude towards the user. In some dictionaries, the introduction serves as a lesson in lexicography due to accuracy of description and clarity with regard to lexicographic policies. On the other hand, Stein observes that other dictionaries do not clarify some points which are essential for understanding the structure of the dictionary. In addition, an introduction may be an important source of information on the policies followed by compilers of old dictionaries and the cultural circumstances that had a role in the making of these dictionaries.

2. Aims of the study

The present study aims to outline the following :

- (a) Similarities and differences between bilingual and EFL dictionaries in the area of dictionary introductions as well as recent developments regarding this aspect in both types of dictionaries.

- (b) The extent to which these dictionary introductions provide users with varied, accurate and sufficient information about these dictionaries.

In order to specify similarities and differences between introductions of bilingual and EFL dictionaries in addition to recent developments, the present study will analyze, compare and criticize the introductory matter in a set of eight dictionaries classified into three categories:

- (a) European bilingual dictionary: *Collins-Robert* (French–English/English–French);
- (b) Bilingual dictionaries of Arabic and English:
 - (i) Bilingual dictionaries compiled by Orientalists:
Dictionary of Modern Written Arabic (Arabic–English),
Wortabet (Arabic–English),
Learner's Dictionary (Arabic–English),
 - (ii) Bilingual dictionaries compiled by Arab lexicographers:
Elias' Modern Dictionary (Arabic–English),
School Dictionary (Arabic–English/English–Arabic),
Al-Mawrid (English–Arabic);
- (c) EFL dictionary: *Oxford Advanced Learners' Dictionary*.

The similarities and differences will be determined by a classification of information categories in the introductions to these dictionaries. Given the wide scope of the subject, the analysis will focus on the first three of the following categories:

- (a) Specification of the intended group of users,
- (b) Indication of the reasons for compiling the dictionary,
- (c) Reference to the sources of the dictionary,
- (d) Outline of new features in the dictionary,
- (e) Specification of the number of entries (or words),
- (f) Practical guidance.

3.1 Specifying the intended group of users

Monolingual and bilingual dictionaries designed for foreign learners of English differ in their content and presentation of information. This is due to the fact that a monolingual EFL dictionary addresses users coming from different cultures whereas the bilingual dictionary is written for a specific linguistic community (Atkins 1985). Lexicographers have agreed on the need to specify a user group such as learners, speakers of the source language or speakers of the tar-

get language, and the impossibility to cater for the needs of all these groups in a single dictionary (Householder and Saporta 1962).

A close examination of the introduction to the *Oxford Advanced Learners' Dictionary (OALD)* shows a clear reference to the foreign learner's needs in both reading and writing. Reference to learners is also made in the bilingual dictionary *Collins-Robert*, where the introduction claims that the dictionary aims to provide the user with the highest standards of proficiency coupled with ease of use and help in areas that cause difficulties to the learner. The introduction also emphasizes its pedagogical role by indicating that one of the important steps for successful language learning is the proper use of a good dictionary. It also addresses the two skills of reading and writing by saying that "this dictionary is indispensable in understanding and self-expression in French".

As for Orientalists' bilingual dictionaries these have been designed to meet the reference needs of Orientalists and students of Arabic in the West. However, the introductions in these dictionaries vary in the extent of accuracy in their reference to the intended user categories. For example, the introduction in Wortabet's dictionary does not specify one category of users but several: "This dictionary is indispensable for students, translators or intellectuals who write in English." Somewhere else, the same introduction specifies the user's language when it refers to the fact that the dictionary was compiled "to serve the Arab youth" although the dictionary was originally written for Orientalists and English-speaking learners of Arabic. Also, in the introduction to the *Learner's Dictionary* there is a reference to the learner, without specifying his native language. This piece of information is given cursorily at the point where the transliteration system is described. A more transparent preface is found in the *Dictionary of Modern Written Arabic* which indicates that the dictionary will be positively received by British and American users as well as Orientalists worldwide. It also refers to the activities in which the dictionary would be most useful, i.e. handling texts written at the beginning of this century. However, it refers cursorily to the Arab user in another section where it describes the microstructure of entries. This is despite the author's awareness that it would be difficult for Arab users to locate the required information in a dictionary primarily designed for Westerners who read or translate from Arabic and not for Arabs writing in or translating into English. Arab users will, in the first place, face the problem of synonym accumulation without enough sense discrimination (Kromann et al. 1984).

In the other two dictionaries, the *School Dictionary* and *Elias' Modern Dictionary*, the Arab compilers address foreign in addition to Arab users for whom these dictionaries were designed. Thus each dictionary contains a pair of introductions, one written in Arabic, the other in English. In the school dictionary, the Arabic introduction says that the dictionary is aimed at students of English at Egyptian and other schools in the Arab East. The English introduction of the same dictionary addresses speakers of English by saying that "this dictionary is primarily aimed at meeting the needs of Arabic speakers in their study of the

English language but in many aspects it will be found useful by English speaking students of Arabic who attained an advanced level in their studies". The introduction of *Elias' Modern Dictionary* clearly indicates both user groups. Its compiler says that his dictionary had been designed for both groups: the Westerner who suffers from expensive and bulky dictionaries and the Arab user for whom no similar dictionary has been written before. He alleges that Arab students of English do not benefit from Orientalists' dictionaries. In his English introduction, he emphasizes the need for an Arabic–English dictionary that helps the foreigner to read modern publications such as newspapers.

The exception here is *Al-Mawrid*, which specifies one user group but without reference to their educational level or the activities in which the dictionary should be used. The compiler specifies Arab users as the targeted group when he says that the dictionary aims "to save the Arab user the trouble of referring to English dictionaries". Somewhere else, at the point where scientific terms are addressed, he identifies this Arab user as the educated Arab.

3.2 Reference to the reasons for compiling the dictionary

Most lexicographers tend to use the introduction for explaining the reasons behind their compilations. However, these explanations vary enormously, from an objective view of EFL learners' linguistic needs to an emphasis on flexibility in acceptance of new and foreign words in Arabic.

The introduction of the *OALD* focuses on the theoretical foundation established by Hornby that EFL learners need dictionaries which differ from those designed for native speakers of English, especially for comprehension and composition. The users' needs are also cited in the bilingual dictionary *Collins-Robert*. Its introduction indicates that the dictionary aims to provide the user with the highest levels of linguistic competence coupled with ease of use.

The introductions of Orientalists' dictionaries usually discuss the increasing interest in Arabic and tend to provide justifications for listing dialect words which are treated as part and parcel of modern Arabic. In the *Dictionary of Modern Written Arabic*, the introduction speaks about the growing interest in bilingual dictionaries of Arabic as the motive behind the compiler's decision to translate his Arabic–German dictionary into English. Also, it provides an analysis of written Arabic which has standard morphology and syntax in all Arab countries, emphasizing the fact that Arabic dialects and foreign words, especially in scientific and technological fields, have had considerable influence on the Arabic language. It also indicates that the dictionary follows a descriptive approach by referring to modern Arabic publications which constitute the corpus. The last dictionary in this group, *Wortabet's Dictionary*, refers in its introduction to the increasing role of English by saying that "the management at Librairie du Liban decided to offer this precious work of reference to the Arab World where the numbers of English students are increasing constantly".

As for bilingual dictionaries compiled by Arab lexicographers, the empha-

sis in their introductions is always on translation and the need to write a dictionary specifically for Arab users, taking into account their different reference skills. The introduction to *Al-Mawrid* contains the compiler's personal observations on his work in English–Arabic translation in which he relied on old dictionaries dating back to the years 1911 and 1913 and the early 1950s. These works were found unsatisfactory vis-à-vis the translator's needs, thus forcing him to refer to English monolingual dictionaries. He also indicates the need to translate scientific terms, saying: "Modern scientific terms are sought by most English–Arabic dictionaries but their absence is more noticeable than other types of words ... Specialized dictionaries provide the technical term without any explanation." The compiler adds another essential aim: writing a comprehensive dictionary that provides the largest number of vocabulary items, thus emphasizing the function of this dictionary as a translation tool.

Elias' Modern Dictionary presents similar reasons in addition to the compiler's opinion about the inclusion of what he calls "contemporary or living vocabulary". The introduction outlines a number of reasons behind the compilation. The first was the desire to meet the Arab learner's need for learning the English equivalents of living Arabic words and the lack of an Arabic–English dictionary. Secondly, Orientalists' dictionaries give explanations which are of interest only to students of classical Arabic, e.g. plurals, verb conjugations, feminine and masculine forms and archaic words, all of which may distract the Arab user. Thirdly, there is the difficulty of using Arabic dictionaries which require their users to know word roots. As fourth reason he mentions the noninclusion of contemporary Arabic vocabulary in older dictionaries, citing the fact that language is similar to a living organism that grows and changes over the years. Accordingly, he calls for flexibility in accepting new words used in newspapers and scientific publications and welcoming those words which find favour with the Arabs.

3.3 Reference to the sources of the dictionary

The introductory matter usually refers to the sources used in the process of collecting the material for the dictionary. This would make users aware of the variety of vocabulary, the extent of lexical comprehensiveness, and the accuracy of definitions or translation equivalents. The analysis of the eight introductions has revealed that three of them do not include any reference to the sources of the dictionary nor the collection of material. The sources of the other dictionaries ranged between computerized corpora for written and spoken forms of language consisting of tens of millions of lexical items to newspapers, periodicals, translated works and dictionaries of different types.

The introduction to the *OALD* informs the user that the lexicographic team relied on the British National Corpus which helped them to determine the relative frequency of words and collocations, to confirm new words, to check the

accuracy of syntactic structures and to rewrite many illustrative examples and add 9 000 new example sentences. This description is obviously meant to tell the users that they can rely on the accuracy and comprehensiveness that characterize the computerized selection and arrangement of dictionary material. Introductions in Orientalists' dictionaries vary with regard to this element, depending on whether the dictionary is the result of team work or individual effort. In the *Dictionary of Modern Written Arabic*, the compiler indicates that he relied on a corpus of primary source materials of selected works by Taha Hus-sain, Muhammad Hussain Haikal, Taufiq al-Hakim and Mahmud Taimur, and numerous newspapers and periodicals in Egypt and Iraq. The secondary sources include different dictionaries such as *Elias' Modern Dictionary* and other bilingual dictionaries of Arabic and French, and Hebrew and Russian for comparison with the dictionary's wordlist. Also, *Webster's International Dictionary* is referred to as an authority on spelling and definitions. However, the introduction does not specify many works used as secondary sources; it only refers to them in general terms, i.e. a number of reference works in European languages, encyclopedias, lexicons, glossaries, technical dictionaries, and specialized literature. These were used to ascertain the correct translation of many technical terms. Reference to these sources occupies a large section of the introduction and shows to some extent the compiler's observance of copyright. Yet, according to modern practice, lexicographers tend to allocate a separate page to list the sources of the dictionary.

Wortabet's Dictionary, the result of an individual effort, refers in its introduction to its sources which are bilingual dictionaries only. It focuses on a comparison with Lane's Arabic-English dictionary. The aim of giving this information might be to imply that the new dictionary is not inferior in quality and comprehensiveness to older Arabic-English dictionaries. In the introduction to the bilingual *Learner's Dictionary* there is a brief reference to the sources of English words. These are translated works such as the *Quran* and *Arabian Nights*, which give the impression that the vocabulary is of a classical nature and on a highly rhetorical level.

Two of the introductions to the bilingual dictionaries compiled by Arab lexicographers do not give any information on the dictionary's sources nor on the process of collecting the material. The compiler of *Al-Mawrid* speaks about the fact that his dictionary was the result of comparing existing English-Arabic dictionaries with English monolingual ones in order to determine the missing vocabulary items, and of referring to specialized dictionaries, the terminology compilations of the Arabic Language Academy in Cairo, and Arabized scientific books. Not only were lexical items taken from these sources but also illustrative examples. A list of sources is given on a separate page, thus conforming to modern practice in dictionaries although no indication is given as to whether permission has been obtained from authors or publishers to use the listed sources.

4. Conclusions

To sum up, the present study has found:

- (a) The dictionaries examined here vary in the extent to which their user groups are specified in the introduction. Some dictionaries refer in broad terms to several groups with different reference needs while others specify one group such as learners but without indicating their mother tongue or the activities for which the dictionary may be used, e.g. writing, reading, or translating from or into a foreign language.
- (b) The European dictionaries (*OALD* and *Collins-Robert*) designed for learners or translators address the reasons for compilation and relate them to users' needs whereas the Orientalists' bilingual dictionaries focus in their explanations on the need for dictionaries that contain dialect Arabic vocabulary in addition to the standard variety. The bilingual dictionaries compiled by Arab lexicographers base their explanations on the needs of translators and learners and the inadequacy of Arabized technical terms.
- (c) Most of these dictionaries (five out of eight) acknowledge their sources which range between computerized corpora and newspapers, periodicals, translated works and dictionaries of different types. However, some dictionaries are not specific in their reference to their source dictionaries and do not indicate whether permission has been obtained from their compilers or publishers.
- (d) The compilers of the dictionaries examined (except for *Collins-Robert* and the *Learner's Dictionary*) claim their dictionaries to be superior to similar dictionaries by focusing on parameters such as accuracy, comprehensiveness of lexical coverage, and ease of use.
- (e) In the most recent dictionaries, e.g. *OALD* and *Collins-Robert*, the introductions concentrate on the new features in the dictionaries, reflecting an increasingly competitive dictionary market.
- (f) The compilers of the modern dictionaries in this study pay special attention to the indication of the number of entries (or words) in their dictionaries to emphasize the lexical comprehensiveness and usefulness of the dictionaries in reading or translating foreign language texts.
- (g) The compilers of *OALD* and *Collins-Robert* emphasize in their introductions the fact that the material was selected with the help of computerized corpora whereas the introductions in the Orientalists' bilingual dictionaries indicate the descriptive approach followed in selecting Arabic

words including dialect vocabulary. The introductions of Arab compilers' bilingual dictionaries stress the role of the dictionary as a translation tool and, therefore, emphasize the use of standard Arabic as the language of translation equivalents, excluding obsolete words and paying special attention to the inclusion of Arabized scientific terms.

Acknowledgments

This study was supported by Kuwait University research grant no. AE01/00.

References

A. Dictionaries

- Atkins, B.T. 1996. *Collins-Robert: French-English/English-French Dictionary*. Glasgow/Paris: D. le Robert-Collins.
- Baalbaki, M. 1998 (1967). *Al-Mawrid: A Modern English-Arabic Dictionary*. Beirut: Dar el-Ilm Lil-Malayin.
- Crowther, J. 1995. *Oxford Advanced Learner's Dictionary*. Oxford: Oxford University Press.
- Elias, A. 1991 (1922). *Elias' Modern Dictionary: Arabic-English*. Beirut: Librairie du Liban.
- Elias, E. 1964. *The School Dictionary: Arabic-English*. Beirut: Dar al-Jeel.
- Steingass, F. 1989. *A Learner's Arabic-English Dictionary*. Beirut: Librairie du Liban.
- Wehr, H. 1995 (1966). *A Dictionary of Modern Written Arabic*. Wiesbaden: Otto Harrasowitz.
- Wortabet, W.T. 1984. *Wortabet's Arabic-English Dictionary*. Beirut: Librairie du Liban.

B. Other references

- Atkins, B.T. 1985. Monolingual and Bilingual Learners' Dictionaries: A Comparison. Ilson, R. (Ed.). *Dictionaries, Lexicography and Language Learning*: 15-24. Oxford: Pergamon Press.
- Béjoint, H. 1981. The Foreign Student's Use of Monolingual English Dictionaries: A Study of Language Needs and Reference Skills. *Applied Linguistics* 2(3): 207-222.
- Berkov, V.P. 1990. A Modern Bilingual Dictionary — Results and Prospects. Magay, T. and J. Zigan (Eds.). 1990. *BudaLEX '88 Proceedings*: 97-106. Budapest: Akademiai Kiado.
- Cowie, A.P. 1987. Syntax, the Dictionary and the Learner's Communicative Needs. Cowie, A.P. (Ed.). *The Dictionary and the Language Learner*: 183-192. Tübingen: Niemeyer.
- Crystal, D. 1986. The Ideal Dictionary, Lexicographer and User. Ilson, R. (Ed.). *Lexicography: An Emerging International Profession*: 72-81. Manchester: Manchester University Press.
- Gimson, A.C. 1973. Phonology and the Lexicographer. McDavid, R. and A. Duckert (Eds.). *Lexicography in English*: 115-121. New York: The New York Academy of Sciences.
- Hartmann, R.R.K. 1986. The Training and Professional Development of Lexicographers in the UK. Ilson, R. (Ed.). *Lexicography: An Emerging International Profession*: 89-92. Manchester: Manchester University Press.

- Hartmann, R.R.K.** 1987. Four Perspectives on Dictionary Use: A Critical Review of Research Methods. Cowie, A.P. (Ed.). *The Dictionary and the Language Learner*: 11-28. Tübingen: Niemeyer.
- Householder, F.W. and S. Saporta.** 1962. *Problems in Lexicography*. Bloomington: Indiana University Press.
- Kromann, H., T. Riider and P. Rosbach.** 1984. "Active" and "Passive" Bilingual Dictionaries: The Ščerba Concept Reconsidered. Hartmann, R.R.K (Ed.). *LEXeter '83 Proceedings*: 207-215. Tübingen: Niemeyer.
- Rossner, R.** 1985. The Learner as Lexicographer: Using Dictionaries in Second Language Learning. Ilson, R. (Ed.). 1985. *Dictionaries, Lexicography and Language Learning*: 95-102. Oxford: Pergamon Press.
- Snell-Hornby, M.** 1987. Towards a Learner's Bilingual Dictionary. Cowie, A.P. (Ed.). *The Dictionary and the Language Learner*: 159-170. Tübingen: Niemeyer.
- Stein, G.** 1984. Towards a Theory of Lexicography: Principles and/vs. Practice in Modern English Dictionaries. Hartmann, R.R.K. (Ed.). *LEXeter '83 Proceedings*: 124-130. Tübingen: Niemeyer.
- Stein, G.** 1989. Recent Developments in EFL Dictionaries. Tickoo, M. (Ed.). *Learners' Dictionaries: State of the Art*: 72-83. Singapore: SEAMEO Language Centre.
- Wiegand, H.** 1984. On the Structure and Contents of a General Theory of Lexicography. Hartmann, R.R.K. (Ed.). *LEXeter '83 Proceedings*: 13-28. Tübingen: Niemeyer.

Lexicography versus Terminography*

Mariëtta Alberts, *Terminology Division, National Language Service,
Department of Arts, Culture, Science and Technology, Pretoria,
Republic of South Africa (vt05@dacst5.pwv.gov.za)*

Abstract: Lexicography and terminography are specialised professions concerned with the compilation and editing of dictionaries. The lexicographer documents the words in the vocabulary of the general language whereas the terminographer documents the terminology of specific subject fields and domains. The terminographer works with a more restricted register than the general lexicographer, but from a holistic point of view, general lexicography and terminography are on a continuum where only the nature of the defined words or terms differs. The conceptual system of the sciences is more systematic and exact than that of the general environment. It is for this reason that nowadays the target group of a particular dictionary determines its format and content. A definition remains extremely important, since it describes the meaning of words and terms and serves to standardise the terms, especially in scientific and technical language. This article concerns itself with the similarities between lexicography and terminography, aspects regarding lexicography as a profession are outlined, followed by a discussion of terminography, and in conclusion attention is drawn to the differences between the two professions.

Keywords: COMPILE, CONCEPTUAL SYSTEM, DESCRIPTIVE APPROACH, DICTIONARY, DOMAIN, EXCERPTION, GENERAL LEXICOGRAPHER, LEXICOGRAPHER, LEXICOGRAPHY, LINGUIST, PRESCRIPTIVE APPROACH, SUBJECT SPECIALIST, SUBJECT FIELD, TARGET GROUP, TERM, TERMINOGRAPHER, TERMINOGRAPHY, TERMINOLOGY, VOCABULARY, WORD

Opsomming: Leksikografie teenoor terminografie. Leksikografie en terminografie is gespesialiseerde beroepe wat hul besig hou met die samestelling en redigering van woordeboeke. Die leksikograaf dokumenteer die woorde in die woordeskat van die algemene taal terwyl die terminograaf die terminologie van spesifieke vakgebiede en terreine dokumenteer. Die terminograaf werk met 'n beperkter register as die algemene leksikograaf, maar vanuit 'n holistiese standpunt beskou, lê die algemene leksikografie en die terminografie op 'n kontinuum waar slegs die aard van die gedefinieerde woorde of terme verskil. Die begrippestelsel van die wetenskappe is meer sistematies en eksak as dié van die algemene omgewing. Dit is om hierdie rede dat die teikengroep van 'n bepaalde woordeboek deesdae sy formaat en inhoud bepaal. 'n Definisie bly uiters belangrik aangesien dit die betekenis van woorde en terme beskryf en dien om die terme te

* This article is a slightly adapted version of a paper presented at the Fifth International Conference of the African Association for Lexicography, held at the University of Stellenbosch, Stellenbosch, 3-5 July 2000.

standaardiseer, veral in wetenskaplike en tegniese taal. Hierdie artikel hou hom besig met die ooreenkomste tussen leksikografie en terminografie, aspekte rakende leksikografie as beroep word geskets, gevolg deur 'n bespreking van terminografie, en ter afsluiting word die aandag gevestig op die verskille tussen die twee beroepe.

Sleutelwoorde: ALGEMENE LEKSIKOGRAAF, BEGRIPPESISTEEM, DESKRIPTIEWE BENADERING, EKSERPERING, LEKSIKOGRAAF, LEKSIKOGRAFIE, LINGUIS, PRESKRIP-TIEWE BENADERING, SAAMSTEL, TEIKENGROEP, TERM, TERMINOGRAAF, TERMINO-GRAFIE, TERMINOLOGIE, TERREIN, VAKGEBIED, VAKSPESIALIS, WOORD, WOORDE-BOEK, WOORDESKAT

1. Introduction

Lexicography and terminography are specialised professions concerned with the compilation and editing of dictionaries. The lexicographer documents the words in the vocabulary of mostly the general language whereas the terminographer documents the terminology of specific subject fields and domains.

Dictionaries serve as aids to the comprehension or to the generation of texts in a language. Ohly (1999: 31) argues as follows:

If the bilingual dictionary is intended to include vocabulary which helps in the full understanding, for example of a novel written in the SL, or to edit an unspecialized text in the TL, then the dictionary's common language has to be saturated with that part of terminology which has sunk into the common language, being taken over from the workshop language and professional language. It may be impossible to separate words from terms. Whether in such circumstances a lexeme still functions as a term or has become a common word has to be seen.

However, the conceptual systems underlying terms belonging to a specific subject field or domain show such a close generic, hierarchical or associative relationship that it is impossible to regard them as common words belonging to the general vocabulary of the layperson.

In its objectives terminology is akin to lexicography which combines the double aim of generally collecting data about the lexicon of a language with providing an information, and sometimes even an advisory, service to language users. Terminology refers to a technical vocabulary, i.e. a collection of terms, which has a certain coherence because the terms belong to a single subject area. It also refers to the compilation of systematic glossaries (Sager 1990: 3).

Some scholars regard the practice of terminology and terminography as a subdiscipline of lexicography. Sager (1990: 1) denies the independent status of terminology as a discipline but affirms its value as a subject in almost every contemporary teaching programme. According to him, terminology is discussed in the context of linguistics, information science or computational lin-

guistics. He sees terminology as a number of practices that have evolved around the creation of terms, their collection and explication and finally their presentation in various printed and electronic media. Wüster (cf. Sager 1990: 2) calls terminology an interdisciplinary field of study, relating linguistics, logic, ontology and information science with the various subject fields, the common element among these disciplines being the concern with the formal organisation of the complex relationships between concepts and terms.

2. Similarities between Lexicography and Terminography

According to Cluver (1992: 32), the difference between general lexicography and technical lexicography (terminography) lies in the variety of language described in each. The assumption can be made that there is no crucial difference between common language and technical language and that it is merely a continuum of registers, where words gradually change to terms and where meanings gradually become more specific. In a holistic sense language workers in both professions of lexicography and terminography use the same basic principles and procedures to document words and terms respectively. In both professions the same basic methods are being used to describe the concept underlying the specific word or term. According to this point of view, the process of lexicography and terminography is placed on a continuous scale. This scale is a reflection of the kind of language usage that extends from common or colloquial language to the language of symbolic logic of science and technology. This scale can further indicate a continuum stretching from context-dependent meanings in lexicography to context-free meanings in terminography (cf. Cluver 1992: 35).

Scheffer (1992: 117-119) suggests the discontinuation of the bipolar situation between general language and terminology. Instead of dividing language into different registers, he would rather speak of a scale on which language functions — from the informal (slang) on the one side to the highly formal registers on the other side with several shades in between. The different shades implicate different grades of standardisation, the need for standardisation as well as possibilities of standardisation. While some registers in terminology are normative, standardisation is in some cases not required or even possible, and in general language it is difficult to regulate matters since emotive connotations are/can be attached to words. According to Scheffer, there is no tension between general language and technical language. The bipolar situation is rather a scale loaded with possibilities.

General lexicography and terminography are methodological facets of the lexicography profession with first objective the systematic description and documentation of the usage of words or terms of a specific language community with its discernible culture and subcultures (cf. Eksteen 1985: 1; Alberts and Mtintsilana 1988: 4; Alberts 1990). The difference is: general lexicography documents the words of a language or languages, while terminography documents the terminology of subject fields, domains and disciplines.

The typology of lexicographical and terminographical products shows

that both professions compile monolingual, bilingual and multilingual dictionaries. The bilingual and multilingual dictionaries can be explanatory or merely translating dictionaries.

3. Lexicography

Lexicography is the activity or profession of writing and editing dictionaries (cf. Cluver 1989: 117). It comprises the recording of the words in the vocabulary of the language into a specific format (e.g. alphabetically).

Lexicography has two basic components (Alberts 2000):

- the theoretical component which consists of (a) the general principles of the applied science, and (b) the theoretical principles that form the basis for the general usage and expert principles; and
- the practical component which deals with the applied science of compiling and editing dictionaries.

Lexicography deals with the compilation and editing of general dictionaries. A general dictionary comprises all aspects of a given source language, i.e. common words, colloquial words, dialects, archaic words, etymology of words, words of literature, slang, vulgar words, deprecated words, sexist words, taboo words, etc. The lexicographer can provide the source-language words with translation equivalents in a target language or target languages.

Lexicography is the process in which linguistic information (that is the result of linguistic research) is being written down, processed and compiled in a specific lexicographical format. The result of the lexicographical process is usually a wordlist, glossary, dictionary, thesaurus or a computerised wordbank.

A general dictionary is an alphabetical list of the words of a given language. A dictionary is "a book containing a selection of words, usually arranged alphabetically, with explanations of their meanings and other information concerning them, expressed in the same (*source language*) or other language (*target language*)" (Barnhart 1966). A dictionary is supposed to comment on a particular language: "The responsibility of a dictionary is to record the language, not set its style" (Gove in Al-Kasimi 1983: 84). According to Béjoint (1981: 208), a dictionary has a two-way function: it can be "a portrait of the vocabulary of a language", or it can be "a tool for more effective communication". A dictionary is therefore an inventory of the vocabulary of a particular language or of specific languages as well as a tool for communicating in the particular language or in the specific languages. But no dictionary would, according to Zgusta (1984: 154), be able to encompass all the linguistic information a dictionary user would need or like to have. The dictionary should nevertheless contain enough information to allow the dictionary user to successfully discover the relevant information.

A general dictionary offers information about some or all of the following: the spelling, pronunciation, definition, syllabification, translation equivalent, derived forms and compounds, grammatical class, meaning, usage, cross reference to other entries, illustrations, etymology, history of words and source. General dictionaries are therefore, as Cluver (1989) emphasises, word-oriented rather than topic-oriented. They are concerned with a variety of linguistic aspects.

Lexicography works with words. The verbal as well as the written vocabulary of a particular language is documented in monolingual dictionaries. Lexicography also compares different languages in bilingual and multilingual dictionaries.

Zgusta (1971: 197) considers Berg's definition to be the best description of a dictionary: "A dictionary is a systematically arranged list of socialised linguistic forms compiled from the speech-habits of a given speech-community and commented on by the author (*lexicographer*) in such a way that the qualified reader (*dictionary user*) understand the meaning ... of each separate form, and is informed of the relevant facts concerning the function of that form in its community."

Keating (1979: 20) defines a dictionary as "a retrieval system in which are stored, against the words of a language, explanations of the meanings of those words and the ways in which they are used".

"'n Woordeboek is 'n konkordans van taalfeite," states Grobler (1978: 29). "Die taalfeite word nie net daarin versamel nie, maar daar word tegelyk ook kommentaar op die taalfeite gelewer. In die woordeboek is daar 'n alfabetiese sensusopgawe van die taal se woordvoorraad sonder grammatiese groepering met hoogstens 'n samehorigheid van woordfamilies."

4. Terminography

Sager (1990: 4) says that terminology is concerned with "the study and use of the systems of symbols and linguistic signs employed for human communication in specialised areas of knowledge and activities":

It is primarily a linguistic discipline — linguistics being interpreted here in its widest possible sense — with emphasis on semantics (systems of meanings and concepts) and pragmatics. It is inter-disciplinary in the sense that it also borrows concepts and methods from semiotics, epistemology, classification, etc.

Terminology is closely linked to the subject fields, of which it describes the vocabulary and for which it seeks to provide assistance in the ordering and use of designations. The International Association of Terminology (1982, in Sager 1990: 4) delineates the scope of terminology as follows:

Although terminology has been in the past mostly concerned with the lexical aspects of specialised languages, its scope extends to syntax and phonology. In its applied aspect terminology is related to lexicography and uses techniques of information science and technology.

"In practice no individual or group of individuals possesses the whole structure of a community's knowledge," stresses Sager (1990: 16). Conventionally knowledge is divided into "subject areas, domains or disciplines, which is equivalent to defining subspaces for the knowledge space".

Terminology (or technical vocabulary) is a collection of terms, which has a certain coherence because the terms belong to a single subject area (Sager 1990: 3). As Sager (1990: 2) puts it, it is "the study of and the field of activity concerned with the collection, description, processing and presentation of terms, i.e. lexical items belonging to specialised areas of usage of one or more languages". In contemporary usage three meanings of the term "terminology" can be distinguished:

- the set of practices and methods used for the collection, description and presentation of terms;
- a theory, i.e. the set of premises, arguments and conclusions required for explaining the relationships between concepts and terms which are fundamental for a coherent (terminographical) activity; and
- the vocabulary of a special subject field.

Terminography is one of the subdivisions of lexicography and deals with the documentation of the terminology of different subject fields, i.e. technical and scientific terms. The terminology of any subject field (physics, mathematics, biology, chemistry, etc.) or domain (sport, music, etc.) can be documented in terminographical dictionaries, called "technical dictionaries". The vocabulary (called "terminology") of a subject is the group of words (called "terms") that are typically used when discussing the specific subject.

A technical dictionary is also known as a subject dictionary or a terminological dictionary. A technical dictionary contains the standardised terms of a particular subject (Cluver 1989). Technical dictionaries are therefore subject-oriented.

Terminology is divided by subject field *before* it is ordered *in any other way*. "In order to deal effectively with large quantities of terms, it is generally considered advisable to introduce a classification of terms by subject areas," says Sager (1990: 147).

The practice-oriented origin of terminology is reflected in its fundamental assumptions and working methods. Examining separate "subject fields", it sees the lexicon as many separate subsystems related to the knowledge structure of each subject field or discipline. It can therefore use as starting point the conceptual structure of an area of knowledge (Rondeau 1981; Sager 1984: 317).

A technical dictionary is a bi- or multilingual (alphabetical) dictionary

which includes the (standardised) terms of a particular subject field, with or without a full definition for each term (Cluver 1978: 91 and 1989a: 154; Alberts 1990: 134). A multilingual technical dictionary includes an alphabetical list of terms in the source language and equivalents in several languages (Cluver 1989a: 121; Alberts 1990: 134).

The terminology in a technical dictionary is a coherent, structured collection of terms. According to Sager (1990: 114), it is "a representation of an equally coherent, but possibly differently structured system of concepts". These are expressed in a distinctive way:

Concepts represented in terminological dictionaries are predominantly expressed by the linguistic form of nouns; concepts which are linguistically expressed as adjectives and verbs in technical languages are frequently found only in the corresponding noun form and some theorists deny the existence of adjectives and verb concepts.

Terms are "the linguistic representation of concepts" (Sager 1990: 57). Each separate meaning of a term is represented by a separate concept and the entries in a terminological glossary are of separate concepts (Sager 1990: 56).

The terminologist orders the collected terms by reference to a conceptual system which is drawn up in consultation with a subject specialist (Sager 1990: 56). Once a conceptual system has been defined, the terminologist is more certain of exhaustively covering the vocabulary (terminology) of the particular subject field that is being described at any one time (Sager 1990: 56).

Sager (1990: 57) explains the difference between general language and special languages as follows:

Unlike in general language, where the arbitrariness of the sign is accepted, the special languages (of subject fields and domains) strive to systematise principles of designation and to name concepts according to pre-specified rules or general (terminographical) principles. General language (and therefore general dictionaries) fully exploits polysemy, metaphor, and adjectival determination; genuine word creation is relatively rare. Where it occurs it is based on the experience of every-day life and thus represents a pre-scientific approach of knowledge. The process of scientific observation and description includes designation of concepts and this in turn involves re-examining the meaning of words, changing designations and coining new terms. ... Designation in special languages (of subject fields and domains) therefore aims at transparency and consistency (standardisation).

This means that terminology (a) is concerned with special subject languages and must therefore be in a position to subdivide the lexicon; (b) has a practical function of guidance in usage for which it needs to examine existing usage; (c) has a teaching function; and (d) is strongly involved in the creation of new designations and is therefore interested in a theory of active, productive term-formation (Sager 1984: 317).

Terminology can claim to be truly interdisciplinary. According to Wüster (see Sager 1990: 2), terminology is vital to the functioning of all sciences, it is concerned with designations in all other subject fields, and it is closely related to a number of specific disciplines. Wüster (see Sager 1990: 2) calls terminology "an interdisciplinary field of study, relating linguistics, logic, ontology and information science with the various subject fields": "The common element among these disciplines is that they are each concerned ... with the formal organisation of the complex relationships between concepts and terms."

"In their advisory function," says Sager (1990: 210), "terminologists can be heavily involved in language planning and in particular in establishing new terminologies to accompany the transfer of science and technology to wider user groups in the same or other linguistic communities."

5. Differences between Lexicography and Terminography (See Table)

In terminology, the conceptual content of a term needs to be conveyed by means of standardised terms to enable exact communication. According to Alberts (1990: 101), "one of the aims of practical terminological work (terminography) is to order the concept systems of subject fields". In general lexicography, the vocabulary of the standard language is documented in dictionaries (cf. Cluver 1989b: 86). In terminography, attention is given to the dictionary user's needs for information on a specific, marked area of human activity. It is usually demarcated in certain subject fields, subdivisions or domains (Alberts 1990: 109).

While general lexicography is in nature descriptive, terminography is primarily prescriptive since terms and the underlying concepts need to be standardised to ensure exact communication (Alberts 1990: 109).

The aim of general lexicography is to document, describe and preserve the vocabulary and its derivations within general language usage. Terminography concentrates on a representation of the terminology (lexicon) of language for special purposes (Alberts 1990: 109).

In its objectives, terminography is akin to lexicography which combines "the double aim of generally collecting data about the lexicon (vocabulary) of a language with providing an information, and sometimes even an advisory, service to language users" (Sager 1990: 3). The justification of considering it a separate activity from lexicography lies in "the different nature of the data traditionally assembled, the different background of the people involved in this work, and to some extent in the different methods used" (Sager 1990: 3).

One of the fundamental differences between lexicography and terminography lies in the attitude to the lexicon, the raw material to be collected and processed. The lexicographer in principle collects "all" the words of a language in order to sort them in various ways. Once he has collected the words, he proceeds to differentiate them by their meanings. His ideal dictionary covers all the words and all their meanings, even though in practice he will produce various types of subsets for diverse uses. The terminologist starts out from a

much narrower position; he is only interested in subsets of the lexicon, which constitute the vocabulary (or terminology) of special languages (subject fields and domains). In order to arrive at these subsets he needs a structure of knowledge which justifies the existence and the boundaries of special languages (subject fields and domains), so that he can attribute words (terms) to separate areas of this structure. Since a word can belong to more than one area of knowledge (the phenomenon of homonymy), the terminologist has to distinguish meanings before he distinguishes words (Sager 1990: 55, 56). The conceptual structures of special subject fields are distinguished not only by special reference, but also by the nature of the concepts that predominate in particular subjects (Sager 1990: 25). The items which are characterised by special reference within a discipline are the "terms" of that discipline, and collectively they form its "terminology"; those which function in general reference works are called "words" and their totality the "vocabulary" (Sager 1990: 19).

According to Wersig (1976), there is a division of labour between lexicography and terminography. Lexicography descriptively examines the current state of the lexicon and its usage by recording existing designations and their meaning. The role of terminography is to establish firm relationships between concepts and designations, by determining, if necessary prescriptively, which designations should be used and how they should be used (Sager 1984: 316, 317).

Lexicography is concerned with language/languages as such, but terminography deals with subject fields, domains and disciplines of special languages.

Although professionals use terminology related to their working environment on a daily basis, they often do not realize how these terms are created, standardised and documented into terminology lists, dictionaries or term banks. A standardised terminology not only plays an important role in their acquisition of knowledge, but also helps to convey their ideas and to facilitate their communication. The terminologist has to excerpt terms from a variety of sources, has to define the concept to clearly depict the meaning of a source term, has to document and standardise this term with its related information on meaning and usage, and has to supply term equivalents in several target languages.

Terminologists are highly dependent on the input and collaboration of subject-field specialists and experts of various occupational domains. They have to liaise with various subject committees and linguists while doing research on terminology and need to systematically document terminological information relating to terms.

Even people practising lexicography are not aware of the difference between lexicography and terminography. For them these are both *just* processes of documenting terms and related information. They are not aware of the intricate difference between the two dictionary professions:

- In lexicography the approach is semasiological, a meaning approach: It starts from the word and looks for its meaning.

In lexicography the point of departure is always language. The words of a given language are documented. The linguistic function of a word in a given language is determined and the word is documented according to its syntactical, semantic, contextual, and other features and functions in the grammar of the particular language. The words of a given language are documented in monolingual dictionaries. These words can then also be compared to words in other languages by means of bilingual or multilingual dictionaries.

Words convey emotions and may have different meanings and connotations in different situations.

- In terminography the approach is onomasiological, a naming approach: It starts from the concepts and creates names for the concepts.

In terminography the point of departure is the concept and its relation to other concepts within a specific field of knowledge. Terminologists are thus focused on the domain or subject field where the terms occur. Terminology concerns the metalanguage of science and technology. It is only when a concept is properly defined within the boundaries of a subject field or domain that the true meaning can be determined. The source-language term needs to be defined and its usage within a subject field or domain needs to be determined. Once this is done, the source-language term can be standardised. A term should have only one meaning — one concept: one term.

Terms with fixed meanings can be standardised and thus used to promote exact communication among subject specialists, and between subject specialists and laypeople.

Terminology can, to a certain extent, be regarded as an artificial language and it uses spelling and orthography principles of the standard variety of a language when coining terms. Terminology is also mainly concerned with written language and consists mainly of nouns and, to a lesser extent, verbs. Terms are exact and should have no emotional connotations attached to them. When emotional connotations are attached to terms, these terms become words and therefore part of the general vocabulary (the terrain of lexicography).

7. Conclusion

Though there are several similarities between lexicography and terminography, it is certain that these are also two very distinct professions with clearly demarcated working areas. The functions of the terminographer, and a terminography unit, cannot be taken over by the lexicographer, and a lexicography office, and *vice versa*. As has been shown in the above discussion, the points of departure and the methods of work of the two professions differ completely.

Bibliography

- Alberts, M.** 1990. *'n Bepaling van Afrikaanse vakleksikografiese behoeftes*. D.Litt. et Phil.-proefskrif. Pretoria: UNISA.
- Alberts, M.** 1992. Behoeftebepaling in die leksikografie. *Lexikos* 2: 1-27.
- Alberts, M.** 2000. *Lecture 1: Vocabulary, Lexicon, Lexicology, Lexicography, Terminology, Terminography*. Class notes, Terminology course. Pretoria: Department of African Languages, University of Pretoria.
- Alberts, M. and P. Mtintsilana.** 1988. *Leksikografiese ontleding van Afrikaanse en Afrikatale-leksikons*. Verslag Lexi-2. Pretoria: RGN.
- Al-Kasimi, A.M.** 1977. *Linguistics and Bilingual Dictionaries*. Leiden: E.J. Brill.
- Al-Kasimi, A.M.** 1983. The Interlingual/ Translation Dictionary. Hartmann, R.R.K. (Ed.). 1983. *Lexicography: Principles and Practices*: 153-162. London: Academic Press.
- Barnhart, C.L.** 1966. *The American College Dictionary*. New York: Random House.
- Béjoint, H.** 1981. The Foreign Student's Use of Monolingual English Dictionaries: A Study of Language Needs and Reference Skills. *Applied Linguistics* 2(3): 207-221.
- Cluver, A.D. de V.** 1978. Die Terminologie. Van Schalkwyk, C. et al. (Comp.) 1978. *Diploma in Vertaling*. Pretoria: UNISA: 67-131.
- Cluver, A.D. de V.** 1989a. *Dictionary of Terminography / Terminografiewoordeboek*. Pretoria: HSRC.
- Cluver, A.D. de V.** 1989b. *A Manual of Terminography*. Pretoria: HSRC.
- Cluver, A.D. de V.** 1992. Die verskille en ooreenkomste tussen algemene leksikografie en vakleksikografie. National Terminology Services. 1992: 32-48.
- Eksteen, L.C.** 1969. Die leksikologie van Afrikaans. *Taalfasette* 9: 30-43.
- Eksteen, L.C.** (Ed.). 1985. *Leksikografie-aangeleentede in Suid-Afrika*. Verslag Lexi-1. Pretoria: RGN.
- Eksteen, L.C. and Alberts, M.** 1985. *Terreinverkenning en Interpretasie*. Eksteen, L.C. (Ed.). 1985: 28-45.
- Gove, P.B.** 1967. The Dictionary's Function. Gove, P.B. (Ed.). 1967. *The Role of the Dictionary*: 5-7. Indianapolis: The Bobbs-Merrill Co.
- Grobler, Hilda.** 1978. 'n Voorlopige toepassing van S.P.E. Boshoff se kriteria vir 'n groot woordeboek op WAT I-VI. *Klasgids* 12(4): 1978: 29-46.
- Hartmann, R.R.K. (Ed.).** 1984. *LEXeter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9-12 September 1983*. Lexicographica Series 1. Tübingen: Max Niemeyer.
- 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. *Wörterbücher. Ein internationales Handbuch zur Lexikographie/Dictionaries. An international Encyclopedia of Lexicography/Dictionnaires. Encyclopédie internationale de lexicographie*: 328-360. Berlin: Walter de Gruyter.
- Keating, P.** 1978. Dictionaries, Technical Dictionaries and the Consumer. *The Language Practitioner* 1: 20-35.
- Keating, P.** 1979. Indeks en sinonieme. *Die Taalpraktisyn* 2: 38-41.
- Landau, S.I.** 1984/1989. *Dictionaries. The Art and Craft of Lexicography*. New York: C. Scribner's Sons/ Cambridge: O.U.P.
- National Terminology Services.** 1992. *Algemene en Vakleksikografie in die Praktijk. / General and Technical Lexicography in Practice*. Pretoria: Department of National Education.
- Ohly, R.** 1999. Term as a Cultural Closure. *NEOTERM – World Specialized Terminology* 35/36: 31-36.

- Rondeau, G.** 1981. *Introduction à la terminologie*. Montréal: Centre éducatif et culturel.
- Sager, J.C.** 1984. Terminology and the Technical Dictionary. Hartmann, R.R.K. (Ed.). 1984: 315-326.
- Sager, J.C.** 1990. *A Practical Course in Terminology Processing*. Amsterdam: John Benjamins.
- Scheffer, C.J.** 1992. Samewerking in die vaktaalomgewing. National Terminology Services. 1992: 115-126.
- Wersig, G.** 1976. Probleme und Verfahren der Terminologiearbeit. *Fachsprachen (Terminologie – Struktur – Normung)*. Berlin: Beuth.
- Zgusta, L.** 1971. *Manual of Lexicography*. The Hague: Academia.
- Zgusta, L.** 1984. Translational Equivalents and the Bilingual Dictionary. Hartmann, R.R.K. (Ed.). 1984: 147-154.

Table

Lexicography	Terminography
The theory and practice of compiling and editing general dictionaries.	The theory and practice of compiling technical/subject/terminological dictionaries.
Uses basic principles and procedures for the compilation and editing of general dictionaries.	Uses basic lexicographical (terminographical) principles and procedures for the compilation and editing of technical dictionaries.
Conducts a process of excerption, documentation, description, processing, and dissemination of information on the general vocabulary of a given language or languages.	One of the subdivisions of lexicography for the excerption, documentation, description, processing, and dissemination of terminological information on various subject fields, domains and disciplines.
Concerns the compilation of general dictionaries.	Concerns the compilation of technical dictionaries.
Point of departure: language (e.g. English, Xhosa, Sepedi, French, German, Tsonga, Polish, Venda, Dutch, Frisian, ...) to promote communication amongst users of the same language or users of different languages.	Point of departure: a subject field or discipline (e.g. physics, zoology, psychology, ...) or a domain (music, art, journalism, ...) to promote communication among subject specialists and between subject specialists and laypeople.
Interested in the spoken and written form of the language.	Mainly interested in the written form of technical language.
Descriptive approach to document, describe and preserve a language in all its facets and registers.	Prescriptive approach to document and describe the concepts of a subject field/domain by means of definitions and terms to standardise the terms.
Documents the words of a language according to the spelling and orthography rules of the given language.	Documents the terms of a subject field according to the spelling and orthography of the standard variety of the given language.
Documents all the words of a given language (e.g. common words, colloquial words, jargon, slang, dialectal usage, archaic words, literature, vulgar words).	Documents the terms of a subject field (e.g. scientific terms (natural and life sciences), technological terms, human sciences, commercial sciences, jargon (e.g. of the workplace, sport, ...) or domain (e.g. journalism, music, art, ...))
Has a semasiological (meaning) approach which starts from the word and looks for its meaning.	Has an onomasiological (naming) approach which starts from concepts and creates names for concepts.
Emotional connotations can be attached to words, resulting in shift in meaning.	Terms are exact. One concept equals one term. No emotional connotations attached to terms.
Typology: monolingual, bilingual or multilingual translating or explanatory general dictionaries.	Typology: monolingual, bilingual or multilingual translating or explanatory technical dictionaries.
Macrostructure: alphabetical, retrograde, unidirectional, bidirectional, back matter, front matter.	Macrostructure: alphabetical, retrograde, unidirectional, bidirectional, back matter, front matter.

<p>Microstructure: Explicit information: part of speech, plural, diminutive form, past tense, degrees of comparison, inflected forms. Implicit information: selected contextual information, collocations. Information classes: spelling; phonology; lemma; word class (parts of speech); flexion morphemes (case, number, degree of comparison, mode, tense, form (active and passive); derivational morphemes; syntactic environment (surface structure); deep case (deep structure); semantic information (semantic definition/meaning of the word), semantically relevant context (context and contextual word); examples of usage; semantic relations (semantic primitives, markings of specialised field).</p>	<p>Microstructure: Explicit information: part of speech (mainly nouns and verbs), plural form. Implicit information: selected contextual information. Information classes: lemma; word class (parts of speech); syntactic environment (surface structure); deep case (deep structure); semantic information (semantic definition/meaning of the word), semantically relevant context (context and contextual word); examples of usage; semantic relations (semantic primitives, markings of specialised field).</p>
<p>Major international lexicography bureaux/publishers: Woordenboek der Nederlandse Taal; Webster; Oxford; Collins; Funk and Wagnall; Longman; Van Dale; COBUILD, Cassell, Duden.</p>	<p>Major international terminology bureaux/termbanks: Infoterm; TermNet; ISO TC/37; WBIT; Phillips; WHO; IOUNT; IFTB; TD of NLS, DACST, National Termbank (RSA); Elsevier'S dictionaries on various subject fields.</p>
<p>National Lexicography Units (NLUs): 11 units for the compilation of general dictionaries for each of the official languages.</p>	<p>National terminology unit: the Terminology Division, NLS, DACST, for the coordination, facilitation and documentation of terminology projects, managing of the National Termbank and dissemination of terms on a variety of subjects.</p>

Beiträge zu einer kognitiv ausgerichteten Lexikographie

Udo L. Figge, *Romanisches Seminar, Ruhr-Universität Bochum,
Bochum, Deutschland (Udo.L.Figge@ruhr-uni-bochum.de)*

Zusammenfassung: Nach einer "kognitiven Wende" in der Psychologie in der zweiten Hälfte der sechziger Jahre haben auch Teile der Linguistik und der Psycholinguistik begonnen, sich kognitiv zu orientieren. Hier wird ein Modell zugrunde gelegt, nach dem sprachliche Einheiten auf Einheiten des Gedächtnisses referieren. Einheiten des Gedächtnisses sind zunächst Merkmale, sodann Konzepte, die sich aus Merkmalen konstituieren, und schließlich Gebilde von Merkmalen, Konzepten und Beziehungen zwischen Merkmalen. Eine solche Konzeption entspricht in den Grundzügen zeitgenössischen neuropsychologischen Erkenntnissen. Konzepte manifestieren sich in substantivischen Satzgliedern. Die Bedeutung eines Substantivs ist lexikographisch als Menge von Merkmalen zu beschreiben, und zwar ohne Beachtung der ohnehin schwer zu rechtfertigenden Unterscheidung zwischen semantischem und enzyklopädischem Wissen. Die Bedeutung eines attributiven Adjektivs ist lexikographisch als Merkmal der Konzepte zu beschreiben, die durch die zugehörigen Substantive ausgedrückt werden. Einem Verb liegt ein konverses Netz von Merkmalen zugrunde; es muß lexikographisch dieser Bedeutung entsprechend beschrieben werden. Bei der Beschreibung von Substantiven ist systematisch zwischen echten Substantiven und solchen Substantiven zu unterscheiden, die, meist als Ableitungen, eine verbale oder adjektivische Bedeutung haben. Insbesondere ist zu beachten, daß solche Substantive in Nebensatzäquivalenten die Funktion substantivischer Verbformen haben, daß sie weiterhin auch substantivische Bedeutung annehmen können und daß sie schließlich häufig Konstituenten von Stützverbgefügen sind. Die Bedeutung von Konjunktionen und von Präpositionen in Nebensatzäquivalenten kann nur nach einer exakten grammatischen Analyse verlässlich beschrieben werden. Texte werden unter anderem durch Textgliederungssignale gegliedert, die meist die Gestalt von Adverbien oder Adverbialen haben. Diese Textgliederungssignale müssen lexikographisch systematisch als solche beschrieben werden.

Schlüsselwörter: KOGNITIVE LINGUISTIK, LEXIKOGRAPHIE, KONZEPT, MERKMAL, GEDÄCHTNIS, SEMANTIK, SEMANTISCHES GEDÄCHTNIS, SEMANTISCHES WISSEN, ENZYKLOPÄDISCHES WISSEN, SUBSTANTIV, DIFFERENTIA SPECIFICA, GENUS PROXIMUM, ADJEKTIV, VERB, DIATHESE, KONJUNKTION, TEXTGLIEDERUNGSSIGNAL

Abstract: Contribution to a Cognitively Orientated Lexicography. After a "cognitive turn" in psychology during the second half of the sixties, linguistics and psycholinguistics have also in part begun to orientate themselves cognitively. Here a model is used as basis in which speech units refer to memory units. Memory units are firstly features, furthermore concepts made up of features, and finally structures of features, concepts and relationships between features. Such an idea basically corresponds to present neuropsychological findings. Concepts are mani-

fested by nominal phrases. The meaning of a noun is lexicographically described as a set of features, and admittedly without taking into account the difference between semantic and encyclopaedic knowledge, which is already difficult to justify. The meaning of an attributive adjective is lexicographically described as a feature of the concepts expressed by the accompanying noun. A converse network of features forms the basis of a verb and must be lexicographically described according to this meaning. In the description of nouns there must be differentiated between true nouns and those which, mostly as derivations, have a verbal or adjectival meaning. It is especially important to note that such nouns in subordinate clause equivalents have the function of nominal verb forms, furthermore that they can adopt nominal meaning and finally that they are often part of supportive verb structures. The meaning of conjunctions and prepositions in subordinate clause equivalents can only be described correctly after precise grammatical analysis. Texts are *inter alia* structured by discourse markers, which usually have the character of adverbs or adverbials. These discourse markers should be lexicographically systematically described as such.

Keywords: COGNITIVE LINGUISTICS, LEXICOGRAPHY, CONCEPT, FEATURE, MEMORY, SEMANTICS, SEMANTIC MEMORY, SEMANTIC KNOWLEDGE, ENCYCLOPEDIA KNOWLEDGE, NOUN, DIFFERENTIA SPECIFICA, GENUS PROXIMUM, ADJECTIVE, VERB, VOICE, CONJUNCTION, DISCOURSE MARKER

1. Einleitung

Wörter sind in zweierlei Hinsicht korrelative Einheiten. Zum einen stehen sie nämlich zu komplexeren Einheiten in Beziehung, und zwar insofern, als sie normalerweise als Teile von komplexeren Einheiten vorkommen. Zum anderen stehen sie zu Bedeutungen in Beziehung, und zwar insofern, als sie Bedeutungen ausdrücken. Das ist eine triviale Feststellung, die jedoch in dem Maße an Trivialität verliert, wie sie präzisiert wird. Die Einheiten, die Wörter als Teile enthalten, bilden eine Hierarchie zunehmender Komplexität: Satzglieder — (einfache) Sätze — Satzgefüge — Texte. Die Bedeutungen, zu denen Wörter in Beziehung stehen, lassen sich als Konzepte und damit als Elemente des Gedächtnisses betrachten. Konzepte sind ebenfalls korrelative Einheiten; sie können miteinander Beziehungen unterschiedlicher Art eingehen. Treten mehrere verschiedene Konzepte miteinander in solche Beziehungen, dann entsteht ein Konzeptgebilde. Was Sätze, Satzgefüge und Texte ausdrücken, sind solche Konzeptgebilde. Auch Konzeptgebilde können von unterschiedlicher Komplexität sein: Texte bezeichnen normalerweise komplexere Konzeptgebilde als Satzgefüge und Satzgefüge stets komplexere Konzeptgebilde als (einfache) Sätze.

Das ist eine kurze Skizze der Grundlagen, auf denen dieser Beitrag fußt. Im Folgenden werden zunächst diese Grundlagen detaillierter dargestellt. Aus dieser Darstellung sollen dann Hinweise für die Lexikographie gewonnen werden, und zwar für eine Lexikographie, die zu einer exhaustiven Beschreibung von Sprache beitragen will.¹

2. Kognitive Semantik

Die Wörter *Konzept* und *Gedächtnis* deuten an, daß ich einer kognitiv orientierten Linguistik gegenüber anderen den Vorzug gebe. Obwohl, wie das seit Jahrhunderten immer wieder abgelegte Bekenntnis zu dem Satz *Sprache dient dem Ausdruck von Gedanken* oder einer seiner Varianten zeigt, kognitive Orientierung ein altes Programm mit Sprache befaßter Disziplinen ist, hat sich eine Linguistik, die dieses Programm wirklich auf ihre Fahnen geschrieben hat, erst entwickelt, nachdem die Psychologie in der zweiten Hälfte der sechziger Jahre eine "kognitive Wende" vollzogen hatte, wie vor allem der Titel eines Buchs von Neisser (1967) verdeutlicht. Diese Wende war eine Abwendung von Verhalten und eine Zuwendung zu kognitiven Prozesse und kognitiven Strukturen als Untersuchungsgegenständen. Zentrale Begriffe dieser neuen Psychologie sind "Wahrnehmung", "Denken", "Gedächtnis".²

In diesem Geiste hat der nordamerikanische Psychologe Quillian damals Untersuchungen durchgeführt, die in eine Theorie und ein Computermodell der Struktur des menschlichen Langzeit-Gedächtnisses münden sollten. Quillian nannte die Grundeinheiten dieses Modells "Konzepte". In diesem Kontext sollte hervorgehoben werden, daß er, um diese Konzepte zu gewinnen, von einem damals gängigen Wörterbuch des Englischen (Ogden 1942) ausging. Einzelne Einträge dieses Wörterbuchs ließ er unter der Kontrolle des von ihm so genannten "semantischen Gedächtnisses" eines Mitarbeiters in sein Computerprogramm eingeben. Dieses semantische Gedächtnis war nach seiner Auffassung diejenige Komponente des menschlichen Langzeit-Gedächtnisses, in der Konzepte gespeichert werden, die gleichzeitig Bedeutungen von Wörtern sind (1967: 410 f.). In seiner Arbeit von 1967 verwendet Quillian den Begriff "semantisches Gedächtnis" noch eher beiläufig. In einer Neufassung dieser Arbeit hat sich der Begriff aber bereits in den Vordergrund geschoben: Sie trägt den Titel "Semantic Memory" (Quillian 1968). "Semantisches Gedächtnis" wurde dann auch der Name eines Gebietes, dem sich im Anschluß eine lebhaft sprachpsychologische Forschung widmete, die selber gern als psychologische Semantik bezeichnet wird (vgl. etwa Hörmann 1976, Le Ny 1979). Das Ergebnis war eine Reihe sich teils ergänzender, teils aber auch widersprechender Modelle des semantischen Gedächtnisses (vgl. etwa Kintsch 1980, Hörmann 1981: 68-82).

Ein Aspekt ist dabei von besonderem Interesse: Öffnet man einen Überblick über die gegenwärtige allgemeine Gedächtnisforschung (etwa Baddeley 1999), so stellt sich heraus, daß die verschiedenen Modelle der Struktur des menschlichen Langzeit-Gedächtnisses — genauer: derjenigen Komponente des Langzeit-Gedächtnisses, die das allgemeine Wissen verwaltet — mit Modellen identisch sind, die für das semantische Gedächtnis entwickelt wurden (Baddeley 1999: 145-167). Das entspricht ja auch einer der Annahmen, von denen Quillian ausgegangen ist.

Eines der Modelle des semantischen Gedächtnisses ist das von Smith, Shoben und Rips (1974) entwickelte. Dieses Modell faßt Wortbedeutungen als Konzepte und Konzepte wiederum als Mengen von Merkmalen. Danach entspricht beispielsweise dem englischen Wort *robin* '(nordamerikanisches) Rotkehlchen' ein Konzept, das sich aus Merkmalen wie 'being a biped', 'having wings', 'having certain distinctive colors', 'perching in trees', 'being undomesticated' zusammensetzt. Dieses Modell ist insofern von kognitiver Relevanz, als es sich mit der Struktur der Wahrnehmung in Beziehung setzen läßt. Ein Organismus wie der Mensch verfügt über eine ganze Reihe von Sinnen. Die Informationen, die er sensorisch aufnimmt, sind daher sehr fragmentarisch. Es sind keineswegs Informationen über komplette Gegenstände, sondern Informationen über äußerst partikuläre Umweltaspekte: Farbinformationen, Gestaltinformationen, Bewegungsinformationen, Schallfrequenzinformationen, Schallintensitätsinformationen, Geruchsinformationen, Temperaturinformationen, verschiedene Arten von Informationen über sein Körperinneres usw. (s. Treisman 1986, 1987). Das Gedächtnis ist aber fähig, einzelne solcher partikulären Informationen, wenn zwischen ihnen deutliche Kontiguitäten, vorwiegend temporaler und lokaler Natur, bestehen, zu einem Konzept etwa von einem Gegenstand oder von einer Person zusammenzufassen. Partikuläre Sinnesdaten, die in ein solches Konzept eingehen, werden dann zu Merkmalen dieses Konzepts. Das soll natürlich nicht heißen, daß sämtliche Konzepte des menschlichen Gedächtnisses und deren Merkmale auf Wahrnehmung zurückgehen. Das Gedächtnis enthält vielmehr viele Konzepte und Merkmale, die auf semiotischer, vor allem sprachlicher Vermittlung beruhen. Die Fragmentierung der Sinne und die Fähigkeit zur Zusammenfassung sensorischer Einzelinformationen liefern aber eine allgemeine Erklärung für den Aufbau des Gedächtnisses aus Konzepten und Merkmalen.

Diese Konzeption ist von der Neuropsychologie im allgemeinen bestätigt und im einzelnen präzisiert worden. Es ergibt sich derzeit folgendes Bild: Werden aufgrund verschiedener, aber zeitlich und räumlich zusammenhängender Wahrnehmungen unterschiedliche Einzelinformationen gebildet, dann oszillieren Neuronenkomplexe in unterschiedlichen sensorischen Rindenbezirken in einem charakteristischen gemeinsamen Rhythmus. Supramodale Konvergenzonen in Assoziationsfeldern des Isocortex, in denen Nervenfasern aus vielen anderen, zum Teil abliegenden Gehirnregionen zusammenlaufen (Assoziationsbahnen), führen gewissermaßen Protokoll über solche gemeinsamen Oszillationen. Erinnerung bedeutet, daß das Protokoll gelesen und dann die betroffenen sensorischen Regionen zu einem erneuten gemeinsamen Oszillieren angeregt werden.³ Das bedeutet, daß Merkmale dort gespeichert werden, wo sie zuerst erarbeitet wurden, daß sie also nicht an irgendeiner Stelle zusammengefaßt aufbewahrt werden. Das bedeutet auch, daß Konzepte immer nur dann existieren, wenn ihre Merkmale aktiviert werden. Es ist überdies anzunehmen, daß von Mal zu Mal unterschiedliche Merkmale ein und des-

selben Konzeptes aktiviert werden. Außerdem können neue Merkmale zu einem Konzept hinzugefügt und alte wieder vergessen werden.

Nach solchen Konzeptionen ist das Gedächtnis lediglich eine Menge von (aus Merkmalen aufgebauten) Konzepten. In der künstlichen Intelligenz, sozusagen einer computerisierten Schwester der kognitiven Psychologie, sind jedoch Modelle entwickelt worden, die auch Beziehungen zwischen mentalen Einheiten vorsehen, die also auf der Idee eines netzartigen Gebildes aus einer Menge von Einheiten und einer Menge von Beziehungen zwischen diesen Einheiten beruhen. Zu diesen Modellen zählen vor allem die semantischen Netze (vgl. Simmons 1973) und die Frame- oder Script-Modelle, die stereotype Situationen wie Restaurantbesuche oder Kindergeburtstage repräsentieren (vgl. Metzger 1980).

Semantische Netze wurden auch zur Modellierung der Sätzen zugrunde liegenden mentalen Strukturen herangezogen. Bei der Modellierung dieser Strukturen konkurrieren jedoch die semantischen Netze mit propositionalen Modellen (Anderson und Bower 1973). Propositionale Modelle lehnen sich an die Prädikatenlogik an und fassen daher die Bedeutung von Sätzen als Prädikat-Argument-Strukturen. Van Dijk und Kintsch (1983) haben die propositionalen Modelle sogar zu Modellen der Textrepräsentation ausgebaut. In ihren Augen liegen Texten bestimmte Makrostrukturen von Propositionen zugrunde.

Ein ganz offensichtliches Problem für propositionale Modelle ist jedoch, daß dem Verb eines Satzes je nach den substantivischen Satzgliedern, die der Satz enthält, verschiedene Merkmale entsprechen. Um ein einfaches Beispiel zu bringen: Der Satz

- (1) Der Gastwirt stellt das Glas in den Schrank.

manifestiert ein Merkmal 'stellen' des *Gastwirt* entsprechenden Konzeptes, jedoch darüber hinaus auch ein Merkmal 'gestellt werden' des *Glas* entsprechenden Konzeptes und ein Merkmal 'Ziel von Stellen sein' des *Schrank* entsprechenden Konzeptes. Dieser Sachverhalt wird von einer propositionalen Darstellung verdunkelt, wenn nicht verkannt. In einer passivischen Formulierung kommt das Merkmal 'gestellt werden' deutlich zum Ausdruck:

- (2) Das Glas wird in den Schrank gestellt.

Beziehungen wie die zwischen 'stellen' und 'gestellt werden' werden konvers genannt; konvers ist aber auch eine Beziehung wie die zwischen 'stellen' und 'Ziel von Stellen sein'. Als Gedächtnis-Beziehungen beruhen konverse Beziehungen darauf, daß derselbe wahrgenommene Umweltaspekt, dasselbe Perzept aus der Perspektive mehrerer verschiedener Konzepte verarbeitet und entsprechend in mehrere verschiedene Merkmale umgewandelt werden kann, die dann eben meist auch zu verschiedenen Konzepten gehören (vgl. Figge 1999: 32-36).

Es sei festgehalten:

- Wörter bezeichnen entweder Konzepte oder Merkmale.⁴
- Einfache Sätze — solche, die nur eine einzige Verbform enthalten — bezeichnen Gebilde aus Merkmalen und konversen Beziehungen zwischen diesen Merkmalen sowie die Konzepte, zu denen diese Merkmale gehören. Eine solche Struktur aus Merkmalen, konversen Beziehungen und Konzepten nenne ich konverses Netz.
- Satzgefüge bezeichnen mehrere konverse Netze und beispielsweise temporale oder kausale Beziehungen, die zwischen diesen konversen Netzen bestehen. Sie bezeichnen also Gebilde aus konversen Netzen und den zwischen ihnen bestehenden Beziehungen.
- Das tun auch Texte. Nur ist die Zahl der konversen Netze, die Texten zugrunde liegen, normalerweise deutlich größer als die Zahl der Satzgefüge zugrunde liegenden konversen Netze. Daher sind die Gebilde, die sich in Texten manifestieren, normalerweise auch deutlich komplexer.

Hinzuzufügen ist:

- Satzglieder sind etwa verbale oder substantivische Satzglieder. Die verbalen Satzglieder machen den Satz zum Satz. Die substantivischen Satzglieder bezeichnen Konzepte.

3. Lexikographie

3.1 Substantivisches Satzglied

Ein substantivisches Satzglied enthält ein Substantiv, zu dem eine Determinante, ein Zahlwort und ein oder mehrere Attribute treten können. Es gibt verschiedene Arten von Attributen, darunter vor allem das adjektivische Attribut, auf das ich mich hier beschränken will.

Doch zunächst zu den Substantiven — Ausdrücken für Konzepte, also für Mengen von Merkmalen.⁵ Die Begriffe Merkmal und Konzept lassen sich in die Nähe eines in der Lexikographie gängigen Begriffspaars stellen, nämlich des Begriffspaars *differentia specifica* und *genus proximum*. Durch Angabe einer *differentia specifica* und eines *genus proximum* werden Bedeutungen von Wörtern beschrieben. Ein *genus proximum* läßt sich als ein Konzept betrachten, dem gedanklich das Merkmal hinzuzufügen ist, als das sich die *differentia specifica* betrachten läßt. Das Konzept, das auf diese Weise entsteht, ist die Bedeutung, die beschrieben werden soll.

Dieses Verfahren der Bedeutungsbeschreibung ist, jedenfalls so, wie es normalerweise betrieben wird, aus der Sicht kognitiver Semantik nicht unproblematisch. Zum einen deshalb, weil es pro Wort (genauer: pro Lesart eines Wortes) durchweg nur ein einziges *genus proximum* vorsieht. Das ist allerdings nur zu gut verständlich, weil nach der aristotelischen Tradition ein Begriff stets eindeutig unter einen einzigen unmittelbaren Oberbegriff fällt. So darf als all-

gemein akzeptiert gelten, daß der Begriff 'Klavier' sich nicht mehr als einem Oberbegriff unterordnet, nämlich dem Begriff 'Musikinstrument'. Daher wird auch die Bedeutung des Wortes *Klavier* normalerweise mit 'Tasteninstrument' oder ähnlich angegeben. Das *genus proximum* ist in diesem Falle 'Musikinstrument', die *differentia specifica* 'mit Tasten'. Nun konzipiert man das Klavier aber nicht immer nur als Tasteninstrument, sondern etwa auch als Einrichtungsgegenstand oder als relativ schweren Gegenstand oder als Gegenstand mit einer plattenartigen Oberfläche oder als Gegenstand mit spezifischen Farben. Merkmale wie die durch diese Beispiele angedeuteten bleiben aber normalerweise außerhalb der lexikographischen Aufmerksamkeit, weil die Fixierung auf ein bestimmtes *genus proximum* den Blick auf sie verstellt.

Es wird allerdings auch argumentiert, daß solchen Merkmale lexikographisch überhaupt nicht Rechnung zu tragen sei. Der Hintergrund einer solchen Argumentation ist eine Unterscheidung zwischen semantischem — d.h. lexikographisch relevantem — Wissen und enzyklopädischem — d.h. lexikographisch irrelevantem — Wissen. So hebt etwa Wotjak "Bedeutung", die er als "kognitive Entität sui generis" betrachtet, und "enzyklopädische Sachwissenrepräsentation", bei der es "um sozialisiertes, usualisiertes Wissen, also bspw. um geteiltes Alltagswissen oder auch geteiltes definiertes Fachwissen" geht, gegeneinander ab (1993: 126; ohne Hervorhebungen zitiert).⁶ Nun wird sich aber einerseits kaum von der Hand weisen lassen, daß auch die Kategorisierung von Klavieren als Musikinstrumenten (mit Tasten) "geteiltes Alltagswissen", also enzyklopädisches Wissen ist. Und auch in einem spezifischen Sinne des Wortes ist diese Kategorisierung enzyklopädisch, denn es wird sich kaum ein Enzyklopädie finden lassen, die im Artikel *Klavier* nicht gleich zu Anfang ausführt, daß das Klavier ein Tasteninstrument sei. Umgekehrt sind die als "enzyklopädisch" eingeschätzten Merkmale semantisch hochrelevant, weil sie das Verständnis von Äußerungen steuern, wie folgende Beispiele zeigen:

- (3) Wo früher der Schrank stand, steht jetzt das Klavier. ('Einrichtungsgegenstand')
- (4) Die Packer konnten das Klavier nur zu dritt das Treppenhaus hinauftragen. ('schwerer Gegenstand')
- (5) Auf unserem Klavier steht eine Beethoven-Büste. ('Gegenstand mit einer plattenartigen Oberfläche')
- (6) Das Klavier soll farblich zu unserem Wohnzimmer passen. ('Gegenstand mit spezifischen Farben')

Daß in unterschiedlichen Kontexten (und natürlich auch in unterschiedlichen Situationen) verschiedene Merkmale desselben Konzepts aktiviert werden, ist ein Phänomen, das mit dem Namen "semantische Flexibilität" bedacht wurde; es ist zum ersten Mal systematisch von Barclay et al. (1974) untersucht worden. Mein Fazit ist jedenfalls, daß die Unterscheidung zwischen semantischen und enzyklopädischen Merkmalen konfus ist und deshalb vergessen werden sollte

(vgl. auch Kintsch 1980: 602 f., 612). Dies gilt allerdings nur unter der eingangs genannten Prämisse, daß Wörterbücher die Aufgabe haben, zu einer exhaustiven Beschreibung von Sprache beizutragen. Die Identifikation von Merkmalen, die nicht mit dem jeweils sanktionierten *genus proximum* zusammenhängen, kann durch eine semantische Analyse der Kontexte geschehen, in der die zu beschreibenden Wörter jeweils vorkommen (vgl. Figge und Job 1987).

Ein weiteres Problem der Bedeutungsbeschreibung durch Angabe von *differentia specifica* und *genus proximum* ergibt sich daraus, daß bestimmte *differentiae specifica* zwar Einheiten entsprechen, die auch allgemein in der Linguistik als Merkmale gelten, die aber keine Merkmale in einem kognitiven Sinne sind. Dazu gehören Angaben wie 'belebt', 'menschlich', 'weiblich', 'männlich', 'erwachsen'. So wird die Bedeutung des Wortes *Frau* in Wörterbüchern gern mit 'erwachsener weiblicher Mensch' umschrieben und die des Wortes *Mann* mit 'erwachsener männlicher Mensch'. Diese beiden Wörter setzen sich gegen die Wörter *Kind* und *Jugendliche(r)* ab, bei deren semantischer Beschreibung die *differentia specifica* 'nicht-erwachsen' etwa so aufgeteilt wird: 'im Alter von der Geburt bis zur Geschlechtsreife', 'im Alter von der Geschlechtsreife bis zum Ende des 18. Lebensjahres'. Nun wird jedoch wohl niemand gern annehmen, daß sich ein Konzept wie 'männlicher Mensch' von einem Konzept wie 'weiblicher Mensch' und ein Konzept wie 'erwachsener Mensch' von einem Konzept wie 'nicht-erwachsener Mensch' jeweils nur in einem einzigen Merkmal unterscheiden. Wir wissen etwa, daß Männer und Frauen sich durch anatomische und physiologische Eigenschaften unterscheiden oder daß Männer normalerweise andere Vornamen haben als Frauen. Wir wissen etwa, daß Kinder sich normalerweise in Körpergröße und -gewicht oder in der Rechtsstellung von Erwachsenen unterscheiden. Offensichtlich verbergen sich hinter Angaben wie 'weiblich', 'männlich', 'erwachsen' und auch hinter den negierten Entsprechungen solcher Angaben ganze Komplexe von Merkmalen. Deutlicher als bei der Bedeutungsbeschreibung eines Wortes wie *Klavier* spielt bei Wörtern wie *Mann* und *Frau* der Unterschied zwischen Merkmalen eine Rolle, die sich auch im Kopf eines — sagen wir — Standardindividuums finden, und solchen, die sich nur im Kopf von Fachleuten finden, etwa, was das Wort *Klavier* betrifft, im Kopf von Klavierbauern und, was die Wörter *Mann* und *Frau* betrifft, im Kopf von Neuropsychologen. Dem wird aber durch eine Unterscheidung zwischen allgemeiner und Fach-Lexikographie Rechnung getragen. Die allgemeine Lexikographie legt, mehr oder weniger bewußt, das Wissen eines Standardindividuums zugrunde. Was nun aber gravierender ist: Konzepte wie 'Mann', 'Frau', 'Kind' können in den Gedächtnissen einer gesellschaftlichen Gruppe Merkmale enthalten, die mit Merkmalen, die sie in den Gedächtnissen einer anderen gesellschaftlichen Gruppe enthalten, nicht kompatibel sind. Man mag sagen, daß es sich dann nicht um Wissen, sondern um Meinungen handelt. Eine feministische Sicht auf den Unterschied zwischen Mann und Frau ist eine andere als eine nicht-feministische Sicht. Über die Frage, wie Kinder zu erziehen sind, streiten sich verschiedene pädagogische Schulen. Dem kann aus

praktischen Gründen kaum durch eine systematische gruppenspezifische Lexikographie — ökologische Lexikographie, liberale Lexikographie, multikulturelle Lexikographie, um drei Beispiele zu nennen — Rechnung getragen werden, obwohl es Ansätze dazu gibt oder gab — etwa eine sozialistische oder eine katholische Lexikographie — oder wahrscheinlich geben wird — eine feministische Lexikographie auf der Grundlage feministischer Linguistik. Die Verfasser allgemeiner Wörterbücher werden sich aber auf jeden Fall fragen müssen, welchen Meinungen oder Ideologien das Standardindividuum anhängt, dessen Wissen sie zu Bedeutungsbeschreibungen heranziehen, was einschließt, daß sie sich fragen müssen, welchen Meinungen oder Ideologien sie selber anhängen, also — um ein aktuelles Beispiel noch einmal zu bemühen — ob sie feministisch eingestellt sind oder nicht.

Attribute manifestieren Merkmale aus dem Konzept, das hinter dem Substantiv steht, zu dem sie gehören. Für bestimmte Klassen von Adjektiven, zu denen etwa die Klasse der Dimensionsadjektive zählt, ist dem schon seit langem durch die Bezeichnung "relationale Adjektive" oder "relative Adjektive" Rechnung getragen worden (vgl. etwa Bierwisch 1970: 172-175). In der Tat hat das Adjektiv *hoch* in *ein hoher Turm* eine andere Bedeutung als in *ein hoher Berg*. Obwohl es hier auf noch mehr Kontexte ankäme, kann man doch sagen, daß die vertikale Ausdehnung, die man sich bei der Verarbeitung des zweiten substantivischen Satzglieds vorstellt, deutlich größer ist als die, die man sich bei der Verarbeitung des ersten vorstellt. Im Prinzip sind aber alle Adjektive auf ihre Relativität hin zu untersuchen. Relativ sind beispielsweise auch Farbadjektive: In *weiße Haare*, *weiße Hautfarbe*, *weißer Flieder* kommen jeweils leicht verschiedene Farbmerkmale zum Ausdruck. Daß hängt damit zusammen, daß es auf der Erde keine Oberflächen mit hundertprozentigem Reflexionsgrad gibt, so daß dem Auge nur annähernd weiße Oberflächen, also nur solche mit irgendeinem Farbanteil, dargeboten werden. Für die Lexikographie bedeutet dies, daß sie bei der Bedeutungsbeschreibung von Substantiven das Merkmal präzise angeben sollte, das zum Ausdruck kommt, wenn dem Substantiv ein bestimmtes Adjektiv attribuiert wird. So umfaßt unser Konzept von einer (Zimmer-)Decke in einem neueren Haus das Merkmal, daß sie etwa 2,6 Meter über dem Fußboden liegt. Hat eine Decke das Merkmal, etwa 3 Meter oder mehr über dem Fußboden zu liegen, dann kann dem Substantiv *Decke* das Adjektiv *hoch* attribuiert werden. Das sollte in der semantischen Beschreibung von *Decke* zur Geltung kommen. Außerdem müßte es einen reziproken Verweis zwischen dem Substantiv-Eintrag und dem Adjektiv-Eintrag geben.

Einfache Sätze — solche, wie gesagt, die nur eine einzige Verbform enthalten — manifestieren ein konverses Netz von Merkmalen und Konzepten. Der bereits zitierte Satz

- (1) Der Gastwirt stellt das Glas in den Schrank.

ist insofern die Standardmanifestation eines konversen Netzes, als die verschiedenen syntaktischen Funktionen seiner Satzglieder in typischer Weise eine

Dynamizitätsabstufung der zugrunde liegenden Merkmale widerspiegeln. Im Subjekt erscheint das Konzept, dessen Merkmal ('stellen') einen hohen Grad an Dynamik repräsentiert, im direkten Objekt das Konzept, dessen Merkmal ('gestellt werden') einen niedrigen Grad an Dynamik repräsentiert, weil es sich um fremdbewirkte Dynamik handelt, und in der lokalen Ergänzung das Konzept, dessen Merkmal ('Ziel von Stellen sein') überhaupt keine Dynamik repräsentiert. Zunächst sei kurz angemerkt, daß Lexikographen eine gewisse Tendenz haben, ihr Augenmerk eher auf das Subjekt zu richten und daher einen Satz wie (1) als Beispiel eher nur unter dem Eintrag des Subjekt-Wortes als unter dem des Objekt-Wortes oder dem des Ergänzungs-Wortes aufzuführen. Wörter wie *Glas* oder *Schrank* kommen aber nun typischerweise seltener im Subjekt als in anderen Satzgliedern vor, so daß der zitierte Satz durchaus repräsentativ für die Verwendung dieser Wörter ist.

Sodann sei aber auf den Sachverhalt hingewiesen, daß im Subjekt auch ein Konzept auftauchen kann, das mit einem weniger dynamischen oder gar undynamischen Merkmal ausgestattet ist:

- (7) Das Glas wird in den Schrank gestellt.

Gleichzeitig tritt eine passive Verbform auf. Im Deutschen und auch in anderen Sprachen mag die Passivbildung so regelmäßig sein, daß ihr lexikographisch nicht unbedingt systematisch Rechnung getragen werden muß. Doch sind Sätze mit passiven Verbformen vielfach gute Belege für die Verwendung der Wörter, die ihr Subjekt bilden:

- (8) Der Briefkasten wird einmal täglich geleert.
(9) Die Schauspieler wurden ausgebuht.
(10) Das Kind wird am Sonntag getauft.

Diese Bemerkungen betreffen den Zusammenhang zwischen Satzsyntax und lexikographischer Beschreibung von Substantiven.

3.2 Verb

Von lexikographischem Belang ist aber auch ein Sachverhalt, der die Beschreibung von Verben betrifft, nämlich der, daß konverse Netze auch in Verbformen in Erscheinung treten können, die zwar eine passivische Funktion haben, aber von geringerer syntaktischer Regelmäßigkeit sind als die regulären Passivformen und die deshalb im Wörterbuch systematisch berücksichtigt werden müssen. Dazu gehören beispielsweise quasi-passivische Verwendungen von Verbalgefügen mit *bekommen* (umgangssprachlicher *kriegen*) im Deutschen oder *se voir* im Französischen:

- (11) Ich habe den Führerschein entzogen bekommen (gekriegt).

- (12) Je me suis vu retirer mon permis de conduire.

Die Formen *entzogen bekommen (kriegen)* und *se voir retirer* müssen auf jeden Fall in einem deutschen Wörterbuch unter *entziehen* bzw. in einem französischen Wörterbuch unter *retirer* aufgeführt und exemplifiziert werden. Es gibt darüber hinaus auch die Verwendung aktiver Verbformen in passivischer Funktion, und das ist noch deutlicher ein Fall für die Lexikographie, weil es sich um stark idiosynkratische Verwendungen handelt. Statt

- (7) Das Glas wird in den Schrank gestellt.

kann auch gesagt werden

- (13) Das Glas kommt in den Schrank.

Ein besonders lebensnahes Beispiel ist der Dialog

- (14) Wohin kommt das Glas? — Das kommt in den Schrank.

In dieser Verwendung ist *kommen* eine Art Suppletivform von *stellen* und muß daher unter *stellen* vermerkt werden. Oder es muß, da *kommen* ein Äquivalent für die Passivformen einer ganzen Reihe von Verben der Verursachung von Bewegung ist, diese seine Verwendung unter seinem eigenen Eintrag exhaustiv beschrieben und dann systematisch zwischen diesem Eintrag und den Einträgen der jeweiligen Verben verwiesen werden. Eine solche passivartige Funktion ist keineswegs nur dem Verb *kommen* eigen, wie etwa die Verwendung von *gehen* in einer möglichen Fortsetzung des gerade angeführten Dialogs zeigt:

- (15) Da geht es aber nicht hinein.

Die Beziehung zwischen *stellen* und *gehen* ist allerdings komplizierter, wie die Umschreibung

- (16) Da kann es aber wegen unzureichenden Platzes nicht hineingestellt werden.

zeigt. Es sei nebenbei bemerkt, daß kein deutsches Wörterbuch eine angemessene Beschreibung dieser Verwendungen von *kommen* und *gehen* gibt.

Aktivische Passiväquivalente können auch komplex sein. So läßt sich statt

- (17) Today, gangsters robbed a bank in Fleet Street of one million pounds.

auch sagen:

- (18) A bank lost one million pounds in a robbery.

- (19) One million pounds was lost in a robbery.
(20) Today, Fleet Street was the scene of a bank robbery.

Daß *to rob, to lose in a robbery, to be lost in a robbery, to be the scene of a robbery* Formen desselben Verbs sind, muß lexikographisch ganz deutlich werden.

3.3 Satzgefüge

Satzgefüge bestehen aus mindestens zwei einfachen Sätzen, die koordinativ oder subordinativ miteinander verknüpft sind. Eine solche Verknüpfung manifestiert eine beispielsweise temporale oder kausale Beziehung zwischen Merkmalen. Die Art der Beziehung kann durch eine Konjunktion verdeutlicht werden. Das ist der erste Punkt, in dem die Bildung von Satzgefügen lexikographisch relevant wird, denn Konjunktionen pflegen in Wörterbüchern verzeichnet zu werden. Dies allerdings häufig stiefmütterlich. Soll das nicht der Fall sein, dann muß die Lexikographie die Grenze zur Grammatik überschreiten. Welche Beziehungen eine Konjunktion im einzelnen bezeichnet, hängt nämlich sowohl davon ab, in welchem Tempus und in welchem Modus das Verb im übergeordneten und das Verb im untergeordneten Satz stehen, als auch davon, ob der untergeordnete Satz dem übergeordneten vorausgeht oder folgt. Ich zeige das ganz kurz an der temporalen Konjunktion *als*. Die zugrunde liegende Beziehung kann sowohl eine solche der Gleichzeitigkeit als auch eine solche der Vorzeitigkeit sein. Ist sie eine Beziehung der Gleichzeitigkeit, dann kann es sich um eine punktuelle oder um eine einrahmende Gleichzeitigkeit handeln:

- (21) Ich ging in das Haus, als Peter gerade herauskam.
(22) Als wir bei Tisch saßen, klingelte das Telefon.

Ist sie eine Beziehung der Vorzeitigkeit, so kann es sich um eine unmittelbare oder um eine nicht-unmittelbare Vorzeitigkeit handeln:

- (23) Als ich das Fenster öffnete, strömte kalte Luft ins Zimmer.
(24) Als ich das Fenster geschlossen hatte, wurde es allmählich wieder warm.

Die semantische Relevanz des Stellungsunterschieds sei an dem Paar

- (25) Als es anfang zu regnen, waren wir auf dem Weg in die Stadt.
(26) Wir waren auf dem Weg in die Stadt, als es anfang zu regnen.

gezeigt. Der Satz (25) gibt tatsächlich nur an, daß das Merkmal 'unterwegs sein' als zeitlicher Rahmen für das (atmosphärische) Merkmal 'einsetzender Regen' konzipiert wird. Der Satz (26) läßt noch etwas Zusätzliches erkennen, nämlich daß das Merkmal 'unterwegs sein' auch einen situativen Hintergrund bildet, so daß seine Rezipienten nun weitere Informationen erwarten, nämlich Informa-

tionen über Ereignisse, die vor diesen Hintergrund passen. Jedenfalls lassen sich Konjunktionen wie *als* lexikographisch nicht angemessen beschreiben, wenn der Beschreibung nicht eine detaillierte syntaktische und semantische Analyse von Satzgefügen vorausgeht, in denen sie vorkommen.

Nebensätze können je nach Sprache eine finite, eine infinitivische, eine partizipiale oder eine gerundiale Verbform enthalten. Hier kommt es aber darauf an, daß ihr Verb auch die Gestalt eines Substantivs annehmen kann:

- (27) Nach Schließung des Fensters wurde es allmählich wieder warm.
- (28) Auf unserem Weg / während unseres Gangs in die Stadt fing es an zu regnen.

Nun ist es nicht so, daß man Wörter wie *Schließung*, *Weg* oder *Gang* in den deutschen Wörterbüchern nicht fände. In nach Nischen organisierten Wörterbüchern würde man *Schließung* sogar im Artikel *schließen* finden. Es fehlt jedoch an Sauberkeit in der Behandlung solcher Substantive. Das ist jedoch keine primäre Schuld der Lexikographie. Vielmehr wird der ungute Zustand dadurch heraufbeschworen, daß die Ableitung von Substantiven aus Verben und Adjektiven als eine Aufgabe der Wortbildungslehre betrachtet wird. Dadurch kommen aber die syntaktischen und die lexikologischen Aspekte der Verbal- und Adjektivsubstantivierung zu kurz. Der wichtigste syntaktische Aspekt besteht darin, daß solche Substantivierungen echte Verbformen sein können, eben weil sie in Konstruktionen eingehen, die als Nebensatz- und damit allgemein als Satzäquivalente gelten müssen. Der lexikologische Aspekt besteht weniger darin, daß es Verbal- oder Adjektivsubstantive gibt, die nicht in einer Wortbildungsbeziehung zu einem Verb oder Adjektiv stehen, wie etwa *Weg* zu *gehen* oder — ein prominentes Beispiel — franz. *cécité* 'Blindheit' zu *aveugle* 'blind' oder *Unfall*, das — wie auch engl./franz. *accident* — zu gar keinem Verb in Beziehung steht. Der lexikologische Aspekt besteht vielmehr hauptsächlich darin, daß Verbal- oder Adjektivsubstantive — soweit ich sehe — drei verschiedene Funktionen haben können: Sie können verbwertig sein — ich spreche dann gern von substantivischen Verbformen —, sie können echte Substantive sein, und sie können Elemente von mehrwortigen Verbalgefügen sein. Am Beispiel von *Gang*: In *während unseres Gangs in die Stadt* ist *Gang* eine substantivische Verbform, die sich auf Merkmale bezieht; in *eine Höhle mit mehreren Gängen* ist *Gang* ein echtes Substantiv, bezieht es sich also auf ein Konzept;⁷ schließlich kommt es in *in Gang bringen* als Teil eines Stützverbgefüges vor, das insgesamt ein Merkmal ausdrückt. Dies alles systematisch und exhaustiv darzustellen, ist eine wichtige Aufgabe der Lexikographie.⁸

3.4 Text

Die mentalen Gebilde, die sich in Texten manifestieren, können sehr komplex sein. Häufig wird ihnen jedoch eine Struktur aufgeprägt, die aus ihnen ein Ge-

bilde von Teilgebilden macht. Das schlägt sich in Texten durch Textgliederung nieder. Ein ganz augenfälliges Verfahren der Textgliederung ist die typographische Gliederung gedruckter Texte in Absätze, Kapitel usw. Es gibt aber auch ein lexikographisch relevantes Gliederungsverfahren, nämlich die Einfügung von Textgliederungssignalen in Texte. Textgliederungssignale sind Adverbien oder adverbiale Wendungen, die ihre semantische Wirkung nicht in dem Satz entfalten, dem sie syntaktisch zugehören, sondern eben in dem Text, in dem der Satz steht. Eine Exemplifizierung ist wegen der Länge der erforderlichen Zitate schwierig (ich verweise auf Figge 1999: 21-25). Im Deutschen handelt es um Adverbien wie *nun, aber, also, unterdessen* oder Wendungen wie *eines Tages*. Solche Textgliederungssignale kommen auch kombiniert vor. *Als X nun aber eines Tages ...* ist eine Formulierung, die in jedem Text, in dem sie vorkommt, eine stark gliedernde Wirkung hat, also den Übergang von einem mentalen Teilgebilde zu einem anderen mit aller Deutlichkeit anzeigt. Die Lexikographie sollte sich der Mühe unterziehen, auch die textgliedernde Funktion von Adverbien und Adverbialen dingfest zu machen.

4. Schlußbemerkungen

Aus den Vorschlägen, die ich vor einem kognitiv-linguistischen Hintergrund gemacht habe, ergibt sich für die lexikographische Praxis außer der Notwendigkeit einer Ausweitung und einer noch stärkeren Systematisierung von Analyse und Darstellung vor allem zweierlei. Zum einen sollten Beispiele und Belege so ausgewählt werden, daß sie sich ganz deutlich auf die unterschiedlichen semantischen Aspekte beziehen, die die Wörter jeweils aufweisen. Zum anderen sollte es eine Vielzahl systematischer Verweise zwischen den einzelnen Artikeln eines Wörterbuchs geben. Das läßt sich am besten mit einem computerisierten Wörterbuch erreichen, das in systematischer Weise auf Korpora zurückgreifen kann und das mit einem dichten Netz von Hyperlinks durchzogen ist.

Endnoten

1. S. auch Figge 1992, 1993, 1996.
2. Ein weiteres Charakteristikum der kognitiven Psychologie ist die Definition von kognitiven Vorgängen als Prozessen der Informationsverarbeitung (Lachman, Lachman und Butterfield 1979).
3. Vgl. etwa Damasio und Damasio 1989: insbes. 55-75, Damasio 1990, Damasio und Damasio 1992, Von der Malsburg und Schneider 1986, Singer 1993. Das Problem des Zusammenhangs zwischen Aktivationen in unterschiedlichen modalen Feldern heißt in der Neuropsychologie "Bindungs"-Problem.
4. Dies gilt allerdings vorwiegend nur für Substantive, Adjektive und Verben.

5. Aus Verben oder Adjektiven abgeleitete Substantive, die eine verbale oder adjektivische Bedeutung beibehalten haben, also die sogenannten Abstrakta, werden unten gesondert behandelt.
6. Vgl. auch Svensén (1993: 2): Ein Wörterbuch "can be said to distil off that part of human knowledge which concerns the language, whereas the encyclopedia deals with the rest".
7. Genauer: eine Menge von Konzepten.
8. Die vor verbwertigen Substantiven vorkommenden Präpositionen sind nach dem Verfahren zu beschreiben, das oben für Konjunktionen vorgeschlagen wurde. Präpositionen, die in einfachen Sätzen vorkommen, sind dagegen im Zusammenhang mit Verbvalenzen zu beschreiben.

Literatur

- Anderson, J.R. und G.H. Bower.** 1973. *Human Associative Memory*. Washington, D.C.: V.H. Winston and Sons.
- Baddeley, A.D.** 1999. *Essentials of Human Memory*. Hove, V.K.: Psychology Press.
- Barclay, J.R., J.D. Bransford, J.J. Franks, N.S. McCarrell und K. Nitsch.** 1974. Comprehension and Semantic Flexibility. *Journal of Verbal Learning and Verbal Behavior* 13: 471-481.
- Bierwisch, M.** 1970. Semantics. Lyons, J. (Hrsg.). 1970. *New Horizons in Linguistics*: 166-184. Harmondsworth, Middlesex, England: Penguin Books.
- Damasio, A.R.** 1990. Synchronous Activation in Multiple Cortical Regions: A Mechanism for Recall. *Seminars in the Neurosciences* 2: 287-296.
- Damasio, H. und A.R. Damasio.** 1989. *Lesion Analysis in Neuropsychology*. New York: Oxford University Press.
- Damasio, H. und A.R. Damasio.** 1992. Sprache und Gehirn. *Spektrum der Wissenschaft* 1992(11): 80-92.
- Figge, U.L.** 1992. Zwei Grundprinzipien der Wortschatzklassifikation. Goebel, H. und M. Schader (Hrsg.). 1992. *Datenanalyse, Klassifikation und Informationsverarbeitung. Methoden und Anwendungen in verschiedenen Fachgebieten*: 31-35. Heidelberg: Physica-Verlag.
- Figge, U.L.** 1993. Konzeptsystem als Grundlage für die Lexikographie. Beckmann, F. und G. Heyer (Hrsg.). 1993. *Theorie und Praxis des Lexikons*: 123-137. Berlin/New York: Walter de Gruyter.
- Figge, U.L.** 1996. Kognitiv orientierte Lexikographie. Weber, N. (Hrsg.) 1996. *Semantik, Lexikographie und Computeranwendungen*: 93-108. Tübingen: Max Niemeyer Verlag.
- Figge, U.L.** 1999. *Gedächtnis — Sprache — Text. Prinzipien und Exempla einer semiotischen Sprachtheorie*. Aachen: Shaker Verlag.
- Figge, U.L. und U. Job.** 1987. Das Gedächtnis in sprachwissenschaftlicher Perspektive. *Romanistisches Jahrbuch* 38: 16-34.
- Hörmann, H.** 1976. *Meinen und Verstehen. Grundzüge einer psychologischen Semantik*. Frankfurt/Main: Suhrkamp Verlag.
- Hörmann, H.** 1981. *Einführung in die Psycholinguistik*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Kintsch, W.** 1980. Semantic Memory. A Tutorial. Nickerson, R.S. (Hrsg.). 1980. *Attention and Performance VIII. 8th International Colloquium on Attention and Performance, Princeton, N.J., 1978, Proceedings*: 595-620. Hillsdale, N.J.: Lawrence Erlbaum.

- Lachman, R., J.L. Lachman und E.C. Butterfield.** 1979. *Cognitive Psychology and Information Processing: An Introduction*. Hillsdale, N.J.: Lawrence Erlbaum.
- Le Ny, J.-F.** 1979. *La sémantique psychologique*. Paris: Presses Universitaires de France.
- Metzing, D.** (Hrsg.). 1980. *Frame Conceptions and Text Understanding*. Berlin/New York: Walter de Gruyter.
- Neisser, U.** 1967. *Cognitive Psychology*. New York: Appleton, Century, Croft.
- Ogden, C.K.** 1942. *The General Basic English Dictionary*. New York: W.W. Norton.
- Quillian, M.R.** 1967. Word Concepts. A Theory and Simulation of Some Basic Semantic Capabilities. *Behavioral Science* 12: 410-430.
- Quillian, M.R.** 1968. Semantic Memory. Minsky, M. (Hrsg.). 1968. *Semantic Information Processing: 227-270*. Cambridge, MA./London: MIT Press.
- Simmons, R.F.** 1973. Semantic Networks: Their Computation and Use for Understanding English Sentences. Schank, R.C. und K.M. Colby (Hrsg.). 1973. *Computer Models of Thought and Language: 63-113*. San Francisco: Freeman.
- Singer, W.** 1993. Synchronization of Cortical Activity and its Putative Role in Information Processing and Learning. *Annual Review of Physiology* 55: 349-374.
- Smith, E.E., E.J. Shoben und L.J. Rips.** 1974. Structure and Process in Semantic Memory. A Featural Model for Semantic Decisions. *Psychological Review* 81: 214-241.
- Svensén, B.** 1993. *Practical Lexicography. Principles and Methods of Dictionary-Making*. Oxford/New York: Oxford University Press.
- Treisman, A.** 1986. Properties, Parts, and Objects. Boff, K.R., L. Kaufman und J.P. Thomas (Hrsg.). 1986. *Handbook of Perception and Performance. Volume II: Cognitive Processes and Performance: 35-1 – 35-70*. New York: John Wiley.
- Treisman, A.** 1987. Merkmale und Gegenstände in der visuellen Verarbeitung. *Spektrum der Wissenschaft* 1987(1): 72-82.
- Van Dijk, T.A. und W. Kintsch.** 1983. *Strategies of Discourse Comprehension*. New York/London: Academic Press.
- Von der Malsburg, C. und W. Schneider.** 1986. A Neural Cocktail-Party Processor. *Biological Cybernetics* 54: 29-40.
- Wotjak, G.** 1993. Semantische Makrostrukturbeschreibung (lexikalisch-semantische Felder) und (enzyklopädische) Wissensrepräsentationen. Lutzeier, P.R. (Hrsg.). 1993. *Studien zur Wortfeldtheorie/Studies in Lexical Field Theory: 121-136*. Tübingen: Max Niemeyer.

The Use of an Improved Access Structure in Dictionaries*

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

Abstract: Dictionary consultation procedures often fail because the structure of a dictionary impedes the user from sufficient access to the data. Modern-day metalexigraphy has identified the access structure as one of the most important structural components of a dictionary. The access structure represents the search route a user follows to reach the required data in order to retrieve the necessary information. This paper focuses on a few aspects of the access structure. It is argued that a frame structure and an improved use of outer texts can enhance access to the lexicographic data. Different macrostructural ordering procedures are discussed and it is suggested that the lexicographers of bilingual dictionaries should avoid multiple niching. The presentation of inner texts in clearly identifiable text blocks can assist the user in rapidly reaching the desired search zone. Although the emphasis is on bilingual dictionaries many of the suggestions also apply to monolingual dictionaries.

Keywords: METALEXICOGRAPHY, LEXICOGRAPHY, DICTIONARY, BILINGUAL, ACCESS STRUCTURE, FRAME STRUCTURE, MULTIPLE NICHING, NESTING, INNER TEXTS, OUTER TEXTS

Opsomming: Die gebruik van 'n verbeterde toegangstruktuur in woordeboeke. Naslaanpogings in woordeboeke misluk dikwels omdat die woordeboekstruktuur die gebruiker van doeltreffende toegang tot die data weerhou. In die hedendaagse metaleksikografie word die toegangstruktuur as een van die belangrikste strukturele komponente van 'n woordeboek beskou. Die toegangstruktuur verteenwoordig die soekroete wat 'n gebruiker volg om die verlangde data te bereik en die nodige inligting te ontsluit. Hierdie artikel fokus op enkele aspekte van die toegangstruktuur. Daar word aangetoon dat 'n raamstruktuur en 'n verbeterde gebruik van buitetekste die toegang tot leksikografiese data kan verbeter. Verskeie makrostrukturele ordeningsmetodes word bespreek en daar word voorgestel dat leksikograwe van vertalende woordeboeke moet probeer om veelvuldige nesting te vermy. Die plasing van binnetekste in duidelik identifiseerbare teksblokke, kan die gebruiker help om 'n verlangde soeksone vinnig te bereik. Alhoewel die klem op tweetalige woordeboeke is, geld baie van die opmerkings ook eentalige woordeboeke.

Slutelwoorde: METALEKSIKOGRAFIE, LEKSIKOGRAFIE, WOORDEBOEK, TWE

* This article is a slightly adapted version of a paper presented at the Second International Conference of the Asian Association for Lexicography, held at Yonsei University, Seoul, 8-10 August 2001.

TALIG, TOEGANGSTRUKTUUR, RAAMSTRUKTUUR, VEELVULDIGE NISTING, NESTING, BINNETEKSTE, BUITETEKSTE

1. Introduction

The success of a bilingual dictionary depends largely on the success, which the typical target user achieves during a normal dictionary consultation procedure. Too often these procedures fail because, although bilingual dictionaries contain all the relevant data, the structure of the dictionary and the presentation of the data is of such a nature that the target user cannot find the data or retrieve the needed information. 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.

Recent research in the field of metalexigraphy has focused on the structure of dictionaries. One of the structural components identified in this research (cf. Hausmann and Wiegand 1989), is the access structure. The access structure can be regarded as the search route a dictionary user follows during a dictionary consultation procedure. A distinction is made between the outer access structure, which leads a user up to the lemma sign introducing the article from which the needed information is to be retrieved, and the inner access structure which guides the user within a dictionary article to the search zone in which the relevant data is presented. In this paper the importance of both the outer and the inner access structures will be emphasised. However, the need will also be stressed for a more sophisticated access structure and for the use of outer texts to enhance the rapid and successful access to the data presented in dictionaries.

2. The frame structure

A user-driven approach in lexicography compels the lexicographer to use all possible means at his/her disposal to improve the quality of dictionary use and to ensure an optimal retrieval of information. Successful information retrieval can only be achieved if the data is accessible. Accessibility of data is not only determined by the use of a well-defined access structure leading the user to a venue within the word list structure of the dictionary but it is also determined by a functional positioning of the data within the word book structure of the dictionary and by employing different texts in the dictionary to explain and explicate the contents of that dictionary (cf. Hausmann and Wiegand 1989: 330, 333).

Utilising a word book approach a dictionary can display a so-called *frame structure* (cf. Kammerer and Wiegand 1998). A dictionary is regarded as a carrier of different text types and the different texts and text types are accommodated in various positions within the dictionary. According to the frame structure approach a dictionary contains three textual focal points, i.e. the front

matter, the central list and the back matter. The central list contains the article stretches, i.e. the presentation of the articles grouped according to the alphabetical ordering system in terms of the different alphabet letters A-Z, in which all the articles are accommodated. This is the section of a dictionary which is the most typical venue for general dictionary consultation procedures. Consequently this textual component has to display a structure which allows the target user quick and unimpeded access to the required data. The outer access structure will guide a user to a specific article stretch and within that article stretch to a specific lemma sign. From there the inner access structure leads the user via an article internal route to the relevant search zone and the specific data type. The front matter section contains all the texts preceding the central list and the back matter section contains all the texts following the central list. The texts in the front and back matter of a specific dictionary constitute the outer texts of that dictionary. Neither the front nor the back matter are functional parts of a dictionary, although both may contain individual texts which have specific functions in the dictionary (cf. Gouws 1999). A frame structure qualifies a dictionary as a text type carrier or a "big text" because the front matter, central list and back matter contain a variety of text types.

The dictionary conceptualisation plan of any new lexicographic endeavour has to make provision for a data distribution programme according to which the data which the lexicographer wishes to include can be distributed to all the focal points of the planned dictionary. Its application results in the introduction of a data distribution structure. Traditional dictionaries have often utilised the front and back matter to present outer texts. However, in many dictionaries these were not functional texts and contributed very little to the genuine purpose of the dictionary.

Due to the functional use of both the central list and the outer texts a dictionary can assist with a much more comprehensive transfer of data than before, e.g. the inclusion of various lists of words in the back matter. The inclusion and distribution of more data in different structural focal areas of a dictionary necessitate clear and unambiguous access to the data. The access structure of a dictionary has to be enhanced to ensure the success of a dictionary consultation procedure performed by a typical target user of the specific dictionary. Access structures are not only needed to guide the user to the article stretches in the central list but also to assist the user in utilising and retrieving information from the texts included in the front and back matter components.

2.1 Outer texts

According to Hausmann and Wiegand (1989: 331) each dictionary should contain two compulsory texts, i.e. the central list which presents the lemmata as macrostructural entries and primary treatment units and an outer text, usually presented in the front matter, with the users' guidelines. In this outer text the editorial system of the dictionary has to be explained in order to familiarise the

target user with the structure and contents of the dictionary and with the presentation of data. This text already has an access function because it not only informs the user of the data collection in the dictionary but also explicates the positioning and arrangement of data types. This allows the user a more rapid dictionary consultation procedure. A typical component of such a users' guidelines text is a sample article taken from the central list to illustrate the default article structure, microarchitecture and search zones. With such an outer text the access possibilities surpass those of the traditional outer and inner access structure.

The notion of a frame structure, as developed by Kammerer and Wiegand (1998), creates the opportunity for a more comprehensive application of front and back matter texts (cf. Gouws 2001), which further improves the accessibility of dictionaries. Frame structures of different levels can be identified. According to Kammerer and Wiegand (1998) the front and back matter texts are not immediate constituents of the front or back matter respectively but rather of the dictionary as a carrier of texts. A front matter text functioning as an immediate constituent of the dictionary as a whole, e.g. a users' guidelines text or a front matter text presenting the table of contents of the dictionary, can be regarded as a primary outer text. Bergenholtz, Tarp and Wiegand (1999: 1777-1778) argue convincingly in favour of a distinction between integrated and nonintegrated outer texts. An integrated outer text is one integrated into and helping to achieve the genuine purpose of the specific dictionary. It quite often has a direct link with the central list, e.g. a dictionary displaying both single and synopsis articles can assist the user by listing the lemmata heading the synopsis articles in a back matter text. Such a text explicates the heterogeneous article structure and creates another access possibility for the user. In this way integrated outer texts, both in the front and the back matter, can be employed to create polyaccessible dictionaries. In a bilingual dictionary the treatment of lemmata representing culturally bound lexical items could best be done in synopsis articles which allows a more comprehensive treatment than that presented in default articles. A back matter text listing all the culturally bound items treated in the central list in synopsis articles enhances the access to these items in the central list.

The relevance and functional value of an outer text does not depend on its integration into the genuine purpose of the dictionary. Gouws (forthcoming) indicates that an outer text can have its own genuine purpose which does not necessarily correspond with the genuine purpose of the dictionary or its integrated outer texts. These unintegrated outer texts can play an important role in the overall transfer of data in a lexicographic text and their selection and presentation may not be done in an arbitrary way. The data distribution structure of a dictionary determines the type of data to be presented in outer texts and the target user of a dictionary needs to be made aware of their contents and gain access to the data in these texts. The average dictionary user is often not in the habit of utilising the outer texts of a dictionary. The functional value and

relevance of outer texts should not only be brought to the attention of the target users but lexicographers should endeavour to promote optimal access to these texts. This demands an innovative approach towards the access structure. The access structure should not be biased by being directed only at the central list. As an integral component of the word book structure, it should establish a search route to all the venues identified by the data distribution programme for the accommodation of lexicographic data. In this regard the table of contents plays an important role in giving a systematic account of the different texts included in the dictionary.

A text presenting the table of contents has a wide scope because it is directed at the dictionary as a carrier of text types and gives information regarding the texts functioning within the different focal areas constituting the word book structure. According to Hausmann and Wiegand (1989: 333) a table of contents functions as a special type of access structure, namely a register-external outer access structure, i.e. an outer access structure component which is not presented as part of the central list. Where the typical outer access structure guides a user to a lemma sign, the table of contents includes entries which guide the user to the different components of the frame structure and to the individual texts contained in these components of the dictionary as a big text. Such a table of contents is part of the primary frame structure of the dictionary and can be regarded as a primary outer text. The functional value of a table of contents which increases accessibility may never be underestimated. This text can assist the intended target user in a substantial way.

Accessibility to the outer texts can be achieved in different ways. One access procedure which has not yet been employed enough is the utilisation of secondary outer texts. Many dictionaries contain primary outer texts with their own outer texts, e.g. an explanatory text, a table of contents or a users' guidelines text directed not at the dictionary as such but only at a specific outer text. These secondary outer texts are not immediate constituents of the dictionary as a big text but rather of the specific outer texts. A primary outer text with its own front and back matter texts constitutes a secondary frame structure (cf. Gouws 2001: 527). A major function of secondary outer texts is to provide access to their primary outer texts and to increase accessibility to data accommodated outside the central list of a dictionary.

3. The external access structure: macrostructural procedures

All dictionaries do not display the same macrostructural ordering procedures. A typical distinction, occurring in dictionaries with an alphabetical ordering, is that between a straight alphabetical ordering and an ordering presenting a sinuous lemma file (cf. Wiegand 1989, Hausmann and Wiegand 1989). These different conventions result in different types of lemmata (cf. Wolski 1989). A dictionary with a straight alphabetical ordering typically displays main lemmata whereas a dictionary with a sinuous lemma file has sublemmata comple-

menting the main lemmata. Sublemmata are ordered in either lemma niches or lemma nests. In monolingual explanatory dictionaries sublemmata often are the so-called "unexplained" complex lexical items. Their semantic transparency motivates their inclusion as sublemmata. This linguistic motivation is complemented by a pragmatic one, i.e. for space-saving purposes. Albeit that they are not allocated paraphrase of meaning entries, sublemmata can still convey certain semantic data, especially if they are entered as nested lemmata.

3.1 Niching and nesting

Niched lemmata display a straight alphabetical ordering and the niche maintains the alphabetical ordering in terms of the preceding and the following main lemmata. The deviation in ordering direction, i.e. horizontal as opposed to vertical, has no influence on the alphabetical ordering principle. This is clear from the following example, taken from the Afrikaans explanatory dictionary *Verklarende Afrikaanse Woordeboek*:

eergie rig. ...
eergister. Die dag voor gister; **eergisteraand**; **eergistermiddag**; **eergistermôre**; ...
eerherstel. ...

Nesting differs from niching in at least one but often two ways. Lemma nests deviate from the straight alphabetical ordering. Although a lemma nest sometimes displays an internal alphabetical ordering similar to a niche, the external ordering, i.e. the alphabetical positioning of the nested lemmata with regard to the preceding and following vertically ordered lemmata, does not display a straight alphabetical ordering. This kind of nesting may be referred to as *first level nesting*. Cf. the following example taken from the *Verklarende Handwoordeboek van die Afrikaanse Taal*:

di-a-mant' (-e) Harde, kosbare edelsteen, bestaande uit kristallyne koolstof: (On)geslypte diamante. [G. *adamas* staal; L. *diamas* onbedwingbaar; G. *damao* mak maak] **diamant**: ~agtig, ~bedryf, ~boor, ~delwer, ~gruis, ~houdend, ~kenner, ~ontginning, ~ring, ~vorming, ~winning.
di-a-mant'brui·lof

A straight alphabetical ordering would have ordered the lemma *diamantbrui·lof* between *diamantboor* and *diamantdelwer*. Internally a nest may also present a deviation from the alphabetical ordering. This is known as *second level nesting* and is usually motivated on morphosemantic grounds. Cf. the following example from *Nasionale Woordeboek*:

broei (ge-) ww. 1. *op eiers sit en hulle warm hou om hulle te laat uitkom*. 2. *voortkom, ontspruit*. Daaruit sal onheil -. 3. *peins, planne maak*. Oor iets -. 4. *ontwikkel, in wording wees*. Daar is iets aan die -. 5. *hitte ontwikkel, warm word*. Die mis, lug -. 6. *warm word deur gisting*. Die hooi -. 7. *deur spesiale verwarming vroeër laat bloei of ryp word*. 8. *warm kry*. In die son sit en -. 9. *kleintjies*

voortbring. Die jakkalse – in die lente. **'broeiery, broeiing; broeieend, -eiers, -gans, -hen, -hok, -kamer, -kolonie, -paar, -proses, -sak, -tent** (by 1); **-mis** (by 5); **-aarde, -bed** (by 6); **-bak, -glas, -huise** (by 7).
'broeiend (-e; -er, -ste) b.nw. 1. *drukkend, wat ...*

The first two lemmata in the nest are derivatives and the rest are compounds. The first semicolon is a structural marker separating members of the nest on morphological grounds. A partial alphabetical ordering prevails but the nest is characterised by different subgroupings of lemmata. Each subgroup is followed by an entry indicating the relevant subcomment on semantics of the main lemma with which the preceding sublemmata can be associated. This deviation from the alphabetical ordering is determined on semantic grounds. Lemma niching and, especially, lemma nesting involve procedures of textual condensation which can impede rapid and easy access to the presented data. Where the niching and nesting also lead to the use of place-keeping symbols in a condensed lemma sign it increases the degree of textual condensation (cf. Wolski 1989a: 964).

A straight alphabetical ordering is much more user-friendly and allows more successful access than an ordering displaying niched and nested lemmata. The dictionary typology and the target user should determine the nature of the presentation, ordering and textual condensation. A sinuous lemma file can impede dictionary consultation procedures in a monolingual dictionary and in a bilingual dictionary, bilingual or multilingual, it can create serious access problems, especially where the lemma signs are the targets of textual condensation. Where a lexical item has a frequent occurrence as first stem of a compound, the lemmatisation of these compounds can impede the access procedure if the lexicographer opts for a system of multiple niching instead of first level nesting. First level nesting can be seen in the following example from *New Words/Nuwe Woorde*:

con-sum-er *n.*: ~ **boycott** verbruikersboikot. **C**~ **Council** Verbruikersraad. ~ **credit** verbruikerskrediet. ~ **demand** verbruikersvraag. ~ **durable** *n.* duursame verbruik(er)sartikel; ~ ~*s* duursame verbruik(er)sgoedere. ~ **friendly** verbruikersvriendelik. ~ **goods** verbruik(er)sgoedere. ~ **protection** verbruikersbeskerming. ~ **research** verbruikersnavorsing. ~ **resistance** verbruikers-, kopersweerstand. ~ **society** verbruikersgemeenskap.
con-sum-er-ism *n.*, (*dikw. neerh.*) verbruikerisme. **con-sum-er-ist** *n.* verbruikerskampvegter; konsumentis, aanhanger/voorstander van die konsumentisme. **consumerist** *a.* verbruiker- (*samelewing ens.*); verbruikers- (*etiek, kultuur, ens.*).

In the following example from *Groot Woordeboek/Major Dictionary* the lexicographer opts for multiple niching:

moon, (n) maan; maand (*poëties*); ... ~**beam**, maanstraal; ~**blind**, nagblind; ~**blindness**, nagblindheid; ~**buggy**, maanbesie, maanbakkie; ~**calf**, gedrog, misgeboorte; uilskuiken ... ~**landing**, maanlanding; ~**less**, donker(maan), sonder maan; *a* ~*less night*, donkermaan, 'n donker aand.

moon`light, (n) maanlig; *a* ~ *NIGHT*, 'n maanligaand; ~ *and ROSES*, rosegeur en maneskyn; (v) bywerk doen, ekstra verdien; ~**er**, nagtrekker; ~**lighting**, nagwerk; vir twee base werk, met twee kakebene kou (*geselst.*), die wors aan altwee kante vreet (*geselst.*).

moon: ~**lit**, maanlig, deur die maan verlig; ~ **probe**, maanreis; maanruimtevaart; maanprojektiel; maantuiglansering; ~**rise**, maansopgang; ~**set**, maansondergang.

moon`shee, *kyk munshi*.

moon: ~**shine**, maanskyn; onsin, kaf, bog; smokkeldrank (*U.S.*); *that's all ~shine*, dis pure bog; ~**shiner**, dranksmokkelaar; ~**shiny**, deur die maan beskyn, maanlig; onsinig; ... ~**watch**, maanwag; ~**watcher**, maanwagter; ~**wort**, maanvaring; ~**y**, maan, maanvormig; maansiek; getik; dronk; ~**year**, maanjaar.

Here the niches are constructed by means of lemma parts (cf. Wolski 1989: 365). In the article clusters **moon: ~lit**, ... ~**set** and **moon: ~shine**, ... ~**year** the niches do not contain a single full lemma. They are introduced by lemma parts in niche external position and the clusters are characterised by condensed lemmata consisting of a place-keeping symbol and the second stem of the respective compound. It is clear that a user looking for translation equivalents of, say, the English words *moonrise* or *moon-watch*, should be quite sophisticated in dictionary using skills to ensure rapid access to the lemmata. Multiple niching decreases the accessibility of the external search route and the dictionary consultation procedure. Where bilingual dictionaries opt for a sinuous lemma file this ordering should be restricted to first level nesting or to single niching. Even second level nesting, although it presents a more comprehensive transfer of linguistic data, can create an impediment in the outer access structure. In order to improve the quality of the external access structure the lexicographer should try to refrain from the use of condensed lemmata.

4. The use of inner texts

Inserted texts can be employed to enhance the access to certain macrostructural entries. An inner text, presented in a clearly identifiable frame, can be regarded as a salient text block and search zone which immediately attracts the attention of a user and highlights the search route to the desired lemma sign. Inserted texts can be presented in different ways, e.g. as inner texts within an article or as a text block at the end of an article. The *Kenkyusha's Lighthouse English-Japanese Dictionary* makes provision for both these procedures. Inserted inner texts, positioned between two subcomments on semantics of a polysemous lemma, present lemma nests, adhering to an internal alphabetical ordering, of compounds in which the main lemma functions as second component. Because the second stem of a component constitutes the semantic core, these compounds are semantically related to the main lemma. In a bilingual dictionary where the user is not familiar with the source language the clustering of semantically

related items as sublemmata in an inserted inner text could be beneficial to ensure access to the data. Cf. the presentation of the sublemmata *hairbrush*, *nail brush*, etc. in the inserted inner text positioned between the first two subcomments on semantics of the article of the lemma sign *brush*¹. The article of the lemma sign *brush*¹ is followed by an inserted text which introduces a cluster of nested articles with sublemmata *brush aside*, *brush down*, etc., representing particle verbs with the main lemma as first component. The frame in which the article cluster occurs marks a specific search zone and forms an integral part of the access structure. If this system is explained clearly in the users' guidelines text and applied consistently throughout the dictionary it will improve the accessibility of the dictionary.

brush¹ /brʌʃ/ (BRŪSH) 図 (~es /-ɪz/) 1 ㊦
 ブラシ, はけ: *Brushes are used for cleaning.* ブラシは掃除をするのに使われる。

brush のいろいろ
 háirbrúsh ヘアブラシ, náil brúsh つめブラシ, páint-brúsh ペンキ用のはけ, 絵筆, sháving brúsh ひげそり用ブラシ, tóothbrúsh 歯ブラシ

2 ㊦ 絵筆, 毛筆: paint with a brush 絵筆で描く。
 3 ㊦ [普通は単数形で] ブラシ[はけ]のひとかけ; 軽く [ちよつと]触れること。 4 ㊦ きつねの尾。

give a brush to ... = give ... a brush [動] ㊦

㊦ ...にブラシをかける。
 ㊦ (brush-es /-ɪz/; brushed /-t/; brushing) ㊦ 1 ...にブラシをかける, (ブラシで)磨く: She brushed her husband's hat. 彼女は夫の帽子にブラシをかけた / I brush my teeth clean every morning. <V+O+C(形)> 私は毎朝歯をきれいに磨く。
 2 ブラシで払いのける, (手などで)ぬく: The girl brushed the tears from her eyes. 少女は目から涙をぬく。 3 かすって通る; 気に留めずに通り過ぎる (against, over, across).

— ㊦ かする; [副詞(句)を伴って] かすって通る: A big dog brushed past [by] me. 大きな犬が私のそばをかするようにさっと通りすぎた。

brush の句動詞

brush aside [away] ㊦ ㊦ 1 無視する。 2 払いのける。

brush down ㊦ (ブラシで)払い落とす。

brush off ㊦ ㊦ 1 払いのける。 2 《略式語》(人のことばなどを)無視する; (人と)手を切る: He brushed off my objections. 彼は私の反対を無視した。

brush up (on) ... ㊦ ㊦ (忘れかけていた勉強などを)やり直す: I'll have to brush up (on) my French before going to France. フランスへ行く前に(忘れかけた)フランス語を新しくやり直さなければならぬ。

brush² /brʌʃ/ 図 1 ㊦ = brushwood. 2 ㊦ 低木の密生地帯。

brush-off /brʌʃɔ:f | -ɔ:f/ 図 [the ~] 《略式語》拒絶; 解雇。 give ... the brush-off [動] ㊦ 《略式語》はなつける; 首にする。

5. The inner access structure: microstructural procedures

Bilingual dictionaries have too often been characterised by a lemmatic bias. This has resulted in an access structure dominated by the external search route and also in an addressing structure where the lemmata function not only as primary but often as the only treatment units. Access to target language entries has often been of an inferior nature. To enhance the access structure a much stronger focus on target language forms such as nonlemmatic addresses and secondary treatment units is needed. This demands a more sophisticated system of typographic and nontypographic structural markers to guide the user on the article internal search route towards secondary treatment units. Lexicographers should consider the use of a well-structured microarchitecture and the employment of unambiguously defined search zones, arranged in clearly distinguishable text blocks to assist the user when negotiating the rapid access structure of a dictionary.

Modern-day lexicography is dominated by the user-perspective and the access structure of any new dictionary has to be user-driven, i.e. it has to be planned in accordance with the user profile of the dictionary and the needs and reference skills of a well-defined target user. Where the speakers of the languages treated in a bilingual dictionary have different cultural backgrounds, the data distribution and presentation will necessarily have to make provision for these differences. The access structure will have to assist users to identify surrogate equivalents rendered in the treatment of culturally bound source language forms. These equivalents will have to be the treatment units of contextual and cotextual items. Surrogate equivalents must be clearly distinguishable from other equivalents by means of structural markers.

6. In conclusion

Access structure models should not be of a language specific nature and must be open enough for the planners of new dictionary projects in various languages and subdomains of bilingual and multilingual lexicography to customise them in terms of the needs of their specific projects. The customisation, however, must ensure an unimpeded access which can result in a dictionary consultation procedure which renders an optimal retrieval of information.

References

- Bergenholtz, Henning, Sven Tarp and Herbert E. Wiegand.** 1999. Datendistributionsstrukturen, Makro- und Mikrostrukturen in neueren Fachwörterbüchern. Hoffmann, L., H. Kalverkämper and H.E. Wiegand (Eds.). *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.
- De Villiers, Meyer, et al.** 1987. *Nasionale Woordeboek*. Cape Town: Nasou.
- Du Plessis, Madeleine.** 1999. *New Words / Nuwe Woorde*. Cape Town: Pharos.
- Eksteen, Louis C. (Ed.).** 1997¹⁴. *Groot Woordeboek / Major Dictionary*. Cape Town: Pharos.
- Gouws, Rufus H.** 1999. Mediostructural Representation, Textual Condensation and User-Oriented in the WAT X. *Lexicographica* 15: 4-37.
- Gouws, Rufus H.** 2001. Der Einfluß der neueren Wörterbuchforschung auf einen lexikographischen Gesamtprozeß und den lexikographischen Herstellungsprozeß. Lehr, Andrea, et al. (Eds.). *Sprache im Alltag*: 521-531. Berlin: De Gruyter.
- Gouws, Rufus H.** Forthcoming. Die Umtexte im DE GRUYTER WÖRTERBUCH DEUTSCH ALS FREMDSPRACHE. Wiegand, Herbert E. (Ed.). Forthcoming. *Perspektiven der pädagogischen Lexikographie des Deutschen II*. Tübingen: Max Niemeyer.
- Hausmann, Franz J. and Herbert E. Wiegand.** Component Parts and Structures of Monolingual Dictionaries. Hausmann, Franz J. et al. (Eds.). 1989-1991: 328-360.

- Hausmann, Franz J., et al. (Eds.).** 1989-1991. *Wörterbücher. Ein internationales Handbuch zur Lexikographie / Dictionaries. An International Encyclopedia of Lexicography / Dictionnaires. Encyclopédie internationale de lexicographie.* Berlin: De Gruyter.
- Kammerer, Matthias and Herbert E. Wiegand.** 1998. Über die textuelle Rahmenstruktur von Printwörterbüchern. Präzisierungen und weiterführende Überlegungen. *Lexicographica* 14: 224-238.
- Labuschagne, F.J. (Ed.).** 1993⁸. *Verklarende Afrikaanse Woordeboek.* Pretoria: J.L. van Schaik.
- Odendal, François F. and Rufus H. Gouws.** 2000⁴. *Verklarende Handwoordeboek van die Afrikaanse Taal.* Midrand: Perskor.
- Takebayashi, S. and Y. Kojima.** 1990². *Kenkyusha's Lighthouse English-Japanese Dictionary.* Tokyo: Kenkyusha.
- Wiegand, Herbert Ernst.** 1989. Aspekte der Makrostruktur im allgemeinen einsprachigen Wörterbuch: Alphabetische Anordnungsformen und ihre Probleme. Hausmann, Franz J., et al. (Eds.). 1989-1991: 371-409.
- Wolski, Werner.** 1989. Das Lemma und die verschiedenen Lemmatypen. Hausmann, Franz J., et al. (Eds.). 1989-1991: 360-371.
- Wolski, Werner.** 1989a. Formen der Textverdichtung im allgemeinen einsprachigen Wörterbuch. Hausmann, Franz J., et al. (Eds.). 1989-1991: 956-967.

The Compilation of Bilingual Dictionaries between African Languages in South Africa: The Case of Northern Sotho and Tshivenda*

Kwena J. Mashamaite, *Department of Northern Sotho, University of the North, Pietersburg, Republic of South Africa*
(*mashamaitek@unin.unorth.ac.za*)

Abstract: Bilingual dictionaries between South African languages have existed for more than three centuries (Smit 1996: 232). These dictionaries have English or Afrikaans as the source language and an African language as the target language. There has never been a case of the opposite except for bi-directional bilingual dictionaries. Still more, there seems to be no record of an attempt ever made to date to compile bilingual dictionaries between African languages.

This scenario illustrates the trend and history of lexicographic practice in South Africa. It is evident from most lexicographic research and products such as the existing kinds of dictionaries that Africans were never considered to be the first target group of users or prospective beneficiaries of such lexicographic products. The absence of bilingual dictionaries between African languages again provides evidence of who the lexicographers were and which population groups they represented.

To fill this void, a model called the hub-and-spoke is proposed in this paper for the compilation of such kind of dictionaries. The model has been chosen for its purported economy of use. Northern Sotho and Tshivenda are the African languages that will be used as examples in the application of the model.

A purposive sampling technique will be used to select lexemes that will constitute the entry words which form the central list of the proposed dictionary.

Keywords: BI-DIRECTIONAL, CENTRAL LIST, CONCEPTUAL EQUIVALENCE, ENTRY WORDS, HUB-AND-SPOKE MODEL, LEXEMES, LEXICAL ITEM, LEXICAL UNIT, LEXICALISATION STATUS, PRAGMATIC CONTRAST, PURPOSIVE SAMPLING TECHNIQUE, SOURCE LANGUAGE, TARGET LANGUAGE, VARIANT STATUS

Opsomming: Die samestelling van tweetalige woordeboeke tussen Afrika-tale in Suid-Afrika: Die geval van Noord-Sotho en Tshivenda. Tweetalige woordeboeke tussen Suid-Afrikaanse tale bestaan vir meer as drie eeue (Smit 1996: 232). Hierdie woordeboeke het Engels of Afrikaans as brontaal en 'n Afrikataal as doeltaal. Daar was nooit 'n

* This paper was presented at the Fifth International Conference of the African Association for Lexicography, held at the University of Stellenbosch, Stellenbosch, 3-5 July 2000.

geval van die teenoorgestelde nie buiten vir tweerigting tweetalige woordeboeke. Nog meer, daar skyn geen rekord te wees van 'n poging tot op hede om tweetalige woordeboeke tussen Afrikatale saam te stel nie.

Hierdie scenario illustreer die verloop en geskiedenis van leksikografiese praktyk in Suid-Afrika. Dit is duidelik uit die meeste leksikografiese navorsing en produkte soos die bestaande woordeboeksoorte dat Afrikane nooit beskou is as die eerste doelgroep gebruikers of waarskynlike voordeeltrekkers uit sulke leksikografiese produkte nie. Die afwesigheid van tweetalige woordeboeke tussen Afrikatale gee ook 'n aanduiding van wie die leksikograwe was en aan watter bevolkingsgroepe hulle behoort het.

Om hierdie leemte te vul, word 'n naaf-en-speek-model vir die samestelling van hierdie soort woordeboeke in dié artikel voorgestel. Die model is gekies vir sy beweerde ekonomiese aard. Noord-Sotho en Tshivenda is die Afrikatale wat as voorbeelde gebruik sal word in die toepassing van hierdie model.

'n Doelgerigte toetsingstegniek sal gebruik word om lekseme uit te soek wat die inskrywings in die sentrale lys van die voorgestelde woordeboek sal vorm.

Slutelwoorde: TWEERIGTING SENTRALE LYS, KONSEPTUELE EKWIVALENSIE, INSKRYWINGS, NAAF-EN-SPEEK-MODEL, LEKSEME, LEKSIKALE ITEM, LEKSIKALE EENHEID, LEKSIKALISASIESTATUS, PRAGMATIESE KONTRAS, DOELGERIGTE TOETSINGSTEGNIEK, BRONTAAL, DOELTAAL, VARIANTSTATUS

1. Introduction

In South Africa there are eleven official languages and other non-official languages such as Khoi and San which in future are also likely, if developed, to acquire official status. South Africa as a multilingual country needs bilingual dictionaries, especially bi-directional learner's dictionaries so that the speakers of these languages can learn each other's language.

The absence of bilingual dictionaries between the African languages can be ascribed to the shortage of trained African lexicographers. Fortunately the Pan South African Language Board, in collaboration with the African Association for Lexicography, has pledged to address this backlog by providing the necessary training.

One of the major tasks bilingual dictionaries in South Africa has to perform, is to assist speakers of various languages to learn to speak each other's language in order to promote multilingualism.

Bilingual dictionaries may be described as dictionaries in which lexical items of the native language are given translation equivalents in a foreign language. The native language may be the source language whereas the foreign language is the target language or vice versa.

Bilingual dictionaries may serve different purposes depending more on the communicative needs of the dictionary users than on the amount of information supplied by the compiler. Mdee (1997: 95) argues that a bilingual dic-

tionary has two functions, the first being that of comprehension, that is, reading and listening, and the second that of production, which means writing and speaking. One may add a third function which is that of translation.

For a bilingual dictionary to be useful to speakers of both languages it has to be bi-directional. *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 translation. This condition circumvents a situation in which four mono-functional mono-directional dictionaries per language of a language pair would be required for the same functions. In a South African context in which there are eleven official languages, one is speaking of 55 different language pairs times four dictionaries per language pair, yielding a total of 220 different kinds of dictionaries. Surely a developing country like South Africa cannot afford to produce such numbers of dictionaries.

It is against this background, as well as because of the need to have such dictionaries, that a more economic solution is sought to achieve the same goal. Consequently, the hub-and-spoke model proposed in this paper seems to be just the kind of solution needed.

According to Mashamaite (1995) and Martin (1995) the main contention of the hub-and-spoke model is that several bilingual dictionaries can be derived by linking the lexical items of the spoke languages, that is, the source languages in a bilingual dictionary, to those of a common hub language which is the target language in a language pair. The condition for the application of the model is that one must have at least two bilingual dictionaries which share a common target language or hub. For instance, in the example used in this paper, Northern Sotho and Tshivenda are the spoke or source languages sharing a common hub or target language which is English.

The application of the model does not require prior collection of data or the building of corpora. It therefore reduces costs normally incurred in the process of hiring fieldworkers and computer experts needed for the building of corpora and lexical databases as well as time spent on this procedure. This is why the model is thought to be cost-effective. It only uses already existing data in the form of dictionaries.

A purposive sampling technique will be used in this study to select lexemes which will form entry words that constitute the central list of the proposed dictionary. Although a non-probability technique, it is, according to Neuman (1997), an acceptable sampling technique used for special situations. One of these special situations is when a researcher wants to identify particular types of cases for in-depth investigation. It has been chosen for this study in particular because it allows one to choose lexemes that might pose an obstacle to the application of the model and consequently affords one the opportunity to suggest a probable solution, something that no other kind of sampling is able to do.

In the application of the model special attention will be paid to the explanation and use of the parameters of the hub-and-spoke model, namely concep-

tual equivalence, pragmatic contrast, variant status, and lexicalisation status.

2. The hub-and-spoke model

The hub-and-spoke model has been explained in the introduction above. This section will look into the parameters of the hub-and-spoke model and the role they play in the application of the model. The parameters have been borrowed from Omkeerbare Bilinguale Lexicale Databanken (OMBI), translated as 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 1995).

The purpose of the parameters is to ensure reversibility of translation equivalence. Translation equivalence according to Svensén (1993: 140) is achieved when a bilingual dictionary is able to "provide words and expressions in the source language with counterparts in the target language which are as near as possible, semantically and as regards style level (register)". It is important to note that in providing translation equivalents one is dealing with meaning equivalence rather than with word equivalence. To support this assertion Martin (1995) argues that in translation relations "it is not words that are translated into other words, but rather words in a specific meaning". He illustrates his argument by the use of the terms *form unit* (FU) to refer to a word form, and *lexical unit* (LU) to refer to a word-specific meaning. One cannot give a precise translation equivalent of an FU because an FU can have more than one meaning. Only meanings or lexical units (LUs) can be given translation equivalents which are also LUs and not word forms in a target language. For example, the Northern Sotho FU *panka* has the following LU: 1. bench 2. desk 3. bank. These three meanings (LUs) have the following translation equivalents which are also the LUs in the target language namely *panka/senno*, *teseke/panka* and *panka* respectively in Northern Sotho.

On the other hand, the English FU *bank* has the following meanings or LUs: 1. *panka* 2. *leriba* 3. *lebopo* 4. *khwiti* (Kriel 1988). It is apparent from the preceding examples that the English FU *bank* and the Northern Sotho FU *panka* have only one common LU which is *bank* (in the sense of money). It is therefore incorrect and misleading too to assume that the FU *bank* has only one translation equivalent in Northern Sotho whereas there are four meanings as indicated. To capture all four meanings one should focus on the translation equivalents of the lexical units attached to the form unit and not the form unit itself. This confirms the assertion that only lexical units and not form units can have translation equivalents in a target language.

3. Parameters of the hub-and-spoke model

To ensure that only meanings and not words are translated, the following conditions in the form of parameters should be met.

3.1 Conceptual equivalence

The term treats of a situation in which the target language translation equivalent refers exactly to the same concept as that referred to by the source language lexical item. It is not always easy to find this kind of translation equivalence. The difficulty lies in the fact that every language has a set of concepts which are perceived differently from or are even nonexistent in another language as a result of historical, geographical, cultural, social and economic differences between countries where these languages are spoken (Svensén 1993).

In a situation where the target translation equivalent refers to the same concept as that in the source language, one would speak of complete equivalence. For example, Northern Sotho *noka* refers to English *river*. It happens sometimes that the conceptual relationship between the target and the source language lexical items is not complete or exactly the same. One would refer to such a situation as partial equivalence. This may occur in the following three ways:

- (a) The source language lexical item may have two or more target language translation equivalents, a state referred to as divergence. It is a situation in which the meaning relation between the source language lexical item is that of a hyponym of the lexical item of the target language. For example, English *tick* refers to Northern Sotho *kgofa*; *patšane*; English *uncle* refers to Northern Sotho *malome*; *ramogolo*; *rangwane* (Kriel 1988).
- (b) Several lexical items in the source language may have one translation equivalent in the target language. This situation is known as convergence. The target language translation equivalent becomes the super-ordinate. For example, Northern Sotho *moeti*, *mosepedi* refers to English *tourist* (Kriel 1988). When the translation equivalents cannot be conceptually regarded as either hyponyms or hyperonyms but show various degrees of overlap between each other, the condition is known as related equivalence. For example, English *shine* refers to Northern Sotho *kganya*, *nyedima*, *phadima* (Kriel 1988).
- (c) The lexical item of the source language may not have any appropriate translation equivalent in the target language. This situation is known as non-equivalence. To overcome this problem the source language lexical item is either substituted by an explanatory equivalent or a phrase, or by a borrowed word, or by another word with a close conceptual meaning, that is, a near-equivalent. Non-equivalents occur very often with the translation of terms relating to culture-specific concepts. For example, Northern Sotho *hlatswadirope* is translated in English as *ancillary wife* (Ziervogel and Mokgokong 1975). The translation equivalent is deficient in the sense that it does not reveal the cultural significance underlying the purpose of having an ancillary wife. An ancillary wife is not any second wife but the principal wife's sister who is married with the specific purpose of bearing children for her barren sister.

3.2 Pragmatic contrast

The term *pragmatic* refers to the contextual adequacy of a word. A translation equivalent should have the same contextual meaning as its source language counterpart. The pragmatic component consists of the subject field, style, connotation, chronology (time), frequency and geography (place). For the translation equivalent to be exact it should have no contrast with the source language counterpart on the basis of the stated pragmatic components. This condition will be specified as **nil** which means that pragmatic components are the same for both words. Sometimes the translation equivalent falls short of one or two of the pragmatic components and, depending on which component is missing, could be described as **obsolete**, **formal**, **neutral**, or **informal**. For example, the translation of the English *window* in Northern Sotho is *letsikangope* which can be labelled as obsolete because of chronology and frequency.

3.3 Variant status

This term refers to a situation where the lexical item of the source language has more than one translation equivalent in the target language. These translation equivalents may be regarded as synonyms although they might have functional restrictions depending on the context in which they are used. The status of the variant may be specified as main native synonym, loan word, hybrid term, and loan translation. Native synonyms are target language lexical items derived from the native language lexicon. Loan words are the adopted lexical items. Hybrid terms are words one part of which is derived from one language and the other part from another language. For example, *condensed milk* is translated into Tshivenda as *mafhi a khondasi* (Wentzel and Muloiwa 1982). *Mafhi* is a Tshivenda native word for *milk* and *khondasi* is a loan word from English *condensed*. A loan translation is a compound word whose constituents are literal translations of the constituents of a corresponding compound in another language (Martin et al. 1992: 43). For example, English *medical examination* is translated into Northern Sotho as *tlhahlobo ya kalafo* (Kriel 1988).

3.4 Lexicalisation status

The concept refers to a situation in which the lexical item of the source language may or may not have a translation equivalent which is a lexical item in the target language. In a situation where the translation equivalent is a lexical item, it will be specified as **lexicalised**. Some lexical items of a given language express culture-bound concepts which may not form part of the conceptual world of the target language. In this case, a translation equivalent may not be a lexicalised word but a description or phrase which explains the conceptual meaning of the source language concept. For example, Northern Sotho *thiswana* means in English *little basin*, *small vessel* (Kriel 1988). The two English transla-

tion equivalents do not appear in the central list as English entry words. This is an indication that they are not lexicalised simply because they have no cultural referent in English. Such non-lexicalised words are specified as either **semi-lexicalised** when they can be inserted in some contexts as equivalents or **non-lexicalised** when they are not insertible in the place of the source language lexical item but only serve to explain or describe its meaning (Mashamaite 1995).

4. Application of the hub-and-spoke model

The process of providing source language lexical items with appropriate target language translation equivalents applying the parameters of the hub-and-spoke model will look as follows:

- (a) (i) Source language: Northern Sotho — **thelebišene**
Target language: English — **television**
Lexicalisation status: lexicalised
Variant status: main
Pragmatic contrast: nil
Conceptual equivalence: complete
- (ii) Source language: Tshivenda — **thelevishini**
Target language: English — **television**
Lexicalisation status: lexicalised
Variant status: main
Pragmatic contrast: nil
Conceptual equivalence: complete
- (iii) Source language: Northern Sotho — **thelebišene**
Target language: Tshivenda — **thelevishini**
Lexicalisation status: lexicalised
Variant status : main, loan word
Pragmatic contrast: nil
Conceptual equivalence: complete
- (iv) **thelebišene**, thelevishini
- (b) (i) Source language: Northern Sotho — **setšidifatši**
Target language: English — **refrigerator**
Lexicalisation status: lexicalised
Variant status: main
Pragmatic contrast: nil
Conceptual equivalence: same

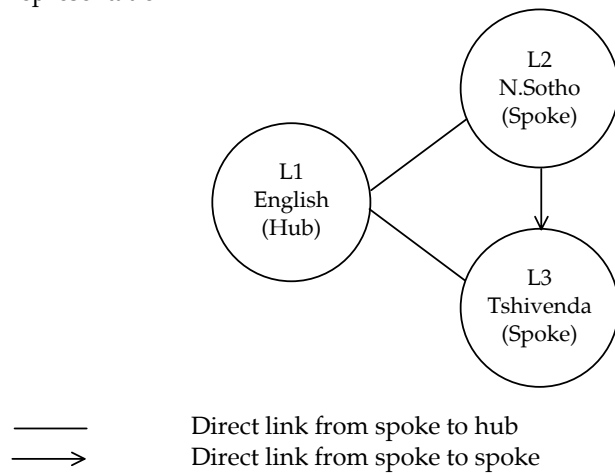
- (ii) Source language: Tshivenda — **tshirothodzi**
Target language: English — **refrigerator**
Lexicalisation status: lexicalised
Variant status: main
Pragmatic contrast: nil
Conceptual equivalence: same
- (iii) Source language: Northern Sotho — **setšidifatši**
Target language: Tshivenda — **tshirothodzi**
Lexicalisation status: lexicalised
Variant status: main
Pragmatic contrast: nil
Conceptual equivalence: same
- (iv) **setšidifatši**, tshirothodzi
- (c) (i) Source language: Northern Sotho — **phasa**
Target language: English — **bring sacrifice to ancestral spirits**
Lexicalisation status: non-lexicalised
Variant status: substitution by explanation
Pragmatic contrast: nil
Conceptual equivalence: same
- (ii) Source language: Tshivenda — **phasa**
Target language: English — **bring sacrifice to ancestral spirits**
Lexicalisation status: non-lexicalised
Variant status: substitution by explanation
Pragmatic contrast: nil
Conceptual equivalence: same
- (iii) Source language: Northern Sotho — **phasa**
Target language: Tshivenda — **phasa**
Lexicalisation status: lexicalised
Variant status: main
Pragmatic contrast: nil
Conceptual equivalence: same
- (iv) **phasa**, phasa

In the above examples (a) and (b) have the same parameters and it is therefore easy to derive the translation equivalents of Tshivenda and Northern Sotho as illustrated by (iv) in both cases.

The difference in one parameter in (c) (ii) indicating a non-lexicalised English translation equivalent does not pose a serious obstacle to the process since the substitution by explanation does not contrast with the pragmatic and conceptual framework of the source language.

What makes the situation even simpler is the fact that the Tshivenda translation equivalent matches all the parameters of the source language lexical item in Northern Sotho thereby becoming a complete equivalent.

The above exercise can be summarised by the following diagrammatic representation.



5. Conclusion

This article, which acknowledges the multilingual nature of the South African population, has expressed a concern about the lack of bilingual dictionaries between African languages. The primary purpose of bilingual dictionaries is to assist the speakers of the various languages spoken in South Africa to learn each others' languages in order to promote multilingualism. One of the reasons mentioned for the lack of such dictionaries is the shortage of trained African lexicographers. Fortunately, the Pan South African Language Board, in collaboration with the African Association for Lexicography, has pledged to address this backlog by providing the necessary training.

Three broad functions of bilingual dictionaries have been identified, namely

- (a) that of comprehension, that is, reading and listening,
- (b) that of production, that is, writing and speaking, and
- (c) that of translation.

The article has again acknowledged the fact that the process of making dictionaries is an expensive one both in terms of personnel, time and funding. This

reason alone has served as motivation to look for a less expensive alternative which is the hub-and-spoke model proposed here.

An attempt has been made to explain and illustrate how the model can be applied. My proposal is that it should be tried.

References

- Kriel, T.J.** 1988. *Popular Northern Sotho Dictionary*. Pretoria: J.L. van Schaik.
- Martin, W.** 1995. *Lexicographical Resources in a Multilingual Environment: An Orientation*. Unpublished paper. Amsterdam: Vrije Universiteit.
- Martin, W.** et al. 1992. Standards for the Terminological Description of Lexical Items. Multilex report. Amsterdam: Vrije Universiteit.
- Mashamaite, K.J.** 1995. *The Hub-and-Spoke Model: A Recipe for Making Bilingual Dictionaries between African Languages in South Africa*. Unpublished doctoral paper. Amsterdam: Vrije Universiteit.
- Mdee, J.S.** 1997. Language Learners' Use of a Bilingual Dictionary: A Comparative Study of Dictionary Use and Needs. *Lexikos* 7: 94-106.
- Neuman, W.L.** 1997. *Social Research Methods*. Boston: Allyn and Bacon.
- Smit, M.** 1996. *Wiegand's Metalexigraphy as a Framework for a Multilingual, Multicultural, Explanatory Music Education Dictionary for South Africa*. Unpublished D.Litt. thesis. Stellenbosch: University of Stellenbosch.
- Svensén, Bo.** 1993. *Practical Lexicography*. Oxford: Oxford University Press.
- Wentzel, P.J. and T.W. Muloiwa.** 1982. *Thalusamaipfi ya Nyambotharu yo Khwiniswaho. Luwenda. Luwuru. Luisimane*. Pretoria: University of South Africa.
- Ziervogel, D. and P.C. Mokgokong.** 1975. *Groot Noord-Sotho Woordeboek*. Pretoria: J.L. van Schaik.

Macro- and Microstructural Issues in Mazuna Lexicography¹

P.A. Mavoungou, *Department of Afrikaans and Dutch,
University of Stellenbosch, Stellenbosch, Republic of South Africa
(13126733@humarga.sun.ac.za)*

Abstract: All the works in Mazuna lexicography have a common denominator: they are translation dictionaries biased towards French and were compiled by Catholic and Protestant missionaries or colonial administrators. These dictionaries have both strong and weak points. The macrostructure although it does not display features of sophistication, i.e. the use of niching and nesting procedures, tends to survey the full lexicon of the language which make these dictionaries real reservoirs of knowledge. The microstructure contains a lot of useful entries. However, no metalexicographic discussion is provided in the user's guide to make it accessible to the target reader. There are also some shortcomings especially in the areas of suprasegmental phonology (absence of tonal indications) and orthography.

Keywords: MACROSTRUCTURE, MICROSTRUCTURE, TARGET USERS, SPECIAL-FIELD TERMS, LEXICOGRAPHIC DATA, PROBLEM OF SOURCES, USERS' GUIDE, RESERVOIRS OF KNOWLEDGE, FAN.

Abstrait: Les problèmes macro- et microstructurels dans la lexicographie Mazuna. Tous les travaux en lexicographie Mazuna ont un dénominateur commun: ce sont des dictionnaires de traduction centrés sur le français et compilés par les missionnaires catholiques et protestants ou les administrateurs coloniaux. Ces dictionnaires ont à la fois des avantages et des inconvénients. Bien que ne présentant pas de caractéristiques de sophistication, par exemple l'usage de procédures de nichification et de nidification, la macrostructure tend à donner une vue d'ensemble du lexique de la langue, ce qui fait de ces dictionnaires de véritables réservoirs de connaissance. La microstructure contient de nombreuses entrées utiles. Mais aucune discussion métalexigraphique n'est présentée dans le guide aux usagers pour les leurs rendre accessible. Il y a également des manquements, spécialement dans le domaine de la phonologie suprasegmentale (absence d'indications tonales) et de l'orthographe.

Mots-clés: MACROSTRUCTURE, MICROSTRUCTURE, PUBLIC CIBLE, TERMES DES LEXIQUES SPÉCIALISÉS, DONNÉES LEXICOGRAPHIQUES, PROBLÈMES DES SOURCES, GUIDE DES UTILISATEURS, RÉSERVOIRS DE CONNAISSANCE, FAN.

Introduction

Mazuna (which is the opening greeting formality "I say that") is according to Kwenzi-Mikala's classification (Kwenzi-Mikala 1998 and 1998a) the language group under which the six dialects of the Fan language are accommodated. Fan² is the home language of 29 percent of the population (circa 427 000 people) of Gabon.

The speech forms of this group are mostly found in four of the nine provinces of Gabon, namely Estuaire, Moyen-Ogooué, Ogooué-Ivindo and Woleu-Ntem. *Fañ-Atsi* is spoken in Ndjolé and Lambaréné, *Fañ-Maké* in Libreville and Kango, *Fañ-Mvai* is found in Minvoul, *Fañ-Ntumu* in Oyem and Bitam, and *Fañ-Nzaman* and *Fañ-Okak* are spoken in the regions of Makokou-Ovan-Booué and Medouneu-Cocobeach-Mitzik respectively. *Fañ* is a so-called cross-border language because it is also spoken in neighbouring countries, namely in Equatorial Guinea, Cameroon, São Tomé, Principé and the Republic of the Congo.

Up to the present, all the dictionaries available for the *Fañ* language have a common denominator: they are translation dictionaries biased towards French and were compiled by Catholic and Protestant missionaries or colonial administrators. These dictionaries reveal a number of shortcomings that need to be remedied, e.g. most of these works neither mark tone, nor indicate the pronunciation of the treated items. To a greater or lesser degree, the earlier dictionaries lack a metalexigraphic foundation. These shortcomings are excusable if one considers that their compilers were neither linguists, nor trained lexicographers (cf. Nyangone Assam and Mavoungou 2000).

In this article the attention will mainly focus on the macro- and microstructure of three contributions, namely Largeau (1901), Martrou (1924) and Galley (1964).

This paper implicitly works with the theoretical framework of Wiegand and Kučera (1981, 1982). According to them, when criticising dictionaries, one has to take the following components into consideration, namely: the history of the publisher, the dictionary basis, and the macro- and microstructure of the dictionary. This will be done and attention will also be given to the treatment of special-field terms before some concluding remarks are given (cf. Smit 1996: 30).

1. The *Encyclopédie pahouine* of Largeau

1.1 The macrostructure of the work

The *Encyclopédie pahouine* was published in 1901 by a colonial administrator. It is an alphabetically arranged dictionary of considerable size, consisting of ±4 996 articles and covering 699 pages.

This dictionary is intended, in the first place, for speakers of *Fañ* who want to translate French into *Fañ*. In the second place, the dictionary is meant for speakers of French. It is a useful aid for foreigners visiting Gabon. It has been an indispensable tool in the economic sector and public administration for French, English and German businessmen operating in Gabon (especially in Woleu-Ntem and Moyen-Ogooué), because it is generally acknowledged that colonial administrations were no longer willing to rely on interpreters. The following is the comment of Courbon (1908) (quoted by Raponda-Walker 1998:

83) as far as this point is concerned: "Ce sera ce dernier (l'interprète) qui jouira de l'influence réelle sur l'indigène, influence que son naturel joint aux circonstances tentatrices lui fera exploiter à son profit, au détriment des indigènes aussi bien que de son maître. Trop souvent, des malentendus, des revirements inexpliqués, des dissentiments, des «palabres» aux conséquences désastreuses ont été dus aux seules intrigues d'un interprète tout d'abord préoccupé de ses propres intérêts."

Because Largeau's work is a monodirectional French-Fan publication, the lexicographer has not been confronted with the problematic aspects that African languages have in common nor with language-specific issues in Fan. The compiler adheres to the lemmatisation principles established in the European tradition, e.g. verbs are lemmatised under their infinitive form or under the first person singular of the present indicative, whereas nouns are entered under their complete form (cf. Hausmann and Wiegand 1989: 329).

In addition to the foregoing, it should be mentioned that different types of lexical units including multiword lexical units such as *tout à coup*, *arc-en-ciel*, *tout à l'heure*, etc. have been lemmatised in the encyclopedia. With regard to the scope of the work, lemmata have been taken from various fields, among others ethnology, religion, oral literature, metallurgy and astrology. As a large reference work it reflects the real language use and/or lexicon of the language.

With regard to the dictionary basis, a large lexicographic corpus was probably used. However, the lexicographer does indicate neither the primary sources (also called lexicographic corpus or dictionary basis which comprises all the sources that are not dictionaries themselves (cf. Smit 1996: 30-31 and 111)) nor the secondary (all the dictionaries consulted during the compilation phase (cf. Wiegand and Kučera 1981: 100ff)) and tertiary sources of his lemma collection. The compiler deserves some credit for providing the user with a relevant discussion on the pronunciation system, the orthography and the punctuation system used in the book as well as some morphological information such as the parts of speech, word-formation and the conjugation system of Fan that is abundantly illustrated by tables. But this presentation in the user's guidelines and the mini-grammar of the work reveals a number of shortcomings. Firstly, the writing system proposed by the compiler is very much along the lines of the orthographic traditions of French. It is a well attested fact that there is a difference between the sound systems of the African languages and those of the European languages. Consequently, the transcriptions proposed are likely to be only approximate. Moreover, the grammatical content is a bit outdated.

1.2 The microstructure of the work

Despite the alphabetical arrangement displayed by the lemmata, the encyclopedic character of the work lies in its focus on extra-linguistic features rather than on the meaning of lexical items. As already mentioned, this encyclopedic

dictionary aims to present data as comprehensively as possible. The interesting fact that has to be mentioned is that the encyclopedia usually provides the user with a lot of contextual data. This point is illustrated in the article of the lemma AISELLE.

- (1) AISELLE (le dessous du bras), *Mvagha*, pl. *me mvagha*: je porterai ce paquet sous le bras (sous l'aiselle), *me ke babe mbôm wele e mvagha'si* (moi irai portant un paquet ce dans aiselle dessous).

Immediately after the lemma sign (in small capitals) appears, in parenthesis, the explanation "le dessous du bras" (the under arm (part of the body)), followed by the translation equivalent (in Italics) and the competence examples in both French (in roman) and Fañ (in italics). The latter is followed by its literal translation, given in parenthesis. The systematic use of parenthesised explanations and word by word translations in the work correspond to the lexicographer's willingness to meet the requirements of pedagogical lexicography. Another feature of the dictionary is the use of typographical markers. For example, in the article of the lemma CHANSON, encyclopedic data is presented in both source language (to the left) and target language (to the right) in two columns set apart typographically, and therefore clearly perceived by the user.

- (2) CHANSON, *za*, pl. *bia*. Voici une chanson que chantaient quelquefois nos anciens compagnons pahouins pendant nos courses à travers le Congo; elle est en dialecte *bedži*.

ZA NGOÏA
Ébon'i ne e Mellañ, ...

CHANSON D'EXPIATION
Le coucher est aux Nénuphars, ...

In (2) the user is provided with the singular and plural forms of the translation equivalent following the item giving the lemma sign. Then the lexicographer, after a short introduction of what is about to follow, gives in extenso a popular Fañ song. A number of articles, MORT, DESCENDANCE, CRÉATION, CONTE, etc. have the same bilingual presentation. The Fañ constituent of all these articles can be expanded and used for the compilation of a monolingual dictionary of Fañ, for example.

In addition to the foregoing, the lexicographer has sometimes thought that more data regarding the cultural context of the lemmata dealt with were necessary for a better understanding of the treated lexical item. To fulfil these anticipated expectations on the part of the users, Largeau has devised special articles divided into two sections or **search areas** (cf. Hausmann and Wiegand 1989: 339). The presentation of these articles is very much the same as that used in synopsis articles. In the first search area, the compiler basically deals with linguistic aspects whereas in the second (introduced by the word *Encyclopédie*), the focus is on the cultural aspect of the lemma treated. This **intrastructural marker** (cf. Gouws 1996: 24-25) (*Encyclopédie*) helps the user with a quick and successful retrieval of the needed information because it clearly indicates the relevant article position. As such it forms part of the rapid inner access struc-

ture (Hausmann and Wiegand 1989: 338). Instead of going through the whole article the user who is merely interested in cultural data can immediately proceed along the inner search path (the outer search path ends at the lemma sign) to the article position *Encyclopédie* and continue from there through the remainder of the article. This point is illustrated by the treatment of the lemma FEMME.

- (3) FEMME: 1° en général, *nga*, pl. *be ya*; — 2° femme accompli, qui donne des enfants, une dame, *Moŋga*, pl. *bonga*;
Encyclopédie: la femme; sa situation dans la famille; ses droits et devoirs. ...

From the data presented in these articles (cf. (2) and (3)) with a strong ethnographical interest, the user can retrieve a lot of information regarding the lifestyle, value system, the beliefs, in brief, the world vision of the *Fañ* people. The strong inclination towards ethnographical data qualifies the encyclopedia to be classified in the subtype of the so-called **ethnographical dictionary**. According to Zgusta (1987: 14) an ethnographical dictionary is "a dictionary that tries to describe a culture in the entries of the single relevant words". Although credit should be given to these cultural explanations, they reveal a number of shortcomings. Firstly, objections can be made with regard to the absence of meta-texts in the user's guidelines explaining the structure of these articles as well as the types of data categories presented in them. Secondly, the cultural analysis of certain treatment units reflects the observation of the colonizing over the colonized. For example, as far as the question of anthropophagy among the *Bafañ* is concerned, there is no doubt that Largeau was aware of the fact that anthropophagy was a very marginal phenomenon in this community. However, quite a significant amount of co-text examples and cultural comments do highlight this practice. This is likely to create a false impression in the mind of the users that the *Fañ* community is anthropophagous. But it has been proven by some scholars of ethnology (Roulet (1866), Deschamps (1962), Du Chaillu (1863) and Metegue N'nah (1984) among others) that anthropophagy was restricted to a very small group of people. Moreover, only one piece of the body of the avowed enemy killed during the war was eaten out of revenge in order to completely destroy any existence of him. Largeau (1901: 337 and 367) is correct when stating that women and children were excluded from these anthropophagous meals. The aspect under discussion and others which are derived from an outmoded and dated Eurocentric vision should be removed from the work. The user profile has to be taken into account in order to achieve this goal. Largeau deserves credit for painting the *Fañ* society through its origin, rituals, mythology, wars, hospitality, etc. However, he could not avoid the trap of cliché and personal judgments. As a result, some of his examples and comments often illustrate his own personal attitude towards the people described. It is well accepted that lexicographic data should illustrate an unbiased, impartial use of the language.

1.2.3 Concluding remarks

Although a large lexicographic corpus was used in Largeau's *Encyclopaedia*, the introductory section fails to indicate the primary as well as the secondary sources of the work. Due to the monodirectional nature of the work (French-Fan), the compiler adheres to the lemmatisation principles established in the European tradition. In the central list, one finds a lot of interesting articles in which medio-structural (cross-referencing), data distribution procedures and typographical exposition (to list but a few) come into play. The latter, especially, the use of different font types (italics, small capitals, etc.), the bilingual presentation of articles divided into search areas set apart typographically, special articles focusing on cultural data introduced by the intra-structural marker *Encyclopédie* as well as the systematic use of literal translations addressed at the translation equivalents and competence examples have drastically improved the quality of the work. Ordering devices are given as a visual aid to the users whereas contextual data in the form of word by word translations help them to disambiguate the different senses of the treated lexical items. Competence examples and comments with a strong cultural load make the *Encyclopédie* a valuable reservoir of knowledge. Unfortunately they sometimes reflect the lexicographer's personal judgement.

2. Martrou's *Lexique fān-français*

2.1 The macrostructure of the lexicon

In his Fan-French lexicon (1924) Martrou presents 3 431 articles with guiding elements arranged in a strict alphabetical order. The compiler of the lexicon deserves credit for explaining his writing system. The orthography employed in the lexicon is based on the principles expounded by Ch. Sarcleux in his book *Essai phonétique*. The fact that the work contains information on stress in the lemmata (according to Martrou tonic accent in Fan is characterised by a global rise and the lengthening of the syllable and is indicated by the stress mark (')) is an indication of the lexicographer's awareness of the necessity of giving an account of the pronunciation of the language.

Martrou does not mention which other dictionaries he consulted. Since no explanation is given as far as the primary sources of the lexicon is concerned, the work seems to have no lexicographic corpus of its own.

The lexicon contains lemmata in Fan and translation equivalents in French. To enhance the identification of entries, the compiler used a lot of typographical structural markers such as roman, italic, bold and small characters. The fact that lemmata have been set in large and small capital letters and the use of running heads have improved the access structure of the lexicon.

Different types of lexical items have been given lemma status in the lexicon. This includes nouns, verbs, adjectives, pronouns, etc. entered in the lexicon according to the word tradition.

With such a word-based approach, the lexicographer does not make provision for the inclusion of word-formation items (affixes (prefixes, suffixes and infixes) and stems).

Unfortunately, the users' guidelines of the lexicon do not answer the question likely to arise: on which ground has the selection of items taken place?

2.2 The microstructure of the lexicon

The microstructure includes some data, e.g. the paraphrase of meaning, translation equivalent (s), and examples. Compare the article of the lemma **Nkor**:

- (4) **Nkor**. Courbe. Adj. Crochu, replié. Ambigu, tortueux. *Nkor osu*. Rivière à méandres. *Nkor ntēm*. Branche recourbée. *Nkor adzu*. Affaire ambiguë.

The translation equivalent paradigm appears immediately after the item giving the form of the lemma sign. Then the user is provided with the relevant examples in the source language (in italics) as well as their translation equivalents in the target language (in roman). The interesting fact to be mentioned here is that the item giving the part of speech (Adj.) is not addressed at the lemma, but has a non-lemmatic address. In other words, it is addressed at the translation equivalent "crochu". The same line of argumentation holds true for the presentation of contextual data. The target language examples are addressed at the source language competence examples. Thus the latter become treatment units or new topics within the article. Here it seems that, as for **Fan**, French has been given the status of language of treatment shifting the focus from the source to the target language (cf. Gouws 1996: 158). When first looking at the microstructure of the work, it seems that the lexicographer has decided not to give grammatical items for the lemma type NOUN. Only when dealing with lemma signs which are interpreted as polyfunctional, i.e. when lemmata can function as a noun or a verb, the grammatical item is given. This corresponds to a practice also followed in other dictionaries. For example, Smit (1996: 192) focusing on the treatment of the lemma **gesalf'de** has pointed out that the *Verklarende Handwoordeboek van die Afrikaanse Taal* (HAT) "does not give the same data types for all lemma types NOUN." This system has of course to be explained in the user's guidelines of the dictionary. However, the treatment of lemma types VERB and ADJECTIVE (to list but a few) reveals inconsistencies. For example, in the articles of the lemma signs **sèroe** and **seroe**, it is not indicated that these lemmata are verbs, whereas in the treatment of the lemmata **sa** and **Nkūnoe** entries indicating part of speech are presented. In addition, why does the lexicographer give data regarding the part of speech in an article like **Ntul** (adj. meaning "old") but fails to do the same for the majority of nouns, some verbs, etc.? Since there is no text explaining the system, it can be assumed that no

fixed criteria were worked out for the different lemma types. The lexicographer too often fails to provide the user with a useful entry: the item giving the part of speech of the lemma sign. This point is illustrated in the article of the lemma **Abum**.

- (5) **Abum**. Abdomen, ventre. *Ekō abum*, avoir mal au ventre. *Abum d'alurœ a*, il a la diarrhée — Grosseur ... — intérieur d'un objet creux ... — devant ... — Grosseur ... — Grosseur moyenne ...

Immediately after the item giving the form of the lemma sign, the users are provided with a mere list of translation equivalents. This is unacceptable because the way in which an item is lemmatised should give an account of its morphological status. As far as this point is concerned (cf. Hausmann and Wiegand 1989: 353), dictionary articles usually display both macro- and microstructural data. The first data types are referred to as the comment on form (e.g. spelling and part of speech) whereas the second data categories are known as the comment on semantics (e.g. monosemy or polysemy, meaning paraphrase). When dealing with the lemma **Abum** the user is at a loss because he/she does not know whether it is a noun, a verb or an adjective. As a matter of fact, **Abum** is a noun. Each cotext example given in this article activates one sense of the lemma **Abum**. The latter also points out an interesting feature of the lexicon: the use of a textual condensation device which decreases the data types in order to save space (cf. Gouws and Prinsloo 1997: 51-55). Employing a textual condensation approach, the lexicographer has used the hyphen to substitute the lemma in the illustrative examples. The question to be addressed here is whether this use of a textual condensation procedure corresponds to the needs and reference skills of the potential users of the dictionary. Since the lexicon does not explicitly identify its target reader, the latter is assumed to be high school pupils, students, scholars and learners. This has a lot of implications as far as the data types to be included in a lexicographic work as well as their presentation are concerned. Lexicographers usually make use of textual condensation devices where the target users of a dictionary have a long tradition of dictionary consultation (cf. Gouws 1999: 33). As a matter of fact, the Gabonese community in general and the Fan people in particular are not well familiar with dictionary using skills. According to the treatment given in the lexicon, the lemma **Abum** can be regarded as a polyfunctional lemma in the sense that it mainly appears with verbs and later with nouns, but can also be used as an adverb.

Therefore, instead of using a system of textual condensation the lexicographer should have employed structural markers (**I**, **II**, **III**, ...) to indicate the different cotexts in which the lemma sign **Abum** can possibly occur. These structural indicators are not to be seen as markers of polysemy but they are part of the inner rapid access structure (cf. Hausmann and Wiegand 1989: 356). They should not be presented in a haphazard way but according to a predetermined set of criteria accounted for in the relevant front matter text. These criteria

should include the use of usage frequency as a presentation asset as well as other principles of arrangement.

2.1.3 Concluding remarks

Like Largeau's encyclopedia, Martrou's lexicon fails to give an account of its primary and secondary sources. The fact that has to be mentioned is that the use of different font types (roman, bold, italic, etc.) as well as the running heads have improved the access structure of the work.

With regard to language-specific issues, lexical items have been entered in the lexicon according to the word tradition. Unfortunately the criteria of incorporation of these lexical items is not accounted for in the front matter of the lexicon. The microstructure of the work provides the users with various data categories: paraphrase of meaning, translation equivalents, competence examples, etc. However the lexicographer too often fails to provide the users with a useful entry: the item giving the part of speech of the lemma sign. The lexicon does not clearly identify its target users as well as their reference skills. Consequently no fixed criteria were worked out for the different lemma types and one finds in some articles a nonmotivated use of textual condensation.

3. Galley's *Dictionnaire fang-français et français-fang*

3.1 The macrostructure of the dictionary

The macrostructure of Galley's *Fan-French and French-Fan* dictionary (1964) contains ±13 925 articles listed in strict alphabetical order. It is a record of the language as it is used by the *Fan* community of Talagouga (around Ndjolé). As far as the scope of the dictionary is concerned, the compiler has included words used in everyday speech, geographic names, plant and animal names, proper names and idiomatic expressions as part of his macrostructural elements. It is really a representative collection of *Fan* lexical items in use in Talagouga. There is no doubt that the compiler wanted his dictionary to be a real container of knowledge (McArthur 1985). However, the compiler or in this case the publisher did not mention the criteria according to which lexical items have been lemmatised in the dictionary. This kind of information is usually found in the user's guide in the front matter of the dictionary. The preface of the dictionary merely states why the Association Française des Amis d'Albert Schweitzer has decided to publish Galley's work. This introductory section also gives a brief outline of Galley's missionary, lexicographic and pedagogical activities. But nothing is mentioned as far as the selection of lemmata is concerned. There are several questions likely to arise:

- What are the primary and secondary sources of the dictionary?
- Who are the target users?

- What are the needs and reference skills of the target user of the dictionary?

The above-mentioned questions are not fully addressed in the front matter of the dictionary. Therefore, one has to guess the sources of the dictionary. As far as the primary sources are concerned, the guide to the dictionary clearly states that the work is based on the translation of the Bible in Fañ that Galley did over a period of 40 years for the Fañ community of Talagouga. Special mention is also made of the help of Galley's interpreter: Obam Mendôme. The secondary sources of the dictionary are not explicitly discussed in the instruction book. The preface does mention the existence of an earlier lexicographic work with restricted macrostructural elements. Most probably it is the *lexicon* Fañ-French compiled by Martrou.

Another aspect of the dictionary is that the compiler has not dealt with Fañ alone but has stated the relationship between Fañ and other Bantu languages as it occurs in the article of the lemma **ABAÑ**.

- (6) **ABAÑ** (b) n. 4, pl. *mebañ*. Arbre à bois dur. En Galwa: mandji. Nom commercial: iroko (*Chlorophora excelsa*).

With the entry "En Galwa: mandji" the lexicographer refers the users to the item denoting the lemma **ABAÑ** in Ghalwa, a dialect of the Omyeng language. This is not surprising if one takes into account that the Fañ community of Talagouga and the speakers of Ghalwa live in harmony in the Moyen-Ogooué province and this language contact also results in language borrowing. However, objections can be made with regard to the lack of explanations in the user's guide as far as the system applied by the lexicographer is concerned.

It seems that the lexicographer took into consideration all existing dictionaries and reference works that were of interest for the building up of its own dictionary corpus. These include all the grammar books, religious studies, and syllabi available in Fañ at the time of the compilation phase.

However, a close look at the central word list, specifically in the article of **Mvep**, leads to one of the secondary sources of Galley's dictionary:

- (7) **Mvep** (h) n. 3, pl. *bemvep* Rongeur, appelé en anglais canerat, rat de canne à sucre ..., Largeau dit le même mot. ...

The treatment of the above-mentioned lemma witnesses the fact that Galley has used the *Encyclopédie pahouine* by Largeau (1901) as a secondary source for the compilation of his dictionary. As a matter of fact, Largeau's work is based on the dialects Atsi and Make. Moreover, it is very possible that Galley did not only consult Largeau's encyclopaedia, but also all earlier lexicographic works in Fañ. As far as the French-Fañ section is concerned, monolingual or bilingual dictionaries of French published during the period concerned are most probably the secondary sources of the work.

Even if the introductory section claims that the dictionary is not aimed at a sophisticated public because Galley has deliberately left out the phonetic tran-

scription of his treatment units, it seems that the work is intended to meet the needs and reference skills of students, scholars and people who wish to learn the language. André-Charles Henry (in the preface to Galley's dictionary) is correct when stating that Galley's work is not a book for well-informed people. It is not reserved for initiates. Indeed, the medium used in the dictionary is everything except hermetic. In fact Galley does not want to sound intellectual. His definitions are simple and concise. But the inclusion of data categories such as class numbers, word-formation, structural markers, etymology, definitions, synonyms and homonyms presupposes a sound linguistic knowledge on the part of the target user, even if one does include a mini-grammar as a back matter text. The target users of the dictionary is likely to be scholars and students. Even so the user still needs some guidelines explaining to him/her all the component parts of the dictionary.

In contrast to earlier dictionaries such as the works by Marling (1872), Lejeune (1892) or Martrou (1924), Galley's dictionary is intended to meet the needs of both speakers of *Fañ* and French because of its bidirectional feature. Another feature of the dictionary concerns the access structure: the use of typographical markers such as the presentation of the lemmata in bold small capitals as well as the running heads of the columns which form the outer rapid access structure (Hausmann and Wiegand 1989: 329) make it easier for users to find the word which they are looking for.

3.1.2 The microstructure of the dictionary

The microstructure of the dictionary includes a lot of data, e.g. the part of speech, class number, translation equivalents, meaning explanation as well as cross-references. The lemma is immediately and systematically followed by a tonal indication. This entry is of great importance to the user. Very often, lexicographers compiling dictionaries for African languages have been guilty of ignoring this basic component of the phonological analysis of the language. Galley distinguishes three tone heights in his dictionary, namely: the high tone (h), the low tone (b) and the medium tone (m). Furthermore, one of the front matter texts gives an account of tonal features. The tonal indication is usually followed by a paraphrase of meaning in the target language (French). As far as this point is concerned, a paraphrase of meaning in the source language (*Fañ*) will have empowered the speakers of *Fañ* to a larger degree. Therefore, the dictionary is more likely to meet the needs of the speakers of French than those of *Fañ*. Compare the article of the lemma **Mvia**:

- (8) **Mvia** (h) n. 2, pl. *mimvia* (vb via h). Femme acquise avec la dot de notre fille. En d'autre termes, femme achetée par le beau-père avec la dot reçu de son gendre. Et sa fille dont la dot a été prise pour cela, c'est sa *mbala* (voir ce mot).

Following the tonal indication the user finds the paraphrase of meaning. Moreover, when dealing with the meaning, the compiler also makes provision for

information regarding distinctions in meaning (which refers to the polysemic values of one lexical item) (cf. *Training Course of the Bureau of the WAT*). Consider the following lemma:

- (9) **ABANA** (h) n. 4, pl. *mebana* (vb *ban h*). *Abana zal*, fétiche enfoui dans la terre et qui garde le village. *Abana mon*, fétiche pour que le bébé ne tette plus sa mère et qu'elle puisse avoir un autre enfant. *Abana nsôm*, fétiche qu'un ennemi fait au chasseur pour l'empêcher de tuer.

The syntagms *Abana mon* and *Abana nsôm* reflect polysemic expansions of the meaning of the lemma **ABANA** ("fétiche enfoui dans la terre et qui garde le village") contained in the syntagm *Abana zal*. Another interesting feature of the dictionary is the treatment of polysemic terms (lexical items with more than one sense). Compare the treatment of the lemma **ABEÑY**:

- (10) **ABEÑY** (b) n. 4, pl. *mebeñle*. 1. Pincettes de forgeron pour tenir le fer rouge. — 2. Mirliton du chanteur de la danse *akôm*, ... — 3. Barre ou plaque de fer ou d'un autre métal. — 4. Museau allongé du crocodile ou de la civette *mvakh*. — 5. Balle de fusil ou cartouche. — 6. *Abeñy ndame*, chevalet à deux X en pente où l'on met en position penchée les lianes à caoutchouc coupées pour en recueillir le latex. — 7. *Abeñy ntsap*, barre de savon. — 8. *Abeñy ebom*, système de fer qui ferme la cangue du prisonnier. — 9. *Abeñy e nžel*, grande barbe.

In (10), the lexicographer provides the users with data categories accommodated in sections of the article, i.e. the "comment on form" (CF) and the "comment on semantics" (CS) (Hausmann and Wiegand 1989: 353). The entries such as the lemma sign (**ABEÑY**), tonal pattern (b), class number (n. 4) and plural indication (*mebeñle*) which appear in the left core structure of the dictionary article are part of the CF, the rest of the entries on the right core structure of the article belong to the CS (cf. Hausmann and Wiegand 1989: 353-354). Furthermore, the user can regard the lemma **ABEÑY** as representing a polysemous lexical item. Each sense is introduced by a structural indicator or marker 1, 2, ... 9 which form part of the inner rapid access structure of the dictionary (cf. Hausmann and Wiegand 1989: 354-356).

In the article of a polysemic lemma such as in (10), the different distinctions in meaning should not be presented on a random basis but according to fixed criteria which should also be explained in one of the front matter texts. For the lemma **ABEÑY**, one has to guess the principles of arrangement. It seems that the compiler has made the arrangement according to primary and secondary distinctions in meaning. Or more probably, for this example, the arrangement has been made according to the principle of conjunctive and disjunctive distinctions in meaning. Sense 1 of the lemma sign seems to be the literal polysemic value of the lexical item, therefore it is placed first in the article. Senses 2, 3, 4, 5, 6, 7, 8 and 9 seem to be metaphorical extensions of the first sense. Moreover, senses 2, 3, 5, 6 and 8 have more or less related meanings. Because they all have something to do with the lexical item *iron* or with metallurgy. Thus they should have been arranged by the lexicographer in the above-

mentioned order. Senses 7 and 9, in which the figurative meaning does not refer directly to the first subcomment on semantics, could have been introduced later in the listing. Of course another principle of arrangement could also have been used by the compiler, namely: the principle of empirical arrangement in which distinctions in meaning are determined on the basis of frequency counts (cf. *Training Course of the Bureau of the WAT* 17-20).

This article of **ABEÑY** also highlights some aspects of the presentation of compounds which reveals some inconsistencies in the dictionary. *Abeñy ndame*, *Abeñy ntsap*, *Abeñy ebom* and *Abeñy e nzel* are compounds with the structure: noun plus (connective) noun and they are all written without hyphen. However, the following lemmata: **MÔR-A-MVAM**, **MVAKH-FÎN** and **AKAM-NYAR** are all presented with hyphens.

The treatment of homonyms also reveals some shortcomings. Consider the following example in this respect:

- (11) **ABI** (b) (lg) n. 4, pl. *mebi*. Cuisse ...
ABI (h) (bf) n. 4. ss pl. Mal, péché, ...
ABI (h) (bf) adj. Sing. ...Mauvais, ...
ABI (h) (lg) n. 4, pl. *mebi*. 1. Sein, ...

The last three treatment units of the above-mentioned paradigm which, according to the compiler, have the same tonal pattern (*h*) have been given lemma status as homonyms. The question arises on what grounds the target user of the dictionary is going to distinguish between those three members of the homonym paradigm? They should have received superscript numbers, or start with a structural indicator, reflecting frequency of use as it is usually done in such cases. For example, ¹**ABI**; ²**ABI**; ³**ABI** or **1. ABI**; **2. ABI**; **3. ABI**. Numerous examples could be listed to illustrate this systematic shortcoming in the dictionary.

Another feature of the dictionary is the treatment of special-field terms and cultural concepts. Compare the treatment of the article of the lemma **AB-YÔM**:

- (12) **ABYÔM** (h) n. 4, pl. *mebyôm*. Chasse à environ 50 hommes, avec grands filets et chiens. Il peut y avoir 300 mètres de filets. 20 hommes restent près des filets, et les 30 autres s'éloignent pour rabattre le gibier qui consiste surtout en antilopes. Syn.: *ézakh* (b). Proverbe: *Wa tô abyôm ñkîâ, ndôkh mveñ za ña ke noñ ndum*, tu as refusé d'aller à la chasse de ton beau-père, mais il va pleuvoir, et tout le monde restera assis au village, et toi comme les autres.

The compiler starts with a meaning explanation of the lemma **ABYÔM**. Then a full explanation is given of how the hunting takes place and the number of the participants. Those explanations which concentrate mainly on extra-linguistic features are indeed not part of the paraphrase of meaning or *definiens*. However, they highlight some cultural aspects of the life of the Bafan. The user can retrieve a lot of useful information from these entries. One of the ways for lexicographers to give an account of cultural explanations without giving too much

encyclopedic information in the paraphrase of meaning is the use of synopsis articles.

It is also important to mention how the compiler deals with plant names. Consider the following example:

- (13) **TÔM** (b) n. 3, pl. *metôm*. Grand arbre parasol. Nom commercial d'après Meniaud: Tsoumbou. (Nom scientifique: *Parkia Klainei* Pierre). Couleur du bois: gris jaune. Arbre magnifique à voir. *Tôlba tôm* (vb *tole h*), se faire verser goutte à goutte une infusion de bois de *tôm* dans l'œil, comme épreuve pour savoir si l'adultère a été commis (ordalie).
- (14) **MVON** (b) n. 3, pl. *bemvon*. Gros singe à long poils noirs (fourrure), colobe (*Colobus Satanus*). ...

(13) contains the following entries: tonal indication (bm), part of speech (n.), plural indication (*metôm*), scientific name, description and uses of the tree and label. In (8), the compiler basically provides the user with a sophisticated entry: the scientific name of the animal. Except for a few examples (not necessary to be mentioned) where the compiler fails to provide the user with entries such as the scientific name (of the plant or animal concerned), uses of the plant as well as the sources consulted, the treatment of special-field words has been done quite satisfactorily as far as the cultural context is concerned. The user is provided with lexicographic data concerning the fauna and the flora of the natural habitat of the *Fan* people. Besides leaves, roots and barks, fruits, seeds and animal ingredients are listed with their uses in the folk medicines. All these data regarding the cultural context of the lemmata treated are very important for the user who is likely to find plants and animal ingredients displayed for sale on markets.

In the French-*Fan* section of the dictionary, macrostructural elements most probably belong to the everyday vocabulary of French.

The microstructure usually provides the target user with the following data: translation equivalent in *Fan*, sometimes preceded by a synonym or synonyms of the lemma in French, and examples (each example in French is immediately followed by its *Fan* counterpart). Articles vary in length. An extreme illustration is the treatment of the article of the lemma **ARBRE** (almost three columns in length):

- (15) **ARBRE** bois, *éli* (h). Boîte en bois, *évora éli* (bh). ...

The length of the article is the result of the compiler's listing of some 200 trees of the Gabonese forest as well as their uses and descriptions. Moreover, the commercial name of the tree appears with its indigenous name. The lexicographer has tried to deal with special-field lexical items on an equal basis as far as the two sections of the dictionary are concerned. Unfortunately, the above-mentioned article has quite an unusual structure and thus deserves some explanation in one of the outer texts of the dictionary.

3.1.3 Concluding remarks

To a greater or lesser degree, the front matter of the dictionary gives enough information regarding the primary sources of the work. However, it fails to give an account of its secondary sources. There is no doubt that the compiler has used a large lexicographic corpus of the language which makes the dictionary a valuable source of reference and a real container of knowledge.

Macrostructural elements have been selected from various fields (oral tradition, religion, forestry, traditional pharmacopoeia among others). Unfortunately, the lexicographer gives no assessment for incorporation. In addition, the macrostructure does not show features of sophistication, i.e. the use of niching and nesting procedures. The microstructure contains a lot of useful entries. But, no metalexicographic discussion is provided in the user's guide to make it accessible to the target reader. Special-field lexical items are treated in a satisfactory way in the sense that the metalanguage used is accessible for non-specialist users. Description of medicinal and medico-magic plants are of importance for the user to understand the richness of Fañ culture.

4. Conclusion

The contributions under discussion represent dictionaries supplying the same kind of lexicographic data. The macrostructure tends to survey the full vocabulary of the language. Special-field lexical items have received some attention to ensure a good retrieval of cultural information on the side of the users. The microstructure of the dictionaries under discussion includes a lot of data, e.g. word-class indications, paraphrase of meaning, translation equivalents, examples, etc. Despite all the good points, there are still some shortcomings especially in the areas of suprasegmental phonology (absence of tonal indications), metalexicography (consistent lack of guidance addressed at the users) and orthography. These dictionaries can hardly be used to solve writing problems on the part of the users because of the lack of uniformity as far as orthographical conventions are concerned. For example Alexandre (1961) (quoted by Kidida Awak 1980: 11) points out that the Fañ language has acquired "trois transcriptions [catholiques] et deux [protestants] pour environ un million de locuteurs, sans parler des différences morphologiques dues à des calques catholiques sur l'allemand, le français et l'espagnol, et protestants, sur l'américain". To be used in the most efficient way, these dictionaries need to be updated according to the new set of orthographic symbols for the Gabonese languages.

Notes

1. I wish to express my gratitude to Prof. R.H. Gouws, my promoter, for encouraging me to write scientific papers and for making invaluable suggestions on the improvement of this article.

2. In this article I use the official Gabonese spelling and not the traditional semiphonetic spelling employed by Guthrie, Jacquot, Kwenzi-Mikala and others. The underlined e and n represent [e] and [ŋ] respectively. The phonetic sound [ə] is represented by the symbol ə in Gabonese orthography. However, for phonological and practical reasons I have chosen to represent [ə] by a, as in Make.

Bibliography

- Bureau of the WAT.** *Training Course of the Woordeboek van die Afrikaanse Taal (WAT)*. Stellenbosch: Bureau of the WAT.
- Deschamps, H.** 1962. *Traditions orales et archives au Gabon*. Paris: Berger-Levrault.
- Du Chaillu, P.B.** 1863. *Voyages et aventures dans l'Afrique équatoriale*. Paris: Michel Levy.
- Galley, S.** 1964. *Dictionnaire fang-français et français-fang, suivi d'une grammaire fang*. Neuchâtel: Henri Messeiller.
- Gouws, R.H.** 1996. Bilingual Dictionaries and Communicative Equivalence for a Multilingual Society. *Lexikos* 6: 15-31.
- Gouws, R.H.** 1999. *A Theoretically Motivated Model for the Lexicographic Processes of the National Lexicographic Units*. Research report submitted to the Pan South African Language Board.
- Gouws, R.H. and D.J. Prinsloo.** 1998. Cross-Referencing as a Lexicographic Device. *Lexikos* 8: 17-36.
- Grimes, B.F.** 1996¹³. *Ethnologue*. Dallas: Summer Institute of Linguistics, Inc.
- Guthrie, M.** 1953. *The Bantu Languages of Western Equatorial Africa*. Oxford: Oxford University Press.
- 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. *Wörterbücher. Ein internationales Handbuch zur Lexikographie / Dictionaries. An International Encyclopedia of Lexicography / Dictionnaires. Encyclopédie internationale de lexicographie*: 328-360. Berlin: Walter de Gruyter.
- Jacquot, A.** 1978. Le Gabon. Barreteau, D. 1978. *Inventaire des études linguistiques sur les pays d'Afrique Noire d'expression française et sur Madagascar*: 493-503. Paris CILF.
- Kidda Awak, Mairo.** 1990. Historical Background, with Special Reference to Western Africa. Hartmann, R.K.K. (Ed.). 1990. *Lexicography in Africa*: 8-18. Exeter: Exeter University Press.
- Kwenzi-Mikala, J.T.** 1998. Parlers du Gabon. Raponda-Walker, A. 1998. *Les langues du Gabon*: 217. Libreville: Éditions Raponda-Walker.
- Kwenzi-Mikala, J.T.** 1998a. Localisation des parlers du Gabon. Raponda-Walker, A. 1998. *Les langues du Gabon*: 215-216. Libreville: Éditions Raponda-Walker.
- Largeau, V.** 1901. *Encyclopédie Pahouine. Éléments de grammaire et dictionnaire français-pahouin*. Paris: E. Leroux.
- Lejeune, L.** 1892. *Dictionnaire français-fang ou pahouin, précédé de quelques principes grammaticaux sur cette même langue*. Paris: A. Faivre et H. Teillard.
- Marling (R.P.).** 1872. *Dictionnaire fang-français*. New York.
- Martrou, L.** 1924. *Lexique fân-français*. Paris: Procure Générale (des Pères du St Esprit)/Abbeville: Imprimerie Paillard.
- McArthur, T.** 1986. *Worlds of Reference*. Cambridge: Cambridge University Press.
- Metegue N'nah, N.** 1984. *Lumière sur points d'ombre (contribution à la connaissance de la société gabonaise)*. Langres: Imprimerie Guéniot.

- Nyangone Assam, B. and P.A. Mavoungou.** 2000. Lexicography in Gabon: A Survey. *Lexikos* 10: 252-274.
- Raponda-Walker, A.** 1998. *Langues du Gabon*. Libreville: Éditions Raponda Walker.
- Roulet, G.** 1866. La rivière Como au Gabon et les populations riveraines. *Annales des voyages, t. IV*. Paris.
- Sarcleux, Ch.** s.a. *Essai phonétique*. Paris.
- Smit, M.** 1996. *Wiegand's Metalexigraphy as a Framework for a Multilingual, Multicultural, Explanatory Music Education Dictionary for South Africa*. Unpublished D.Litt. thesis. Stellenbosch: University of Stellenbosch.
- Wiegand, H.E. and A. Kučera.** 1981. Brockhaus-Wahrig: Deutsches Wörterbuch auf dem Prüfstand der praktischen Lexikologie. I. Teil: 1. Band (A-BT); 2. Band (BU-FZ). *Kopenhagener Beiträge zur germanistischen Linguistik* 18: 94-217.
- Wiegand, H.E. and A. Kučera.** 1982. Brockhaus-Wahrig: Deutsches Wörterbuch auf dem Prüfstand der praktischen Lexikologie. II. Teil: 1. Band (A-BT); 2. Band (BU-FZ); 3. Band (G-JZ). Wiegand, H.E. 1982. *Studien zur neuhochdeutschen Lexikographie II. Germanistische Linguistik* 3-6/80: 285-373.
- Zgusta, L.** 1987. Translational Equivalence in a Bilingual Dictionary. *Dictionaries* 9: 1-43.

The Compilation of Electronic Dictionaries for the African Languages

D.J. Prinsloo, *Department of African Languages, University of Pretoria, Pretoria, Republic of South Africa (prinsloo@postino.up.ac.za)*

Abstract: Lexicographers increasingly acknowledge the enormous potential of electronic dictionaries. The great capacity and speed characteristic of electronic products, combined with enhanced query and data retrieval technology, pave the way to a new generation of dictionaries unimagined in the paper-dictionary era. It is amazing to see how many of the lexicographer's greatest obstacles disappear in the electronic dictionary. This article will, firstly, attempt to give a perspective on typical features of electronic dictionaries. Secondly, electronic-dictionary entries will be designed as a solution to some of the most burning lemmatization problems encountered by lexicographers for African languages in paper dictionaries.

Keywords: ELECTRONIC DICTIONARY, LEXICOGRAPHY, DATA RETRIEVAL, LEMMATIZATION, CD-ROM, ACCESS ROUTE, POP-UP FUNCTIONALITIES, POP-UP SCREENS, EDUTAINMENT, CROSS-REFERENCING, INFORMATION RETRIEVAL, ENCODING, DECODING, AFRICAN LANGUAGES, NAVIGATION BAR.

Opsomming: Die samestelling van elektroniese woordeboeke vir die Afrikatale. Leksikograwe erken in toenemende mate die enorme moontlikhede van elektroniese woordeboeke. Die groot vermoë en spoed wat kenmerkend is van elektroniese produkte, tesame met die bykomende tegnologie van soektogte en dataopsporing, berei die weg voor na 'n nuwe generasie woordeboeke wat ondenkbaar was in die era van gedrukte woordeboeke. Dit is verbasend om te sien hoeveel van die leksikograaf se grootste struikelblokke verdwyn in die elektroniese woordeboek. Hierdie artikel sal eerstens probeer om 'n oorsig te gee oor die kenmerkende eienskappe van elektroniese woordeboeke. Tweedens sal inskrywings vir die elektroniese woordeboek ontwerp word as oplossing vir sommige van die dringendste lemmatiseringsprobleme wat leksikograwe van Afrikatale by gedrukte woordeboeke teëkom.

Sleutelwoorde: ELEKTRONIESE WOORDEBOEK, LEKSIKOGRAFIE, DATAOPSPORING, LEMMATISERING, CD-ROM, TOEGANGSROETE, OPSPRINGFUNKSIONALITEITE, OPSPRINGSKERMS, OPVOEDKUNDIGE VERMAAK, KRUISVERWYSING, INLIGTINGSHERWINNING, ENKODERING, DEKODERING, AFRIKATALE, NAVIGASIESTAAF.

Introduction

The electronic dictionary, whether on CD-ROM, online, or hand-held¹, will supersede the paper dictionary in ways unimaginable in the paper-dictionary

dimension, just as the computer has completely superseded the typewriter. Many lexicographers are of the opinion that, generally speaking, the paper dictionary has attained its maximum potential. In reference to learners' dictionaries, Bolinger (1990: 144) says that hard-copy dictionaries have reached their capacity and that "any really dramatic advance would burst the covers". Nesi (1999: 56) emphasizes the fact that many useful features such as indexes and cross-reference symbols have been added in the paper-dictionaries to assist the user in the navigation of multiword searches and that coding systems have been refined. However, she (Nesi 1999: 55-56) concludes:

The more information the paper-based dictionary contains, the harder (and more time-consuming) it will become for learner users to find exactly what they need to know, without first having to negotiate a quantity of information that they do not need to know, or cannot process.

Meijs (1992: 152, as reported by Nesi 1999: 65) predicts "the imminent demise of the dictionary as a book":

In a decade or so, on-line dictionaries on disk or CD-ROM will no doubt be the norm rather than the exception.

Atkins (1996: 515-516) takes a less subtle stand:

At last we are liberated from the straitjacket of the printed page and alphabetical order.

It is amazing how many of the lexicographer's major problems of the past centuries simply disappear in the electronic dictionary dimension. For example, the selection problem, what to put in the dictionary and what to leave out, traditionally believed to be one of the lexicographers greatest difficulties, is solved. Limitations on the size of the dictionary are no longer major concerns, especially in online dictionaries. No difficult decisions on lumping or splitting, one of the greatest vexations of the African language lexicographer, have to be made. The dictionary comes alive in a way that is unimaginable in the paper-dictionary dimension. For the first time the user can control data accessing from the dictionary. He/She can decide, for example, to access only those information types that are relevant to him/her at that very moment, for example, definitions, translation equivalents, etymology, examples of use, collocations, citations, and so forth, by simply switching them on or off (as in Table 3).

In exploring the electronic version of the *Grote Van Dale*, Geeraerts (2000: 76) distinguishes three basic functions of that electronic dictionary: *semasiological*, *onomasiological* and *edutainment* functions. He points out that these functions are combined in a single dictionary and that this fact illustrates the enhanced multifunctionality that can be achieved in an electronic dictionary.

It could be argued that the power of an electronic dictionary is rooted in its *storage capacity*, the *speed of information retrieval* and *sophisticated search and*

querying strategies.

The storage capacity of a CD-ROM is about 600 MegaBytes — enough to contain, for example, the 44 million-word database of the Britannica CD™. The twenty-volume Oxford English Dictionary, and the twelve-dictionary compilation of The Sanseido Wordhunter (Sharpe 1995: 48) are both sold on single CD-ROMs. (Nesi 1999: 59)

No need to *compress text* into cramped paragraphs — each definition and example sentence can start on a new line. No need for *obscure codes* — fuller descriptions become possible (e.g. parts of speech and grammar descriptions no longer need to be abbreviated to save space), with hyperlinks to more detailed information. No need to *exclude possible inclusions or extra example sentences* on grounds of space. (Harley 2000: 85)²

A computer database is *almost infinitely extensible*, and so there is more scope for the *inclusion of extra material on any item*. ... Far more examples can be stored and presented to the user. ... The ... dispute about the inclusion or exclusion of encyclopaedic information also disappears. These data can be retained in the database without a balancing reduction in the space used for "purely" dictionary material. (Dodd 1989: 91)

As for speed, regular users of electronic dictionaries will know that the average retrieval time of an entry is but a second or two.

One should also not think of an electronic dictionary as a fixed object or as fixed as a printed dictionary is. Rather one should think of an electronic dictionary in terms of an *organic, changing database* through which a variety of searches and even artificial intelligence queries may be conducted, or, in the words of Dodd, as a "dictionary service":

We are not far from the point at which the *dictionary will cease to be merely a product*, such as a book, or a somewhat more sophisticated substitute for a book, for example, a CD-ROM, which remains as fixed in its contents as a book is, and *will also become a service*. ... Instead of multiple identical copies of a dictionary, sold to users, there would be a *single version of a database, from which clients of the dictionary service obtained the information they required*, much as professionals of various sorts already get abstracts and similar data "on-line". (Dodd 1989: 87)

Consider in this regard the online version of the Oxford English Dictionary (*OED Online*) in which efforts are made to update the dictionary continuously "offering online subscribers a unique opportunity to chart the course of the English language on an ongoing basis" (*OED Online*, Introduction).

Towards multiple access routes to the dictionary³

In contrast to the paper dictionary a variety of sophisticated search and querying strategies become possible in the electronic dimension. In fully exploiting

such strategies the compiler of an electronic dictionary can offer the user an exciting new range of data access routes to the dictionary and at the same time can add a new dimension to traditional paper-dictionary access routes.

Access to and between sources on CD-ROM is quickly *made by a variety of routes. Dictionary, grammar and usage information in COBUILD, LIED and LIAD may be accessed directly, by clicking on a section menu.* OALD on CD-ROM likewise provides a choice of five menus, for the A-Z dictionary, pictures, maps, appendices and games. In LIED and LIAD the user searching one section of the database will also be alerted if extra information about a search word is contained elsewhere, while in COBUILD all sections of the database are accessed during an initial word search; the number of 'hits' in each area are then displayed on the screen and the user can choose whether to access the Dictionary, usage, Grammar or Word Bank. All the dictionaries also allow searches by 'chaining' or 'hyperlinking', a search mechanism by which a double click on a word on screen will call up a dictionary entry for that word. (Nesi 1999: 61)

The major advantage of a computer-based dictionary is that it *permits a whole range of new routes to the data stored in it.* A printed dictionary is condemned to a single method of gaining access to the information it holds, usually an alphabetic route. ... In a truly dynamic dictionary, it should be possible to gain access to any entry by means of any of the pieces of information composing it. *Potential routes are thus limited only to the frontiers of what is contained in the dictionary, combined with possible manipulations or intersections of these items of data.* (Dodd 1989: 87, 88)

Dodd (1989: 89) envisages a series of typical potential search routes such as:

"sounds like A"; "rhymes with B"; "is spelt like C"; "has an etymology of D"; "dates from year/century E"; "is used in style of F"; "is used in technical field G"; "is an antonym of H"; "is a synonym of I"; "is a hyponym of J"; "is a superordinate of K"; "includes the word(s) L in its definition"; "is of grammatical class M", "has syntactic valency or pattern N".

Perhaps the single most promising new search route/strategy currently being developed in electronic dictionaries is one that enables users to find words based on their semantic specifications in contrast to the traditional way of looking up alphabetically ordered lexemes. Formulated differently it means that an information retrieval access route meaning/definition → lemma has been developed in contrast to the traditional route, lemma → meaning/definition.

Disk-based dictionaries are beginning to cast off the constraints of hardcopy and *to move away from a linear approach based on form rather than meaning.* Rogers (1996: 84) points out that electronic replications of paperbased publications are *still 'word-based rather than meaning-based'* even though they offer better search and retrieval facilities. She proposes a *semantically organized dictionary which would take the user from the definition to the word, and which would deal with such queries as 'find me*

the name of the thing which is a kind of boat and which is flat-bottomed and travels on canals and rivers'. (Nesi 1999: 63)

Prospective electronic-dictionary compilers should, however, realize that an electronic dictionary is not merely "a paper dictionary on computer", just as television is not a "picture-radio" or a computer is not an electronic typewriter. The user would not look up the word on-screen in the same way as he/she would look it up in a paper dictionary. Sharpe (1995: 48) rightfully complains that most electronic bilingual dictionaries "seem to use the content of printed dictionaries as their database without making any additions or alterations" thus not utilizing the potential of the electronic dictionary to the full. Atkins (1996: 515-516) agrees:

Dictionaries of the present ... may even come to you on a CD-ROM rather than in book form, but *underneath these superficial modernizations lurks the same old dictionary*. ... It is up to us to take up the real challenge of the computer age, by asking not how the computer can help us to produce old-style dictionaries better, but how it can help us to create something new.

Consider the following example which is an extract from *SeDiPro 1.0* (Prinsloo and De Schryver 2000), a monodirectional Northern Sotho – English dictionary compiled with a basic word-processing package. The paper version can only be used in the direction Northern Sotho → English. However, the basic built-in search function of a typical word processor is useful in looking up Northern Sotho words, and this function can even be used to look up English words, thus simulating the existence of an English → Northern Sotho dictionary.

thuša help, assist, ~go useful, handy, serviceable, profitable, who helps
thušanō mutual aid/assistance, co-operation
thušō help, assistance, aid, ~ya potlakō/tlhaganēlō/pele first aid
thušopele first aid
thušwa be helped/assisted/aided
thušwe, thušwê must be helped/assisted/aided; ..ga/sa/se..~ not be helped/assisted/aided

The user can type any English word (or part of it) into the "Find/ Search" box, for example, "help", "assist", "co-operation", and so on, and find the Sepedi lemma-sign in which article the requested English word is included. However useful this "added value" to the user might be, it does not mean that the dictionary can be regarded as an electronic dictionary. It still offers little more than searchable versions of the printed text.

Dodd (1989: 87, 88) emphasizes the limitations in this regard in reference to the *Oxford English Dictionary*. Also the very useful *Pharos Dictionaries 5 in 1* (2000) which is marketed as an electronic dictionary, is to a large extent merely five paper dictionaries on CD-ROM. A true electronic dictionary should offer

much more than such "value-added" features. It should reflect a totally new design that exploits all available electronic features.

Electronic formats have big advantages over paper formats, a fact that has too often been sadly ignored when electronic dictionaries have provided little more than searchable versions of the printed text. ... The important thing to note is that it is a new design process. (Harley 2000: 85)

In the following paragraph a brief overview of typical functions of electronic dictionaries is given especially for the benefit of the reader who might not be so familiar with electronic dictionaries. For a detailed discussion of the numerous advantages of electronic dictionaries and their functions, see De Schryver (forthcoming).

Enhanced information retrieval processes

Table 1 represents the typical layout of an entry in the electronic Portuguese dictionary *Diciopédia, Grande Dicionário Enciclopédico Multimédia* (1997). It reveals a neatly integrated layout of pronunciation, two clickable camera icons, a full treatment of the lemma, an alphabetical scroll bar, a general navigation bar, and so on.

Table 1

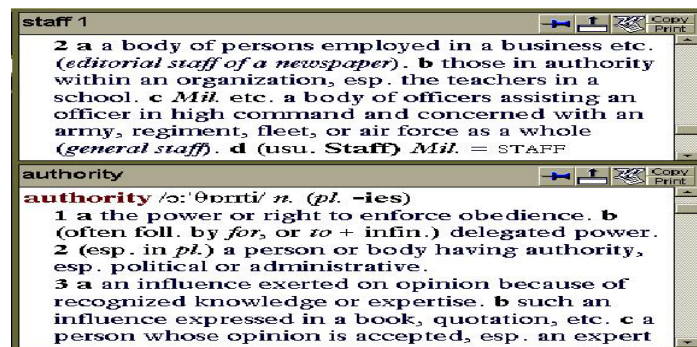


Probably the most exciting feature for a first-time user who is learning to master the pronunciation of words in the target language, is the function that

enables the user to actually hear the pronunciation of words through the speakers attached to his/her computer. This is normally achieved by simply (double-)clicking on the phonetic transcription of the lemma, or, as in this case, on a specially designed icon. No phonetic orthographical presentation, be it a formal orthography such as IPA or an informal respelling of the word, can even come close to hearing the lemma pronounced by live male or female voices. Paper dictionaries have probably reached their maximum potential as far as guiding the user in pronunciation is concerned and the electronic pronunciation function is a good example of what is meant by "the electronic dictionary superseding the paper dictionary".

The second exciting innovation is user control over cross-referencing in the electronic dictionary. In the past the user had to page manually from an entry at the beginning of the dictionary to a cross-reference at the back, meanwhile forgetting along the way what he/she was looking for. In the electronic dictionary dimension the user can now read a definition and if any word is unclear he/she can immediately access a full cross-reference by simply double-clicking on the unclear word. Consider the following example from the electronic version of the *Concise Oxford Dictionary* (ninth edition on CD-ROM).

Table 2



While looking up the word *staff* in this electronic dictionary, the user is confronted with the word *authority* which is also unknown to him/her. However, by immediately double-clicking on *authority* a definition of this word is presented. In this dictionary the user has the added advantage that both definitions can be displayed on screen simultaneously, thus bringing both reference entry and reference address visually together.

The user can even adjust the font size of the different information types relative to each other in order to easily retrieve information.

Some electronic dictionaries such as the *Oxford English Dictionary* (second edition on CD-ROM, OED II), allow the user to customize the dictionary by offering him/her limited options of switching information categories on or off.

In OED II the user has the choice to switch the categories *etymology*, *definition* and *quotations* on or off. Compare, for example, four different information offerings which can be achieved in this way for the word *Miliola*.

Table 3

1	<i>etymology:</i> OFF <i>definition:</i> OFF <i>quotations:</i> OFF	Miliola (mɪ'lɪəʊlə). <i>Zool.</i> Pl. -æ.
2	<i>etymology:</i> ON <i>definition:</i> OFF <i>quotations:</i> OFF	Miliola (mɪ'lɪəʊlə). <i>Zool.</i> Pl. -æ. [mod.L., dim. of L. <i>milium</i> millet.]
3	<i>etymology:</i> ON <i>definition:</i> ON <i>quotations:</i> OFF	Miliola (mɪ'lɪəʊlə). <i>Zool.</i> Pl. -æ. [mod.L., dim. of L. <i>milium</i> millet.] An important genus of imperforate foraminifera; an animal of this genus.
4	<i>etymology:</i> ON <i>definition:</i> ON <i>quotations:</i> ON	Miliola (mɪ'lɪəʊlə). <i>Zool.</i> Pl. -æ. [mod.L., dim. of L. <i>milium</i> millet.] An important genus of imperforate foraminifera; an animal of this genus. 1836 Buckland <i>Geol. & Min.</i> (1837) I. 385 The <i>Miliola</i> , a small multilocular shell, no larger than a millet seed, with which the strata of many quarries in the neighbourhood of Paris are largely interspersed. 1879 Carpenter in <i>Encycl. Brit.</i> IX. 376/2 The shells of the <i>Miliolæ</i> ..are at present found in the shore sands of almost every sea.

Electronic dictionaries such as the OED II enable the user to run full text searches of the lemma which amount to the generation of concordance lines. Consider the following selection from a full text search of *run* in OED II.

Table 4

aback, and run the hazard of being dismayed. 1870 Daily News Sept. 16 This pr
bandon to run at one's own discretion, without restraint, impetuously. See also
they have run it down, and are closing round it. to stand at abay, said of the dog
under 16, run away from home together. Neither abducts the other. To draw aw
the long run renders impossible, should be introduced and carefully fostered. 15
go, about run, about stand; these have sometimes been connected by hyphens, t
he might run the chance of 'getting a bit above hissself', as horsedealers graphica
3 Horses run best when they are above themselves. 1937 V. Woolf The Years 282
those who run down the incognito, nor those who speak it fair, have ever condes
makes Sammy Run? x. 190, I know we're going to knock them for a row of Acade
principles which run directly counter both to the theory and the practice of the
to be run. 1971 Daily Tel. 18 Feb. 2/3 Formal acceptance trials with the 1906A wi
not yet run eight yeares accomplished. 1726 Vanbrugh Relapse i. iii. (1730) 17, 11
account has run..to eighty one. 1844 Lingard Hist. Anglo-Saxon Ch. (1858) II. 3

Learning the language while using the electronic dictionary can be great fun for children. Electronic dictionaries offer a lot in terms of *edutainment* by allowing the user to click on words and pictures and play spelling or word games. Consider the following example from *The Dorling Kindersley Children's Dictionary* (1996).

Table 5



In this example taken from the spelling game, the user has been presented with the picture of cargo on a ship but has given the incorrect final consonant. This mistake is then clearly marked and he/she is given the opportunity to correct it. Correct and incorrect efforts are reacted to by a variety of sound effects.

A perspective on crucial lemmatization problems in the African languages

Lexicographers generally agree that lemmatisation, especially of nouns and verbs, in the African languages is extremely problematic. The lexicographer is, in the words of Prinsloo and Gouws (1996: 103), the *mediator between linguistics and the everyday dictionary user*. For the African languages it means that the lexicographer must find lemmatisation strategies that result in a user-friendly end product. However, as will be indicated in terms of Van Wyk (1995) below, instances even occur where neither the user nor the lexicographer knows for sure what the exact lemma-sign should be, especially in so-called "stem dictionaries" where lexical items are lemmatized by their stem forms, e.g. **mosadi** 'woman', under the stem **-sadi**. The core of the problem lies in a complicated derivational system and a multitude of irregular forms in the language. Such problems are multiplied if the language has a conjunctive orthography.

The treatment of lexical entries in the dictionary has been one of the main difficulties in Swahili lexicography, as in that of many other Bantu languages. *The central problem is particularly the method of arranging the nominal and verbal items of the language, emanating from the complex morphological structure common to Bantu languages, of a nominal classification system categorising nouns by means of prefixes and a verbal derivation system forming new verbs by means of derivational affixes.*

u	subject
me	present tense perfect aspect
tu	object
kut	verbal base ("meet")
an	reciprocal
ish	causative
a	declarative

The main task for the lexicographer will be to decide on which lexical forms should be listed in the dictionary and in what manner. (Bwenge 1989: 5, 6)

Furthermore, lexicographers report a constant struggle in the paper dimension against the redundancy factor when trying to increase user-friendliness.

Het punt is dat "een connectief + **balume**" lemma-status heeft en dus gelemmatiseerd zou moeten worden in de macro-structuur. Dit zou echter betekenen dat men deze informatie ofwel op 15 (aantal *verschillende* connectieven) plaatsen herhaalt; ofwel 14 verwijzingen voorziet naar één bepaald connectief. Daarbij komt nog dat men deze werkwijze dan ook zou moeten herhalen voor alle lemma's beginnend met een connectief! Een *ontoelaatbare redundantie* dus. (De Schryver and Kabuta 1998: xiii, original italics)

Such complexities can lead to major problems in a normal printed dictionary, obliged to be of much greater bulk than otherwise, through including the varying forms at least as cross-references to the normal headword. The alternative is to *force the unsophisticated user to wearisome plodding through potential entries, with no certainty of success.* (Dodd 1989: 90)

As far as the lemmatization of *nouns* on the macrostructural level is concerned, compilers generally fail in their efforts to lemmatize them satisfactorily within the physical limitations of a printed dictionary while still producing a user-friendly product, or they err by including nouns unlikely to be looked for by the target user at the expense of essential ones.

Physical limitations on volume (generally based on the number of pages and therefore on the amount of entries that can be accommodated in a specific dictionary or subdictionary) have a far greater impact on lemmatization in African languages than one would expect. It amounts to the need for a *strategy of word selection* (that is, a strategy to determine which words are to be chosen and which words can be left out of the dictionary) and the absence of such a strategy.

One of the basic problems of lexicography is to decide what to put in the dictionary and what to exclude. (Tomaszczyk 1983: 51)

Selection is guided by usefulness, and usefulness is determined by the degree to which terms most likely to be looked for are included. (Gove 1961: 4a)

The decision what to include in the dictionary still has to be made by the lexicographer himself, however, and this depends in turn upon the nature and size of the dictionary and its intended users. In this respect lemmatised frequency-lists can be a further help ... We have reached a stage where co-operation between man and machine is useful and perhaps indispensable in making better dictionaries. (Martin et al. 1983: 81-2, 87)

Lexicographers constantly have to make pragmatic decisions on what to include

in a dictionary to conform to the dictates of space available. (Walter 1996: 640)

This fact, namely the need for selection in the sense of which nouns to include or exclude from the dictionary, whether to lemmatize singular and plural forms or singular forms only, whether to lemmatize noun stems, and the procedure for handling irregular forms, is an impediment for many compilers of African language dictionaries. (See Prinsloo and De Schryver (1999) for a detailed analysis of problematic aspects regarding the lemmatisation of nouns and an evaluation of different lemmatisation strategies.)

Van Wyk (1995: 89) emphasizes the fact that the prefix morphology of nouns, apart from being irregular, is also subject to fairly complex morpho-phonological rules. He elaborates on this issue in great detail and analyses existing dictionaries where nouns are lemmatised on their stem forms. It is clear that in many instances (such as stems containing the *nasal prefix* of class nine or *aspirated* and *non aspirated* noun stems) it is simply not possible for *neither the user or the lexicographer to determine what the stem is*. This results in great inconsistency and even the abandonment of the dictionary's editorial policy. This immense difficulty for lexicographers can be summarised in terms of the following quotation from Van Wyk (1995: 90):

This forces the lexicographer to choose between four options. (1) He may lemmatize all such stems with aspirated consonants, irrespective of whether they occur in this form or not, e.g. *impala* 'impala', *intaba* 'mountain', *impilo* ['health'] and *inkosi* ['king'] as **-phala**, **-thaba**, **-thombi** and **-khosi** respectively. (2) He may enter all stems with unaspirated consonants, i.e. **-pala**, **-taba**, **-tombi** and **-kosi**. (3) He may enter those stems which also occur with aspirates under the aspirates, and stems which do not under the unaspirated explosives, e.g. **-thombi**, **-khosi**, **-pala** and **-taba**. (4) Or he may enter stems with corresponding aspirates under the aspirates, and those without under the relevant nasal compounds, e.g. **-thombi**, **-khosi**, **-mpala** and **-ntaba**. ... DV [Doke and Vilakazi 1948], and virtually all other stem lexicographers, opt for the fourth option. This puts the onus on the user to know which nouns have stems to which the de-aspiration rule applies and which not. The result is quite confusing, as the following examples show.

NOUN	LEMMA
<i>impala</i> 'impala'	-mpala (impala)
<i>impilo</i> 'health'	-philo (impilo) [<phila]
<i>intaba</i> 'mountain'	-ntaba (intaba)
<i>intombi</i> 'girl'	-thombi (intombi) [<thomba]
<i>ubuntombi</i> 'maidenhood'	-ntombi [<intombi]
<i>inkosi</i> 'king'	-khosi (inkosi amakhosi)
<i>inkabi</i> 'ox'	-nkabi (inkabi)

In the case of verbs, numerous derivations have to be considered by the lexicographer and will be illustrated by the Sepedi verb *reka* 'to buy'.

In the case of verbs, Ziervogel and Mokgokong (1975) offer a complicated

tiered layout of 18 modules, each representing the root, with or without one or more suffixes. For each module, "standard modifications" (as Prinsloo (1994: 96) calls them) are added. These "standard modifications" (the *perfect*, *passive* and *perfect plus passive*) are given in the third column of Table 6. Finally numerous nominal derivations within each of the modules are given. These are listed in the fourth column of Table 6.

Table 6

#	structure	derivations	deverbatives
1	root + standard modifications	<i>reka, rekile, rekwa, rekiwe</i>	<i>direkarekane, lereko, mareko, moreki, bareki, sereki, direki, sereko, direko, theko, ditheko</i>
2	root + reciprocal + standard modifications	<i>rekana, rekane, rekanwa, rekanwe</i>	<i>barekani, thekano, dithekano</i>
3	root + reciprocal + causative + standard modifications	<i>rekantšha, rekantšhitše, rekantšhwa, rekantšhitšwe</i>	<i>morekantšhi, barekantšhi, serekantšhwa, direkantšhwa, thekantšho, dithekantšho</i>
4	root + alternative causative + standard modifications	<i>rekanya, rekantše, rekanywa, rekantšwe</i>	<i>morekanyi, barekanyi, serekanywa, direkanywa, thekanyo, dithekanyo</i>
5	root + neutro passive + standard modifications	<i>rekega, rekegile</i>	
6	root + applicative + standard modifications	<i>rekela, reketše, rekelwa, reketšwe</i>	<i>borekelo, morekedi, barekedi, morekelwa, barekelwa, serekelo, direkelo, thekelo, dithekelo</i>
7	root + applicative + reciprocal + standard modifications	<i>rekelana, rekelane, rekelanwa, rekelanwe</i>	<i>barekelani, thekelano, dithekelano</i>
8	root + causative + standard modifications	<i>rekiša, rekišitše, rekišwa, rekišitšwe</i>	<i>morekiši, barekiši, serekišwa, direkišwa, thekišo, dithekišo</i>
9	root + causative + reciprocal + standard modifications	<i>rekišana, rekišane, rekišanwa, rekišanwe</i>	<i>barekišani, thekišano, dithekišano</i>
10	root + causative + neutro passive + standard modifications	<i>rekišega, rekišegile</i>	
11	root + causative + applicative + standard modifications	<i>rekišetša, rekišeditše, rekišetšwa, rekišeditšwe</i>	<i>borekišetšo, morekišetši, barekišetši, thekišetšo, dithekišetšo</i>
12	root + causative + applicative + reciprocal + standard modifications	<i>rekišetšana, rekišetšane, rekišetšanwa, rekišetšanwe</i>	<i>barekišetšani, thekišetšano, dithekišetšano</i>
13	root + reversion transitive + standard modifications	<i>rekolla, rekološše, rekollwa, rekološšwe</i>	<i>morekolli, barekolli, serekollwa, direkollwa, thekollo, dithekollo</i>
14	root + reversion transitive + reciprocal + standard modifications	<i>rekollana, rekollane, rekollanwa, rekollanwe</i>	<i>barekollani, thekollano, dithekollano</i>
15	root + reversion transitive + applicative + standard modifications	<i>rekollela, rekolletše, rekollelwa, rekolletšwe</i>	<i>morekolledi, barekolledi, thekollelo, dithekollelo</i>
16	root + reversion transitive + applicative + reciprocal + standard modifications	<i>rekollelana, rekollelane, rekollelanwa, rekollelanwe</i>	<i>barekollelani, thekollelano, dithekollelano</i>
17	root + reversion transitive + causative + standard modifications	<i>rekolliša, rekollišitše, rekollišwa, rekollišitšwe</i>	<i>morekolliši, barekolliši, thekollišo, dithekollišo</i>
18	root + reversion transitive + causative + reciprocal + standard modifications	<i>rekollišana, rekollišane, rekollišanwa, rekollišanwe</i>	<i>barekollišani, thekollišano, dithekollišano</i>

Ziervogel and Mokgokong (1975) deserve some credit for their exhaustive treatment of verbs, but they produced an extremely unfriendly product in the end. (See Prinsloo (1994) for a detailed discussion.)

Towards a solution in the electronic dimension of lemmatization problems in the African languages

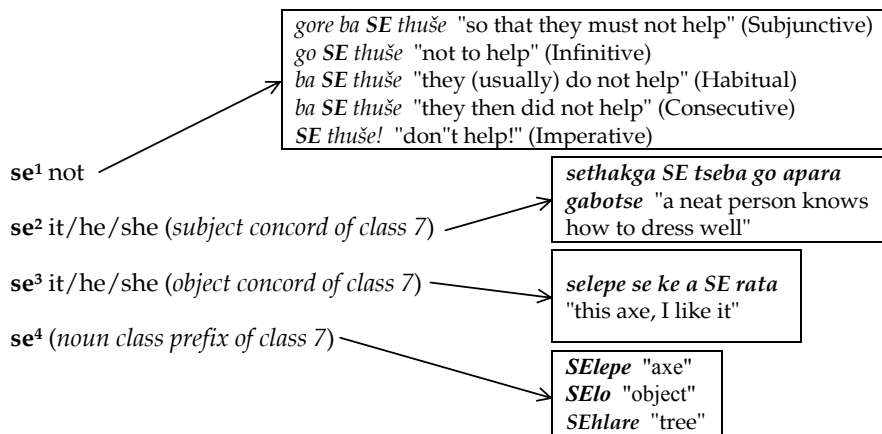
After the presentation of the principal problems with lemmatization in the African languages, it will now be attempted to design typical prototypes of (mainly) pop-up screens to solve these "major" lemmatization problems that are encountered in the paper dictionary.

Adding information is facilitated by the pop-up functionalities of the electronic product. Information that would take up too much space in the classical dictionary, can now be made accessible from within an entry without disturbing the layout of the article. (Geeraerts 2000: 77)

Prospective lexicographers could adapt the following prototypes for any African language or use them as models for lemmatization problems other than those for nouns and verbs considered here. Most of these designs are deliberately proposed within the *capabilities and limitations of an ordinary word processor*. This means that sophisticated multimedia capabilities are not, per definition, a prerequisite for "adding electronic value" to the dictionary.

In principle a variety of approaches is possible. The lexicographer could for instance decide to compile a dictionary entry which resembles the typical layout in a conventional paper dictionary as in Table 7, linking the pop-up screens directly with selected words in the article.

Table 7



In this example the inexperienced user has to deal with different kinds of *se* in Northern Sotho which can vary in function from being concords and negation morphemes to being affixes. Text boxes are *directly linked to the lemma-sign* in each case. In moving the cursor over *se*¹ to *se*⁴ different boxes open up, each containing vital semantic, morphological and syntactic information. In the case of *se*¹ the user is guided towards distinguishing between different moods of the verb in which this negative morpheme occurs. The information given for *se*² and *se*³ are short and straightforward examples of the typical use of the subject and object concords respectively. In the case of *se*⁴ a few examples of nouns in Class 7 consisting of this *se* plus a nominal stem are given. The only effort required of the user is to rest the cursor momentarily on the lemma-sign; he/she does not even have to click the mouse!

Such dictionary articles can be short but multifunctional in that they simultaneously serve the decoding needs of the more experienced user, as well as the encoding needs of the less experienced user. For decoding purposes the skeleton entry might suffice and the user does not have to consider additional information. For the inexperienced user the text boxes offer encoding capabilities which simply cannot be offered within the physical limitations of the paper dictionary. One of the major advantages is that the gap between dictionary and grammar which is generally believed to be "unbridgeable" is starting to close.

In this way, the *differences between dictionary and grammar begin to diminish*: the dictionary entries are linked to a grammatical description of the language that offers more detail than the grammatical compendium that is sometimes included with paper dictionaries. (Geraerts 2000: 77)

Instead of, or in addition to, linking text boxes directly to the lemma-sign a *navigation bar*, marked with a special symbol, **■**, can be introduced as shown in Tables 8 and 10.

Table 8

leratô love, affection, passion; passionate; ~ng in love
■ structure; pronunciation; combination; frequency; concords; idioms; expressions;

↓ picture

Class 1 <i>monna</i>	Class 7 <i>selepe</i>
Class 2 <i>banna</i>	Class 8 <i>dilepe</i>
Class 3 <i>moswe</i>	Class 9 <i>nku</i>
Class 4 <i>meswe</i>	Class 10 <i>dinku</i>
Class 5 <i>lerato</i>	Class 14 <i>bogobe</i>
Class 6 <i>masogana</i>	

In the case of nouns, the *noun class system* could be presented in an innovative way. In Table 8 the user looks up the word *lerato* and finds the translation

equivalents "love", "affection", "passion", and so on. If the user now puts the cursor on *structure* in the navigation bar, a text box opens, not only reflecting the *total scope of the noun class system*, but also *putting the word itself within its appropriate position in the noun class system*, namely class five. In addition, a variety of information regarding *structure, pronunciation, typical combinations and collocations, frequency of use, concords, idioms, expressions* and so on, can be accessed from the navigation bar (see also Table 10).

All this is achieved by moving the mouse over different sections of the navigation bar. Thus, information boxes only appear if the user wants to see them. This innovation is impossible in the paper-dictionary dimension where the lexicographer has to decide whether to include or exclude information, and where information is normally excluded since space is severely limited and all presentations in a paper dictionary are at surface level — there is no opportunity to switch information categories on and off.

COBUILD takes matters further, by *excluding all overt grammatical classification from entries, relegating it to a special margin along with synonyms and antonyms*, although a careful choice of definition formats permits the reader to deduce the word class of an entry. *What is true of word class and syntactic information also applies to the synonyms and antonyms just mentioned, to phonetics, to illustrative citations, in brief to all the potential elements of information that constitute a dictionary. In a sense this is an attempt to achieve the dream of the perfect dictionary, which should have in it one entry — just the word the user is seeking at the moment in question, with just the information needed and no other.* (Dodd 1989: 92)

In the case of verbs, we can recall the lemmatisation problems that resulted from the complicated derivation structure discussed in the previous paragraph. In Table 6, 18 different modules were distinguished for every verb (module one is repeated for the ease of reference):

Table 9

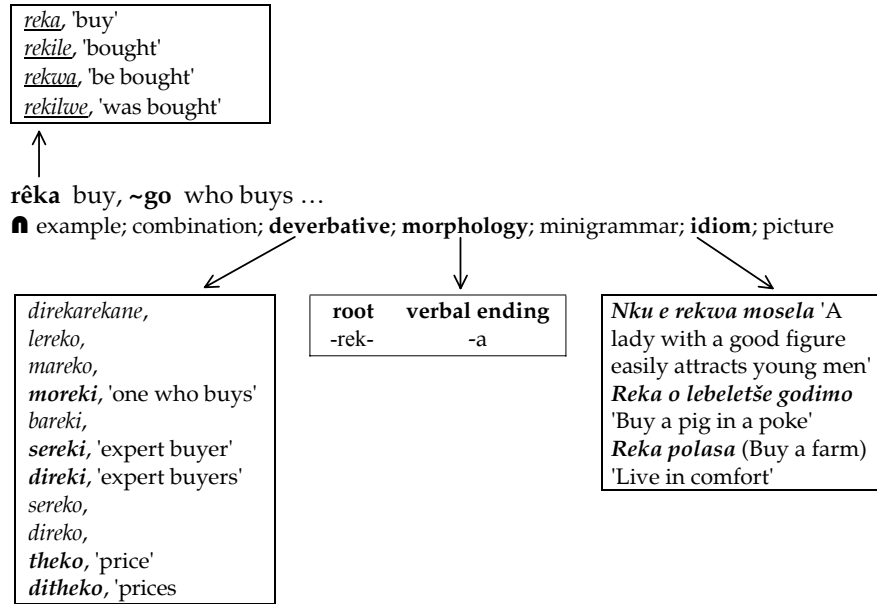
1	root + standard modifications	<i>reka, rekile, rekwa, rekilwe</i>	<i>direkarekane, lereko, mareko, moreki, bareki, sereki, direki, sereko, direko, theko, ditheko</i>
---	-------------------------------	-------------------------------------	---

A suggested treatment for module 1 can be found in Table 10.

Here the lexicographer could opt for linking the "standard modifications" (*perfect, passive* and *perfect plus passive*) to the lemma-sign. The user can also access all the nominal derivations of *reka* from the reference marker *deverbative* in the navigation bar. The value of this information to the user is threefold. Firstly, it gives him/her a quick summary of the different deverbatives with their basic meanings. Secondly, words with a high frequency of use are presented in boldface. Thirdly, boldface is also an implicit cross-reference to a reference address where *full treatment* of these frequently used derivations can be found. Depending on the sophistication of the software, the full treatment can

be accessed directly, by simply clicking the mouse, in the same way that normal cross-referencing is done.

Table 10



As far as grammatical information is concerned a distinction could be made between giving a basic morphological analysis and giving a cross-reference option to a full minigrammar. Constant reference to the minigrammar can be so crucial to some users that De Schryver and Kabuta (1998) repeat it as a footnote on every other page of their dictionary.

Table 11

Affix	1pe	1pm	2pe	2pm	1	2	3	4	5	6
NP	mu-	ba-	mu-	ba-	mu-	ba-	mu-	mi-	di-	ma-
SC	"N-	tu-	"u-	nu-	ù/"à-	bà-	"ù-	"i-	dì-	"à-
OC	-N-	-tù-	-ku-	-nù-	-mu-	-bà-	-"ù-	-"i-	-dì-	-"à-
PP	u-	`bà-	u-	`bà-	u-	`bà-	`ù-	`i-	`dì-	`à-

Affix	7	8	11	12	13	14	15	16	17	18
NP	ci-	bi-	lu-	ka-	tu-	bu-	ku-	pa-	ku-	mu-
SC	ci-	bi-	lù-	kà-	tù-	bù-	kù-	pà-	kù-	mù-
OC	-ci-	-bi-	-lù-	-kà-	-tù-	-bù-	-kù-	-pà-	-kù-	-mù-
PP	`ci-	`bi-	`lù-	`kà-	`tù-	`bù-	`kù-	`pà-	`kù-	`mù-

From *idiom* in the navigation bar the user can activate typical idioms and expressions in which the verb being treated occurs. What has been done for module one can then be repeated for the other modules, producing in the end a well-structured but simple layered representation.

Optimized microstructural representation is achieved through the *layered representation of entries*. In a dictionary of the size of the GVD [Grote van Dale], entries may be quite long, with a complex internal structure. In order to facilitate finding one's way through the entries, the entries may be accessed at different levels. There are four levels:

- the headword level, containing information about spelling, pronunciation, hyphenation, grammatical and morphological characteristics, and etymology
- the level of senses, i.e. the definitions constituting the semantic backbone of the dictionary
- the level of nuances and phraseological units (collocations, idioms, proverbs etc.) that belong with a given sense
- the quotations and examples that illustrate senses, or nuances, or phraseological entities.

(Geeraerts 2000: 78)

The great virtue of the layout is that the article is a simple and fixed point of reference. The user does not have to leave the article to find more information.

Finally, an attempt will be made to solve the problem of identification of the lemma especially in conjunctively written languages, as discussed by Van Wyk (1995). An inexperienced learner of isiZulu could easily be confronted by one or more of the following derivations of the word *-hamba* "go":

ihambe, **ukuhamba**, **kayihambi**, **ayehamba**, **sebehamba**, **ngangilihamba**, **ngingahamba**, **hambani**, **ekuhambeni**, **ubehambele**, **ngizihambela**, **owayehambele**, **wamhambisa**, **ayengasahambeli**, **zihambayo**, **ngihambile**, **kabahambanga**

In most cases he/she will be unable to isolate the stem and will thus not be able to look up the word in a paper dictionary. However, in an electronic dictionary the whole paradigm for *-hamba* could be incorporated and linked to the lemma-sign. Depending on the sophistication of the software and the level of markup of the words, a variety of options is possible. In sophisticated applications the user can be automatically rerouted from the orthographic word to the lemma-sign and thus find a full treatment. This goal can even be achieved in a basic word processor by entering the full range of derivations as running text with manual or automatic pointers⁴ to the lemma-sign.

Within this area of new routes to information could easily be included a grammatical analyser that would be able to take any given morphologically marked word and find the corresponding headword, yielding, in addition, an "audit trail"

that would demonstrate what the morphological markings that were detected indicated in terms of case, number, tense, or whatever. (Dodd 1989: 89)

Even the perennial problem of what should constitute a headword, in particular with regard to the thorny problems of polysemy and homonymy, ceases to be relevant, as it no longer matters whether the set of letters "ROW", to give an example, demands one entry, two, or five, on the basis of its single spelling, two pronunciations, and five senses (as verb meaning "propel with oars", noun meaning "voyage in a boat propelled with oars", verb meaning "argue", noun meaning "argument" and noun meaning "file, rank, or series"). (Dodd 1989: 91)

Conclusion

The *multimedia dimension* opens new horizons for dictionary compilation especially for languages with complicated morphological structures such as the African languages. This enormous potential coupled with the power of *electronic corpora* should be maximally *exploited* to enhance traditional access routes for information retrieval and to develop new accessing strategies which are only possible within the electronic environment.

It is important that the newly established South African National Lexicography Units for the African languages should plan ahead in terms of electronic dictionaries. One could predict that the end product for each African language (a 20 to 30 volume dictionary) may never appear on paper but will only be accessible in electronic media because it is estimated that a 20 volume dictionary might cost R20 000 to R30 000 in the year 2010. Thus, if we speak about dictionaries for the African languages beyond 2000 we must maintain a clear perspective on the advantages of electronic dictionaries versus paper dictionaries, even at this relatively early stage of lexicographic planning.

Some argue that electronic sources of reference are playthings of the rich and out of reach of the poor. However, this situation is changing dramatically since the emphasis is no longer on ownership of expensive computer equipment, but on *access* to such equipment. Today even pupils in some primary schools in the poorest rural areas of South Africa have *access* to computers at school.

The printed dictionary will not disappear overnight, nor perhaps ever, given the durability and relative inexpensiveness of books, but the increasing advantages of a dictionary in a dynamic machine form will swing the balance in favour of words retrieved from a constantly changing database and displayed rather than words fixed for ever on paper. (Dodd 1989: 87)

The past is print dictionaries; the present is print dictionaries with some electronic versions of the same text; the future must be print dictionaries and truly electronic dictionaries, compiled afresh for the new medium, enriched with new types of information the better to meet the needs of the multifarious users. (Atkins 1996: 515)

Endnotes

1. Electronic dictionaries can be stored and accessed in a number of different ways. They can be built into a hand-held device, or inserted in a hand-held device via an 8cm CD-ROM or an IC (Integrated Circuit) card. Alternatively, they can be stored on a hard disk or a 12cm CD-ROM for use with a desktop computer. An electronic dictionary may be purchased and used by just one individual, but it may be possible for a dictionary on disk to be accessed from all the computers on a local area network, while a dictionary linked to a World Wide Web site can be accessed by users all around the world. (Nesi 1999: 55, 56)
2. Unless indicated otherwise, emphasis in all quotations is added.
3. Detailed discussions of access structures in paper dictionaries are given in Hausmann and Wiegand (1989: 337-339), Gouws (1996: 19-25) and Louw (1999 and 2000). In this article it will not be attempted to formulate access routes to the electronic dictionary in terms of such criteria since a different set of rules apply which cannot necessarily be described in existing theories.
4. Depending on the sophistication of the program, the user, on entering a search word or phrase, can either be prompted to *type in* the suggested lemma or, if sophisticated enough, the software can *directly display* the suggested entry in full.

References

- Atkins, B.T.S. 1996. Bilingual Dictionaries: Past, Present and Future. Gellerstam, Martin et al. (Eds.). 1996. *Euralex '96 Proceedings: Papers Submitted to the Seventh EURALEX International Congress on Lexicography in Göteborg, Sweden*: 515-546. Göteborg: Göteborg University.
- Bolinger, D. 1990. Review of the *Oxford Advanced Learner's Dictionary of Current English*. *International Journal of Lexicography* 3(2): 133-145.
- Bwenge, Charles. 1989. Lexicographical Treatment of Affixal Morphology: A Case Study of Four Swahili Dictionaries. James, G. (Ed.). *Lexicographers and their Works*: 5-17. Exeter: University of Exeter.
- Concise Oxford Dictionary, Ninth Edition. 1996. Oxford: Oxford CD-ROM, Oxford University Press.
- De Schryver, Gilles-Maurice. Forthcoming. Lexicographers' Dreams in the Electronic Dictionary Age.
- De Schryver, Gilles-Maurice and Ngo S. Kabuta. 1998. *Beknopt woordenboek Cilubà – Nederlands en Kalombodi-mfundilu kàà Cilubà (Spellingsgids Cilubà): Een op gebruiksfrequentie gebaseerd vertalend aanleerderslexicon met decodeerfunctie bestaande uit circa 3.000 strikt alfabetisch geordende lemma's en Mfundilu wa myakù ìdì itàmba kumwènèka (De orthografie van de meest gangbare woorden)*. Ghent: Recall.
- Diciopédia, Grande Dicionário Enciclopédico Multimédia. 1997. Porto Codex: Priberam. Porto Editora Multimedia.
- Dodd, W.S. 1989. Lexicomputing and the Dictionary of the Future. James, G. (Ed.). *Lexicographers and their Works*: 83-93. Exeter: University of Exeter.
- Doke, C.M. and B.J. Vilakazi. 1948. *Zulu-English Dictionary*. Johannesburg: Witwatersrand University Press.

-
- Dorling Kindersley Children's Dictionary (The).** 1996. London: Dorling Kindersley Multimedia.
- Geeraerts, D.** 2000. Adding Electronic Value: The Electronic Version of the *Grote van Dale*. Heid, Ulrich et al. (Eds.). *Proceedings of the Ninth EURALEX International Congress. EURALEX 2000. Stuttgart, Germany, August 8th–12th, 2000*: 75–84. Stuttgart: Stuttgart University.
- Gove, Philip B. (Ed.).** 1961. *Webster's Third New International Dictionary of the English Language*. Springfield: Merriam-Webster.
- Gouws, R.H.** 1996. Bilingual Dictionaries and Communicative Equivalence for a Multilingual Society. *Lexikos* 6: 14–31.
- Harley, Andrew.** 2000. Software Demonstration: Cambridge Dictionaries Online. Heid, Ulrich et al. (Eds.). *Proceedings of the Ninth EURALEX International Congress. EURALEX 2000. Stuttgart, Germany, August 8th–12th, 2000*: 85–88. Stuttgart: Stuttgart University.
- Hausmann, F.J. and H.E. Wiegand.** 1989. Component Parts and Structures of Monolingual Dictionaries. 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*. Berlin: Walter de Gruyter.
- 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.
- Louw, Phillip Adriaan.** 1999. Access Structures in a Standard Translation Dictionary. *Lexikos* 9: 108–118.
- Martin, W.J.R., B.P.F. Al, and P.J.G. van Sterkenburg.** 1983. On the Processing of a Text Corpus from Textual Data to Lexicographical Information. Hartmann, R.R.K. (Ed.). 1983. *Lexicography: Principles and Practice*: 77–87. London: Academic Press.
- Meijs, W.** 1992. Computers and Dictionaries. Butler C. (Ed.). *Computers and Written Texts*: 141–65. Oxford: Blackwell.
- Nesi, Hilary.** 1999. A User's Guide to Electronic Dictionaries for Language Learners. *International Journal of Lexicography* 12(1): 55–66.
- Oxford English Dictionary, Second Edition (OED on CD-ROM).** 1989. Oxford: Oxford University Press.
- Oxford English Dictionary (OED Online).** 2001. Oxford: Oxford University Press. <http://www.oed.com>.
- Pharos Woordeboeke/Dictionaries 5 in 1.** 2000. Johannesburg: Pharos and Logos Information Systems.
- Prinsloo, D.J.** 1994. Lemmatization of Verbs in Northern Sotho. *S.A. Journal of African Languages* 14(2): 93–102.
- Prinsloo, D.J. and Gilles-Maurice de Schryver.** 1999. The Lemmatization of Nouns in African Languages with Special Reference to Sepedi and Cilubà. *S.A. Journal of African Languages* 19(4): 258–275.
- Prinsloo, D.J. and Gilles-Maurice de Schryver (Eds.).** 2000. *SeDiPro 1.0, First Parallel Dictionary Sepedi–English*. Pretoria: University of Pretoria.
- Prinsloo, D.J. and R.H. Gouws.** 1996. Formulating a New Dictionary Convention for the Lemmatization of Verbs in Northern Sotho. *S.A. Journal of African Languages* 16(3): 100–7.

- Rogers, M.** 1996. Beyond the Dictionary: The Translator, the L2 Learner and the Computer. Anderman, G. and M. Rogers (Eds.). 1996. *Words, Words, Words: The Translator and the Language Learner*: 69–95. Clevedon: Multilingual Matters.
- Sharpe, P.** 1995. Electronic Dictionaries with Particular Reference to the Design of an Electronic Bilingual Dictionary for English-Speaking Learners of Japanese. *International Journal of Lexicography* 8(1): 39–54.
- Tomaszczyk, J.** 1983. On Bilingual Dictionaries: The Case for Bilingual Dictionaries for Foreign Language Learners. Hartmann, R.R.K. (Ed.). 1983. *Lexicography: Principles and Practice*: 41–51. London: Academic Press.
- Van Wyk, E.B.** 1995. Linguistic Assumptions and Lexicographical Traditions in the African Languages. *Lexikos* 5: 82–96.
- Walter, Elizabeth.** 1996. Parallel Development of Monolingual and Bilingual Dictionaries for Learners of English. Gellerstam, Martin et al. (Eds.). 1996. *Euralex '96 Proceedings: Papers Submitted to the Seventh EURALEX International Congress on Lexicography in Göteborg, Sweden*: 635–41: Göteborg: Göteborg University.
- Ziervogel, D. and P.C. Mokgokong.** 1975. *Groot Noord-Sotho Woordeboek*. Pretoria: J.L. van Schaik.

Dictionary Quality and Dictionary Design: A Methodology for Improving the Functional Quality of Dictionaries

Piet Swanepoel, *Department of Afrikaans, University of South Africa, Pretoria, Republic of South Africa (swaneph@alpha.unisa.ac.za)*

Abstract: Although recent dictionaries for the ESL market have been praised for their innovative design features, the prime concern of users, lexicographers and metalexicographers is the functional quality of the dictionary products provided for the market. The functional quality of dictionaries and the scientific assessment thereof forms the topic of this paper. The functional quality of dictionaries is defined in section 2. In section 3 the current methodological approach to assessing the functional quality of texts in the fields of web design, instructional design and document design is discussed and its relevance for dictionary design is indicated. Section 4 explicates in more detail how this methodology can be used to design functional dictionaries and to assess the functional quality of the design features of existing dictionaries.

Keywords: FUNCTIONAL QUALITY, DICTIONARY DESIGN, WEB DESIGN, INSTRUCTIONAL DESIGN, DOCUMENT DESIGN, RESEARCH METHODS, THEORETICALLY MOTIVATED EMPIRICAL RESEARCH

Opsomming: Woordeboekkwaliteit en woordeboekontwerp: 'n Metodologie vir die verbetering van die funksionele kwaliteit van woordeboeke.

Alhoewel die jongste woordeboeke vir die Engelse aanleerdersmark geprys word vir hulle vernuwend ontwerpkenmerke, is die hoofbelang van gebruikers, leksikograwe en metaleksikograwe die funksionele kwaliteit van die woordeboeke op die mark. Die funksionele kwaliteit van woordeboeke en hoe dit wetenskaplik bepaal kan word, is die hooftema van hierdie artikel. In afdeling 2 word die funksionele kwaliteit van 'n woordeboek nader omskryf. Afdeling 3 gee 'n uiteensetting van die metodologie wat op die gebied van webgebaseerde onderrig en teksontwerp gevolg word om die funksionele kwaliteit van tekste te bepaal en die relevansie daarvan vir woordeboekontwerp word aangedui. Die gebruik van hierdie metodologie vir die ontwerp van funksionele woordeboeke en vir die vasstelling van die funksionele kwaliteit van die ontwerpkenmerke van bestaande woordeboeke word in afdeling 4 bespreek.

Sleutelwoorde: FUNKSIONELE KWALITEIT, WOORDEBOEKONTWERP, WEBONTWERP, INSTRUKSIONELE ONTWERP, TEKSONTWERP, NAVORSINGSMETODES, TEORETIES-GEMOTIVEERDE EMPIRIESE NAVORSING

1. Introduction: The Problem of Dictionary Quality

English lexicographers have responded to the mounting international demand for English by providing the market with a wide range of lexicographic works

to assist in the use and acquisition of English as a second or a foreign language (ESL/FL). This makes the ESL/FL lexicographic market one of the most competitive dictionary markets in the world, and one in which, as Herbst (1999) notes, publishing houses are continually under pressure to come up with innovative products to keep an edge on their competitors and to protect their share of the lexicographic market.

The existing English monolingual advanced learner's dictionaries (MLDs) are a case in point. No less than four of them compete in the market, viz. CIDE, COBUILD2, LDOCE3 and OALD6. All four of them are intended as multifunctional lexicographic tools to assist as wide as possible a target group in whatever SL communicative and learning activities they engage. They have been labelled the "big four", and praised by a number of reviewers as the pinnacle of monolingual learner lexicography, especially so, on grounds of their innovative (re)design features.

The big four MLDs have been praised, but then with a number of provisos, for the innovative way in which the problems of learners have been approached by redesigning various aspects of the content, structure, style, etc. of these dictionaries (cf. Herbst 1999, Rundell 1999 and Scholfield 1999). The three most common problems learners experience when using these dictionaries are:

- finding the relevant information,
- having found it, comprehending it,
- applying what has been comprehended to the specific lexical problem that triggered the dictionary look-up (cf. Bogaards 1998, Scholfield 1982).

The findability problem within dictionary entries is addressed, for example, by basing the ordering of senses on corpus frequency data and the use of so-called "signposts" and/or advance-organizers.

The comprehension problem, on the other hand, is addressed by a number of innovative design features, some of course, older than others, e.g.

- the use of a controlled/limited defining vocabulary in definitions,
- the elimination of all kinds of dictionarese (symbols, labels, some abbreviations, parentheses, etc.) and arcane expressions in definitions and incorporating some of the information traditionally conveyed by these means in the dictionary definitions themselves,
- the use of a full sentence definition format that imitates the style and structure of "folk definitions" of native speakers (cf. especially COBUILD),
- the use of corpus-driven contextual paraphrases as a defining technique to help learners match generic abstract definitions with the specific senses of target words in their contexts of use,

- the use of definitional schemata to ensure comprehensiveness and systematicity in defining the meaning of headwords that belong to the same grammatical and/or semantic class,
- extended information on the paradigmatic sense relations of target words (hyponyms, synonyms, and antonyms),
- the extended use of authentic example sentences to illustrate a target word's collocational features, selectional restrictions and stylistic characteristics in addition to its meaning-in-use,
- the use of extensive (nonverbal) illustrations to support definitions and to clarify a wide range of the semantic features of target words, and
- the use of extended usage notes to elucidate the meaning and use of target words, and especially to disambiguate semantically related words.

Although innovativeness might be a crucial commercial concern, dictionary users, including teachers of English, place a higher premium on the functional quality of the dictionaries they eventually purchase, *i.e.*, on the degree to which a dictionary successfully provides them with the kind of lexical support they need for the various kinds of SL communicative and learning activities in which they engage.

When they state that "lexicography as practice and the theory of lexicography have a common goal, namely to foster the effective use of dictionaries" (Hausmann, Reichmann, Wiegand and Zgusta 1989: XVII), the editors of the comprehensive *International Encyclopaedia of Lexicography* underline the fact that the functional quality of dictionaries should be one of the main issues with which we concern ourselves.

A crucial question for learners, teachers, lexicographers, and metalexicographers alike, should then be:

- How functionally effective are the available ESL/FL English dictionaries in providing learners with the kind of lexical support they need in the various receptive, productive and learning activities in which they engage?

For lexicographers and metalexicographers a second crucial question is:

- If not functionally effective, what kind of dictionaries, both with regard to dictionary type and qua their design (*i.e.* with regard to content, structure, style, presentation and integration in instructional activities and materials, etc.), would *optimize* lexical support for different kinds of users' needs in different contexts of use and acquisition?

In Swanepoel (2000), I argue and demonstrate that, despite the commitment to improving the functional quality of dictionaries, practical lexicography and metalexicographic research are currently troubled by two major methodological problems, *viz.* that

- the evaluation of the functional quality of dictionaries does not feature prominently in either the dictionary design process, nor in current meta-lexicographic research, and
- when it does feature, it is most often based on impressionistic and anecdotal evidence or common-sense heuristics, not on the application of generally accepted scientific methodologies and techniques, such as theoretically driven experimental research.

The crucial point is that the introduction of all kinds of innovative features in monolingual learner's dictionaries does not in itself guarantee an improvement of their functional quality. The functionality of each of these innovative features has to be tested empirically as there is simply no mechanical transfer of the content of dictionaries to the mental lexicon of the learner/user. Any dictionary consultation constitutes a complex cognitive process in which a number of crucial mediating variables determine the success with which various kinds of learners/users infer, process, acquire and use information from dictionaries of various types and designs in different contexts of use.

The second important point is the fact that lexicographic and pedagogic experience or "informed opinion" can guide hypothesis formation in theory construction and empirical testing, but on its own it is not enough to elucidate the real problems learners experience with the functionality of dictionaries. For one, these assumptions may simply be wrong, misguided, or have little empirical support beyond the limits of the lexicographer's, teacher's or even dictionary reviewer's own subjective experience (cf. Sims-Knight 1992).

A telling example of this is the empirical experiment reported on in Cumming, Crop and Sussex (1994) in which the authors question the commonly held belief that the full-sentence definition format (cf. the list above) is superior to the phrasal definition format. The full sentence definition is an innovative feature introduced into modern learners' lexicography by COBUILD, and has now become a standard feature of a number of MLDs. What these authors then demonstrate in their research is that production and performance measures do not in fact vary over the use of the phrasal definition format and the sentence definition format.

Despite the apathy towards questions concerning the real functional quality of dictionaries, the need to address the questions posed above and for appropriate methodologies and techniques to answer them, has not diminished. In fact, it has required a renewed urgency with the advent of electronic dictionaries. As Nesi (2000) points out in her overview of the development of electronic dictionaries, the obsession with the technical possibilities of the new media has given way to a concern with how these technical possibilities can, via the design of such dictionaries, best support the needs and capabilities of the end-user (cf. also K hler Simonsen 2000, Laufer 2000, M ller-Landmann 2000, Sato 2000 and Tono 2000).

Nesi (2000: 846) concludes:

Electronic dictionaries are certainly capable of offering the user choices of source material, search routes, and level of detail ..., yet we still know little of the benefits of different potential options. An exciting new medium demands creative new approaches to dictionary design, and more thorough exploration of dictionary users' wants and needs.

Given the priority the functional quality of dictionaries should have in the field of lexicography, the main aim of this article is

- to explicate how the functional quality problem is approached in three closely related fields, viz. that of document design, web design and instructional design, and
- to indicate the relevance and applicability of the methodologies and techniques of these fields not only for tackling the functional quality problem in the field of (E)SL lexicography, but also for various other kinds of dictionaries.

In section 2, I focus on the problem of defining the notion of "functional quality" with regard to dictionaries, i.e. what has to be measured. In section 3, the current methodological approach to assessing the functional quality of texts in the fields of web design, instructional design and document design are discussed and its relevance for dictionary design are indicated. Section 4 explicates in more detail how this methodology can be used to design functional dictionaries and to assess the functional quality of the design features of existing dictionaries.

It must be stressed though, as pointed out in Swanepoel (2000), that the apathy in the field of lexicography towards empirical research on the real functional quality of dictionaries, does not imply that no relevant research has been done on the topic. As will be indicated in the rest of this article, various researchers in the field have addressed questions relating to the functional quality of dictionaries. The methodological framework outlined in the rest of this article provides a coherent framework within which such contributions can be evaluated and integrated into a coherent methodological programme.

2. Defining "Functional Quality"

2.1 Exploring the Concept of "Dictionary Effectivity"

As indicated in section 1, it is generally accepted that practical lexicography and metalexicography have a common goal, namely to foster the effective use of dictionaries.

However, when it comes to the use of dictionaries, "effectivity" or "efficacy" is a slippery concept, desperately in need of clarification if it is to serve as basis for the design of effective dictionaries.

The editors of *An International Encyclopaedia of Lexicography* fail to define or explicate what they mean by "effectivity", or to provide answers to the such important questions as, for example,

- whether effectivity is a monolithic or complex construct and what its defining parameters would be,
- whether effectivity is a function of the dictionary user, the dictionary itself, or both, or even of a larger number of variables under which the dictionary and the user are included,
- how effectivity translates to concrete design principles for various kinds of dictionaries, for different target users, in different contexts of use, and
- how effectivity should be operationalized and measured.

However, if we pretheoretically define the effectivity of a dictionary as the degree or measure in which a dictionary supports its users in accomplishing certain communicative and learning activities, then it should be obvious that various lexicographers and metalexicographers have touched on the topic. A survey of the literature reveals that it is discussed under various labels and in various guises, as for example, in research on dictionary use (user research: who uses what kind of dictionary in what contexts of use and with what effect). It also features prominently in dictionary reviews and the study of dictionary criticism, and in articles that suggest innovative ways for the improvement of the design of dictionaries. Research that highlights the kinds of problems users experience with monolingual learner's dictionaries (e.g., the findability, comprehensibility and applicability problems referred to by Bogaards (1998)) have direct bearing on the kinds of variables that determine the effective use of dictionaries.

Furthermore, improving the effectivity of the dictionary lies at the heart of the various principles or heuristics for the design and compilation of effective dictionaries expounded in the literature, especially handbooks on practical lexicography. What one misses, however, is an assessment of the validity of these principles, i.e. whether they have been proven empirically to improve the functional quality of dictionaries.

Secondly, from a survey of the literature it should also be obvious that effectivity itself is a complex concept, and that the effectivity of a dictionary is a function of a number of variables. These would include, for example, a number of

- *user variables*, such as users' referential skills, their ability to detect and determine their lexical needs in specific communicative situations, their linguistic competency, their ability to infer and process information provided in dictionaries and their ability to apply this knowledge in various communicative and learning tasks, their learning and cognitive styles and their preferences for certain dictionaries when experiencing gaps in

- their lexical knowledge,
- *dictionary variables*, of which the design elements of dictionaries are amongst the most prominent, i.e. their content, structure, style, layout, and medium (e.g. printed vs. electronic dictionaries),
- *functional and contextual variables*, i.e. the different kinds of functions for which dictionaries are used (e.g. encoding, decoding, translating and learning functions) in different contexts of use (in the classroom, using a CALL application, i.e. electronic format), and
- *external limitations*, i.e. variables that are imposed on the design of dictionaries (e.g. constraints of space, developmental cost, person power).

There may, of course, be a number of other variables not subsumed under those mentioned above. What determines the effectivity of a dictionary is, after all, an empirical question — one that has to be explored through systematic research with the aid of existing research methodologies and techniques. To this end, we are in a dire need of some coherent framework in which all these variables can be subsumed and systematically organised, and their content, effectivity measures (e.g. relevance, efficiency, clarity, accessibility, etc) and their internal relationships (e.g. correlational or causal) can be explored.

What we need, more specifically, is research within this framework, that can provide lexicographers with some very concrete answers to such questions as the following:

- What kind of content should a dictionary have to adequately support the lexical needs of a user for a specific task?
- How does one organize/structure this content (both at the macro- and microstructural level) to optimize its accessibility for various kinds of users?
- To what kind of stylistic criteria should the presentation of dictionary content adhere in order to optimize comprehensibility?
- How does one have to adapt the content, structure, style and layout of a dictionary if it is provided as part of a CALL application to optimize its use in various kinds of decoding, encoding and SL learning activities?

2.2 Redefining "Dictionary Effectivity"

In section 1 of this article, I subsumed the concept of dictionary effectivity under the concept of the functional quality of a dictionary. This concept I defined for heuristic purposes as the degree to which a dictionary successfully provides dictionary users with the kind of lexical support they need for the various kinds of SL communicative and learning activities in which they engage. In the light of the numerous variables discussed above that have an impact on the effectivity of a dictionary we can refine this definition to the following:

The functional quality of a dictionary is the degree to which a dictionary, qua its design features (i.e. its content, structure, style, layout, and medium), within the limits of its intended function(s), and the competencies of its (target) users, successfully provides them with the kind of lexical support they need for the various kinds of SL communicative and learning activities in which they engage in different contexts of use.

This definition aims at providing a framework against which a number of the variables that bear on the effective use of dictionaries can be explicated.

As stated, it is a definition that has been narrowed down to the functional qualities of dictionaries. The concept of "functional quality" is borrowed from research on the communicative quality of document design (cf. Van Waes, Woudstra and Van den Hoven 1994) where it forms part of three explicit criteria — also applicable to the evaluation of dictionaries — on which readers/users assess the quality of a document, viz.

- its functional quality, as defined above (i.e. the quality of help it provides for specific tasks, goals or functions),
- its technical quality (quality of technical features, such as binding, size, portability, etc.), and
- its aesthetic quality (aesthetically pleasing appearance) (cf. Shelby 1994).

Although I will not pursue the nature of the other two criteria/heuristics here, they will also have to feature in more comprehensive and systematic empirical research on the range of variables on which users base their decisions on the quality of a dictionary.

3. Mapping and Measuring Functional Quality

Having determined what constitutes the functional quality of a dictionary, one has to deal with the question of how to go about designing a dictionary that is optimally functional with regard to its intended functions, target users and different contexts of use.

In this section, I will give an overview of how this question with regard to different kinds of documents or texts is approached in three other related fields, viz. document design, web design and instructional design. These three fields share a common generic set of methodological assumptions, models, procedures and techniques, but with specific adaptations according to the end product (documents, web sites, instructional materials) and the processes involved in designing and developing the end product.

In all three these fields the problem of optimizing the functional quality of the end product is approached from two closely related perspectives:

- the process that should be followed in designing the product, and

- the design features the end-product has to possess to ensure its optimal usability/functional quality.

3.1 The Design Process

In all three these fields adaptations of the so-called ADDIE model (Analyse — Design — Develop — Implement — Evaluate) (cf. figure 1) is followed in the design process (cf., e.g., Gaffney 2000, Multimedia Services 2000, Nielsen 1994, 1999, Schellens and Steehouder 1994, Wilson, Jonassen and Cole 1993).

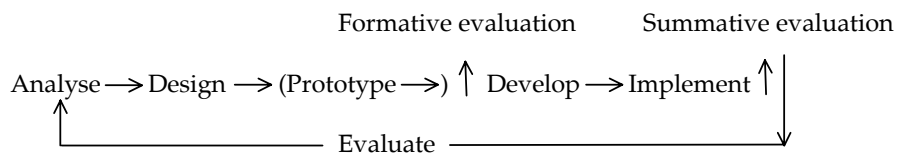


Figure 1: The ADDIE model

Underlying this model is the assumption that the optimal design of an end-product should be approached according to a problem-solving strategy. The model specifies a number of, often overlapping, and cyclic steps that should be followed in the design process to ensure the best design. The main activities and the kind of result of each, can be described as follows:

- Analyse: Determine and analyse all the variables that have been proven (experimentally) to mediate the successful use/functional quality of the document. (Cf. the discussion below.)
Outcome: An analysis report, indicating, for example, the existing need for a document (needs analysis), usage trends, features of competing products, etc.
- Design: Design an outline for the product, i.e. make specific choices with regard to genre, content, structure, style, layout, format and delivery medium, as dictated/motivated by the relevant variables analysed in the analysis phase/plan.
Outcome: Usually a project plan with a style guide in which each of the design elements of the document are specified. Very often this includes a prototype of the document, with detailed descriptions of subsections of the document. In a detailed project plan there are also often specifications with regard to production processes to be followed, specification of sources to be used and, especially in the case of collaborative writing projects, a specification as to who will be responsible for what tasks.
- Develop: During this phase a first/concept draft of the final document is produced according to the specifications of the project plan and style guide, and the final draft after editing.
Outcome: The final document.

- Implement: In the case of a web design and instructional design this refers to the phase in which the product is released for use by the target users in (specified) contexts.
- Evaluation: Although evaluation is stipulated in most versions of the model as a phase that follows on implementation, it forms the cornerstone of all activities/phases in the model specified in figure 1.

Before elaborating on the various phases involved in the design process, the focus will first of all be on the process as a whole and especially on the role of evaluation within the process.

3.1.1 Evaluating Design

According to system theory from which this model originates, any design process requires constant monitoring. Evaluation has to act like a thermostat, always monitoring the system's effectiveness and pointing out what revisions are needed to adapt the system to differences in different contexts of use (Wilson, Jonassen and Cole 1993). Design and evaluation, therefore, go hand in hand.

Specific formative and summative evaluation checks are specified for the design process. Formative evaluation involves quality control of the product in the stage before (mass) production, i.e. in a stage in which adjustments can still be made to a draft, prototype or concept text. Summative evaluation, on the other hand, involves control of the functional quality of a product after (mass) production within its intended contexts of use with its intended users. Results of both these quality checks are fed back into the design process: formative evaluation to the improvement of a concept text, summative evaluation into the redesign of an existing text.

Rapid prototyping is itself a specific form of quality control. Given the high cost of extensive analysis and project planning, a small-scale prototype is constructed that exhibits all the key features of the intended end-product. The prototype is tested (with its intended users) in an effort to get a clearer picture of the requirements to which the end-product will have to adhere. The revised prototype is then used as a model for the construction or development of the final product.

Within the three fields under scrutiny, a number of instruments have been developed to test the functional quality of a text before final production. Of these, expert- and user-orientated methods seem to dominate. In the case of expert-evaluation design experts are asked to determine whether a prototype or end-product adheres, in their opinion, to its intended functions and the needs of the target users. In web design, for example, this form of evaluation — known as "heuristic evaluation" — has become one of the most popular formative evaluation procedures (cf. Nielsen 1994, 1999 and De Jong and Van der Geest 2000). The evaluation itself, proceeds on the basis of how well a design adheres to a set of recognized usability principles (cf. Van der Geest and Spyri-

dakis 2000, and De Jong and Van der Geest 2000 for an overview).

Expert-evaluation has much in common with the text-based approach developed in the field of document design. A number of well-motivated checklists and models have been developed to improve the functional quality of specific text types. As in the case of the heuristics that are suggested to improve interface and web design, these checklists are often meant to serve a double purpose: either as guidelines to be followed in the design of a product, or as instruments to evaluate an existing product.

In the case of user-orientated methods, a sample of the intended target group itself is selected and their response to the functional quality of a text is elicited and evaluated. The main aim again is to pinpoint all the problems users experience with a text, given its intended functions, the users' information needs and their competencies. Think-aloud protocols have proven to be valuable instruments for this purpose. The user-orientated approach dominates in the field of instructional design and the full set of empirical research methods are used, including the tools and techniques offered by CALL applications, such as tracking technology (cf. Lomicka 1998, and Schriver 1989, 1993 for an overview of the user-approaches in document design).

Some researchers have tried to determine which of these methods provide the designer with the best kind of feed-back for improving the design of a product. Although user-evaluation is often given priority, it is well-established that expert evaluation is an adequate instrument to detect the most, and most severe, problems of a text. However, given the fact that the two methods often focus on different kinds of problems with the functional quality of a text, a combination of the two methods more than often gives the best results.

3.1.2 Process Approaches in Dictionary Design

What processes are followed in dictionary (re)design?

Given the high costs involved in publishing commercial dictionaries and given the complexity of the task of coordinating the various processes in producing and publishing a dictionary, one could safely assume that publishing houses do some form of planning for the production of a dictionary, and that they go at least through some of the stages referred to in the ADDIE model. However, given the current state of research, it is not at all sure what of these stages are involved, and, especially, if and what form(s) of assessment of the functional quality of the dictionaries are involved. Given the competition in the field, publishing houses are rather discreet about their trade secrets and very few of the (in-house) documents that are generated in the design process reach the public arena.

Although some publishing houses claim in their sales materials that some form of research underlies the design of their dictionaries, they seldom elaborate on what kinds of research were conducted, what theoretical assumptions and methodologies were involved, or on how the results of such research were fed into the design process. In short, we have very little knowledge of how lexi-

cographers themselves go about the design process and how the process approach is utilized to enhance the functional quality of dictionaries. In this regard the ADDIE model provides a framework to guide such research and for synthesizing existing lexicographic approaches and practices as discussed below.

Within the field of lexicography, proof of the effectiveness of the process approach to dictionary design is provided in De Schryver and Prinsloo (2000). The authors outline a procedure for enhancing the functional quality of a dictionary by eliciting feedback from users on crucial design issues during the various stages in the production of a dictionary.

However, there does not yet exist a comprehensive set of heuristics for either the formative or summative evaluation of dictionaries. The evaluation of dictionaries is, however, the prime concern of dictionary criticism — a clearly demarcated field of research within the field of lexicography. What needs to be done, is

- a systematic description and assessment of the criteria or heuristics that dictionary critics themselves employ in evaluating dictionaries, and
- a description and assessment of the functional quality of the design features that are incorporated into dictionaries of various kinds.
(Cf. Swanepoel 2000 for a discussion of how these issues are currently dealt with in dictionary criticism and in research on the content and structure of dictionaries as texts.)

A general outline of how such evaluation/design criteria and specific design proposals for improving the functional quality of dictionaries can be systematically described and evaluated is provided by the template proposed in De Jong and Van der Geest (2000) for the description and evaluation of web heuristics.

3.2 Analysis

According to the ADDIE model, the first two steps in the design process consists of analysis and then of design informed by the results of the analysis.

The main aim of the analysis stage is to determine and analyse all the variables that have been proven experimentally to have an impact on the functional quality of a document. Given this input, a document is then designed with regard to the choice of its content, structure, style, layout and medium in a way that will optimize its functional quality within the range of requirements set by the mediating variables that impact on its functional quality.

In the field of document design, for example, four crucial variables have been identified that have a direct impact on the functional quality of a document — variables that, therefore, crucially determine the design features of a functional document (cf. figure 2), viz.

- the intended function(s) of a document (What is the document going to

- be used for?),
- the target users (Who is going to use the document?),
- the context of use (In/under what circumstances will the document be used?), and
- the marginal restrictions on the design of the document (What restrictions of time, financial and human resources apply?).

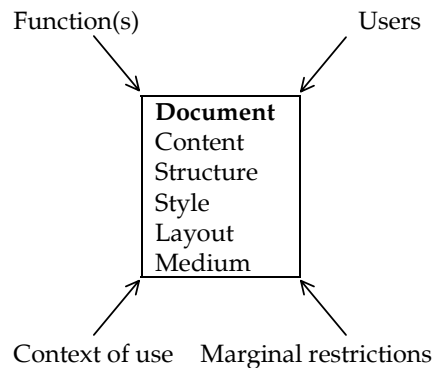


Figure 2: Variables that determine a document's design

In the literature of document design, attention is given to the way in which each of these variables correlate with/determine the functional quality of a document qua its design elements, and to the most appropriate methods to analyse each of these variables (cf., e.g., Hoeken 1995).

3.2.1 Lexicographic Approaches

Although not always explicitly stated as such, much of the descriptive research done within in the field of lexicography in fact also falls within the parameters defined in figure 2. For example:

- User research focusses on the question of who (users) use what kind of dictionary (document) for what purpose (functions) under what conditions of use (context of use). Wiegand (1998) has recently developed a fully fledged, theoretically motivated framework for precisely analysing these variables (users, functions, contexts of use).
- Dictionary criticism, as a form of expert evaluation, focusses on the problems that various kinds of users do or could in principle experience with specific design elements of particular dictionaries.
- Research on the structural elements of dictionaries aims at providing a comprehensive taxonomic description of the design elements of dictionaries which can provide lexicographers with various design options to explore in an effort to optimize the functional quality of a dictionary.

What is crucially missing from dictionary research, however, is that there is hardly any systematic effort to determine experimentally how each of the variables listed above (user, functional, usage variables) in fact impact on the design of a functionally effective dictionary and, thus, what requirements they set for designing a dictionary of a specific kind (cf. Swanepoel 2000). The result is that 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. As indicated in the ADDIE model, such designs of necessity have to be experimentally assessed by way of formative and summative techniques. What is needed, therefore, is experimental research to determine how each of the variables listed above, separately and in conjunction with each other, determine the functional quality of each of the design elements of a dictionary.

3.3 Functional-cognitive Design

In all fields under consideration, it is accepted that user variables have a crucial role to play in the design of a document/dictionary (cf. also the discussion in section 2). In cognitive approaches to document design, it is assumed more specifically that the functionality of a document is crucially determined by the degree to which it supports — qua its design elements — the cognitive processes and representations required of users to successfully consult a document for a specific purpose (function). (Cf., e.g., Hoeken 1995 for a discussion of how cognitive factors impact on the design of documents.)

Given the importance of this approach to the design of effective lexical support in SL instructional design (cf. Groot 2000, Hegelheimer and Chapelle 2000, Plass 1998, and Swanepoel and Van de Poel 2000), the functional-cognitive approach to document design is explicated in more detail below.

In the functional-cognitive approach the design of a document starts off with an analysis of the cognitive processes and representations required for the successful completion of a specific decoding, encoding or acquisition task (e.g. reading or writing in a SL or acquiring the vocabulary of a SL). The second step is to determine how each of the design elements of a document could support this process. To this end, designers rely on what is available by way of empirically supported theories of information processing, and, more specifically, by way of theories of SL processing in different contexts of use. These theories provide the rational basis for the design of a document in as much as the design of the document (i.e. its content, structure, style and layout) is determined by the cognitive processes and representations involved.

Given the hypothetical nature of these processing theories, the design phase is always followed by empirical evaluation of the functionality of such a document by evaluating the success with which the user is able to complete a specific task with the aid of the document under consideration. The results of the evaluation of the functional quality of the document is then fed back into the (re)design process. This approach is presented schematically in figure 3:

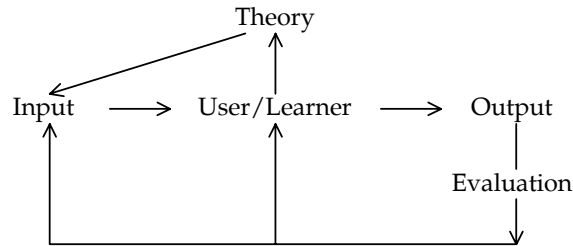


Figure 3: The functional-cognitive approach to dictionary design
(Cf. also Hegelheimer and Chapelle 2000.)

Within the field of SL instructional design, a number of research instruments have been developed to probe online how users process the information provided in documents and what knowledge stores, skills and strategies they employ to this end. These include, *inter alia*:

- think-aloud protocols (i.e. verbal reports by the user while interacting with the document, which can be followed up by interviews).
- constructive-interaction protocols — a variant of the think-aloud protocol — in which pairs of users use a document together and report aloud on how they access and use the information in the document to achieve their specified task.
Transcripts of these interactions are used to infer how users reason during processing of input and what hypotheses they generate while interacting with the content of a document.
- tracker technology which is incorporated in CALL applications and which records every action a user makes while interacting with the input material.
These files are used to determine what material in a document is selected by the user for processing, how often it is consulted and what amount of time is spent on it. Correlations are then determined between this information, the user's output performance, and specific design elements of the input document.

These auxiliary research instruments are used to support the more traditional methods of empirical evaluation, such as attribute-interaction testing (AIT research designs), in which the strength of correlational links between specific design elements of a document and the user's output performance is established. (Cf. Hegelheimer and Chapelle 2000 for a discussion of the methodological issues involved.)

However, an analysis of the cognitive processes, strategies and mental representations required for the successful completion of any task is not enough. As indicated in section 3.1, various user, text, task and contextual variables have been shown to mediate the successful execution of the processes and

representations required for the processing of the information provided in a document. Therefore, in designing a functional document for a specific purpose, the influence of these variables on the successful completion of any task also have to be accounted for in the design of a document.

In line with the assumptions and methodology outlined above, the cognitive-functional approach to document design can be operationalized in the following steps:

- (i) Give a model-theoretic account of the linguistic and other kinds of knowledge, cognitive processes, skills and strategies and mental representations that are required for the successful completion of the task for which the user consults a document.
- (ii) Determine the mediating user, textual, task and contextual variables that have been shown empirically to mediate the successful use of a document for a specific task.
- (iii) Deduce hypotheses from (i) and (ii) with regard to the appropriate design of the input document that would optimize the processing of the required information and design the document accordingly.
- (iv) Empirically evaluate the success of the design of the document and determine if and how its success or failure can be attributed to the underlying theory of information processing.
- (v) If not successful or only partially successful, devise an alternative design and empirically test its functional quality with regard to the intended users, functions and specific contexts of use.
(Cf. also Chun and Plass 1997.)

Obviously, the cognitive-functional approach to design requires explicit, comprehensive and empirically tested theoretical models of all the mediating cognitive variables that underlie the successful processing of the visual and verbal information provided in documents. Despite the affinity in the field of document design and instructional design for Levelt's processing model (cf. Levelt 1989), we have as yet no single comprehensive model to account for all the variables involved in processing information from all types of texts. What we do have, though, is a number of partial theories, each focussing on at least some of the variables involved.

3.3.1 Cognitive Variables and Dictionary Design

Turning now to the field of lexicography, it must be noted that very few researchers, or for that matter lexicographers, have explicitly and/or in a comprehensive fashion linked the design of a dictionary to the cognitive processes and skills required for successful dictionary consultation. Scholfield (cf. Scholfield 1982) was probably one of the first to sensitize lexicographers and researchers to the fact that a set of complex cognitive processes and skills underlie dictionary consultation in any context of use. However, since this seminal article of Scholfield, hardly any theoretical research has been forthcoming that

probes the cognitive complexities involved in successful dictionary consultation, and very few attempts have been made to systematically and in a theoretically motivated way link the design of a specific dictionary to specific theoretical models. As noted in Swanepoel (2000), the same situation holds with regard to experimental testing of the functional quality of the (re)design features of existing dictionaries.

Given the above state of affairs within the field of lexicography, the cognitive-functional approach with its concomitant set of assumptions, methodologies and techniques could be used as a guideline to redress the current problems with functional dictionary design. In the next section, we outline some of the key issues that will have to be addressed in such an approach and relate it to current approaches and experimental findings in SL instructional design, specifically current research on how to design functional lexicographic support for CALL materials.

4. Designing Functional Dictionaries

4.1 Determining Functional and Cognitive Variables

In the prototypical case, dictionaries and other word books have the primary function of providing the user with information on the linguistic — often "encyclopaedic" — characteristics of the lexemes of a language, should such a need arise in the course of executing specific tasks (e.g. decoding, encoding, and acquisition tasks).

In line with the cognitive-functional approach, the first step in designing a dictionary would require that the designer should get answers to all, or at least some of the following questions depending on the kind of major task users engage in:

- (i) Precisely for what kind of tasks (reading, writing, translating, vocabulary acquisition, etc.) are various kinds of dictionaries used?
- (ii) What cognitive processes, skills, strategies and mental representations are involved in the successful execution of such a task?
- (iii) What role does lexical knowledge play in the successful execution of the task?
- (iv) What constitutes lexical knowledge?
- (v) How is lexical knowledge stored and accessed in the mental lexicon?
- (vi) What kinds of gaps could arise in the user's lexical knowledge and what implications does that have for the successful execution of the task?
- (vii) What strategies — other than consulting a dictionary — do/can users employ to compensate for these lexical deficiencies?
- (viii) Under what circumstances will users revert to dictionary consultation to address a gap in their lexical knowledge?

The first set of questions is aimed at elucidating the kinds of lexical needs that could prompt dictionary consultation in specific contexts of use, and, by exten-

sion, the kind of lexical information that a dictionary of a specific kind should include. However, the set of questions above do not only aim to give a designer an answer as to the kind of lexical information that has to be included in dictionary entries for specific tasks, but also as to how users could contextually (i.e. in terms of the local task involved and their assessment of their specific lexical needs) frame their search questions when reverting to a dictionary for help.

A number of the answers to the first sets of questions the designer needs are (in more or less detail) exactly those that user-research is aimed at providing or, at least, have provided partial answers to: what kinds of users use what kinds of dictionaries for what kinds of activities. Wiegand (1998) provides an overview and a critical assessment of existing studies. Dictionary typologies, including those that focus on mixed categories such as the set of electronic dictionaries (cf. Nesi 2000), are an important supplementary source in answering the questions above.

In most user-studies users are differentiated into broad categories on the basis of their linguistic competence in a language, for example, a distinction is made between first, second and foreign language users, and, with regard to the latter, distinctions are made, for example, between beginners, intermediate and advanced users. These distinctions are more often used as the sole basis on which to predict the kinds of lexical knowledge users will/ought to have and, consequently, what kinds of lexical problems they will/could encounter when engaging in specific tasks. However, there are a number of other mediating variables that the designer has to keep in mind when assessing how the information provided by this first set of questions should be evaluated in terms of the implications they have for the design of a dictionary. Most of these are user variables and task variables. For example, users do not only differ individually from one another in terms of their lexical competence, but also in terms of their ability to perform certain tasks (reading, writing, etc.), their ability to detect and determine their lexical needs in specific communicative situations, and their mastery of compensatory skills when confronted with lexical gaps. Likewise, the kinds of tasks users engage in, may also show great variety in the degree of lexical competency they demand.

There are a number of theoretical models that can guide the dictionary designer in analysing the cognitive processes, strategies, skills, and representations that underlie each of the various kinds of tasks users engage in and the way lexical knowledge features in the execution of those tasks. Given its empirical support, Levelt's model for L1 encoding and decoding (cf. Levelt 1989) and adaptations thereof for explaining L2 encoding and decoding activities, lexical gaps and compensatory strategies feature the most prominently (cf. Swanepoel and Van de Poel (forthcoming) for a discussion).

Of the activities that users engage in, reading and the role that lexical knowledge plays in this process has probably received the most attention. Reading in a L2 is a complex skill (cf. Urquhart and Weir 1998), which requires

not only adequate lexical knowledge of the L2, but also knowledge of the topic and world knowledge, plus a variety of inferencing and synthesizing skills. Users have a variety of compensatory mechanisms they can revert to when their lexical knowledge fails them, and, depending on their knowledge of the topic and their skills in inferring the meaning of unknown words from context, are able — in different degrees — to make good what they lack in lexical knowledge.

4.2 Dictionary Consultation

We now turn to the complexities of the dictionary consultation process itself.

In the field of lexicography, the knowledge, skills and strategies that a user has to master for successful dictionary consultation are usually analysed under the rubric of "dictionary reference skills". Broadly speaking, these can be divided into the following composite knowledge, skills and strategies, each of which has also been shown to be those aspects of dictionary consultation with which users have problems (cf. Bogaards 1998 and Scholfield 1982):

- finding the relevant information in the appropriate dictionary,
- comprehending the information provided,
- evaluating the information provided, and
- applying the acquired knowledge to the task at hand.

Each of these can be further broken down into a number of variables.

Finding the relevant information in the appropriate dictionary requires

- knowledge of the various types of dictionaries available, and specifically
 - of the one(s) best suited to provide the necessary information, and
 - given the various formats in which electronic dictionaries provide lexical information, also of the best format for the presentation of the information for the search at hand,
- knowledge of how the information is organized in the dictionary, i.e. of the principles underlying its macro- and microstructure and its cross-referencing system, as well as of the contents and structure of the information contained in its appendices,
- the ability to frame the lexical information required in the (most appropriate) search format required by the dictionary, and
- practical skills in using the above knowledge to locate the required information in a specific dictionary.
(Cf. also Scholfield 1982.)

No uniform macro- and microstructural organization has yet been proposed or adopted by the designers of printed dictionaries, and very little research has in fact been done on what would constitute the most effective kind of macrostructural presentation with regard to such variables as

- dictionary type,
- target users,
- tasks, and
- language.

With regard to printed dictionaries, a number of new design features have been incorporated in learner's dictionaries to assist the learner in locating the correct information on the microstructural level, such as ordering the senses of lexemes according to corpus frequency data, the use of so-called "signposts" and/or advanced organizers. Again, however, the functionality of these features has not been tested empirically, and we have little knowledge of the mediating variables that impinge on their effectivity.

Much of the knowledge and skills referred to above have become redundant in consulting electronic dictionaries. As both Geeraerts (2000) and Nesi (2000) point out, the search facilities provided for in electronic dictionaries eliminate the need for most of this knowledge and skills. Even the ability of formatting appropriate queries is alleviated by the ability of the computer to execute searches (and in various databases) on partially specified strings of a lexeme, if available, by "sound alike" searches, and wildcards. The crucial question, however, is not what is technologically possible, but what users want and need, and this has to be determined empirically.

The first three variables above address the skills and knowledge a user has to possess to locate the information in a dictionary. Once the user has located the required information, interpreting/understanding the information provided in a dictionary article becomes the following hurdle. The complexities faced by both the user and the designer, can be explicated in terms of the following schematic representation:

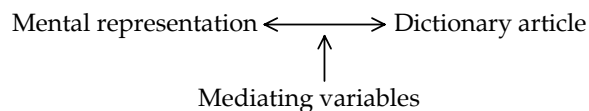


Figure 4: Building a mental representation from dictionary information

To be able to understand (and finally acquire) the information provided in a dictionary on the grammatical features of a lexeme, the user

- has to have a knowledge of the lexicographic principles that underlie the encoding of the lexical information, i.e. the lexicographic codes (symbols, layout and typography) used to identify and differentiate the various information categories (i.e. information on a lexeme's morphology, syntax, semantics, stylistics, usage, etc.),
- must be able to identify which information is encoded explicitly and which only implicitly, and

- must be able to interpret/understand the information provided (explicitly and implicitly) on the grammatical features of a lexical item, i.e. build a mental representation of the lexical information provided in a dictionary article.

For the dictionary designer the crucial questions are:

- What constitutes the grammatical characteristics of a lexeme, i.e. what information should be included in a dictionary/article?
- What cognitive processes, knowledge, representations, skills and strategies do users employ in trying to understand (and finally to acquire) the grammatical features of a lexical item, i.e. to build a mental representation of the information provided?
- What kind of presentation (encoding) and layout of the grammatical characteristics of a lexeme in a dictionary article will optimize the construction of a mental representation of these characteristics?
- What mediating (user, task, text) variables have been shown experimentally to influence this process?

4.2.1 Theoretical Models

With regard to the first question, dictionary designers have relied for the most part on what linguistics has to offer with regard to the characterization of "lexical knowledge". Laufer and Paribakht (1998: 366-367) note, however, that there is no clear and unequivocal consensus as to what exactly constitutes lexical knowledge. Most researchers accept, though, that lexical knowledge is not an all-or-nothing phenomenon, but a continuum consisting of several layers and dimensions of knowledge, skills and strategies that develop over time. Most researchers accept, furthermore, that acquiring a lexical item involves acquiring a mental representation of all its grammatical characteristics (broadly defined). These include its orthographic, phonemic, morphologic, syntactic (including its collocational properties) and semantic and stylistic/use characteristics. Lexical semantic knowledge includes not only the lexical item's intraconceptual characteristics (primary and secondary senses, polysemic structure, etc) but also its interconceptual structure, its relation to other knowledge structures (frames and schemata) and its reference. Lexical comprehension does not only involve an understanding of the semantics (intra- and interconceptual structure) of a lexical item, but also of how its semantics interacts with its syntactic and morphologic properties, and *inter alia*, with all other knowledge structures mentioned above. In short, as proposed in most lexical semantic theories, it is accepted that the mental representations of lexical items form intricate and multidimensional networks in the mental lexicon of the learner (cf. Aitchinson 1994, Geeraerts 1989 and Schmitt 1997).

Although this whole range of information categories may constitute the knowledge users have of lexical items, no dictionary includes all of these for

every lexical item. In choosing what information to include in a dictionary, designers are guided by a number of pragmatic and functional considerations, viz. the type of dictionary and the lexemes chosen for lexicographic treatment, the information needs of the end-user and the constraints imposed by the limitations of space and the linearity of the printed dictionary (cf. Geeraerts 1989).

With regard to the presentation (encoding and layout) of the lexical information provided in dictionary entries, very few dictionary designers have explicitly linked their dictionary design to models of the knowledge, processes and strategies underlying the understanding and acquisition of the grammatical information of a lexeme. In the field of SL instructional design, however, a number of researchers have reverted in this regard to Levelt's model of the lexicon and to adaptations of it to account for SL lexical knowledge acquisition and use (cf. De Bot, Paribakht and Wesche 1997, and Dörnyei and Kormos 1998).

According to Levelt's model, the mental representation of a lexical item is organized into three distinct representational levels: the conceptual level, the lemma level and the lexeme level (morphonologic word form). The lemma specifies the semantic/conceptual features (propositional content and structure) and syntactic-semantic features (syntactic category, argument structure, grammatical functions of arguments, complementation structure and diacritic parameters). De Bot, Paribakht and Wesche (1997: 313) give the following example of a verb lemma in which the semantic and syntactic information and their interaction (mapping) are explicated:

<i>give</i> :	
conceptual specification:	CAUSE(X,(GOposs(Y,(FROM/TO(X,Z))))))
conceptual arguments:	(X,Y,Z)
syntactic category:	V
grammatical functions:	(SUBJ,DO,IO)
relations to COMP:	none
lexical pointer:	713
diacritic parameters:	tense
	aspect
	mood
	person
	number
	pitch accent

In terms of this model, building up a mental representation of the grammatical features of a lexeme is a process of construction in which the learner uses available linguistic input to infer the grammatical features of a lexical item to fill in the slots at the various representational levels of a lexical template and to establish links between the various levels. A learner will have acquired a lexical item only when he/she has acquired all the relevant grammatical features as specified at the various levels, has established links between the various levels, and — as is assumed in some definitions of lexical knowledge — is also able to activate the required lexical knowledge in decoding and encoding tasks. A

learner's knowledge of a lexical item could, however, also be only of a partial nature, i.e. include any one or more but not all of the lexical characteristics specified at the different levels of representation and be the result of the fact that characteristics at the various levels of presentation are not linked.

Although this model provides a theoretical basis for understanding what the content and structure of the mental representations of lexical items could be like, it obviously is also rather vague on a number of aspects. The first is the fact that it does not explicate what knowledge, processes, strategies and skills underlie the comprehension process, i.e. how a user would interpret, extract and synthesize the lexical information provided in a dictionary entry for *give* in the form of the mental representation. Nor does Levelt's model indicate what variables could have an impact on the comprehension process itself, such as the knowledge the user has to have of the lexicographic conventions used to encode grammatical information in various kinds of dictionaries.

Secondly, the model is underspecified with regard to all the information categories that are taken to constitute lexical knowledge (cf. the discussion above). For example, it does not provide for stylistic and pragmatic information and it is not clear how the model provides for the multidimensional structure of lexical meaning. With regard to the latter, Geeraerts (1989) indicates that most polysemic lexical items do not have a set of clearly delineated senses (as is often provided in dictionary entries) but — as is the case with prototype categories — a number of mostly overlapping senses that relate to each other in a variety of ways. The challenge for the dictionary designer, of course, is to come up with a representation format/formats that would reflect the complexities of the internal structure of the semantics of a lexeme and thereby to optimize the comprehension process.

Optimizing the encoding of information on the meaning of lexemes has certainly received the most attention, especially in SL lexicography, as it has also proven to be the most problematic. Besides the lexicographic definitions of various kinds, the other major tools lexicographers use for this purpose are verbal examples and illustrations. The user has the task of building up a mental representation of the meaning of a lexeme from the information provided by these descriptive devices, separately and in conjunction with each other. This can be schematically depicted as follows:

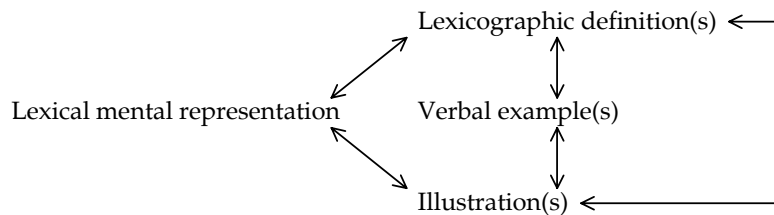


Figure 5: Building-up a mental representation of lexical meaning

Again, such a schematic depiction merely captures the information sources that the user can utilise in building up the mental representation of the meaning of

a lexeme, but not the complexities of the knowledge and cognitive processes involved in inferencing/extracting and synthesizing the information from these three sources to construct a mental representation of the meaning of a lexeme.

With regard to the latter, one could assume that, although unspecified, learners/users employ the same cognitive resources, strategies and skills as those that they employ when learning from any other contexts of language use. But lexicographers need more detailed information on

- how and what semantic information users extract, process and store and finally integrate from definitions, verbal examples and illustrations, and
- how this information is finally organized in the mental lexicon of the user.

Most theories are rather vague with regard to these two aspects and theoretical research to elucidate these aspects are crucially needed. Despite this, a number of new design features have been incorporated into dictionaries, especially learner's dictionaries, based on (although often not explicitly stated) assumptions of the knowledge, processes, strategies and skills involved in building up a mental representation of the meaning of a lexeme from definitions and verbal examples, and of the variables that effect this process.

4.2.2 Design Innovations in Lexicography

The assumption that the user first has to understand the lexicographic definition to be able to construct a representation of a lexeme's meaning, motivates the use of a controlled/limited defining vocabulary in definitions and the elimination of all kinds of dictionarese (symbols, labels, some abbreviations, parentheses, etc.) and arcane expressions in definitions. With regard to the former, the assumption is that elimination of dictionarese takes the burden off the user of first having to translate the meaning of these symbols, labels, etc. and of determining how their meanings relate to that provided by the rest of a definition. Likewise, restricting the defining vocabulary goes a long way to ensuring that the user will have a command of the lexical items that are used to define the meaning of a lexeme. As some dictionary critics have pointed out though, a limited defining vocabulary also has a number of drawbacks, especially that it often leads to vague definitions, which thwarts the goal of comprehensibility.

As indicated in section 1, the so-called full-sentence definition format has become a standard feature of most learner's dictionaries. Besides incorporating information on the syntactic features of the defined lexical item, its use is motivated by the assumption that it has the format of folk definitions, and would therefore be easier to comprehend. However, the superiority of this definition format still has to be supported experimentally (cf. Cumming, Crop and Sussex, 1994). Furthermore, the full-sentence definition and phrasal definitions are only two of the various kinds of lexicographic definition formats. In line with the functional-cognitive methodology, one would expect that the design of diction-

ary definitions should be guided by explicit models of the acquisition and mental storage of lexical meaning and by research on the variables that mediate their effectivity for different kinds of users.

With regard to the use of verbal examples, Harras (1989), following Wiegand, explicitly states that we are in need of a lexicographic theory of the use of verbal examples. Such a theory would have to explicate

- how verbal examples support lexicographic definitions by clearly indicating what relationship(s) exist between examples and other elements of a dictionary article (including the definition), and
- what makes a good lexicographic example.

The second normative component of the theory would no doubt be linked to the first component in that good examples would be those that either illustrate the use-aspects of the other elements of a dictionary article, add to the information these other elements already convey (cf. Harras 1989 for further discussion), or independently contribute information to the mental representation of a lexeme (as spelt out in the lexical model of Levelt above). Again, however, whatever norms may be deduced for "good" examples, these norms still have to be tested empirically, and the mediating variables that impinge on the effectiveness of examples with regard to the three aspects mentioned above, have to be determined.

Levelt's model does not provide in an explicit way for the use of illustrations as a means for explaining the meaning of lexemes. Given the fact that illustrations are widely used in printed dictionaries (cf. Hupka 1989) and that electronic dictionaries provide for multimedia input (illustrations, graphics, animation, video) an adequate theoretical model for designing adequate dictionaries should also account for the processing of visual information.

To this end a number of researchers in the field of SL instructional design revert to Paivio's dual coding theory (cf. Plass 1998, Chun and Plass 1997, and Plass, Chun, Mayer and Leutner 1998 for a discussion). According to this theory (Paivio 1986) the information presented in verbal and visual form are processed differently and stored in two different systems in the cognitive component. Verbal input represents information in the symbolic structures of a language and it is processed sequentially to form, first of all, propositional mental representations of the meaning of the linguistic input. The propositional representations are then mapped onto a mental model of the information that is presented.

Visual input, on the other hand, conveys information by means of visuo-spatial structure (i.e. spatial arrangements of the components of an image) and the information is encoded in parallel or simultaneous fashion. The comprehension of an image requires establishing an analogy between the picture and a corresponding mental model. Image comprehension thus bypasses the propositional representation of information and is, therefore, assumed to be

language independent (cf. Chun and Plass 1997).

When represented with verbal input and visual input that relates to the verbal input, the learner has to make referential connections between the two mental representations that are established by the interpretation process. In the case of the acquisition of unknown words, this implies that the learner will have to match his/her mental representation of the semantic/conceptual features of a lexical item with the features and their relations to each other in the mental model of the visual input.

It is now generally accepted that the comprehension of the meaning of lexemes can be greatly enhanced by supporting lexicographic definitions with illustrations, especially where very complex concepts have to be defined. As in the case of lexicographic examples, however, we are still in need of

- theoretically motivated analyses of how the information provided in illustrations relate to and are linked to the information provided in lexicographic definitions, and
- experimental evidence on what constitutes a good lexicographic illustration (for what aspects of the meaning of lexemes and for what word classes).

(Cf., e.g., Chun and Plass 1997, Hupka 1989, Mayer 1999 and Schriver 1997.)

With regard to the latter, Chun and Plass (1997) indicate that a number of variables have been shown empirically to either enhance or impede the mediating effect of visual input in building up a mental representation of the meaning of a lexeme. The most important of these are the kind of visual material used, the integration of visual and verbal materials, and the cognitive learning styles of users (visualizers vs. verbalisers).

4.3 Applying Acquired Lexical Knowledge

The last aspect of dictionary consultation that will receive attention, is the ability of the user to apply the lexical knowledge that has been acquired from the dictionary to the task that prompted the consultation process in the first place. In this respect, the information provided by dictionaries on the meaning of lexemes again illustrates the problems users can experience. Dictionary consultations are often triggered by highly context specific meanings of lexical items, while dictionaries in most cases only provide for the most general, prototypical meanings of lexemes. The user, therefore, has to be able to contextualize the more general meaning to the specific meaning required by the context of use of a lexeme — a skill that SL learners, for example, often lack.

Scholfield (1981) proposes the use of contextual paraphrases as dictionary definitions to help the user overcome this problem. Another, and very old design solution to this problem is to annotate/gloss difficult/unknown words in texts intended for the SL learner. Text glosses provide users with only the

relevant contextual meaning of an unknown lexical item. Glosses thus relieve students from the arduous task of finding the relevant senses of unknown words in printed dictionaries, and, if given as a contextual paraphrase of an unknown word, of having to integrate some generic meaning from a dictionary in the word's context of use. In electronic texts the type of glosses are furthermore expanded to include a variety of information types (e.g. translations, definitions in the L1 and the L2, encyclopaedic information, visual glosses (graphics, pictures, animation and video) and sound), thereby providing for the different learning styles of different dictionary users. The layout and design possibilities of computer screens also allow for the fact that glosses can be presented and integrated into textual (multimedia) input in such a way that they do not negatively impact on the comprehension process. Glossing has proven to be rather beneficial, both in the case of printed texts and electronic texts, and a growing body of research has been devoted to this topic (cf. Chun and Plass 1997, Lomicka 1998, Roby 1999, and Yeung, Putai and Sweller 1997).

In such cases, however, the printed dictionary functions merely as a lexical resource for the teacher/materials designer who has to select the relevant senses, translations, appropriate visual materials, etc. from existing dictionaries, and come up with a contextually relevant presentation of the information in glosses that can support the learner. Research on how such glosses can be automatically generated in CALL applications incorporating fully fledged electronic dictionaries is still in its infancy. Furthermore, a number of deleterious effects have been shown to influence the functional quality of multimedia glosses (cf. Chun and Plass 1997 and the references cited above).

No doubt, other possibilities than contextual paraphrases and glosses could be devised, but again the question of which one of these is the most effective for what kinds of users for what kinds of tasks, is an empirical one.

5. Conclusion

In section 1 of this article, I argue that, despite the commitment to improving the functional quality of dictionaries, two major methodological problems currently trouble practical lexicography and metalexigraphic research, viz. that

- the evaluation of the functional quality of dictionaries does not feature prominently in either the dictionary design process, or in current metalexigraphic research, and
- when it does feature, it is most often based on impressionistic and anecdotal evidence or commonsense heuristics, not on the application of generally accepted scientific methodologies and techniques, such as theoretically driven experimental research.

Subsequently, an outline is given of the methodology used in the fields of web design, document design and instructional design to tackle the functional qual-

ity problem. As I have indicated throughout, this methodology also provides a conceptually coherent framework for the design of dictionaries, either in printed or electronic format, and for assessing the functional quality of their design features. Secondly, the framework provided by the cognitive-functional approach to design not only explicates how current research in lexicography conceptually relates to dictionary design, but also what areas are in dire need of further theoretical elaboration and experimental research.

No doubt, not all the relevant variables that optimize the functional quality of dictionaries have been identified and explicated in this article. That would be part of an ongoing research programme within the field of lexicography — a programme in which the traditional fields of lexicographic research is also supplemented by a concern with functional quality.

If lexicographers are really concerned about user needs and have the real goal of improving the functional quality of dictionaries, they simply have to take the methodological leap that is required of them. The methodology outlined in this article is precisely aimed at achieving this. It provides for the design of theoretically and empirically motivated dictionaries, but then in a way that cognisance is taken of what has already been done in the field of lexicography on the functional quality of dictionaries and what areas are in need of further research.

References

A. Dictionaries

- OALD5: **Crowther, J. (Ed.)**. 1995. *Oxford Advanced Learner's Dictionary of Current English*. Oxford: Oxford University Press.
- CIDE: **Procter, P. (Ed.)**. 1995. *Cambridge International Dictionary of English*. Cambridge: Cambridge University Press.
- COBUILD2: **Sinclair, J. (Ed.)**. 1995. *Collins COBUILD English Language Dictionary*. London: HarperCollins.
- LDOCE3: **Summers, D. (Ed.)**. 1995. *Longman Dictionary of Contemporary English*. Harlow: Longman.

B. Other Sources

- Aitchinson, J.** 1994. *Words in the Mind: An Introduction to the Mental Lexicon*. Oxford: Blackwell.
- Boogaards, P.** 1998. Scanning Long Entries in Learner's Dictionaries. Fontenelle, T. et al. (Eds.). 1998: 555-562.
- Chun, D.M. and J.L. Plass.** 1997. Research on Text Comprehension in Multimedia Environments. *Language Learning and Technology* 1: 60-81.
- Cumming, G., S. Crop and R. Sussex.** 1994. On-line Lexical Resources for Language Learners: Assessment of Some Approaches to Word Definition. *System* 23: 369-377.

- De Bot, K., T.S. Paribakht and M. Bingham Wesche.** 1997. Toward a Lexical Processing Model for the Study of Second Language Vocabulary Acquisition: Evidence from ESL Reading. *Studies in Second Language Acquisition* 19: 309-329.
- De Jong, M. and T. van der Geest.** 2000. Characterizing Web Heuristics. *Technical Communication* 47: 3001-322.
- De Schryver, G.-M. and D.J. Prinsloo.** 2000. Dictionary-Making Process with 'Simultaneous Feedback' from the Target Users to the Compilers. Heid, U., S. Evert, E. Lehmann and C. Rohrer (Eds.). 2000: 197-210.
- Dörnyei, Z. and J. Kormos.** 1998. Problem-Solving Mechanisms in L2 Communication: A Psychological Perspective. *Studies in Second Language Acquisition* 20: 349-385.
- Fontenelle, T. et al. (Eds.).** 1998. *EURALEX '98 Proceedings*. Liège: University of Liège Press.
- Gaffney, G.** 2000. Gerry Gaffney's 'Introduction to Web Usability'. Online: <http://www.infodesign.com.au> Retrieved: 2000-08-28.
- Geeraerts, D.** 1989. *Wat er in een woord zit. Facetten van de lexicale semantiek*. Leuven: Peters.
- Geeraerts, D.** 2000. Adding Electronic Value. The Electronic Version of the *Grote van Dale*. Heid, U., S. Evert, E. Lehmann and C. Rohrer (Eds.). 2000: 75-84.
- Groot, P.J.M.** 2000. Computer Assisted Second Language Vocabulary Acquisition. *Language Learning and Technology* 4: 60-81.
- Harras, G.** 1989. Zu einer Theorie des lexicographischen Beispiels. Hausmann F.J., O. Reichmann, H.E. Wiegand and L. Zgusta (Eds.). 1989: 607-613.
- Hausmann, F.J., O. Reichmann, H.E. Wiegand and L. Zgusta (Eds.).** 1989. *Wörterbücher. Ein internationales Handbuch zur Lexikographie. / Dictionaries. An International Encyclopedia of Lexicography. / Dictionnaires. Encyclopédie internationale de lexicographie*. Berlin/New York: Walter de Gruyter.
- Hegelheimer, V. and Carol A. Chapelle.** 2000. Methodological Issues in Research on Learner-Computer Interactions in CALL. *Language Learning and Technology* 4: 41-59.
- Heid, U., S. Evert, E. Lehmann and C. Rohrer (Eds.).** 2000. *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000*. Stuttgart: Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.
- Herbst, T.** 1999. On the Way to the Perfect Learner's Dictionary: A First Comparison of OALD5, LDOCE3, COBUILD2 and CIDE. *International Journal of Lexicography* 12: 321-355.
- Hoeken, H.** 1995. *The Design of Persuasive Texts: Effects of Content, Structure, and Style on Attitude Formation*. Unpublished Ph.D. thesis. Tilburg: Catholic University of Brabant.
- Hupka, W.** 1989. Die Bebilderung und sonstige Formen der Veranschaulichung im allgemeinen einsprachigen Wörterbuch. Hausmann F.J., O. Reichmann, H.E. Wiegand and L. Zgusta (Eds.). 1989: 704-726.
- Köhler Simonsen, H.** 2000. Design, Development and Compilation of a Bilingual Multifunctional Intranet-based Differential Telecom Lexinome at a Major Danish Telecoms Group. Heid, U., S. Evert, E. Lehmann and C. Rohrer (Eds.). 2000: 89-96.
- Laufer, B.** 2000. Electronic Dictionaries and Incidental Vocabulary Acquisition: Does Technology Make a Difference? Heid, U., S. Evert, E. Lehmann and C. Rohrer (Eds.). 2000: 849-854.
- Laufer, B. and T.S. Paribakht.** 1998. The Relationship between Passive and Active Vocabularies: Effects of Language Learning Context. *Language Learning* 48: 365-391.
- Levelt, W.J.M.** 1989. *Speaking: From Intention to Articulation*. Cambridge, MA.: MIT Press.

- Levelt, W.J.M.** 1994. *Onder woorden brengen: Beschouwingen over het spreekproces*. Haarlem: Nederlandsche Maatschappij der Wetenschappen.
- Lomicka, L.L.** 1998. "To Gloss or Not to Gloss": An Investigation of Reading Comprehension Online. *Language Learning and Technology* 1: 41-50.
- Mayer, R.** 1999. Research-based Principles for the Design of Instructional Messages: The Case Of Multimedia Explanations. *Document Design* 1: 7-19.
- Multimedia Services.** 2000. Computer-based Training. ADDIE Model. Online: <http://www.orci.com/~vanzant/cbt.html> Retrieved: 2000-02-10.
- Müller-Landmann, S.** 2000. Design eines Internet-Lexicons zwischen Recherche und Rezeption. Heid, U., S. Evert, E. Lehmann and C. Rohrer (Eds.). 2000: 97-106.
- Nesi, H.** 2000. Electronic Dictionaries in Second Language Vocabulary Comprehension and Acquisition: The State of the Art. Heid, U., S. Evert, E. Lehmann and C. Rohrer (Eds.). 2000: 839-848.
- Nielsen, J.** 1994. Heuristic Evaluation. Nielsen, J. and Mack (Eds.). 1994: 25-62.
- Nielsen, J.** 1999. *Useit.com: Jacob Nielsen's Website*. Online: <http://www.useit.com> Retrieved: 2000-09-28.
- Nielsen, J. and R.L. Mack (Eds.).** 1994. *Usability Inspection Methods*. New York: John Wiley and Sons.
- Paivio, A.** 1986. *Mental Representations. A Dual Coding Approach*. Oxford: Oxford University Press.
- Plass, Jan L.** 1998. Design and Evaluation of the User Interface of Foreign Language Multimedia Software: A Cognitive Approach. *Language Learning and Technology* 2: 35-45.
- Plass, J.L., D.M. Chun, R.E. Mayer and D. Leutner.** 1998. Supporting Visual and Verbal Learning Preferences in a Second-Language Multimedia Learning Environment. *Journal of Educational Psychology* 90: 25-36.
- Roby, W.B.** 1999. "What's in a Gloss?" *Language Learning and Technology* 2: 94-101.
- Rundell, M.** 1999. Dictionary Use in Production. *International Journal of Lexicography* 12: 35-54.
- Sato, H.** 2000. Multi-Functional Software for Electronic Dictionaries. Heid, U., S. Evert, E. Lehmann and C. Rohrer (Eds.). 2000: 863-870.
- Schellens, P.J. and M.F. Steehouder.** 1994. Tekstontwerp: Schrijven als ontwerpproces. *Tijdschrift voor Taalbeheersing* 16: 161-172.
- Schmitt, N.** 1997. Vocabulary Learning Strategies. Schmitt, N. and M. McCarthy (Eds.). 1997: 199-227.
- Schmitt, N. and M. McCarthy (Eds.).** 1997. *Vocabulary; Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press.
- Scholfield, P.** 1981. Vocabulary Explanation by Paraphrase in Context. *Studia Anglica Posnaniensia* 15: 103-121.
- Scholfield, P.** 1982. Using the English Dictionary for Comprehension. *TESOL Quarterly* 16: 185-194.
- Scholfield, P.** 1999. Dictionary Use in Reception. *International Journal of Lexicography* 12: 13-34.
- Schrivier, K.A.** 1989. Evaluating Text Quality: The Continuum from Text-focused to Reader-focused Methods. *IEEE Transactions in Professional Communication* 32: 238-255.
- Schrivier, Karen.** 1993. Quality in Document Design: Issues and Controversies. *Technical Communication* 40: 239-257.
- Schrivier, Karen.** 1997. *Dynamics in Document Design*. New York: John Wiley.

- Shelby, A.N.** 1994. Communication Quality as Metacommunication: A Conceptual Analysis. Van Waes, L., E. Woudstra and D. van den Hoven (Eds.). 1994: 5-17.
- Sims-Knight, J.E.** 1992. To Picture or Not to Picture: How to Decide. *Visible Language* 29: 325-388.
- Swanepoel, P.** 2000. Providing Lexicographic Support for SL Vocabulary Acquisition: What Kind, under What Conditions, for Whom, and Why? Heid, U., S. Evert, E. Lehmann and C. Rohrer (Eds.). 2000: 403-419.
- Swanepoel, P. and K. van de Poel.** Forthcoming. Designing Effective Lexical Support for Second-Language Acquisition in Computer-Assisted Language Learning: The LINC Approach.
- Tono, Y.** 2000. On the Effects of Different Types of Electronic Dictionary Interfaces on L2 Learners' Reference Behaviour in Productive/Receptive Tasks. Heid, U., S. Evert, E. Lehmann and C. Rohrer (Eds.). 2000: 855-862.
- Urquhart, S. and C. Weir.** 1998. *Reading in a Second Language: Process, Product and Practice*. London: Longman.
- Van der Geest, T. and J.H. Spyridakis.** 2000. Developing Heuristics for Web Communication. *Technical Communication* 47: 301-311.
- Van Waes, L., E. Woudstra and P. van den Hoven (Eds.).** 1994. *Functional Communication Quality*. Amsterdam: Rodopi.
- Wiegand, H.E.** 1998. *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie*. Berlin: Walter de Gruyter.
- Wilson, B., D. Jonassen and P. Cole.** 1993. Cognitive Approaches to Instructional Design. Online: <http://www.cudenver.edu/~bwilson/training.html> Retrieved: 2000-02-17.
- Yeung, A.S., J. Putai and J. Sweller.** 1997. Cognitive Load and Learner Expertise: Split-attention and Redundancy Effects in Reading with Explanatory Notes. *Contemporary Educational Psychology* 23: 1-21.

Supplying Syntactic Information in a Quadrilingual Explanatory Dictionary of Chemistry (English, Afrikaans, isiZulu, Sepedi): A Preliminary Investigation*

Elsabé Taljard (*etaljard@postino.up.ac.za*) and Rachéle Gauton (*rgauton@postino.up.ac.za*), Department of African Languages, University of Pretoria, Pretoria, Republic of South Africa

Abstract: In response to a number of recent government publications on the promotion of science and technology training, the *Suid-Afrikaanse Akademie vir Wetenskap en Kuns* (South African Academy for Science and Art) initiated the compilation of a multilingual explanatory dictionary for chemistry. The need for such a dictionary is especially urgent in learning environments where learners receive tuition through a medium other than their mother tongue. This is particularly relevant for learners whose mother tongue is an African language. Taking into account the target user's fragmented knowledge of the subject field and low second language competence, the planners decided that certain syntactic information should be specified for the isiZulu and Sepedi entries; specifically the part of speech to which a lemma belongs. This poses certain problems for the lexicographer, since the issue of word categorisation is, especially in Sepedi, an unresolved one. A second problem that presents itself to the lexicographer is the lexicographic convention which is used to indicate the word class to which a specific lemma belongs. Abbreviations referring to the parts of speech are normally used for this purpose. No standardised abbreviations for the different parts of speech exist in isiZulu or Sepedi. Principles for the formation of abbreviations have also not been formulated for the official orthographies of these languages. The only solution to this problem is to study existing abbreviations in order to abstract the principles of abbreviation formation. Only then can abbreviations indicating the parts of speech be formulated for use in the dictionary.

Keywords: SPECIAL FIELD DICTIONARY, CHEMISTRY, ISIZULU, SEPEDI, WORD CATEGORISATION, LEXICOGRAPHICAL CONVENTION, ABBREVIATIONS, MULTILINGUAL DICTIONARY

Opsomming: Die insluiting van sintaktiese inligting in 'n viertalige verkla- rende chemiewoordeboek (Engels, Afrikaans, isiZulu, Sepedi): 'n Voorlopige ondersoek. In reaksie op 'n aantal onlangse regeringspublikasies rakende die bevordering van wetenskap- en tegnologieopleiding, het die Suid-Afrikaanse Akademie vir Wetenskap en Kuns die

* This paper was presented at the Fifth International Conference of the African Association for Lexicography, held at the University of Stellenbosch, Stellenbosch, 3-5 July 2000.

inisiatief geneem om 'n viertalige verklarende chemiewoordeboek saam te stel. Die behoefte aan so 'n woordeboek is veral dringend in leeromgewings waar leeders onderrig in 'n ander taal as hul moedertaal ontvang. Dit is veral van toepassing op leeders wat 'n Afrikataal as moedertaal het. Met die teikengebruiker se gefragmenteerde kennis van die studieveld en swak tweedetaalvaardigheid in ag geneem, het die beplanners besluit dat bepaalde sintaktiese inligting ook ten opsigte van die isiZulu- en Sepedi-inskrywings gespesifiseer moet word; veral die woordklas waartoe 'n lemma behoort. Dit veroorsaak sekere probleme vir die leksikograaf, aangesien die aangeleentheid van woordkategorisering veral in Sepedi, 'n onopgeloste een is. 'n Tweede probleem waarmee die leksikograaf gekonfronteer word, is die leksikografiese konvensie wat gebruik word om die woordkategorie waartoe 'n spesifieke lemma behoort, aan te dui. Afkortings wat na die woordklasse verwys, word normaalweg vir hierdie doel gebruik. In nóg isiZulu, nóg Sepedi bestaan daar egter gestandaardiseerde afkortings vir die verskillende woordklasse. Geen beginsels vir die vorming van afkortings is ook nog in die amptelike ortografieë van hierdie tale geformuleer nie. Die enigste oplossing vir hierdie probleem is om bestaande afkortings te bestudeer ten einde die beginsels vir die vorming van afkortings daaruit te abstraheer. Slegs dan kan afkortings wat die woordklasse aandui, geformuleer word vir gebruik in die woordeboek.

Slutelwoorde: SPESIALISWOORDEBOEK, CHEMIE, ISIZULU, SEPEDI, WOORDKATEGORISERING, LEKSIKOGRAFIESE KONVENSIE, AFKORTINGS, VEELTALIGE WOORDEBOEK

1. Introduction

The compilation of the *Quadrilingual Explanatory Dictionary of Chemistry* (QEDC) is a project which was initiated by the *Suid-Afrikaanse Akademie vir Wetenskap en Kuns* (South African Academy for Science and Art) in response to a number of recent Government policy publications in which strong emphasis is placed on education and training in the field of science, engineering and technology. Especially relevant in this regard are the *White Paper on Reconstruction and Development* of 1994, *South Africa's Green Paper on Science and Technology* (s.a) and the *White Paper on Science and Technology* of 1996. The key concept in these papers is the development of human potential through optimal access to information. Carstens (1997: 2) argues "that multilingual, explanatory special-field dictionaries can be implemented to realise these objectives by providing easy access to new or incompletely learnt concepts".

2. Role players

The initial planning of the dictionary project was to a large extent done by Prof. A. Carstens of the Department of Afrikaans and Prof. D.J. Prinsloo of the Department of African Languages, both from the University of Pretoria, in collaboration with Dr Van Zyl de Villiers of the Atomic Energy Corporation and Dr M. Alberts of the National Language Services. An article providing

detailed information on the planning of the project "Issues in the Planning of a Multilingual Explanatory Dictionary of Chemistry" was published by Carstens in 1997 in *Lexikos* 7.

In the aforementioned article, it is indicated that a so-called "multispectral" approach to data-collection for the QEDC should be followed, involving numerous role players from different fields. Since it is a chemistry dictionary, consultation with experts in the various fields of chemistry is a prerequisite. The multilingual aspect necessitates the involvement of mother-tongue speakers of Afrikaans, isiZulu and Sepedi; not only as lexicographers and terminologists, but also as teachers, lecturers and experts in the chemistry field. Carstens (1997: 11) states that the involvement of mother-tongue speakers in various capacities serves a number of purposes. Firstly, it can legitimise the dictionary amongst the speakers of the various language groups; secondly, it keeps the lexicographer/terminologist in touch with the linguistic reality, thus preventing him/her from coining terms where terms already exist; and thirdly, it ensures continuous monitoring of the relevance and usefulness of the dictionary with regard to the communicative functions it has to fulfil.

3. Target users

The heterogeneous nature of the target users of the QEDC impacts on all aspects of the dictionary and the planning thereof. The target user of this dictionary has been identified as either a learner in the senior secondary school phase (grades 10-12), or a pregraduate university or technikon student in the physical sciences. This user group is characterised by varying levels of encyclopaedic knowledge of chemistry and second language competence. One could safely assume that the average target user would have a fragmented knowledge of the conceptual system of the subject field and, generally speaking, an average to low second language competence. Learners who have one of the African languages as mother-tongue are further disadvantaged by the fact that they receive their tuition through the medium of English and not, as is mostly the case for Afrikaans-speaking learners, through the medium of their mother-tongue.

4. Function of the dictionary

Carstens (1997: 10) refers to three possible communicative functions which are to be fulfilled by a dictionary, viz. **production** (native language or foreign language), **reception** (native language or foreign language) or **translation** (into or from a foreign language). In the case of the QEDC, the reception function would be of primary importance, since the majority of target users are taught in English, which is their second or even third language. The appearance of both term and concept definition in Afrikaans, isiZulu and Sepedi would assist

in the cognitive function of conceptualisation. Production would only come at a later stage, with translation being a long-term possibility. Bergenholtz and Tarp (1995: 25) do however warn that where a special field dictionary is designed for users with different mother-tongues, an integration of the above functions presents problems.

5. Format of the dictionary

After an extensive needs assessment, it was decided that English should be the language of lemmatisation in the primary word list. All the encyclopaedic and linguistic information supplied for the English lemma would then also be supplied for the translational equivalents in Afrikaans, isiZulu and Sepedi. Compare example (1)¹ by way of illustration:

(1) **English**

acid (*n*) A substance that produces an excess of H₃O⁺ (hydronium) ions when it dissolves in water, e.g. HCl (hydrochloric acid); HNO₃ (nitric acid); CH₃COOH (acetic acid). *Opposite* **base**. ♣ **acidic** (*a*); **acidify** (*v*).

Afrikaans

suur (*n*) 'n Stof wat 'n oormaat H₃O⁺-ione (hidroniumione) vorm wanneer dit in water opgelos word, bv. HCl (soutsuur); HNO₃ (salpetersuur); CH₃COOH (asynsuur). *Teenoorgestelde* **basis**. ♣ **suur** (*a*); **aansuur** (*w*).

IsiZulu

i-asidi (*bz*) Isiq̣a noma into ekhiq̣iza ama-ayoni ehayidroniyamu (H₃O⁺) ngokweq̣ile uma sinyibilika emanzini, izib. HCl (i-asidi yehayidroklorikhi); HNO₃ (i-asidi initric); CH₃COOH (i-asidi emuncu). *Umq̣ondophika* **isiq̣u**. ♣ **-neasidi** (*isich.*); **-ba yiyasidi** (*sbnj.*) / **-muncisa** (*sz*)

Sepedi

esiti (*l*) Selo se se tšweletšago diayone tša haedroniamo (H₃O⁺) tše di feleletšego ge di tologa ka meetseng, mohl. HCl (esiti ya haedrokloriki); HNO₃ (esiti ya naetriki); CH₃COOH (esiti ya asetiki). *Lelatodi* **peisi**. ♣ **-nago esiti** (*tlha*); **esitifatša** (*led*).

Reverse word lists for the other three languages will then be supplied after the primary word list. The QEDC thus shares some features with translation dictionaries. The reason for selecting English as the language of primary lemmatisation is that lexical gaps abound in especially isiZulu and Sepedi, and to a lesser extent in Afrikaans. Swanepoel (1989: 244) points out there tend to be lexical gaps in languages of developing communities and/or languages which have only recently become standard languages. He contends that in such cases it is advisable to compile a translation dictionary in which the language(s) containing the lexical gaps is (are) used as the target language(s). The language

from which such languages obtain their scientific and technical information, in this case English, then serves as the source language. In this manner the lexical gaps in the so-called "younger language(s)", or "languages of limited diffusion (LLDs)" as they are also called, can be identified more easily.

6. Two phases of dictionary compilation

The dictionary is compiled in two phases, the first of which has been completed. During this first phase ± 500 chemistry terms and their conceptual definitions were terminologically and lexicographically processed for English and Afrikaans and entered in an electronic database. The second phase consists of adding the isiZulu and Sepedi equivalents of both terms and definitions. It was decided during the planning stages that all encyclopaedic and linguistic information supplied for the English lemmas would also be supplied for the translation equivalents. With reference to the inclusion of linguistic information in a chemistry dictionary, Carstens (1997: 15) indicates that traditional monolingual science dictionaries provide very little linguistic information, the reason being that "many dictionaries of science and technology have been designed by experts in the subject-field without any linguistic background or interest in grammar". Experts in the natural sciences may consider grammar as less important and not crucial for communication in a scientific or technical domain, thus putting a higher premium on economy, conceptual precision and accessibility. Carstens (1997: 15), however, argues convincingly that a dictionary of the type in question is not adequate without linguistic information such as part of speech, morphological formation, contextual information and information on linguistic and stylistic usage. The decision to include this type of linguistic information was apparently motivated by the low second language proficiency of the target user.

Furthermore, as was stated earlier, the QEDC, being a multilingual term dictionary, shares certain features with translation dictionaries. With reference to translation dictionaries, Swanepoel (1989: 234) argues that an indication of the part of speech to which the main lemmas belong is of the utmost importance, especially where no examples providing information on the syntactic behaviour of the lemma are supplied. This would seem to be the case with the QEDC and other technical dictionaries where the minimum of syntactic information is provided. He further contends that there is a direct correlation between the amount of grammatical and syntactic information supplied, and the envisaged functions of the dictionary. If the aim of the dictionary is to assist the user in **decoding** texts, then less syntactic information is supplied than would be supplied if the aim of the dictionary is to enable the user to **generate** texts. Since the ultimate aim of the dictionary is to enable the target user to generate texts in the target language, the inclusion of syntactic information is an essential aspect of the QEDC.

Although the usefulness of such an approach cannot be questioned, providing linguistic information does pose certain practical problems regarding the entering of Sepedi and isiZulu equivalents. For the purpose of this discussion, it was decided to focus on one aspect of linguistic information to be provided in the dictionary, viz. specifying the part of speech to which every lemma belongs. It should be noted that the problems which have been identified are primarily of a **linguistic nature**, but they do have terminological and/or lexicographical implications not only for the compilation of the QEDC, but perhaps even to a larger extent for the compilation of future monolingual dictionaries.

7. Specification of the part of speech to which a lemma belongs

Two problems regarding the specification of the part of speech to which a lemma belongs have been identified. The first is one with which linguists studying the indigenous South African languages have long been grappling, viz. that of word categorisation. The second concerns the lexicographic convention used to indicate to which word class a specific lemma belongs.

7.1 Word categorisation in Sepedi and isiZulu

Firstly, specifying the part of speech to which a specific lemma and/or its derivatives belong, is a relatively simple exercise in Afrikaans and English. Unfortunately, this is not the case in the African languages, especially in Sepedi. Due to the different approaches to word identification and categorisation, two sets of word classes or parts of speech are currently recognised in grammars of Sepedi. The one, with which most Sepedi linguists are probably familiar, is that formulated by Van Wyk (1961: 69 et seq.), and adapted by Lombard (1985). This classification has dominated the Sepedi linguistic scene in South Africa over the past four decades, since it represents the first attempt at scientifically identifying the word classes of especially the Sotho languages. In explaining the theoretical principles on which his identification of the parts of speech are based, Van Wyk (1961: 69 et seq.) emphasises the fact that all linguistic features of words must be taken into account when words are to be classified into word categories. Consequently, he distinguishes four principles according to which the words of a language should be classified, namely syntactic, morphological, phonological and semantic principles. He regards the syntactic and morphological principles as dominant principles of classification, since they take the word into account as a complete linguistic sign, i.e. as a unit of sound and meaning. By making use of the dominant principles of classification, the linguist can identify the essential features of a specific word category. Phonological and semantic principles are termed subsidiary principles of classification, the use of which leads to the identification of the additional features of word categories. Van Wyk distinguishes the following eight word classes

based on these principles:

- (2) Van Wyk's word classes (Kosch 1994: 59, 60):
- I. Words with normal phonological structure
 - A. Substantives
 - 1. Nouns
 - 2. Pronouns
 - B. Predicatives
 - 3. Verbs
 - 4. Locative copulative demonstratives
(Interjective demonstratives)
 - C. Morphologically heterogeneous words
 - 5. Adverbs
 - 6. Particles
 - II. Words with paranormal phonological structures
 - 7. Ideophones
 - 8. Interjections

A ninth class "conjunctions" was added at a later stage, presumably by Lombard (1985). This system of word categorisation is not without its problems, as has been pointed out by Louwrens (1994: 133, 134). The recognition of a word class "particles", which implies that these linguistic units enjoy autonomous word status, is especially problematic and has been questioned by a number of scholars. In 1994, in their publication *A Linguistic Analysis of Northern Sotho*, Poulos and Louwrens (1994: 9, 10) propose an alternative set of word classes. Where Van Wyk's classification is based mainly on the structural features of words, Poulos and Louwrens' classification is based on the function of words in sentences, thus showing a strong Dokeian influence:

Words perform various functions when they are used in sentences, and on the basis of these functions, they are classified into the different word categories, or to use the traditional term, "parts of speech" (Poulos and Louwrens 1994: 9).

The parts of speech they distinguish for Sepedi are the following:

- (3) Poulos and Louwrens' word classes (Poulos and Louwrens 1994: 10):
- | | | |
|------|-------------------|--|
| i) | The noun | |
| ii) | The pronoun | Absolute
Quantitative |
| iii) | The demonstrative | |
| iv) | The qualificative | Adjective
Possessive
Relative
Enumerative |
| v) | The verb | |
| vi) | The copulative | Identifying
Descriptive |

- vii) The adverb
 - viii) The ideophone
 - ix) The interjection
 - x) The conjunction
 - xi) The interrogative
- Associative
Locational
Numerous subtypes (formed in many different ways)

The problem of two systems of word-class categorisation in Sepedi is illustrated in (4) below, using the term "constant" as an example:

(4)

English	Part of speech	Afrikaans	Part of speech	Sepedi	Part of speech
constant	Adjective	konstant(e)	Adjective	-sa fetogego (NST&O: 134)	Relative verb (Van Wyk) or qualificative (Poulos and Louwrens)

Due to the different classificatory criteria being used, the contents of the word classes of Van Wyk differ from that of Poulos and Louwrens. An example such as *-sa fetogego* 'constant' is regarded as a verb within the Van Wyk framework, based on the fact that its structure consists of *inter alia* a subject concord and a verb stem *-fetog(a)*. Based on its qualificative function however, Poulos and Louwrens (1994: 90) classify examples such as these as qualificatives, since they serve to qualify or describe a noun. It is also clear from example (4) that word categories in Afrikaans and English do not correspond with word categories in Sepedi.

The problem of word-class categorisation does not occur to the same extent in isiZulu as in Sepedi, as Doke's (1927) word-class classification is generally accepted for isiZulu.³ Compare (5) which shows his slightly revised classification:

(5) Doke's word classes (Doke 1945⁴: 34):

- I. Substantive: (a) Noun 1.
(b) Pronoun 2.
- II. Qualificative: (a) Adjective 3.
(b) Relative 4.
(c) Enumerative 5.
(d) Possessive 6.
- III. Predicative: (a) Verb 7.
(b) Copulative 8.
- IV. Descriptive: (a) Adverb 9.
(b) Ideophone 10.
- V. Conjunctive 11.
- VI. Interjective 12.

From the lexicographer's point of view, a more pressing problem regarding word-class categorisation is similar to that found in Sepedi, namely that the word categories of isiZulu do not correspond to those of Afrikaans and English. Compare for example:

(6)

English	Part of speech	Afrikaans	Part of speech	IsiZulu	Part of speech
constant	Adjective	konstant(e)	Adjective	-ngaguqukiyo (ZT&O ⁺ : 60)	Verb

As can be deduced from the Sepedi and isiZulu examples (4) and (6), this problem is especially relevant in the case of adjectives (and this also applies to adverbs), since there is very little correspondence between these parts of speech in English and/or Afrikaans and their equivalents in Sepedi and isiZulu.

The problem concerning the specification of word categories is further compounded by the extensive use of paraphrases in isiZulu and Sepedi to supply translational equivalents for English terms. Zgusta (1971: 319), as well as Al-Kasimi (1983: 60), terms such paraphrases "explanatory" or "descriptive" equivalents, as opposed to "translational" or "insertible" equivalents. Whereas a translational equivalent can be immediately inserted into a target-language (TL) sentence, the explanatory equivalent cannot always be directly inserted into such a TL sentence.

7.1.1 The use of paraphrases

As was stated earlier, the QEDC and translation dictionaries have certain features in common, which means that the compilers of the QEDC experience problems similar to those of the compilers of translation dictionaries. The underlying cause of these problems is the lack of equivalence or "anisomorphism"⁵ between languages. Swanepoel (1989: 205) points out that a consequence of the basic anisomorphism between languages is that few multilingual dictionaries have been compiled. He argues that the problems relating to establishing equivalence between languages increase proportionally with the number of languages being incorporated into such multilingual dictionaries. It must be pointed out, however, that there are far fewer problems caused by the lack of equivalence between languages in the case of multilingual scientific term dictionaries such as the QEDC, than is the case with (multilingual) general language dictionaries. Pinchuck (1977: 165) points out that scientific language is specialised and tends to become more so, in contrast with the versatility of ordinary language. Scientific language seeks the most economic use of linguistic means in order to achieve standardisation of terms and usage. It also seeks to avoid ordinary language associations and endeavours to define its terms accurately. It is therefore far easier to fit a term directly into a target-language text than would be the case with a word from the general vocabulary.

Languages such as isiZulu and Sepedi — so-called languages of limited diffusion or LLDs — make extensive use of paraphrases to fill onomasiological gaps that occur because scientific and technological terminology in the source language (in this case English), do not exist in the target languages, i.e. isiZulu and Sepedi. Consider (7) and (8) in this regard:

(7) **IsiZulu**

- (a) **Activated complex:** Ukuphithana okubukhuphekhuphe / okuxwayile (lit. complex (lit. entanglement) that is active)
- (b) **Activation energy:** Umdlandla wobukhuphekhuphe (lit. enthusiasm/keenness of activity)
- (c) **Acid anhydride:** I-anidrayidi ye-asidi (lit. anhydride of acid)
- (d) **Chemical:** (adj): -kwesayensi yekhemistri (lit. of (pertaining to) the science of chemistry) (ZT&O: 51)

(8) **Sepedi**

- (a) **Activation energy:** Mafofolo a maatla (lit. enthusiasm of power)
- (b) **Chemical:** (adj): sa khemisi (lit. of (pertaining to) the science of chemistry) (NST&O: 134)
- (c) **Constant:** (adj): -sa fetogego (lit. which does not change) (NST&O: 134)

The problem for the compilers of the QEDC regarding word-class categorisation is that these paraphrases consist of more than one (linguistic) word and can therefore not be said to belong to a single word category.

7.1.2 A possible solution for the problems regarding word-class categorisation

As was indicated earlier, two problems regarding word-class categorisation were identified. The first of these problems concerns the existence of two competing word-class categorisations for Sepedi. The lexicographer must take cognisance of this issue and its linguistic implications and will therefore have to make a choice as to the word categorisation to be used. The most logical option would be to select the system of word categorisation with which the target user is familiar. This will in all probability be the categorisation which is used in the teaching of Sepedi in secondary schools. It will therefore be the responsibility of the lexicographer to ascertain which of the two systems enjoy preference in the teaching of Sepedi at secondary school level in order to enable him/her to make an informed choice in this regard.

Secondly, the extensive use of paraphrases in Sepedi and isiZulu in order to supply translational equivalents for English terms, creates difficulties for the lexicographers of the QEDC in that a term consisting of more than one linguistic word does not easily lend itself to satisfactory categorisation. The compilers of the English and Afrikaans sections of the dictionary were clearly also con-

fronted with similar problems. Compare the examples taken from the QEDC given in (9) below:

(9) **English**

homologous series (*n*) A series of compounds containing common structural elements, but differing in the number of atoms making up the molecule, e.g. alkanes such as C_nH_{2n+2} .

Afrikaans

homoloë reeks (*n*) 'n Reeks verbindings wat dieselfde struktuurelemente het, maar verskil ten opsigte van die aantal atome waaruit die molekule bestaan, bv. alkane soos C_nH_{2n+2} .

In both the English and Afrikaans entries, the lemma consists of an adjective followed by a noun. However, the part of speech is simply indicated as (*n*), apparently since the lemma refers to a single concept.

A similar solution can be applied in the case of the isiZulu and Sepedi paraphrases as listed in examples (7) and (8) earlier. The isiZulu translational equivalents in examples (7)(a), (7)(b) and (7)(c) could all be labelled *noun*, whilst the example in (7)(d) could be labelled *qualificative*, even though the lemmas themselves consist of more than one linguistic word. This would be acceptable since the referent of the lemma in question is a single concept in each case. The same principle can be applied to the Sepedi examples in (8). Even though, linguistically speaking, the Sepedi term in (8)(a) consists of three linguistic words, it can be labelled *noun*, since it refers to a single concept. The examples in (8)(b) and (8)(c) can be labelled as *qualificatives*, since they serve to qualify whichever noun might precede them. This underscores the difference between words as found in general language and terms used in languages for special purposes. The lemma in general language dictionaries consists of a single linguistic word which lends itself easily to categorisation, whereas a term is often complex or may even constitute a phraseological unit as illustrated in (7) and (8) above. It is therefore clear that in terms of categorisation, words and terms cannot be handled in the same manner.

Earlier, we identified two problems regarding the specification of parts of speech in the QEDC. Whereas the first problem (as discussed in the previous sections) dealt with word categorisation in Sepedi and isiZulu, the second problem concerns the lexicographic conventions to be used in indicating the word-class affiliation of a particular lemma. This issue will be discussed in detail in the next section.

8. Lexicographic conventions used to indicate to which word class a specific lemma belongs

The second problem to be addressed in this article concerns the use of abbreviations to indicate the part of speech to which a specific lemma belongs. A standardised list of abbreviations for word categories does not exist in either

Sepedi or isiZulu. Furthermore, no rules on the formation of abbreviations are explicitly stated in dictionaries or in official orthographies of these languages. This makes the formation of new abbreviations very difficult.

Therefore, the only viable option available to us was to abstract principles for the formation of abbreviations from existing examples of abbreviation. The first step was to obtain as many examples of isiZulu and Sepedi abbreviations as possible. The official terminology and orthographies of these languages, as well as various dictionaries, were consulted.⁶ It must be noted, however, that this was by no means an exhaustive study, and that it should be seen as constituting a preliminary investigation only. After these examples had been studied, a number of rules determining the formation of abbreviations could be formulated. These principles were then implemented in drawing up lists of proposed abbreviations for the various word categories of Sepedi and isiZulu. Although a number of abbreviations referring to the isiZulu parts of speech are used in the three monolingual explanatory dictionaries (Nkabinde 1982 and 1985, and Nyembezi 1992), it was found that there are word categories which are not referred to in abbreviated form in these dictionaries. By identifying the principles underlying abbreviation formation, it was possible to devise abbreviations for all the word categories of isiZulu.

The suggested principles for abbreviation formation appear as appendixes A (isiZulu) and B (Sepedi). The proposed abbreviations for the word classes/parts of speech for isiZulu and Sepedi are listed in (10) and (11) respectively.

In example (10), Doke's (1927: 34-35) so-called "fundamental or basic parts of speech" (identified according to the work that they do in the sentence) are given in upper case characters. The so-called "real parts of speech" (identified according to the form in which they appear) are written in lower case characters. Also note that in example (10), the parts of speech for which abbreviations already exist, appear in bold type. These existing abbreviations have been taken from the isiZulu monolingual dictionaries of Nkabinde (1982 and 1985) and Nyembezi (1992). Lastly, note that in the formation of the proposed abbreviations in (10), **any** of the identified principles of abbreviation formation (as set out in appendix A) could have been used. The suggested abbreviations have therefore been chosen relatively arbitrarily, but cognisance was taken of (a) the shape of those abbreviations already in use in Nkabinde's and Nyembezi's dictionaries, and (b) the frequency of occurrence of each principle of abbreviation formation as set out in appendix A. Consider:

(10) **Proposed Abbreviations for the Word Classes/Word Categories of isiZulu**

ENGLISH	ISIZULU	ABBREVIATION	PRINCIPLE
SUBSTANTIVE	USOBIZO	usbz.	1
noun	ibizo	bz.	1
pronoun	isabizwana	sbz.	1
QUALIFICATIVE	ISICHASISO	isich.	4
adjective	isiphawulo	ph.	1

relative	isibaluli	bl.	1
enumerative	inani	nan.	3
possessive	ubumnini	ubumn.	4
PREDICATIVE	ISILANDISO	slds.	1
verb	isenzo	sz.	1
copulative	isibanjalo	sbnj.	1
DESCRIPTIVE	ISIKHANYISO	skhny.	1
adverb	isandiso	sd.	1
ideophone	isenzukuthi	szk.	1
CONJUNCTIVE	ISIHLANGANISO	isihl.	4
INTERJECTIVE	ISIBABAZO	bbz.	1

Note that in the case of the suggested Sepedi abbreviations in (11), word categories not distinguished by Van Wyk (1961), but listed by Poulos and Louwrens (1994), are marked with an asterisk:

(11) *Proposed Abbreviations for the Word Classes/Word Categories of Sepedi*

ENGLISH	SEPEDI	ABBREVIATION	PRINCIPLE
Noun	Leina	Lna. / L.	3 / 1
Pronoun	Lešala	Leš.	2
Verb	Lediri	Led. / Ldr.	2 / 1
Demonstrative-copulative	Lešupileba	Ll.	1
Adverb	Lehlathi	Lehl.	2
Particle	Sekantšu	Sek. / Sn. / Skn.	2 / 1 / 1
Conjunction	Lekopanyi	Lkp.	1
Ideophone	Leekiši	Lki.	3
Interjection	Lelahlelwa	Lla.	3
Demonstrative	*Lešupi	Lpi.	3
Qualificative	*Tlhaodi	Tlha.	2
Copulative	*Leba	Leb.	2
Interrogative	*Lebotšiši	Ltši.	3

Clearly, not all of the listed abbreviations will be relevant for use in the QEDC, but we are of the opinion that they could be of use in future, specifically when compiling monolingual dictionaries for isiZulu and Sepedi. This study will therefore contribute not only to lexicography in the African languages, but also to isiZulu and Sepedi linguistics. The principles that have been formulated could also be helpful to other grammarians and linguists for the formation of abbreviations in these languages.

9. Conclusion

It was found that the inclusion of syntactic information in the isiZulu and Sepedi entries of the QEDC, led to certain practical lexicographic problems. These problems are caused by the unique linguistic structure of the African

languages and cannot be solved without considering the grammatical structure of these languages as well as the unresolved linguistic issues present in these languages.

Notes

1. The isiZulu and Sepedi equivalents of this entry are provisional.
2. *Northern Sotho Terminology and Orthography* (1988).
3. This does not mean, however, that other word-class categorisations have not been suggested for this language, *inter alia* those of Van Eeden (1956), Cope (1957), Ziervogel (1964), Van Wyk (1958, 1961 and 1967) and Nkabinde (1975), to name the most important classifications.
4. *Zulu Terminology and Orthography* (1980).
5. Anisomorphism < **an-** 'not' + **-iso-** 'equivalent' + **-morph** 'form' (Swanepoel 1989: 205).
6. These works are listed under (a) in the bibliography.

Bibliography

A. Dictionaries, terminologies and orthographies

- Dent, G.R. and C.L.S. Nyembezi. 1977. *Scholar's Zulu Dictionary*. Pietermaritzburg: Shuter and Shooter.
- Department of Education and Training. 1980. *Zulu Terminology and Orthography*. Pretoria: The Government Printer.
- Department of Education and Training. 1988. *Northern Sotho Terminology and Orthography*. Pretoria: The Government Printer.
- Doke, C.M., D.M. Malcolm, J.M.A. Sikakana and B.W. Vilakazi. 1990. *English-Zulu/Zulu-English Dictionary*. Johannesburg: Witwatersrand University Press.
- Nkabinde, A.C. 1982. *Isichazamazwi 1*. Pietermaritzburg: Shuter and Shooter.
- Nkabinde, A.C. 1985. *Isichazamazwi 2*. Cape Town: Oxford University Press.
- Nyembezi, S. 1992. *AZ Isichazamazwi sanamuhla nangomuso*. Pietermaritzburg: Reach Out Publishers.

B. Other literature

- Al-Kasimi, A.M. 1983. *Linguistics and Bilingual Dictionaries*. Leiden: Brill.
- Bergenholtz, H and S. Tarp. 1995. *Manual of Specialised Lexicography*. Amsterdam: Benjamins.
- Carstens, A. 1997. Issues in the Planning of a Multilingual Explanatory Dictionary of Chemistry for South African Students. *Lexikos* 7: 1-24.
- Cope, A.T. 1957. The Grammatical Structure of Zulu. *African Studies* 16: 210-220.
- Doke, C.M. 1927. *Text Book of Zulu Grammar*. Johannesburg: The University of the Witwatersrand Press.
- Doke, C.M. 1945⁴. *Text-Book of Zulu Grammar*. London: Longmans, Green and Co. Ltd.
- Kosch, I. M. 1994. *A Historical Perspective on Northern Sotho Linguistics*. Pretoria: Via Afrika.
- Lombard, D.P. 1985. *Introduction to the Grammar of Northern Sotho*. Pretoria: J.L.van Schaik.
- Nkabinde, A.C. 1975. *A Revision of the Word Categories in Zulu*. Unpublished D.Litt. et Phil. thesis. Pretoria: University of South Africa.

- Pinchuck, I.** 1977. *Scientific and Technical Translation*. London: Deutsch.
- Poulos, G. and L.J. Louwrens.** 1994. *A Linguistic Analysis of Northern Sotho*. Pretoria: Via Afrika.
- South African Government.** s.a. *Green Paper on Science and Technology*. Pretoria: Government Printer.
- South African Government.** 1994. *White Paper on Reconstruction and Development*. Pretoria: Government Printer.
- South African Government.** 1996. *White Paper on Science and Technology*. Pretoria: Government Printer.
- Swanepoel, P.H.** 1989. *Only Study Guide for LEKPER-Q*. Pretoria: University of South Africa.
- Van Eeden, B.I.C.** 1956. *Zoeloe-grammatika*. Stellenbosch/Grahamstad: Die Universiteitsuitgewers en -Boekhandelaars.
- Van Wyk, E.B.** 1958. *Woordverdeling in Noord-Sotho en Zoeloe*. Unpublished D.Litt. thesis. Pretoria: University of Pretoria.
- Van Wyk, E.B.** 1961. Die woordklasse van Noord-Sotho. *Feesbundel vir Prof. Dr. Jan Antonie Engelbrecht*: 69-85. Johannesburg: Afrikaanse Pers Beperk.
- Van Wyk, E.B.** 1967. Word Classes: Northern Sotho. *Lingua* 17: 230-260.
- Zgusta, L.** 1971. *Manual of Lexicography*. The Hague: Mouton.
- Ziervogel, D.** 1964. Bantoetaalstudie in Suid-Afrika. Van der Merwe, H.J.J.M. (Ed.). 1964. *Studierigtinge in die Taalkunde*: 208-224. Pretoria: Van Schaik.

Appendix A The Formation of Abbreviations in isiZulu

Some 73 abbreviations were identified in the dictionaries and terminology and orthography listed under (a) in the bibliography, and the following principles of abbreviation formation were abstracted from these examples:

PRINCIPLE 1: (Approx. 38% of all examples.)

The first letter of one or more syllables of the word is used. This principle also applies to abbreviations of compound nouns and where more than one word is abbreviated. Where two or more separate words are abbreviated, the first letter of every *stem* is used, e.g.:

ENGLISH	ISIZULU	SOURCE
C.O.D. (cash on delivery)	K.F. (Khokha i-fika)	ZT&O: 55
C.W.O. (cash with order)	K.O. (u-kheshe ne-oda)	ZT&O: 67
G.C.M. (greatest common measure)	F.V.K. ((i)-fektha (i)m-vama-en-kulu)	ZT&O: 99

It is interesting to note that in the case of syllables containing nasal compounds, the nasal in the compound is never used in the abbreviation, but rather the consonant that follows the nasal. (However, this does not seem to be the case with the so-called "vocative forms", as is illustrated further on.) See the following examples in this regard, where the nasal compounds concerned have been highlighted:

ENGLISH	ISIZULU	SOURCE
cwt. (hundredweight)	ct. (i- ce-nta -li)	ZT&O: 67
lb. (pound)	pd. (i- pha-wu-ndi)	ZT&O: 120
L.C.M. (least common multiple)	P.V.C. ((isi) phindo- (i)mvama- ngci) Take note of the nasal compound in the last syllable where the consonant (not the nasal) is used, i.e. n + c > ngc	ZT&O: 120
aux. (auxiliary)	sgsz. (i- si-nga-se-nzo)	Nyembezi 1992

PRINCIPLE 2: (Approx. 34% of all examples)

The abbreviation is borrowed from English, i.e. the same abbreviation as in English is used, e.g. a.m., p.m., i.e., cm., P.S.

PRINCIPLE 3: (Approx. 5% of all examples)

The first syllable of the stem is combined with the first letter of the next syllable, e.g.:

ENGLISH	ISIZULU	SOURCE
doz. (dozen)	daz. (i- da-zi -ni)	ZT&O: 78
h (height)	phak. (uku- pha-ka -ma)	ZT&O: 103
sec. (second)	sek. (i- se-ke -ndi)	ZT&O: 178

PRINCIPLE 4: (Approx. 5% of all examples)

The (full) class prefix is followed by the first letter of the stem. These abbreviations all seem to be of words of a general type commonly used in textbooks, i.e. "for example", "answer", "exercise" and "figure". See:

ENGLISH	ISIZULU	SOURCE
e.g.	izib. (izi-bonelo)	Dent and Nyembezi 1977: ix; ZT&O: 82
ans. (answer)	umph. (um-phumela)	ZT&O: 27
ex. (exercise)	ums. (um-sebenzi)	ZT&O: 87
fig. (figure)	umd. (um-dwebo)	ZT&O: 91

Vocative Nouns / Terms of Address:

IsiZulu vocative nouns / terms of address seem to form a class of their own. They only conform to the general principles in the broadest of terms. Rather, they seem to follow their own set of rules as determined by their morphological structure, e.g.:

ENGLISH	ISIZULU	SOURCE
Mr.	Mnu. (Mnumzane)	Dent and Nyembezi 1977: ix
Rev. (Reverend)	Mfu. (Mfundisi) Mf. (Mfundisi)	ZT&O: 171 Dent and Nyembezi 1977: ix; ZT&O: 133
Miss	Nks. (Nko-sa-za-ne)	Dent and Nyembezi 1977: ix
Mrs.	Nkz. (Nko-si-ka-zi)	Dent and Nyembezi 1977: ix

No further general principles could be abstracted from the remainder of the examples.

Punctuation:

It would seem that in isiZulu a full-stop is generally used after every abbreviated word. This is also the case with (conjunctively written) compound nouns that consist of more than one word, e.g.:

ENGLISH	ISIZULU	SOURCE
H.C.F. (highest common factor)	F.V.K. (i-fektha + (i)-mvama eli-khulu)	ZT&O: 106
cub. yd. (cubic yard)	yd. kb. (i-ya-di- + (i)-kyu-bhu)	ZT&O: 66
sq. ml. (square mile)	ml. sk. (i-ma-ye-la + (i)si- kwele)	ZT&O: 189

Where the Zulu abbreviation has been borrowed from English, however, the same punctuation as in English is used, e.g. a.m., i.e., cm, g, ha, P.S.

Appendix B The Formation of Abbreviations in Sepedi

The only Sepedi source in which abbreviations are listed, is *Northern Sotho Terminology and Orthography* (1988). It must be pointed out that in this specific source there is no consistency regarding the use of full-stops after abbreviations. The original punctuation as it appears in the said work has thus been retained, pending further research into this matter.

PRINCIPLE 1: (Accounts for approx. 57% of all examples)

The first letter(s) of all of/some of the syllables of the word is (are) used:

ENGLISH		SEPEDI	
Full form	Abbreviation	Full form	Abbreviation
Wednesday	Wed.	Laboraro	Lbr.
post meridiem	p.m.	thapama	tp.
height	ht.	bogodimo	bg.

In the case of compounds, the first letter(s) of the words forming the compound should be used. This is also the case where two or more separate words are abbreviated:

ENGLISH		SEPEDI	
Full form	Abbreviation	Full form	Abbreviation
etcetera	etc.	bjalobjalo	bj.bj
cash on delivery	C.O.D.	kešhe kamogedišong	k.k.

PRINCIPLE 2: (Accounts for approx. 33% of all examples)

The first syllable only, or the first syllable and the first letter(s) of the following syllable are used:

ENGLISH		SEPEDI	
Full form	Abbreviation	Full form	Abbreviation
Monday	Mon.	Mošupologo	Moš
dozen	doz.	tosene	tos.
Reverend	Rev.	moruti	mor.
figure	fig.	seswantšho	ses.

PRINCIPLE 3: (Accounts for approx. 10% of all examples)

First and last letters of the word and in some cases, the first letter(s) of other syllables should be used:

ENGLISH		SEPEDI	
Full form	Abbreviation	Full form	Abbreviation
Mister	Mr.	Morena	Mna
Mistress	Mrs.	Mohumagadi	Mdi

Note: When a syllable starts with hl, the use of h in the abbreviation should be avoided. A better option for the abbreviation of Labohlano would therefore be Lno or possibly Lbo, which would be in accordance with principle 3.

From Business Corpus to Business Lexicon*

Li Lan (*egllan@polyu.edu.hk*) and Grahame T. Bilbow
(*eggrabil@polyu.edu.hk*), *Department of English, Hong Kong Polytechnic University, Hong Kong*

Abstract: Language corpora are now indispensable to dictionary compilation. They help broaden the role of the dictionary from standardizing the vocabulary to recording a language. The trilingual corpus generated by the Hong Kong Polytechnic University gives a record of business languages used in Hong Kong. It differs from other corpora in that (1) it includes English, Chinese and Japanese; (2) it shows local characteristics; and (3) it focuses on a specific area (financial services, including banking, accounting, auditing, insurance and investment). The paper discusses various issues of setting up a tricorpus, and how to make full use of the data to generate a trilingual lexicon.

Keywords: MULTILINGUAL, SPECIAL PURPOSE, CORPUS, LEXICON

Opsomming: Van sakekorpus tot sakeleksikon. Taalkorpora is tans onontbeerlik vir die samestelling van woordeboeke. Hulle help om die rol van die woordeboek uit te brei vanaf die standaardisering van die woordeskat tot die optekening van 'n taal. Die drietalige korpus wat deur die Hongkongse Politegniese Universiteit ontwikkel is, verskaf 'n opgawe van die sake tale wat in Hongkong gebruik word. Dit verskil van ander korpora deurdat (1) dit Engels, Chinees and Japanees insluit; (2) dit plaaslike eienskappe vertoon; en (3) dit op 'n spesifieke gebied (finansiële dienste, insluitende bankwese, rekeningkunde, ouditering, versekering en belegging) fokus. Die artikel bespreek verskillende aspekte van die totstandbrenging van 'n drietalige korpus, en hoe om volle gebruik te maak van die data om 'n drietalige leksikon te genereer.

Sleutelwoorde: MEERTALIG, SPESIALE DOEL, KORPUS, LEKSIKON

1. Introduction

There is a large variety of dictionaries to satisfy different users worldwide. Among them, English monolingual learner's dictionaries undoubtedly occupy one of the most competitive markets in the publishing world. In Hong Kong, as in many other ESL/EFL countries, the "Big Four" of British publishers have taken a large market share, not only with monolingual but also with bilingual versions. However, room still exists for culture specific dictionaries for local users. Tom McArthur (1998: 206) predicted that in the new millennium there

* This paper was presented at the Fifth International Conference of the African Association for Lexicography, held at the University of Stellenbosch, Stellenbosch, 3-5 July 2000.

will be four trends in compiling dictionaries: globalization, localization, bilingualization and semi-bilingualization.

Hong Kong is one of the most important international financial centres in the world. According to government statistics, there are 172 internationally or locally licensed banks, 209 insurers, and 930 CPA (Certified Public Accountant) firms conducting business in Hong Kong. It is generally accepted that good communication is central to any successful business. Investigating business languages in Hong Kong enables one to form a clear picture of how languages function in business. The current research project, started in 1999, has the following objectives:

- to establish a trilingual corpus of business languages used in Hong Kong (English, Standard Written Chinese and Japanese),
- to compile a trilingual content-based lexicon, and
- to carry out linguistic research into and between languages in business in Hong Kong.

2. Generating the business corpus

The Business Lexicon started with the setting up of a business corpus. This procedure is not new, of course. Since the 1970s the field of dictionary making has been influenced by empirical and corpus-based methods. "All the major UK dictionary-publishers now have access to large and diverse corpus resources which provide the raw materials for a far more reliable description of English" (Rundell 1998: 315). Large corpora of general English, such as the British National Corpus (BNC) and the Bank of English, are regarded as standard corpora in that they include both spoken and written texts with a balanced mixture of different types. The texts are basically monolingual, produced by native English speakers only. These corpora are normally the result of a collaboration between linguists, lexicographers and computer scientists.

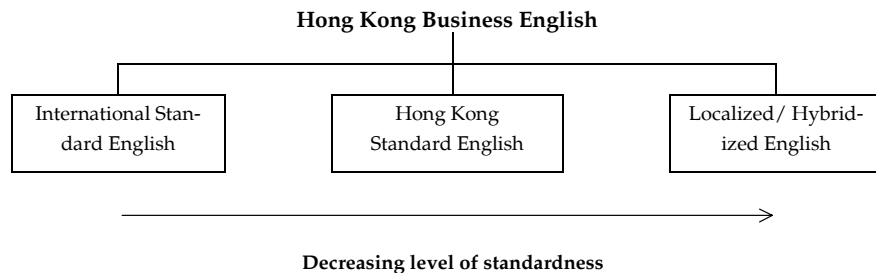
In addition to Standard English corpora, modern technology has also enabled the creation of many free selection corpora — involving guided text collection from the World Wide Web, newspapers and book CDs, and other machine-readable sources. Each corpus is unique in its own right and for its own purposes. Many free corpora are also multilingual and serve the purpose of keeping a record of languages in everyday use. The Hong Kong PolyU Business Corpus (PUBC) is a corpus which is compiled for special purposes in a specific region, and bears vocational, geographical, national and regional features.

2.1 Delimiting the languages of the tricorpus

The languages used in Hong Kong are commonly known as "two languages in three tongues". However, the extent to which they are used and the situations

in which they appear, vary considerably. English and modern standard Chinese are official languages used in formal situations, e.g. in law, in business, in education and by government institutions. Spoken local Chinese (Cantonese) functions as the spoken lingua franca among 97% of the ethnic Chinese population in Hong Kong. In any working environment, oral communication tends to be dominated by Cantonese or a mixed code (English/Cantonese) whereas formal internal and external written communication such as through letters, memoranda or notices tend to be in English. When working on the International Corpus of English in Hong Kong, Bolt and Bolton assumed that the particular range and quality of the Hong Kong data would be affected by a matrix of sociolinguistic relationships. They (Bolt and Bolton 1998: 198) found that "these relationships include those between the ethnic and linguistic background; between the local educational system and the linguistic community; and between the linguistic backgrounds of local English speakers and salient features of the type of English found in Hong Kong". Interestingly, they (Bolt and Bolton 1998: 212) comment that "more people than ever are speaking 'good English', and more people than ever are speaking 'bad English'".

Written English business texts in Hong Kong fall into three categories. The first involves internationally shared materials, such as financial news reports by Reuters, Associated Press, and Bloomberg. Within the branches of large international companies such as HSBC, AXA and AIA, there are also a fairly large number of shared documents (such as regulations and annual reports) from parent companies. The second type of text are those written locally by expatriates and Chinese educated in English-speaking countries, possessing a proficiency in English comparable to that of native speakers. The third type of text are those written by local employees for daily internal and external communication. This trajectory can be presented in the following chart.



The first block can arguably be regarded as an international standard, which provides norms for the local language users. "Hong Kong Standard English" refers to English with local norms accepted in the region. Users of HKSE commonly act as standardisers in companies producing publicly distributed texts in English. "Localized/Hybridized English" refers to a mixture of English and Cantonese. These three types are linked together by certain shared features;

there is no clear division between them. However, few would dispute the striking differences between the two ends of the spectrum. The data in the PUBC reflects the first two types, as all the collected texts are publicly distributed and demonstrate a high level of standardisation.

Chinese is another official language in Hong Kong. Although its written form is based on standard Putonghua from the Mainland, certain localized features can be clearly identified. Since the corpus is limited to languages used in Hong Kong, all the Chinese texts have been collected from local institutions. Apart from geographical and linguistic limits, there also exists a technical barrier. "Big5" is the coding system for complex Chinese characters used in Hong Kong and Taiwan, not compatible with the "GB" system for simplified characters used in Mainland China.

The language contained in the PUBC is Japanese. However, very little written Japanese appears to be used in Hong Kong. The majority of Japanese companies in Hong Kong tend to use English as medium of written communication. Unlike the varieties of English and Chinese, Japanese exhibits few formal differences at home and abroad, and very little variety appears to be acceptable. Therefore, in the PUBC, Japanese texts have been collected from resources in Japan as well as from Hong Kong. Since English has been made the second language in Tokyo, and the websites of big financial institutions in Japan tend to be bilingual, the temptation to include parallel texts (Japanese/English) in the tricorpus was great. However, this temptation was resisted for the following reasons. First, Japan has never been an English-speaking territory and in Japan English is almost entirely a "learnt" language. Second, the principal focus of the project has been defined as "English used in Hong Kong". The inclusion in the corpus of English from Japan might cause problems in its description. It is hoped that with texts from the same domain and of the same type such as business news, company reports and government policy statements, the Japanese texts will be thematically parallel to the English and Chinese texts.

2.2 Delimiting business language

It has been noticed that, in particular domains, the use of language is more predictably structured and subject to less ambiguity. Using this observation, computational linguists are now able to postulate the existence of special languages or "sublanguages". A sublanguage in the definition of Grishman and Kittredge (1986: ix) is a variety of language used in a given science or technology that is "not only much smaller than the whole language, but is also more clearly systematic in structure and meaning:

... [It is] a subsystem of language that behaves essentially like the whole language, while being limited in reference to a specific subject domain. In particu-

lar, each sublanguage has a distinctive grammar, which can profitably be described and used to solve specific language-processing problems.

The concept of business can encompass many sectors. A business dictionary may cover terms concerning all business life, from office to stock exchange, and from international trade fair to classroom. With a limited budget, the project needed to be kept manageable. Since financial services are commonly viewed as the most significant for Hong Kong, business language has been limited to the language used in financial sectors, including banking, auditing, accounting, investment and insurance. During the process of data collection, the boundaries between the sectors were naturally found to be blurred: banks provide services of investment and insurance, accounting companies offer auditing and investment consultancy, and insurance companies cover investment services — therefore differences between individual sectors tend to be de-emphasised.

2.3 Defining text types

The design of the corpus focuses on a range of internal and external categories of business writing including business news, corporate annual reports, news releases, newsletters, minutes, posters, notices, leaflets, letters, faxes, memoranda and emails. We aimed at as even a distribution as possible between each of these types of texts. Our resources included the World Wide Web, printed materials from companies, government institutions and library on-line databases. The bilingual situation in Hong Kong has made it possible for us to get many parallel texts and thematically linked news reports.

We had much difficulty in obtaining internal administrative and regulatory writing and correspondence. The issue of confidentiality, as also found by Bolt and Bolton, made it very difficult to obtain texts of business transactions. The ICE (International Corpus of English) in Hong Kong settled on a broad conception of business transactions, using two-thirds business texts from the education sector (Bolt and Bolton 1996: 207). The fact that such texts exist but are unavailable to us, was a matter of some concern in obtaining a "balanced" corpus. However, to produce a business lexicon, it is not necessary to give priority to any particular text type. Oostdijk (1998: 169) argued that while general purpose corpora need to be balanced and represent a wide range of styles and registers, corpora representative of a single variety are already defined a priori on the basis of their specific domain of use and topic, "issues of corpus design only play a very minor role".

2.4 Setting up the database

The project started in June 1999 and 1.2 million words in each language have been collected, gleaned and put into three parallel subcorpora: English, Chinese and Japanese. The files are stored in ASCII format and the demographic

details of each file are kept in Microsoft ACCESS to create an index system for future retrieval. The records include thirteen items: Title, Author, Origin of the author, Text type, Source, Date, Language, Sector, Keywords, Length, Filename, Parafilename and Text.

3. Towards the business lexicon

3.1 The target user group

In the competitive dictionary market, the key to success rests in shaping the dictionary to meet the needs of its users. We positioned our lexicon as a reference tool for the professional discourse community in Hong Kong, i.e. for those in the business sector and those planning to join this sector.

3.2 Vocabulary control

Vocabulary control is always a central concern in dictionary compilation. A context-based lexicon starts from a wordlist generated from a corpus. Of the three subcorpora, the English subcorpus was the base or starting point of the lexicon. Using WordSmith Tools (Oxford University Press), we produced a preliminary wordlist, giving the frequency of each word form. The 1.2 million word texts generated a list of 22 600 types, with a type/token ratio 1.80.

It is clear that a wordlist from a concordance can in no way be used directly for the entries of a lexicon. Such a wordlist includes different forms of a word: word stem, inflected forms, run-on forms, derivatives, subderivatives and even nonsense data. For instance, the base word *act* produced 20 tokens.

No.	Word	Freq.	%
250	act	168	0.01
251	acted	30	
252	acting	133	0.01
253	action	213	0.02
254	actions	97	
255	activate	2	
256	activated	3	
257	active	169	0.01
258	actively	82	
259	actives	1	
260	activism	2	
261	activist	4	
262	activists	6	
263	activities	537	0.04
264	activity	1	
265	activity	156	0.01

266	actor	2
267	actors	1
268	acts	31
268	actvities	1

This list needs thorough trimming. The word *act* should be approached carefully. Simple words in English are normally polysemic and problematic. Without tagging, there is no way of knowing how many of the occurrences are nouns, and how many verbs. *Act* can be a noun or a verb; *acts* can be a third person singular form or a plural form of a noun. We also need to classify this group into lexical morphemes and grammatical morphemes.

- Grammatical / inflectional morphemes: acts, acted, acting/action, actions, ...
- Lexical / derivational morphemes: act, action, active, activate, actively, ...
- Nonsense words: activity, actvities, ...

Lexical morphemes are normally taken as entries in general dictionaries. For a special purpose dictionary, entries should be relevant to the subject area. More factors need to be taken into consideration. The words we finally put into the business lexicon as entries are *act, action, acting, active, activate, activity*.

3.3 New words or non-words

The PUBC has recorded several instances of recent language change: borrowing words from other languages; creating new words; using the names of people or places to refer to a related object; making shifts and conversions where meanings of words or their parts of speech change.

The introduction of modern technology in the business sector causes the emergence of many new words. Words with prefixes *cyber-*, *e-*, *elec-* and *i-* are now commonly used as proper nouns such as names of websites or publications, and many also shift to proper lexical words. The following are some new words and their occurrences in PUBC.

Word	Freq.
cyberbanking	10
cyberbase	8
cybercash	1
cyberforce	2
cybermall	1
cyberport	25
cyberspace	28
cybertrading	2
cyberworks	193

The word *dotcom* which appears six times in the corpus, is widely utilized by Internet companies. There are *dotcom connections*, *dotcom stores*, *dotcom consulting*, *dotcom marketing*, *dotcom industry* and so on. The following concordance is from the PUBC.

But he said the fast-growing	<u>dotcom</u>	companies helped improve the
hnology stock prices soar and	<u>dotcom</u>	companies rush in to the mark
in. "No doubt some of our	<u>dotcom</u>	companies will do better than
the new economy, we own no	<u>dotcom</u>	companies. What we've been lo
ong STORY: WITH the raging	<u>dotcom</u>	fever, local investors are fr
o switch from previously hot	<u>dotcom</u>	issues to quality technology

Another new word identified is *bancassurance*, which means an insurance service provided by a bank. It was first used as the name of a project or programme and has become a special term in business English. However, even though it appears eight times in the PUBC, there is little evidence to suggest it is sufficiently popular to be recorded in a dictionary.

As the project started immediately before the new millennium, it bears strong features of this period. The Y2K threat was a major concern then due to the fact that modern business relies heavily on computers. The word Y2K has a frequency of 380 in the PUBC, occurring in business news reports, company reports and internal and external documents. In contrast, no occurrence of Y2K is found in the BNC. We therefore have to consider carefully whether or not to include Y2K in the lexicon, for it may be more a general English word than a genre-specific technical term.

3.4 Statistical analysis

Although a frequency list provides basic statistical information from a corpus, we cannot use frequency as the sole criterion for the choice of words. The top 20 most frequent words in the PolyU Business Corpus are mainly grammatical words, very much similar to other English corpora. A comparison with the British National Corpus (BNC) is shown in Table 1.

Next in order are common lexical words which depend strongly on the design of the corpora, the topic and textual sources. Evidences of the register collected in the project can easily be seen from the list: the words *market*, *company*, *financial bank* and *insurance* appearing within the top 50 words, and occurring more frequently than grammatical words like *we*, *he*, *up* and *if*. There are also proper nouns with high frequencies, such as names of people, places, companies, and countries. The basic decision was to exclude grammar words, general English words and proper names. The names of important financial institutions and foreign currencies were put in appendices.

As many linguists have noticed, "a statistical model may not necessarily represent the use of a particular word in a particular context" (Oakes 1998: 43).

If a comparatively small corpus is used, frequency should not be the only criterion for word selection. For instance, although the word *millennium* appears 132 times, *hotel* 137 times and *university* 122 times, they are not relevant to a lexicon of financial terms. On the other hand, about 45% of the words in the list appear only once. Even with very low frequency, however, they can be highly relevant to the subject area. Although the words *debenture* and *abate* had two instances and *overdraw* had only one in the 1.2 million words, they are all recorded in the lexicon.

PUBC				BNC			
No.	Word	Freq.	%	No.	Word	Freq.	%
1	the	91,845	6.79	1	the	109,524	5.15
2	of	45,984	3.40	2	and	54,545	2.56
3	to	39,315	2.91	3	of	50,741	2.38
4	and	32,225	2.38	4	to	49,433	2.32
5	in	28,924	2.14	5	a	42,256	1.99
6	a	23,753	1.76	6	I	38,787	1.82
7	for	13,195	0.98	7	in	35,229	1.66
8	is	11,651	0.86	8	it	34,243	1.61
9	on	10,726	0.79	9	you	30,562	1.44
10	be	9,271	0.69	10	that	30,134	1.42
11	that	9,171	0.68	11	s	27,977	1.31
12	as	9,162	0.68	12	is	20,400	0.96
13	by	8,834	0.65	13	for	17,548	0.82
14	with	7,575	0.56	14	was	16,138	0.76
15	or	7,185	0.53	15	on	15,394	0.72
16	said	6,972	0.52	16	n't	14,368	0.68
17	will	6,901	0.51	17	we	14,132	0.66
18	Hong	6,814	0.50	18	they	13,015	0.61
19	it	6,553	0.48	19	be	12,777	0.60
20	at	6,387	0.47	20	he	12,446	0.58

Table 1 Comparison of the most frequent words in the PUBC and the BNC

The other method of word selection is to check the keyness of a word. The British National Corpus (sample, 1.2 million words on CD-ROM) was used as a reference corpus. Although it is composed of both spoken and written texts (as opposed to the PolyU written corpus), it is the only standard English corpus currently available. The table of keyness was obtained by Wordsmith Tools. The top 23 words are shown in Table 2.

Of the words obtained from the corpus, some are highly specific, some are semispecific, and others are general English words. Keyness indicates which words recur consistently in texts of a given genre. For example, the word *consolidate* was found to occur in many of a set of business annual reports. It did not occur very often in each of them, but did occur much more consistently in

the business reports than in a mixed set of texts. The table of keyness helped us to make decisions in terms of word selection. The higher the number, the more relevant the word is to the lexicon. Words with high frequency may have very low keyness. For instance, the word *university* appears 122 times in PUBC, but its keyness is very low, -305.5, and it was excluded from the list of the business lexicon. Other words with high frequencies but low keyness are *problem* (f = 309, KN = -29.20), *few* (f = 387, KN = -24.8), and *always* (f = 159, KN = -315.3). By contrast, some words, although with low frequencies, have comparatively high keyness and are therefore retained in the lexicon, for example, *deflation* (f = 67, KN = 99.4), *inflow* (f = 60, KN = 91.9), and *liquidation* (f = 62, KN = 85.1). In addition to morphological analysis, frequency count and keyness study, we will also seek expert advice in order to minimize undue personal judgement in word selection.

Keywords (keyness)							
No.	Word	Freq.	All.Lst %	Freq.	BNCS.Lst %	Keyness	P
1	Hong	6,814	0.50	92		12,017.5	0.000000
2	Kong	6,299	0.47	94		11,040.1	0.000000
3	market	5,493	0.41	693	0.03	6,740.0	0.000000
4	per	5,266	0.39	587	0.03	6,732.2	0.000000
5	cent	4,521	0.33	358	0.02	6,349.3	0.000000
6	securities	2,951	0.22	45		5,160.6	0.000000
7	HK	2,591	0.19	0		4,902.0	0.000000
8	financial	3,362	0.25	240	0.01	4,833.7	0.000000
9	insurance	2,953	0.22	146		4,553.6	0.000000
10	bank	3,249	0.24	283	0.01	4,454.4	0.000000
11	company	3,736	0.28	599	0.03	4,176.0	0.000000
12	the	91,845	6.79	109,524	5.15	4,040.3	0.000000
13	exchange	2,331	0.17	135		3,495.6	0.000000
14	China	2,192	0.16	126		3,291.6	0.000000
15	billion	1,945	0.14	60		3,198.9	0.000000
16	companies	2,399	0.18	246	0.01	3,143.3	0.000000
17	of	45,984	3.40	50,741	2.38	3,095.6	0.000000
18	stock	2,264	0.17	239	0.01	2,940.7	0.000000
19	million	2,653	0.20	446	0.02	2,903.9	0.000000
20	investors	1,698	0.13	47		2,825.5	0.000000
21	shares	2,274	0.17	276	0.01	2,824.8	0.000000
22	banks	1,715	0.13	75		2,695.0	0.000000
23	investment	1,875	0.14	142		2,658.9	0.000000

Table 2 Keyness of PUBC against BNC 1.2m sample

3.4 Concordance in the lexicon

The idea of a corpus-based lexicon is to provide users with words in their context. Since the Hong Kong PolyU Business Lexicon has been positioned as a pedagogical dictionary, the established conventions of dictionary microstruc-

ture will still be used to meet the needs of English learners. An entry in the lexicon includes the following items:

1. Head word/Entry
2. Pronunciation
3. Part of speech
4. Definition
5. Phrases
6. E-concordance
7. Derivatives
8. Concordance of derivatives
9. Synonyms and antonyms
10. Chinese equivalent
11. Chinese concordance
12. Japanese equivalent
13. Japanese concordance

The special feature of this lexicon is the limiting of the definitions of a word and the examples the concordance provides for the business sector. The word *margin* normally has more than six senses in a general English dictionary, but we only included three in our lexicon: "1. a permissible difference; 2. security deposit; 3. gross profit". The definitions are directly related to a financial context and the senses are supported by evidences from the PUBC.

Semitechnical words are normally polysemic and have different meanings in different registers. Research into dictionary users has revealed that such words cause more problems to foreign language learners than either general or highly technical terms (Li 1998: 72). The concordance from the business corpus can provide clear examples of how they are used and what they mean in a specific domain. The word *cushion* in the corpus does not mean "a soft pillow to make sitting and resting more comfortable or something soft to decrease a collision"; rather, it means "to lessen the adverse effect of".

<p>n.</p> <p>is ahead and have a firm <u>cushion</u> against any new contingencies the banks have a large capital <u>cushion</u> and a drop in profits (or ev content with a weak yen as a <u>cushion</u> for its economy, but the US a ng also proposes creating a <u>cushion</u> for the management of interba ese H-shares to gain a bit of <u>cushion</u> from US interest rate worries se of their relatively strong <u>cushion</u> of capital and liquidity." n adequate level to provide a <u>cushion</u> of security for depositors AMRO Asia (Holdings) Ltd will <u>cushion</u> such provisions which the Boa with FRR has provided a sound <u>cushion</u> to both investors and market liquidity so as to give them a <u>cushion</u> to guard against any unforese he LAF has already provided a <u>cushion</u> to prevent sharp interest rat</p> <p>v.</p> <p>table. But companies can best <u>cushion</u> the blow by establishing an</p>

t rise for BP Amoco helped to *cushion* the FTSE's decline. Defensi
mittee. This should help to *cushion* them against the adverse impa

The concordance provides evidence of how the word *cushion* is used in financial services rather than in everyday speech. Other examples are *flow*, *liquid*, *curb*, and *lobby*. Such metaphors have lost their figurative content and have become genre-specific, essential to be included in a specialised lexicon.

Color words in the business-oriented corpus also have extended meanings, e.g. *blue chips*, *red chips*, *gray market*, *green shares*, and *in the black*. Many examples of this kind can be found with semitechnical words. Showing the range of their use in a certain field in a specialized dictionary can therefore be more convenient for language learners than a general English dictionary.

The trilingual project continues to develop. "Because all humans have the same basic perceptual apparatus and share many other experiences, there would be some strong similarities in the structuring of semantic space across languages" (Hatch and Brown 1995: 116). However, not only do languages differ in the number of terms they use for a concept, but also the range of meaning of each term may cover the concept in different ways.

Conclusion

A corpus-based lexicon is extremely useful to users in a professional discourse community. However, it is essential to have a clear model before building up a corpus. Decisions need to be made at an early stage regarding language(s), text type, database structure and potential end users. With the help of a self-selected corpus, it is possible to compile a dictionary or a lexicon targeting a special user group, which benefits from a high level of specialty and currency.

Acknowledgment

The project is sponsored by the Hong Kong PolyU Postdoctoral Fellowship Scheme.

References

- Biber, D., S. Conrad and R. Reppen. 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Bolt, P. and K. Bolton. 1996. The International Corpus of English in Hong Kong. Greenbaum, S. (Ed.). 1996. *Comparing English Worldwide*: 197-214. Oxford: Clarendon Press.
- Grishman, R. and R. Kittredge. 1986. *Analyzing Language in Restricted Domains: Sublanguage Description and Processing*. Hillsdale, N.J.: Earlbaum.
- Hatch, E. and C. Brown. 1995. *Vocabulary, Semantics, and Language Education*. Cambridge: Cambridge University Press.

- Li, L.** 1998. Dictionaries and their Users at Chinese Universities: With Special Reference to ESP Learners. McArthur, T. and I. Kernerman (Eds.). *Lexicography in Asia*: 61-79. Tel Aviv: Password Publishers.
- McArthur, T.** 1998. *Living Words: Language, Lexicography and the Knowledge Revolution*. Exeter: Exeter University Press.
- McArthur, T.** 1999. *The English Languages*. Cambridge: Cambridge University Press.
- Oakes, M.** 1998. *Statistics for Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Ooi, V.B.Y.** 1998. *Computer Corpus Lexicography*. Edinburgh: Edinburgh University Press.
- Oostdijk, N.** 1998. Language Use in a Restricted Domain. Renouf, A. (Ed.). *Explorations in Corpus Linguistics*: 170-179. Amsterdam: Atlanta.
- Rundell, M.** 1998. Recent Trends in English Pedagogical Lexicography. *International Journal of Lexicography* 11(4): 314-342.
- Svensén, B.** 1993. *Practical Lexicography — Principles and Methods of Dictionary Making*. Oxford/New York: Oxford University Press.

Working on an Historical Dictionary: The Swedish Academy Dictionary Project

Per Stille (*per.stille@svenskaakademien.se*) and Bo-A. Wendt (*bo.wendt@svenskaakademien.se*), Dictionary Staff of the Swedish Academy, Lund, Sweden

Abstract: The Swedish Academy Dictionary is one of the world's largest dictionary projects. Work on it was started in 1884 and it will be completed by 2017. The dictionary describes the written standard language of Swedish from the beginning of the sixteenth century until the present. It is therefore both historical and contemporary. In this article the background and scope of this huge project are presented with a description of what an entry contains, and how it is constructed. The problems of describing a language simultaneously from both an historical and a contemporary aspect are also dealt with.

Keywords: CHRONOLOGY, CONTEMPORARY DICTIONARY, DIACHRONY, ENTRY, HISTORICAL DICTIONARY, LANGUAGE HISTORY, LEXICOGRAPHIC DESCRIPTION, STANDARD LANGUAGE, SYNCHRONY.

Opsomming: Om aan 'n historiese woordeboek te werk: Die woordeboekprojek van die Sweedse Akademie. Die woordeboek van die Sweedse Akademie is een van die wêreld se grootste woordeboekprojekte. In 1884 is begin om daaraan te werk en dit sal teen 2017 voltooi wees. Die woordeboek beskryf die geskrewe standaardtaal van Sweeds vanaf die begin van die sestiende eeu tot die hede. Dit is gevolglik sowel histories as kontemporêr. In hierdie artikel word die agtergrond en omvang van hierdie enorme projek aangebied met 'n beskrywing van wat 'n inskrywing bevat, en hoe dit saamgestel is. Die probleme om 'n taal gelyktydig van sowel 'n historiese as 'n hedendaagse gesigspunt te beskryf, word ook behandel.

Slutelwoorde: CHRONOLOGIE, HEDENDAAGSE WOORDEBOEK, DIACHRONIE, INSKRYWING, HISTORIESE WOORDEBOEK, TAALGESKIEDENIS, LEKSIKOGRAFIESE BESKRYWING, STANDAARDTAAL, SINCHRONIE.

Introduction

To write a combined historical and contemporary dictionary covering more than five centuries is an immense task. A large amount of material has to be handled. It is inevitable that a project of such huge scope will take a considerable time to complete. In order to keep the same format and standard throughout this whole period, thorough planning is essential. In this article the project of the Dictionary of the Swedish Language published by the Swedish Academy

(SAOB) is presented with an account of how it was established, what its guidelines are, and how the content of the entries is constructed.

The Swedish language

Swedish is spoken by about nine million people in Sweden and some 300 000 in Finland. It belongs to the Germanic family of the Indo-European languages and is closely related to the other Scandinavian languages, Danish, Norwegian, Icelandic and Faroese. A thousand years ago, these languages appear to have constituted one fairly uniform language. Swedes, Norwegians and Danes are still, with some effort, able to understand each other. In the Viking Age (A.D. 700–1000) Scandinavian had some impact on English. Nouns such as *birth*, *egg*, *fellow*, *husband*, *leg*, *sister*, adjectives such as *flat*, *ill*, *weak*, verbs such as *get*, *give*, *take*, *want*, and even a pronoun such as *they* (with *them* and *their*) are borrowings from Scandinavian. The importance of the Scandinavian languages today is more modest. A rare exception is the Swedish word *ombudsman*, which has found its way into legal terminology in many languages.

The oldest stages of Swedish are known only from runic inscriptions. Swedish literature developed only in the thirteenth century. At that time Latin was the main language of written texts and Swedish only slowly made its way into new domains. This situation also inspired the first attempt to create a Latin-Swedish dictionary in order to enable people to translate from Latin into Swedish or *vice versa*, the *Latinskt-svenskt glossarium efter Cod. Ups. C 20* from the fifteenth century.

It is customary to distinguish between Swedish before 1526, Old Swedish, and after 1526, Modern Swedish. The shift which of course was not so abrupt, had to do with several factors. In medieval times, there was no complete translation of the Bible in Swedish, but the Reformation resulted in a translation of the New Testament in 1526 and of the whole Bible in 1541, which had a great impact on the written language. Equally important was the centralized bureaucracy that was built up by King Gustavus Wasa. It used a modernized language, heavily influenced by Low German.

During the period of Early Modern Swedish (1526–1732), several bilingual or multilingual dictionaries were written to facilitate the learning of other languages, such as Latin, German, English and French. It also brought the first attempt at a descriptive monolingual dictionary of Swedish, *Glossarium-Sveo-Gothicum Eller Swensk-Ordabok* (1712), by the bishop Haqvin Spegel. It was a rather small dictionary with less than 10 000 entries. The eighteenth century thus saw much linguistic work in the fields of grammar and lexicography, but the need for a great standard dictionary was yet to be filled.

The Swedish Academy Dictionary

In 1786, the Swedish Academy was inaugurated by King Gustavus III. The main objective was to work for the "purity, vigour and majesty" of the Swedish

language, which in modern language would be its clarity, expressiveness and prestige. To attain this, the Academy was to produce a dictionary and a grammar (the grammar appeared in 1999). What the king had in mind was surely a normative dictionary such as *Le dictionnaire de l'Académie française* (1694). At first the Academy did not have the staff or economic resources for such a huge undertaking, so in spite of several attempts, nothing really substantial could be achieved. Not until 1883 was a feasible plan drawn up and the work begun. Three other large dictionary projects served as model: *A New English Dictionary* (reissued as *The Oxford English Dictionary*), *Deutsches Wörterbuch* and *Woordenboek der Nederlandsche Taal*. They all are a combination of an historical and a contemporary dictionary and illustrate the definitions with plenty of authentic quotations.

The dictionary, *Ordbok över svenska språket utgiven av Svenska Akademien* (SAOB), or the Swedish Academy Dictionary of the Swedish Language, covers the period of Modern Swedish. It does not try to cover all periods of the language as *The Oxford English Dictionary* does, because there is already a dictionary of Old Swedish, *Ordbok över svenska medeltids-språket*, by K. F. Söderwall (1884–1918, suppl. 1925–73). The scope is more or less limited to the written standard language. This means that spoken language, dialects and technolects are included only to a limited extent.

The work started with the collecting of a large corpus of authentic quotations from all kinds of Swedish literature. A large part of this work was done before the editorial work began, but the collecting continued until last year. Then it was decided that the staff should discontinue the collecting of quotations in order to accelerate the work. The text of the SAOB is based on these roughly eight million quotations selected from more than 20 000 titles (Hast 1983: 160, Lundbladh 1996: 94). During compilation work the editors also consult existing dictionaries and encyclopedias. The present editors furthermore take advantage of modern computer-based corpora, mainly from newspapers and novels, to describe modern language use (cf. Stille and Wendt 2000). Since 1991, all text production within the project has been computerized.

From the start of the project the editorial staff has resided in the university town of Lund in southern Sweden. There are about twenty staff members working on the project.

The first volume of the dictionary appeared in 1898, and the latest (vol. 32, *stå an-svåpa*) in 1999. Seven, or perhaps eight, volumes still remain. The Swedish Academy has decided that the dictionary must be completed by the year 2017. It will then probably comprise nearly 640 000 lemmas, and will be one of the largest dictionaries in any language (cf. Allén et al. 1996: 8).

The content of an entry

The SAOB describes the vocabulary of written Swedish from 1521 until the present — up to the time of the editing of an article. The lexicographic description emphasizes the semantics of a word including shifts in meaning, but the

account of other aspects of word history is also detailed and precise. It is, however, much more condensed. At the beginning of an entry (the "head") various categories are dealt with in a comparatively small space, whilst the semantic description is allowed to take up much more space, not least because of the illustrative text samples. Initially the following are given: by-forms (if there are any), pronunciation, flexion, variation of word forms and spelling, and finally etymology.

That pronunciation is given in a dictionary of the written language only is something of an anomaly. It is due to the fact that at the start of the project the first editors did not systematically exclude spoken language as object of description — it only became the natural consequence of the circumstance that the empirical material only consisted of written texts, which in the course of time was turned into an explicit principle. It is the medium itself that is the conclusive criterion. There is no (further) stylistic restriction beyond the exclusion of spoken language — no matter how commonplace or colloquial an utterance in a novel may be, in the view of the SAOB it is still considered as written language, i.e. language recorded in writing.

As far as the etymology of Indo-European words, whether inherited words or loanwords from other Indo-European languages, is concerned, it is the object of the SAOB to trace them back to their original stage, although no reconstructed forms are normally given. For non-Indo-European words the SAOB is content with stating the words in the language from which they originate.

With these quite concisely described categories the initial part of the entry is complete, and the semantic description begins. If there are different sections, these are arranged etymologically — as far as the course of the semantic development is known or possible to reconstruct. The main sections are numbered 1, 2, 3 etc., and the first always deals with the sense that equals or comes closest to the etymologically original one. Here the SAOB maintains the principle that this should apply to the whole history of the word, even if it is a loanword in Swedish with the original sense recorded later than another one having entered the language at an earlier stage. In some, although rare, cases section 1 can therefore deal with a sense not recorded until recently, while the oldest sense from a purely Swedish perspective is given in a later section. This is an heritage of the comparative and neogrammarian all-Indo-European view of linguistic material. In practice, however, it very seldom has the consequences sketched above — and moreover, the semantic chronology of the word in Swedish is clear enough from the dates of the text samples. In every section the oldest attested example of the sense (within the SAOB material) is given as the first of the text samples. If the present-day usage is restricted, the most recently attested sample is also given (as the last one).

At the end of the entry, after the completed semantic description, follows an extensive list of compounds in which the word in question forms the initial component, as well as of derivatives of the word. In an article about a verb, there is also a section of what in the SAOB is called "separate combinations"

(Sw. *särskilda förbindelser*), which are roughly to be understood as verb-particle collocations, e.g. *springa före* 'run ahead', *komma undan* 'get away'. These compounds, derivatives and "separate combinations" are in a way little entries in their own right, although subordinate and within a bigger entry, but they are very summary and include (mostly) only a semantic description, although derivatives are described somewhat more extensively. Flexion and other kinds of form variation of the latter component of compounds are to be found in the entry for the simplex word. As Swedish, like most other Germanic languages, admits the formation of compounds to an almost endless degree, the number of compounds can sometimes be quite large, so that this part of the entry occupies far more space than does the account of the simplex word. Yet, the SAOB, at least in its current editorial practice, records only the well-established and more or less lexicalized compounds; the occasional compounds (which can be formed in almost indefinite numbers as the need arises) are normally not accounted for, although they may be attested in the SAOB material.

Hierarchic levels in the semantic description

As already mentioned, the main sections are numbered from 1 onwards. Sometimes, if there are many of these, they are semantically grouped under **A**, **B**, etc. If necessary, a superordinate level of **I**, **II** etc. is also introduced, often reflecting a more grammatical division, e.g. into separate parts of speech or into transitive and intransitive uses of a verb. More extensive articles are often subdivided into an intricate system of subordinate sections below the levels of **1**, **2** etc. (labelled *a*, *b* etc.), sub-subordinate sections (*α*, *β* etc.) and even sub-sub-subordinate sections (*α'*, *β'* etc.). (When there is only one subordinate section, no introductory letter is used.) A fundamentally important difference in the status of the subordinate sections is whether they form an exhaustive account of subordinate senses or whether they are mere examples of certain (e.g. especially frequent, peculiar or lexicalized) instances of a superordinate sense. In the former case the semantic description of the main section (or superordinate subordinate section) is not followed by any text samples; instead, the subordinate sections appear immediately after the superior definition.

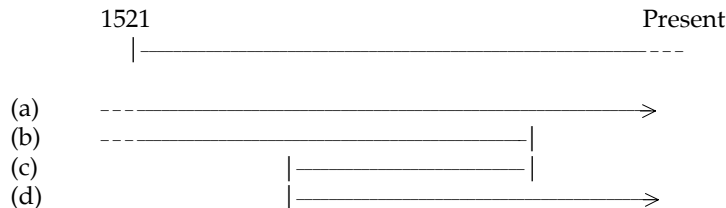
If the subordinate sections are only special instances of the superordinate sense, the definition of this sense (with its own text samples) is followed by "especially" (Sw. *särskilt*, abbr. *särsk.*). This indicates a sense which for some reason or other the editor has chosen to distinguish — it may be because it is a very frequent, perhaps even the most frequent sense, but it may equally be because it is the most unusual (and perhaps equally rare) one. The meaning of *särsk.* in the SAOB thus roughly equals "among other things". (This SAOB use of *särsk.* may be compared to those cases where *The Oxford English Dictionary* labels its first subordinate section with a *b* instead of an *a*.) Normally a superior section that ends with *särsk.* has text samples of its own, but sometimes its definition may be followed immediately by *särsk.*, whereby the editor has chosen to ignore other (attested or just theoretically imaginable) usages than those ac-

counted for in the subordinate sections. This *särsk.* might as a matter of fact be used almost anywhere in the descriptive language, not least within definitions, denoting in the same way that something has been mentioned separately without any assertion about its frequency.

Both an historical and a contemporary dictionary

Because the SAOB is simultaneously an historical and a contemporary dictionary, it always has to ascribe usage restrictions (from a modern point of view) to each of its lexicographic categories. Naturally there may be stylistic, social or regional restrictions in the use of a form or a sense. It can also be a matter of the degree of lexicalization (not least for compounds which, although attested in earlier language use, may at present not exist otherwise than as occasionally formed). Most importantly, however, the SAOB is concerned with usage — whether a form or a sense is in current use or whether it is more or less obsolete (within a three-graded scale: less used (Sw. *mindre bruklig*), seldom used (Sw. *föga bruklig*) and unused/obsolete (indicated by †). The lack of a label of this kind is then to be understood as: commonly used (at the time of publication).

That the SAOB in a way has a double mission, at the same time describing an historical development and the present-day language, has been extensively discussed by the former editor-in-chief Hans Jonsson (1983), and the following exposition is partly based on his views. Within the description span of the SAOB (1521–today) the following different extensions of a certain sense or usage (or for that matter form) are possible:



A certain sense of a word may have existed from the very beginning of the SAOB time span, that is to say, it may be attested already in the medieval period, in Old Swedish, and be current till this day. (Even for a sense of undisputable Old Swedish origin the SAOB gives the first record in its own — Modern Swedish — material, from whatever year it comes. That the word already existed in Old Swedish can only be seen in the etymological parenthesis, where only less exact specifications of the different senses are given.) In this case, as in (a) above, we therefore have a sense that is (from the point of view of the description period) original and current. A sense could likewise be older than 1521 but could have become unused during the SAOB period: original, but at present obsolete, as in (b) above. The sense may also have appeared for the first time — through spontaneous sense development or foreign impact — after

1521, whether early in the sixteenth century or quite recently, perhaps just a few years before the time of publication, and thereafter has either disappeared again or is still in use: recent and obsolete as in (c) or recent and current as in (d) above.

Synchronic and diachronic perspectives

When it is a question of distinct senses, these naturally appear in separate main sections, each with its own usage label. In principle this account gives a purely synchronic description in separate layers: a nowadays obsolete sense **1** is described in its own right, assigned to an older era of the language (with the boundaries given (or more accurately: indicated) by the oldest and the most recent text sample), then a still current sense **2** (whether original or new) is described in its own right and assigned to the prevailing state of the language (with the past boundary at (or indicated by) the oldest text sample). The senses separated into different sections are thus synchronic descriptions of one time period each, although usually partly overlapping each other. The development from sense **1** to sense **2**, the diachronic dimension, is often only implicit. Sometimes, however, it can be made explicit through information in an introductory etymological parenthesis in the latter section, e.g. "**2** [properly metaphoric use of **1**] ...". On rather rare occasions, it can be part of the semantic description proper, e.g. "**2** in metaphoric use of **1**, about ...". In general, however, the historic and the present-day dictionary are in these cases combined by merging synchronically described layers, in so far as they are not common for the objects of both dictionaries.

The combination becomes more complicated when the changes of usage concern nuances of a sense which are hardly divisible into separate sections — when an historic state which is somewhat differentiated from the present-day state is described in the same section as the latter. This opposition within one and the same section can be handled in two fundamentally different ways: either with an introductory usage label or with an indication of restricted usage later in the definition text. In both cases there is a diachronic dimension in the manner of description, but most evidently so in the latter.

In the first case an explicit usage label may be lacking (equal to full usage), and then older, somewhat extended or special uses of the sense are accounted for after the formula "earlier also". Here the synchronic point of view is still dominating: a still prevailing state is described as it is, and then a more peripheral addition is also made to cover a somewhat wider usage that existed in earlier times. The description becomes more truly diachronic when restricted use is introduced by specifying explicit exceptions of common use in certain detached subordinate sections, i.e. of the type "**1** († except in *b-d*)", then followed by the general sense which existed earlier in the language but is nowadays only found in those sense nuances, usages with lexical or syntactic restrictions or more or less lexicalized phrases that are listed in *b*, *c* and *d* (which in a case such as this are all introduced with the label "commonly

used"). These isolated present-day usages of a word are thereby anchored in an older general and once unrestrictedly productive sense. They are put into a diachronic context within which they are present-day remainders of an earlier state.

The diachronic perspective becomes even more evident when the present-day restrictions are given only further on in the section, as an integrated part of the definition text. Without any introductory usage label an older and more general semantic description is given, then followed by "nowadays only" and the restriction that applies to this sense of the word in present-day language, whether concerning sense nuances, types of referents, lexical or syntactic contexts or something else. Thus, in this case, there is an exception to the principle that sections in the SAOB without any introductory usage label implicitly indicate unrestricted present-day usage; the whole of the definition has to be read to be certain of that. In this case, the definition itself includes the sense development that has taken place: from a more general former state to a more restricted modern state: "this was the meaning of the word originally (within the SAOB period of description), but nowadays it only applies to the following cases".

Since every section has only one collection of text samples in strict chronological order, the reader, on the other hand — in the latter case — is left without any answer as to when the change(s) took place. The most recently attested sample that shows the older sense nuance or usage is not necessarily the very youngest of those in the total SAOB material. Thus, when the descriptive language of the SAOB is the most diachronic — as we maintain it is when an only gradually given restriction is explicitly introduced within the definition — the diachrony is the least unambiguous as far as text sample chronology is concerned. Instead this chronology is most unambiguously presented when the sense development is stratified into each (partial) sense described synchronically by itself. Compare e.g. section 6 with sections 7 and 8 of SIMPEL (the same word as Eng. *simple*) (in translation):

- 6) [...] as an adverb: simply [...]; nowadays only in the phrase *helt (enkelt och) simpelt*, quite simply [...]. [Text samples in chronological order, the last ones displaying the only phrase in use nowadays.] [...]
- 7) (†) straightforward, naive, unspoiled [...]. [Text samples from 1526 to 1885.]
- 8) plain, ordinary [...]. [Text samples from 1620 to 1967 (the year of publication).]

Which of the solutions sketched above is preferred in the individual case is dependent on what the empirical material actually looks like. The editor often has to accept that the most sensible and for that matter most practical and convenient solution for a certain word, does not do full justice to every aspect of the complex web of sense development.

Concluding remarks

In the SAOB — without comparison the largest project within the humanities

in Sweden (cf. Jonsson 1993) — a large number of text samples have been dealt with since the work began, and, as we hope to have shown, also a great mass of information is given in the 32 volumes published, available for those who can assimilate the often very concentrated presentation. Since its start in 1884, more than four generations of editors have been working on the project and a great deal of work still remains, but the present editors will probably see the project completed.

Postscript

The version of the dictionary on Internet can be visited at: <http://g3.spraakdata.gu.se/saob/konk/>. The modern corpora of Swedish can also be consulted on Internet at: <http://spraakbanken.gu.se/>. More information about the Swedish Academy can be found at: <http://www.svenskaakademien.se/>.

References

A. Dictionaries

- Deutsches Wörterbuch* by Jacob Grimm and Wilhelm Grimm. Leipzig 1852-1960.
Le Dictionnaire de l'Academie françoise dédié au Roy. Paris 1694.
Glossarium-Sveo-Gothicum Eller Swensk-Ordabok by Haqvin Spegel. Lund 1712.
Latinskt-svenskt glossarium efter Cod. Ups. C 20 ed. by Erik Neuman. Uppsala 1918-42.
A New English Dictionary on Historical Principles. Oxford 1883-1928.
Ordbok över svenska medeltids-språket by K.F. Söderwall. Lund 1884-1918.
Ordbok över svenska språket utgiven av Svenska Akademien (SAOB). Lund 1893-.
The Oxford English Dictionary. Oxford 1933-1986.
Woordenboek der Nederlandsche Taal. s'Gravenhage 1852-1998.

B. Other sources

- Allén, Sture, Yvonne Cederholm, Sofie Johansson Kokkinakis, Lena Rogström, Rudolf Rydstedt and Lars Svensson. 1996. *Om svar anhålles. Rapport från projektet OSA*. Göteborg: Department of Swedish, Göteborg University.
- Hast, Sture. 1983. Om SAOB:s material. *Nysvenska studier* 63: 159-192.
- Jonsson, Hans. 1983. Om historisk lexikografi typ SAOB. *Nysvenska studier* 63: 127-158
- Jonsson, Hans. 1993. Sveriges största humanistiska projekt. *Kungl. Vitterhets-, Historie- och Antikvitetsakademiens årsbok* 1993: 119-128.
- Lundbladh, Carl-Erik. 1996. Värdering av SAOB:s korpus och språkprovssamling. *LexicoNordica* 3: 91-103.
- Stille, Per and Bo-A. Wendt. 2000. Användning av Språkdatas konkordanser i det redaktionella arbetet med SAOB. Byrman, Gunilla, Hans Lindquist and Magnus Levin (Eds.). *Korpusar i forskning och undervisning. Rapport från ASLA:s höstsymposium i Växjö, 11-12 november 1999*: 266-272. Uppsala: ASLA.

Foreign Words as a Problem in Standardisation / Lexicography: English and Afrikaans Loan-words in isiXhosa*

Anja Drame, *Department of African Studies, University of Leipzig, Leipzig,
Germany (anjadrame@hotmail.com)*

Abstract: Languages are not static systems. They develop and change, add new items while others become outdated. These changes can be clearly observed in the lexicon especially. No language can afford to ignore or neglect foreign influence. Due to globalisation, especially English gains more and more influence on other (also European) languages. In developing countries, the languages of the former colonisers also still have an enormous impact on the indigenous languages. Some of these nations are slowly heading towards endogenous language policies which demands the modernisation of the technical vocabulary. This is however a costly and time-consuming process. In this regard language planners often prefer borrowing from foreign sources as a quick and therefore cheap method.

The first part of this paper deals with the discussion amongst linguists, sociolinguists and lexicographers about the extent to which foreign words should be allowed in an indigenous language. The second part looks at the example of isiXhosa, one of South Africa's eleven official languages, which is strongly influenced by foreign words, especially English and Afrikaans, and shows problems and methods of the integration of foreign words into the isiXhosa grammatical structure.

Keywords: FOREIGN WORDS, ISIXHOSA, ENGLISH, AFRIKAANS, BAHASA INDONESIA, RUSSIAN, ESTONIAN, GERMAN, LANGUAGE POLICY, LANGUAGE PURISM, LSP, MORPHOLOGY, SEMANTICS, PHONOLOGY

Zusammenfassung: Fremdwörter als Problem in der Standardisierung/ Lexikographie: Englische und afrikaanse Lehnwörter in isiXhosa. Sprachen sind keine statischen Systeme. Sie entwickeln und verändern sich, fügen neue Bestandteile hinzu, während andere veralten. Diese Vorgänge lassen sich besonders deutlich im Lexikon einer Sprache beobachten. Keine Sprache kann es sich leisten fremde Einflüsse zu ignorieren oder zurückzuweisen. Aufgrund von Globalisation gewinnt vor allem das Englisch immer mehr Einfluss auf andere (auch europäische) Sprachen. Aber auch in Entwicklungsländern beeinflussen die Sprachen der früheren Kolonisatoren die einheimischen Sprachen überaus viel. Einige dieser Nationen bewegen sich nach und nach in Richtung einer endogenen Sprachpolitik, was die Modernisierung des

* This paper was presented at the Fifth International Conference of the African Association for Lexicography, held at the University of Stellenbosch, Stellenbosch, 3-5 July 2000.

technischen Vokabulars dringend nötig macht. Dies ist jedoch ein kosten- und zeitaufwendiges Unterfangen. Daher bevorzugen Sprachplaner oftmals die relativ schnelle und somit billige Methode der Entlehnung aus fremden Sprachen.

Der erste Teil dieses Aufsatzes behandelt die Diskussion unter Linguisten, Soziolinguisten und Lexikographen, zu welchem Anteil Fremdwörter in einer einheimischen Sprache aufgenommen werden sollten. Der zweite Teil betrachtet das Beispiel des isiXhosa, eine von Südafrikas elf offiziellen Sprachen und stark von Fremdwörtern, besonders Englisch und Afrikaans, beeinflusst und zeigt Probleme und Methoden der Integration von Fremdwörtern in die grammatische Struktur des isiXhosa.

Stichwörter: FREMDWÖRTER, ISIXHOSA, ENGLISCH, AFRIKAANS, BAHASA INDONESIA, RUSSISCH, ESTONISCH, DEUTSCH, SPRACHPOLITIK, SPRACHPURISMUS, FACHSPRACHEN, MORPHOLOGIE, SEMANTIK, PHONOLOGIE

1. Foreign words as a problem in language planning

In this paper, I define a foreign word as a word whose phonic shape does not conform to the native phonemic pattern of the language. In its common sense it is "a word, taken from another language more or less unchanged", or, "a word of which you can tell by pronunciation, stress or orthography that it is not of native origin" (Wahrig 1968, translated and slightly changed by the author). Foreign words play an important role in language planning processes in developing nations.

Language planning is a deliberate, planned language change. It is a problem-solving, future-oriented activity by organisations, established for this purpose (Rubin and Jernudd 1971). It is an activity closely connected with lexicography. Dictionaries not only describe a language. Many modern dictionaries are prescriptive works. They function as reference works and therefore play an important role in the implementation of new terminology. Loan-words, i.e. foreign words which are phonologically and morphologically adapted and integrated into another language, and foreign words exist in almost every European language and have at different times caused different reactions pro and contra.

Language purism predominated German lexicography for a long time. The first German *Fremdwörterbuch* was published as early as 1571. In 1801 Campe's *Wörterbuch zur Erklärung und Verdeutschung der unserer Sprache aufgedrungenen fremden Ausdrücke* was published. Campe did not only explain foreign words in his dictionaries but also substituted them with German equivalents. Distinction was made between the origin of the words. English and Dutch words were considered as foreign but the main focus was on words from French and other non-Germanic languages (Kirkness 1976). In the new Duden Anglicisms and words from other languages form a great part of the entries.

Nationalistic purism is, however, losing ground in many countries. With

Israel as example, Fellman and Fishman (1977) warn that if a language is too purist, especially in science, it hampers the natural language change and development in language structure which implies that phoneme and morpheme inventories are to a certain degree closed systems and that the lexicon is open but very limited. It also hampers standardisation because in many instances of higher education students are already familiar with English terms.

The fact that a word is understood by several nations enhances international communication. Internationalisms make it easier to learn the other's language. But internationality cannot be the only principle to be considered when evaluating foreign words, because a word may be understood by intellectuals but not the great mass of the people (Tauli 1968).

The most common arguments against foreign words are on the one hand the already mentioned purism. On the other hand they are criticised for their lack of transparency. Laypeople do not understand them, because they are unmotivated and etymologically not clear. A word is transparent if the concept it represents is understood without further explanation, for instance: *Blinddarm-entzündung* v. *Appendicitis*.

But a compound or a derived word is only clear if it literally refers to the designated concept. The fact that people do not understand foreign words often depends on the circumstance that people do not understand the concept. Empirical tests prove this to be true. When one encounters a new word whose meaning one does not know, it matters little whether the word is a foreign or a native word. For example, the German *Hubschrauber* is not clearer than *Helicopter* if one does not know the thing it describes. Nevertheless, motivated words are easier to memorise than unmotivated words.

A very common statement is that foreign words obstruct communication between intellectuals and ordinary people. One can argue against this that it is more a question of style and audience. If precise learned terms, be they foreign or native words, can be avoided in a popular presentation, they should be. However, if precise terms are needed, for instance in a scientific text, and they are lacking in a particular language, foreign words are unavoidable. This speaks against having different terms for the different levels of special subject discourse such as jargon or scientific language, as is the case in all European languages. In German, for instance, especially in scientific language, foreign words play an important role. Therefore dictionaries for special purposes as well as terminological databases should include both jargon and scientific expressions and provide clear, detailed definitions as well as information about the context of use of these lexical items connected with examples of their typical use in a text.

Abdulaziz (1989) argues that borrowing of international terms should be mixed with coining of local terms so as to preserve the identity of the African languages. Although many terms have been coined by means of borrowing, there are also a great number of very resourceful modern native terms which sometimes go hand in hand with the European expressions.

From the point of economy, foreign words have some disadvantages. Foreign phonemics and stress patterns cause problems and are inconvenient for the speakers. This, of course, differs strongly among different language groups. Foreign phonemic elements should not be integrated into international words in native languages. Either the phonic structure should be adapted to the native language or an adequate native word should be used.

Foreign words are also often very long and cumbersome and thus more difficult to pronounce. But this is not only a problem with foreign words, sometimes a native word also is a long compound or circumscription (Tauli 1968). Swahili compounding is, according to Batibo (1992), a very transparent method of term creation which can be applied to isiXhosa as well: *umabonakude* (*bona see, kude afar*) as opposed to *iTV/itelevizhini*.

Long words and circumscriptions are contrary to the tendency towards modern, short words and are therefore uneconomical. The speakers of the discourse community may prefer an abbreviation if it is often used. But this again is contradictory to transparency: abbreviations of long words or circumscriptions lose transparency and are therefore probably harder to memorise. In this case adapted foreign words are to be preferred.

From the aesthetic point of view, there are two arguments to be considered:

1. The use of foreign words is to a certain degree a question of style, particularly in science. In scientific literature the foreign word can be preferred. If one language has two variants of an expression (a native word and a foreign word), it serves the purpose of variety.
2. Acceptance must be considered as most important. If a large part of the population does not accept a foreign word for emotional reasons, it must not be insisted that it should be used. A country which was under the rule of a colonial power for long, might feel a certain reluctance to the colonial language. Although necessary because of a lack of native terminology, it might not appreciate terms from the colonial language and replace them with either native words or words from another more acceptable language. Tauli (1968) gives as an example that Estonian in 1920 substituted the word *tānav*, originally an old Germanic loan-word, for the Russian loan-word *uulits*, street. Linguists rejected *tānav* as imprecise (it originally referred only to a village street) but the ordinary people were unaware of its origin and therefore accepted it.

In the area of the former Soviet Union, all the national languages borrowed about 70-80% of their lexicon from Russian as imposed by the Soviet government. This uniformity facilitated the preparation of educational materials and thus promoted education in those languages as well as literacy in Russian as a second language. At the same time Russian was strengthened.

To give another example: Bahasa Indonesia was promoted since Indonesia's independence in 1949. The strategies were not like those in the case of Russian in Estonia, but borrowings for the growing demands in the fields of industry, government, education, commerce, science and technology came from different sources. They were either taken from existing words or morphemes in Bahasa, or other Indonesian languages, or loan-translations from or an influx of foreign terms (English and Dutch). The controversy mainly concerned the source of the borrowings: European languages, mainly English for the fields of modern science and technology, and the languages from which borrowings traditionally come (mainly Sanskrit, the language of Indonesia's height of political power in South East Asia from the 7th to the 14th century, used especially in the fields of literature, culture, and scholarship). Although Sanskrit loan-words are used by military leaders and politicians to arouse nationalistic sentiments in support of their politics and English by other sections of the society to express modernity, it is a good example of a compromise between two poles of corpus planning: modernity and authenticity, which, according to Fishman (1983), must both be present for successful modernisation (Cooper 1989).

Language planning does not take place in a vacuum. A wide range of social factors must be considered and in the end the speakers of a certain language or discourse group alone decide if they prefer and want to use a foreign or a native word.

Not much research has been done on the acceptance of foreign words or on language inherent coinages in isiXhosa, but one can assume that they are relatively well accepted. The South African Broadcasting Corporation (SABC) extensively uses foreign loan-words for the news in the African languages. These are understood by the young urban population and cause only a few complaints by language purists. Vice versa, if the SABC were to use more language inherent versions, people would have problems to understand these. Dictionaries should therefore include those foreign expressions in the way they are used by the speakers and give information about the source language. *The Greater Dictionary of Xhosa* which is compiled at the University of Fort Hare, works according to a language descriptive rather than a prescriptive way. This implies that Anglicisms and words of Afrikaans origin are included and their origin is indicated as can be seen from the examples *i-sekile* < Afr. sekel and *i-sekisi* < Eng. circus. (Tones are not considered in this paper.)

2. Afrikaans and English words in isiXhosa

IsiXhosa is a Nguni language spoken in the Eastern Cape, especially in the areas of the former Transkei and Ciskei. IsiXhosa is not a homogeneous language, but has a multitude of dialects (e.g. Ngqika, Gcaleka, Thembu, Mfengu, Mpondo, Xesibe, and Bomvana) which are largely intercomprehensible despite variation that might occur in their lexicon. Apart from this, the written stan-

dard differs considerably from the spoken varieties. The Ngqika and Gcaleka dialects are the closest to the standard. The development of an orthography by John Bennie in 1824 was based on these two dialects.

The typical feature of isiXhosa nouns is that they also, like those in other Bantu languages, consist of a stem and a class marker. The stem contains the basic semantics of the word while the class markers represent the categorical class.

Example:	umfazi (/u- / - /-m- / - /-fazi/)	woman
	abafazi (/a- / - /-ba- / - /-fazi/)	women
	ubufazi	womanhood
	isifazi	typical of women as a group

The language is influenced by other Bantu languages (especially seSotho), English and Afrikaans. It was formerly largely influenced by Khoisan languages — the origin of the clicks — but today there is no living contact anymore. Only little is known about the influence of the Khoisan languages on the isiXhosa lexicon (Jafta 1987), but it is assumed that approximately 15% of the isiXhosa vocabulary is of Khoisan origin.

Only 22% of the language standardisation strategies in isiXhosa are language inherent strategies, the remaining 78% are loan-words or foreign words. Of the European languages, English is about 57% more important than Afrikaans (Britz 1983). Reasons for this are that education of isiXhosa speakers is mainly through English as medium of instruction and therefore English loan-words are integrated with a higher frequency. English is also seen as a prestige language, the language of the ANC. Afrikaans had and still has some pragmatic value. Other Bantu languages have their influence mainly in the big cities where people from different language groups live together and in the North-Eastern Cape where the isiXhosa area borders on Lesotho.

The usage of loan-words depends mainly on different social factors like age, geographical area, and education. For example, the loan-words *ukuprakhithiza* or *ilayibrari* are never or seldom used in rural areas where isiXhosa exists in its most natural form and bilingualism is limited to teachers or maybe nurses at hospitals.

2.1 Morphology

The isiXhosa noun structure consists of two or three morphemes (pre-class-markers), basic class-marker and the stem.

English or Afrikaans nouns which consist only of one morpheme do not cause problems since they can easily be transformed into isiXhosa by adding the class markers. If the English or Afrikaans word consists of more than one morpheme it will still be treated the same way, i.e. the whole polymorphemic foreign word forms the stem and the class-markers are added to it.

Examples: sweets > iswitisi or iiswitisi
paper > i-phepha
lekkers > ilekese or iilekese
sussie > usisi
heiden > u-m-hedeni

We can conclude that the foreign structure of the English or Afrikaans loan-words does not cause a problem for integration into isiXhosa.

The structure of an isiXhosa verb consists of three morphemes: core (basis, semantics), subject-marker and final marker (mood and other categorical functions).

Examples: hamb-a! go!
ba-fund-e of that you learn

Here again, as with nouns, there are no difficulties with verbs of only one morpheme. Afrikaans and English verbs consisting of more than one morpheme are treated like monomorpheme words.

Examples: saag > -sarh-
pass > -pas-

We can conclude that the transmorphology from English or Afrikaans into isiXhosa is more or less regular. Foreign words can be completely integrated into the morphological structure of isiXhosa.

2.2 Semantics/Transsemantics

Transsemantics is the process in which a loan-word is integrated into isiXhosa semantics. We can assume that there are at least four transsemantic processes possible:

- (a) complete integration (the concept is taken over unaltered from English or Afrikaans),

Examples: koffie [+ warm drink + made from coffee beans] > ikofu [+ warm drink + made from coffee beans]
receipt [+ written + acknowledgement that something is received] > irisithi [+ written + acknowledgement that something is received]

- (b) narrowed meaning (the concept includes more constituents in English or Afrikaans),

Examples: skool [+ educational institution + teachers and pupils + group of

artists/scientists following the same principle + large crowd + meeting] > isikolo [+ educational institution + teachers and pupils]
beer [+ European kind of beer + African kind of self-brewed beer] > ibhiya [+ European beer] (in contrast to umqombothi [+ African beer])
order [+ request (to supply something) + (proper/usual) sequence + command] > iodolo [+ request (to supply something) + sequence]

- (c) broadened meaning or semantic extension (the original meaning of the loan-word from English or Afrikaans has one or more additional meanings), or

Examples: balloon [+ form of a ball + filled with air] > ibhaloni [+ form of a ball + filled with air + flight instrument]
gas [+ element + airlike substance + heat source] > igesi [+ element + airlike substance + heat source + electricity]

- (d) semantic shift (the meaning of the loan-word from English or Afrikaans changes in the process of integration into isiXhosa, sometimes considerably, resulting in false friends).

Examples: kandelaar [+ candlestick] > ikhandlela [+ candle]
railway [+ tracks + public institution] > uloliwe [+ train]

The number of the completely integrated words is by far bigger than the number resulting from the other processes, which means that complete integration is the most common process.

2.3 Phonological assimilation

Lexical borrowing is closely connected to phonological assimilation. To a large extent, foreign sounds undergo regular changes.

Examples: /r/ > /l/
rok > ilokhwe
dinner > idinala
supper > isaphala
lorry > iloli

Especially in modern adoptive words, Zulu or isiXhosa use /r/ more and more as in the example of *dairy* > ideri.

The phoneme /r/ is becoming a regular phoneme although restricted to

lexical items from English and Afrikaans. Sometimes, even /rr/ is used, especially by modern urban-educated speakers. A certain regularity can be observed, but there are also examples of irregular assimilation of foreign phonemes:

Examples:

[θ] > [th]	theology	> ithiyoloji
[θ] > [f]	thimble	> imfimbolo
	bath	> ibafu

The syllable structure in a Bantu language word is the skeletal or canonical structure of CVCV, which means almost every word ends in a vowel, whereas English and Afrikaans words usually do not employ vowels as final phonemes. For integration into the isiXhosa phonological structure, this means that a vowel is added. A tendency can be observed to solve this problem but again there is no absolute regularity. Loan-words which end on a labial phoneme usually take a back vowel, in most cases /u/.

Examples:	lab	> ilebhu
	pump	> impompo
	lap	> ilaphu
	waskom	> ivasikomu

If the consonant is nonlabial, it will be a front vowel like /i/.

Sound clusters are uncommon in the isiXhosa grammatical structure. If there is a consonant combination in Afrikaans or English which is not allowed in or clashes with the isiXhosa phonetic system, a vowel is usually placed in between. This vowel insertion is also known as vowel epenthesis. It includes the insertion of a vowel between two consonants as well as after a consonant in a syllable final position (Batibo 1996).

Examples:	broek	> ibhulukhwe
	cricket	> ikhilikithi

Sometimes this foreign combination is kept, although it clashes the isiXhosa system.

Examples:	flag	> ifleki
	dribble	> -dribula
	drink	> idrinki

2.4 Morphological assimilation

With verbs, loan-words in their basic form take the final vowel /a/ to form the

verbal stem.

Examples: bore > -bhola
dance/dans > danisa
phone > fowuna

To this stem the normal pre- and suffixes are added according to mood, tense, aspect and other additional semantic aspects.

Examples: om te jaag > -dyarha (to chase)
jaag > dyarha
> udyarha (he chases)
> udyarhile (he has chased)

Noun loan-words undergo morphological assimilation as well. According to other Bantu language noun specific patterns, a class marker will be added.

Examples:
Class 1(a) sussie > u-sisi
boetie > u-bhuti
Class 5 esel > i-esile
band > i-bhanti
Class 9 pincers > i-mpintshisi
coffee > i-kofu

In most cases loan-words are integrated in one of these three classes. Exceptions are loan-words with an initial consonant combination beginning with /s/, such as Afrikaans *stoel*. They are usually integrated in class 7. A vowel, mostly /i/, is inserted between /s/ and the other consonant.

Examples: stoel > i-situlo
spieël > isipili
straat > isitalato

Like verbs, nouns too can be derived by pre- and suffixation.

Examples: dorp > i-dolophu (village)
> iidolophu (villages)
> idolophana (small village)

3. Conclusion

The discussion on foreign words has not been terminated yet. Their influence will increase with the growth of globalisation. In South Africa, with its tenden-

cies towards English monolingualism, there will be an even greater increase of Anglicisms in the African languages and probably in Afrikaans as well. Will it be useful to develop and standardise native terminology and words people are not likely to use? Or will language awareness increase if more projects such as tertiary education in the native languages are initiated? And which role will lexicography play in the implementation of modernised terminology?

The aim of languages, especially languages for special purposes, is to function as a vehicle for communication between people and enable the exchange of ideas and information among speakers of a discourse group. If this goal is reached by the use of foreign words, why should a purist language policy be forced? The German of today would not have existed if it had not been for foreign influences, as can be illustrated by the following sentence: *Am vergangenen Freitag nahm der Grossvater des Herzogs, mit Rücksicht auf die Beschwerden der Untertanen, an einer Sitzung in der Hauptstadt teil.* In this sentence only the articles and prepositions are German, the rest are assimilated loan-translations and loan-words, which are found in every dictionary, not specially marked as foreign, but treated like native German words. It is important to be aware of one's own language and to find a middle way between the original language and foreign influences.

Bibliography

- Abdulaziz, M.H.** 1989. Development of Scientific and Technical Terminology with Special Reference to African Languages. *Kiswahili* 56: 32-49.
- Batibo, Herman M.** 1992. Term Development in Tanzania. Stevenson, I.A. (Ed.). 1995: 27-38.
- Batibo, Herman M.** 1996. Loanword Clusters Nativization Rules in Tswana and Swahili: A Comparative Study. *South African Journal of African Languages* 16(2): 33-41.
- Britz, R.M.J.** 1983. *Vreemdetaalbeïnvloeding op die leksikon van isiXhosa*. Unpublished D.Litt. thesis. Port Elizabeth: University of Port Elizabeth.
- Cooper, Robert L.** 1989. *Language Planning and Social Change*. Cambridge: University Press.
- Fellmann, J. and J.A. Fishman.** 1977. Language Planning in Israel: Solving Terminological Problems. Rubin, J. et al. (Eds.). 1977: 79-95.
- Fishman, J.A.** 1983. Modeling Rationales in Corpus Planning: Modernity and Tradition in Images of a Good Corpus. Cobarrubias, Juan and J.A. Fishman (Eds.). *Progress in Language Planning: International Perspectives*: 107-118. The Hague: Mouton.
- Jafta, N.** 1987. The Development of Terminology in Xhosa — A Case Study. *Logos* 7(2): 127-138.
- Kirkness, A.** 1976. Zur Lexikologie und Lexikographie des Fremdworts. Moser, H. 1976. *Probleme der Lexikologie und Lexikographie*. Düsseldorf: Schwann.
- Rubin, J. and Jernudd, B.H.** (Eds.). 1971. *Can Language be Planned?* Hawaii: University Press.
- Rubin, J. et al.** (Eds.). 1977. *Language Planning Processes*. The Hague: Mouton.
- Stevenson, I.A.** (Ed.). 1995. *Sociolinguistics for Applied Linguistics*. Reader for HSOPER-U. Pretoria: UNISA.
- Tauli, Valter.** 1968. *Introduction to a Theory of Language Planning*. Uppsala: Almqvist and Wiksell.
- Wahrig, Gerhard.** 1968. *Deutsches Wörterbuch*. Gütersloh: Brockhaus.

Problems of Equivalence in Shona-English Bilingual Dictionaries

Nomalanga Mpfu, *African Languages Research Institute, University of Zimbabwe, Harare, Zimbabwe (nomalanm@yahoo.com)*

Abstract: In compiling bilingual dictionaries, lexicographers are mostly concerned with semantic equivalence. As a result, the practice of bilingual dictionary compilers is usually that of giving one-word equivalents. However, this equivalence is at times difficult to arrive at because of the disparities and incommensurability between languages and cultures. According to Sapir (1921) and Whorf (1956), one cannot expect an exact match between two languages that express different cultural realities. Given this scenario, compilers of bilingual dictionaries end up bridging the gap between languages by giving translational equivalents rather than relying solely on one-word equivalents. The Shona-English bilingual dictionaries, namely Hannan (1974) and Dale (1981), like other bilingual dictionaries, also display this characteristic. This article will discuss the problems of translation equivalents in Shona-English dictionaries where lexicographers will be dealing with divergent languages and cultures, traditional practices of lexicography and the absence of reliable corpora.

Keywords: BILINGUAL LEXICOGRAPHY, CORPUS, CULTURE, CULTURE-BOUND, CULTURAL GAP, DICTIONARY, EQUIVALENCE, INCOMMENSURABILITY, LEXEME, LEXICOGRAPHY, SHONA, SOURCE LANGUAGE, TARGET LANGUAGE, TRANSLATION, UNTRANSLATABILITY

Opsomming: Ekwivalensieprobleme in Shona-Engelse tweetalige woordeboeke. By die opstel van tweetalige woordeboeke, is leksikograwe meesal bedag op semantiese ekwivalensie. Gevolglik is die gebruik by opstellers van tweetalige woordeboeke gewoonlik om enkelwoordekwivalente te gee. Soms is hierdie ekwivalensie egter moeilik om te bereik vanweë die ongelykheid en nieooreenstemming tussen tale en kulture. Volgens Sapir (1921) en Whorf (1956) kan 'n mens nie 'n presiese ooreenstemming verwag tussen twee tale wat verskillende kulturele werklikhede uitdruk nie. Gegee hierdie scenario, oorbrug die opstellers van tweetalige woordeboeke die gaping tussen tale uiteindelik deur vertalende ekwivalente aan te gee eerder as om uitsluitlik op enkelwoordekwivalente staat te maak. Die Shona-Engelse tweetalige woordeboeke, naamlik Hannan (1974) en Dale (1981), soos ander tweetalige woordeboeke, vertoon ook hierdie kenmerk. Hierdie artikel bespreek die probleme van vertalende ekwivalente in Shona-Engelse woordeboeke waar leksikograwe te doen het met uiteenlopende tale en kulture, tradisionele leksikografiese gebruike en die afwesigheid van betroubare korpusse.

Sleutelwoorde: BRONTAAL, EKWIVALENSIE, KORPUS, KULTURELE GAPING, KULTUUR, KULTUURGEBONDE, LEKSEEM, LEKSIKOGRAFIE, NIEOOREENSTEMMING, ONVERTAALBAARHEID, SJONA, TEIKENTAAL, TWEETALIGE LEKSIKOGRAFIE, VERTALING, WOORDEBOEK

1. Introduction

The analysis of meaning is the basic activity in lexicography. The meaning of a lemma or expression is a concept related to the thing or idea that it refers to or that it represents which can be described with other words. In monolingual dictionaries, only one language is used for both the lemma and the glosses. On the other hand, in bilingual dictionaries, the meaning of a lexical item is derived through the process of translation of the source language (SL) word into that of the target language (TL). A bilingual dictionary, then, is one where two languages are used, one for the lemma and the other for the glosses. In the case of the Shona-English bilingual dictionaries, the lemmas are in Shona and the glosses in English.

According to Bassnett-McGuire (1988: 2), a minimally acceptable translation is one where the surface meanings of the SL and TL are approximately similar and also one that preserves the structures of the SL. In other words, the ideal in translation is to achieve structural and semantic equivalence. Absolute equivalence between any two languages, however, is rare. The reason for this is that one is dealing with languages with different structures and sociocultural settings. The semantic problems involved in bilingual dictionaries are more complicated than those in monolingual ones because the latter are written for people who participate in and understand the culture being described, whereas the former describe a culture which differs from that of the users (Al-Kasimi 1983: 58).

The major task of a bilingual lexicographer is to find appropriate equivalents (Al-Kasimi 1983: 58) but in some cases, the entries take the form of an explanatory or a descriptive equivalent. Such an equivalent is one which cannot be inserted into a sentence in the target language (Al-Kasimi 1983: 60). Explanatory or descriptive equivalents are necessitated by the absence of a translational equivalent in the target language. A translational equivalent is a lexical unit, which can be immediately inserted into a sentence in the target language (Al-Kasimi 1983: 60).

Much has been written on the problems of bilingual lexicography in other languages. Al-Kasimi (1983) discusses the problems of bilingual dictionaries from a linguistic point of view and with particular focus on English dictionaries. Landau (1984) and Zgusta (1971) concentrate on methods of dictionary making and general problems encountered in the process of compiling monolingual and bilingual dictionaries in English. Otani (1996) looks at the problems of equivalence in English-Japanese dictionaries with particular focus on content words, while Neubert (1990) writes on what one can expect a bilingual dictionary to achieve and for what, in fact, it should be consulted. His examples are drawn from German-English bilingual dictionaries. The problems these writers discuss, do not focus on the problems specific to Shona-English bilingual dictionaries, and this is the aim of this article.

The two best known bilingual Shona-English dictionaries are by Hannan, published in 1959 with a revised and expanded edition in 1974, and by Dale,

published in 1981. Hannan and Dale were missionaries who compiled their dictionaries for evangelical purposes. Their choice of headwords seem to be of simple Shona words. Hannan's style of defining consists in giving equivalents for the main entries. Phrasal entries, such as idioms, are given as run-ons and are defined in the form of citations. Dale, on the other hand, first explains the main entries in Shona and then gives suitable explanations in English.

In Shona-English bilingual lexicography, there are instances where it is difficult to give exact equivalents, and this is where the lexicographer is left with little choice but to use explanations. The lexicographer, in trying to make his dictionary useful, will then resort to giving explanations, perhaps in the form of a paraphrase (Smit 1990: 30). These instances are: (1) lexical items with several designative functions, and (2) culture-bound words. The rest of this article will discuss these two topics.

2. Problems of lexical items with several designative functions

A word in a language may have more than one lexical function. A lexical function is the naming of the basic concept to which a word refers. According to Zgusta (1971: 27), a word can have several functions, one of which is the designative function. In Shona, certain lexical items have several designative functions which are not obvious to the dictionary user, and for which equivalents may be absent in the target language. A word's designative function is the thing or idea that it represents in the extralinguistic world. The other designative functions that a word may carry are sense extensions, which are usually figurative. A lexicographer, with the help of a comprehensive corpus, should be able to capture such senses so that the user is given the full range of meanings of any lexical item and their frequency of occurrence in the corpus.

The one problem of the existing Shona-English bilingual dictionaries seems to be the absence of a reliable corpus as well as the choice of literature the compilers thought relevant. Both Hannan and Dale used a corpus in the form of cards on which the words and the contexts were written. They extensively made use of biblical literature and Bible translations of which the language is now archaic. The corpus is therefore unbalanced and unreliable, and inaccurate in many respects, in comparison with an electronically processed corpus which is comprehensively and representatively built and in which frequency counts can be carried out. That is why, in Hannan for example, one finds unusual senses and word forms, such as *hafubhaki* (half-back), *angere* (angel) (instead of the familiar form *ngirozi*) and *endekesi* (a volume of the Bible, i.e. the Old or New Testament printed as one volume). Such uncommon forms would be met only once in an electronically processed corpus, and they would have been recognized as nonrepresentative of current usage, and might have been entered for historical interest only. The other problem has been that of reliance on traditional practices of lexicography. The relevant traditional prac-

tices of dictionary making are: (1) giving no encyclopaedic information, and (2) focussing on structural and semantic equivalence.

The following examples refer to lexical items where the secondary designative functions are specific to Shona. The term *-rara* (sleep) can be used with sense extensions which refer to the outstanding characteristics of a slow or stupid person. Thus, a statement like *Tendai akarara* can mean that "Tendai is asleep", "Tendai is passive" or "Tendai is behaving foolishly". Both Hannan and Dale give the meaning of "being asleep". Because of the absence of a reliable source for all possible word senses for particular entries, neither of them gives the extended sense of the lemma, as is the case in *Duramazwi ReChiShona* (DRC) (Chimhundu (Ed.) 1996), a monolingual Shona dictionary compiled by mother-tongue Shona speakers, who, however, used a corpus of about a million running words. In this dictionary, a second sense of *-rara* is given as *Kana munhu akarara, anenge akapusa* (If a person is asleep, it means that he/she is passive/foolish). According to DRC's definition, the term *-pusa* can also have several sense extensions other than that of being passive or foolish. It can also mean that the person is slow or stupid or that he/she is thickheaded and unable to think clearly, and further that something is easy or even light in weight. Dale just gives the sense of being passive or foolish, while Hannan gives the meaning of a thing or task being easy.

The word *mai* refers to one's own mother, but it is also used to address one's mother's sisters and one's uncles' (father's brothers') wives, and furthermore as a respectful term of address for any female who is assumed to be married. Dale gives the senses of "mother" and *maiguru* (mother's elder sister) and *mainini* (mother's younger sister), but does not give the sense of respect. Hannan gives the sense of "mother" and that of the title "Mrs". Both are correct, but their coverage is deficient in that Dale mentions the senses of *maiguru* and *mainini* but not that of "Mrs", while Hannan does not give the senses of *maiguru* and *mainini*.

The word *mukadzi* refers to a woman but it can also be used figuratively to refer to men in a negative way. Thus, if one says *John mukadzi*, it means that "John is a woman", but for John, being a man, such a statement would mean that John is either a coward or a good-for-nothing. The word *mukadzi* can also be used in a positive way to refer to a woman who is regarded as being good in everything she does. Thus, if one says *Chenai mukadzi*, one is saying that "Chenai is a woman", but what is actually meant is that Chenai is a good woman on whom one can always rely.

Similarly, the word *murume* refers to a man, but it is also used to refer to an economically independent woman who has managed to move into the domain usually occupied by men. The word can also be used to refer to a man who perseveres through hard times. If someone says *John murume*, it literally means that John is a man, but it can also mean that he is a man worth his salt. None of these sense extensions is incorporated in Hannan and Dale. The senses

of *mukadzi* they include, are those of "woman" and "wife". For *murume*, they have the senses "man" and "husband".

Hannan's and Dale's omissions of some senses can be attributed to the lack of comprehensively and representatively built corpora for Shona and English. Reliable corpora would have given them examples of sentences where the extended senses of the words come out clearly. They may have relied solely on the intuitions of their mother-tongue Shona informants, but, even for mother-tongue speakers, some senses only become distinctive when the different occurrences of a word in a corpus are studied. That is why, even for monolingual dictionaries, mother-tongue speakers rely heavily on corpora. For bilingual dictionaries, reliably built corpora for both languages are needed.

3. Culture-bound words

Language is at the core of culture and no culture can exist which does not have at its centre a natural language. A language thus reflects a specific culture, and languages are different because they reflect different cultures. As such, Shona has a wide range of terms that refer to traditional practices and artefacts and to cultural norms and rites. Because of this, there is a translation problem between Shona and English due to the cultural gap between the speakers of the source and target language. If the speakers of a certain language refer to a particular concept which does not exist in another language, it is almost impossible to translate it with a word or short phrase from that language. This problem where functionally relevant features in a source language do not exist in the target language (Smit 1990) can be referred to as cultural untranslatability. Culture, in this discussion, will be taken to mean whatever a person must know in order to function in a particular society (Wardhaugh 1998: 215). This knowledge is acquired socially. Two well-known linguists, Edward Sapir (1921) and Benjamin Lee Whorf (1956), wrote extensively on relationships existing between language and culture. Their findings came to be referred to as the Sapir-Whorf hypothesis which postulates that language and culture are inextricably related, so that one cannot understand or appreciate one without a knowledge of the other. According to this hypothesis, language cultures proceeding from different bases are fundamentally different from and incommensurable with each other (Hewson and Martin 1991: 25). This problem of the difference and incommensurability between two languages and cultures can be seen in any bilingual dictionary, particularly when it comes to terms that deal with culture.

The following illustrative examples will show some of the areas where the phenomenon of culture-bound words is frequent in Shona.

(a) Traditional ceremonies and games

The Shona have a number of traditional ceremonies to mark different events, mostly rites of passage. One such ceremony is that of *kurova guva* (the bringing

home ceremony). This is a ceremony for bringing home the spirit of a person one year after his/her death. There is no equivalent term in English because such a ceremony does not exist in English culture. Hannan describes the ceremony as "the ceremony for uniting a *mudzimu* (ancestor) with his fellow spirit elders". Dale, on the other hand, defines it as "settle and incorporate spirit of dead person into the spirit world". Here, both Hannan and Dale concur that the ceremony has to do with a dead person's spirit. The point of deviation between them is that Dale gives the explanation in Shona first and then translates it into English:

-rova guva *t. -gadzira t. Kuchenura mweya womunhu akafa (nenhumbi dzawo) kuti ave mudzimu. vb settle and incorporate spirit of dead person (at kurova guva ceremony) into the spirit world.*

Hannan's explanations of idioms such as *-rova guva* are usually run-ons under a main verb, in this particular example under *-rova* (strike, beat). They take the form of citations:

Hatiiti zvokurova guva kana mushakabvu aiva tsvimborume: we do not carry out the ceremony for uniting a mudzimu (ancestral spirit) with his spirit elders when the deceased was unmarried.

This citation does not explain what *kurova guva* is beyond observing that it is a "ceremony". All the user gets is a statement that makes no attempt to define the concept. What Hannan manages to bring out is the point that the ceremony is only held if the deceased was married. But this piece of information can only be useful to someone who knows what the ceremony of *kurova guva* involves. The ceremony is one that English speakers do not practice and that non-Shona speakers do not fully understand. The concept is "untranslatable" in the sense that it can be explained to an English speaker, but not in a word or phrase. It can only be done by a long definition that contains encyclopaedic information. Let us look at a possible definition when we do not have to rely on equivalents and are allowed to give encyclopaedic information:

-rova guva ceremony held after a year of a married person's death to bring home his/her spirit so that he/she may guide his/her family. (Bringing home ceremony). SEE ALSO *mudzimu* 3, *-gara nhaka, sarapavana* 1a/9.

One reason perhaps why bilingual lexicographers have not been free to give definitions that contain encyclopaedic information, is space constraints. According to Landau (1984: 137), the need to save space in dictionaries naturally leads to an injunction for brevity. But, at times, the need to save space produces definitions that fail to bring out the meaning of a word. Good bilingual dictionaries cannot always give brief definitions and still be useful. Cultural words can best be handled by giving encyclopaedic information.

Like ceremonies, traditional games are often also culture-bound. An example is the game of *nhodo*. This is a game that girls usually play with pebbles, whereby a player will toss one into the air and before she catches it, she should have managed to pick up another one out of the hole. Both Hannan and Dale give the definition as "game similar to 'Jacks'". The explanation serves to inform the user that it is a game close to the one he/she could be familiar with, but it is not an exact equivalent of that game. This is one of the words where a little encyclopaedic information would be useful. It would help users more if bilingual lexicographers explain the concept fully, in this case enough of how the game is played so that the user gets an idea of it. Only then will he/she be able to link it to the one he/she knows better.

(b) Customs

The Shona people observe many customs. Most of these customs pertain to rites of passage. One example is that of *roora* (lobola). *Roora* is the money and/or cattle that a man pays to his in-laws as a way of seeking permission for him to live with their daughter as his wife. It can be loosely translated into English as "bride price". There is an extensive vocabulary in Shona for terms pertaining to this custom. There are no exact equivalents for most of these terms in English because the English have a different system of marriage.

One such term is *pwanyaruzhowa*. It is a nominal construction made up of the verb *pwanya* (break) and the noun *ruzhowa* (hedge). Literally it means that the man who will have come to marry one of the daughters, has actually broken the protective hedge around the homestead in order to gain entrance to her. The term refers to the money or a beast that the man has to pay for disturbing the peace of the homestead by wanting to take one of the daughters. Hannan defines it as "that one of the beasts of the *danga* (to do with cattle) part of *roora* that is not returnable in the case of divorce". While this piece of information may be important, what Hannan does not bring out is the purpose behind the paying of the beast in Shona culture.

The term *matsvakirai kuno* is another term that has no equivalent in English. It refers to the money that a man sends via his *munyai* (go-between) to the parents of the girl he has made to elope as a way of informing them that he is the one who has their daughter. Hannan defines it as "one of the impositions of the father-in-law". Hannan's definition lacks specificity and cultural content. In the *roora* custom, the father-in-law makes a lot of impositions that are regulated by the culture within which the custom is enshrined. With more information about these customs, Hannan would have been able to give more detail in his definitions.

While it may be acceptable for bilingual dictionaries to give explanations where there are no equivalents, these explanations have to be clear and helpful to the target users. The explanations given for the terms *pwanyaruzhowa* and *matsvakirai kuno* are not clear in existing bilingual Shona-English dictionaries

and will not help either the Shona user who wants such terms explained well in English or the English user who wants to know the meanings of these Shona terms. When Shona speakers consult either Hannan or Dale, they find that their language and culture have not been properly described. Let us look again and see what a bilingual dictionary can be capable of achieving were it permitted to give encyclopaedic information for *pwanyaruzhowa*:

pwanyaruzhowa customary money or beast that is not returnable in the event of divorce, paid by the husband-to-be to his in-laws for disturbing the peace of the homestead by wanting to take one of their daughters.

(c) **Musical instruments**

What has been discussed of customs is also relevant for musical instruments. Some musical instruments are culture-bound. An instrument that is found among the Shona may not necessarily be found in other cultures and likewise, an instrument found in the English culture, may not be found among the Shona. The *mbira* is one instrument that is peculiar to Shona music. It is an instrument with many keys fixed to a board which is then fitted to a gourd that has been scraped so that it resonates well, and it is played with the fingers. The instrument has no English equivalent. Hannan describes it as a "small many-keyed musical instrument", while Dale gives the same definition, adding "Shona piano" as further explanation. In both these cases, the descriptions are imprecise because they focus on equivalence. As a consequence of culture-boundness, this equivalence is not possible.

The *marimba* is another instrument played by the Shona. This instrument is made of oblong blocks that are fitted onto a frame that has resonators such as gourds of different sizes attached underneath. The player beats the blocks with a padded stick and, depending on the size of the instrument, he/she can either sit or stand when playing it. Hannan describes it as a "many-keyed musical instrument with a resonator", while Dale just says that it is a "musical instrument". The *hwamanda* is defined by Hannan as a "horn (musical)", and by Dale as a "horn used as kind of bugle". It is, in fact, a horn, usually of a kudu, that is blown to accompany drums for dances.

Dale's and Hannan's explanations are correct, but they are not precise. Simply to say that an instrument is many-keyed presupposes that the user knows something about the instrument's make-up, shape, size and composition. Where there are no equivalents in the target language, as is the case here, then the explanations should be clear enough to create a precise image in the mind of the user. A suggested sample definition for *hwamanda* would be:

hwamanda musical instrument made out of animal horn, e.g. kudu, that is scraped and made hollow inside, and blown by the player to accompany drums for dances.

All the examples discussed have no one-word equivalents in English, therefore the dictionary should describe or explain them.

4. Conclusion

The shortcomings of earlier Shona-English bilingual dictionaries are that glosses were made as brief and as short as possible. These bilingual dictionaries relied on the traditional norms of bilingual lexicography, that is, of giving equivalents and not adequate explanations or encyclopaedic information. However, more is needed, especially now that corpora have been developed. The use of a corpus would have yielded a better treatment of words. Pictures or sketches could also have been employed, as in Dale, to show the referents of some words in order to make the concepts clearer. Traditional lexicographic practices should be violated to produce a better product. As authoritative sources of reference and research, the bilingual dictionaries cited in this article fall short of their expectations because of the way they handle a variety of lexemes.

When users consult dictionaries, they will in most cases be looking up meanings. Examples such as those cited in this article, immediately show that the Shona-English bilingual dictionaries give inadequate explanations because of the lack of detail that characterizes these dictionaries. The users will thus fail to get the correct meaning of words and concepts found in a language, and, in turn, they will get limited information about that language from the available dictionaries.

These problems can be overcome by changing the general practice in bilingual Shona-English dictionaries of simply giving lexical equivalents, and, at least for some lexemes, shifting more in the direction of giving phrases or complete explanations with encyclopaedic information. There is also a need for electronically processed corpora to reveal extended senses so that, where necessary, the entry shows a wide variety of uses.

Hannan and Dale relied on what they thought were informed experts and a few books for headword selection and definitions. The absence of an electronically processed corpus can account for the lack of specificity in their definitions. With some of the cultural terms, a corpus would have been useful and they could have used it for adding encyclopaedic information as well as words for cross-referencing some of their entries.

References

- Al-Kasimi, A.M. 1983. *Linguistics and Bilingual Dictionaries*. Leiden: E.J. Brill.
Bassnett-Mcguire, S. 1988. *Translation Studies*. London: Routledge.
Chimhundu, H. (Ed.). 1996. *Duramazwi ReChiShona*. Harare: The College Press.
Dale, D. 1981. *Duramazwi: A Basic Shona-English Dictionary*. Gweru: Mambo Press.

- Hannan, M.** 1959. *Standard Shona Dictionary*. London: Macmillan
- Hannan, M.** 1974. *Standard Shona Dictionary*. Harare: The College Press.
- Hewson, L. and J. Martin.** 1991. *Redefining Translation: The Variational Approach*. London: Routledge.
- Landau, S.I.** 1984. *Dictionaries: The Art and Craft of Lexicography*. Cambridge: Cambridge University Press.
- Neubert, A.** 1990. Fact and Fiction in the Bilingual Dictionary. *Euralex '90 Proceedings*: 29-42. Barcelona: Biblograf.
- Newmark, P.** 1988. *A Textbook of Translation*. New York: Prentice Hall.
- Otani, Y.** 1996. Cross Language Equivalence: Between Lexical and Translation Equivalents in the Case of English-Japanese Dictionaries. Gellerstam, Martin et al. (Eds.). *Euralex '96 Proceedings*: 609-614. Göteborg: Göteborg University.
- Sapir, E.** 1921. *Language: An Introduction to the Study of Speech*. New York: Harcourt, Brace and Co.
- Singleton, D.** 2000. *Language and the Lexicon*. London: Arnold.
- Smit, M.** 1990. *Translating Culture-bound Words: An Analysis in Terms of Semantic Voids*. Unpublished M.A. Thesis. Stellenbosch: University of Stellenbosch.
- Wardhaugh, R.** 1998. *An Introduction to Sociolinguistics*. Oxford: Basil Blackwell.
- Whorf, B.L.** 1956. *Language, Thought and Reality*. Edited with an introduction by J.B. Carroll. Cambridge, Mass.: Massachusetts Institute of Technology / New York: John Wiley and Sons.
- Zgusta, L.** 1971. *Manual of Lexicography*. The Hague: Mouton.

"Nonmorphological Derivations" and the Four Main English Learner's Dictionaries

Geart van der Meer, *Department of English, University of Groningen,
The Netherlands (vdmeer@let.rug.nl)*

Abstract: This article addresses the problem of "nonmorphological derivations" in English and its consequences for pedagogical lexicography, i.e. learner's dictionaries. The term "nonmorphological derivation" refers to cases such as *sun — solar, moon — lunar, cat — feline*, and in general to all cases where instead of, or in addition to, native derivational processes (*sunny, moony, catty*) English frequently uses borrowings especially from Romanic sources (Latin/French) and from Greek. The paper suggests that for such cases learner's dictionaries should give more active guidance and support, in view of the fact that many foreign learners will be from linguistic backgrounds where different, more native morphological processes are more common. An appendix lists the major examples of such derivatives.

Keywords: LEARNER'S DICTIONARIES, PEDAGOGICAL LEXICOGRAPHY, DICTIONARY DESIGN, FOREIGN LANGUAGE TEACHING, MORPHOLOGY, MORPHOLOGICAL OPACITY AND TRANSPARENCY, STRUCTURE AND CHARACTER OF THE LEXICON.

Opsomming: "Niemorfologiese afleidings" en die vier vernaamste Engelse aanleerderswoordeboeke. Hierdie artikel bespreek die probleem van "niemorfologiese afleidings" in Engels en die uitwerking daarvan op die pedagogiese leksikografie, d.w.s. aanleerderswoordeboeke. Die term "niemorfologiese afleiding" verwys na gevalle soos *sun — solar, moon — lunar, cat — feline*, en oor die algemeen na alle gevalle waar in plaas van of ter aanvulling van, natuurlike afleidingsprosesse (*sunny, moony, catty*) Engels dikwels ontleenings gebruik veral aan Romaanse bronne (Latyn/Frans) en aan Grieks. Die artikel doen aan die hand dat aanleerderswoordeboeke in sulke gevalle meer aktiewe leiding en bystand behoort te gee, gesien die feit dat baie van die vreemde aanleerders van linguïstiese agtergronde sal wees waar ander, meer inheemse morfologiese prosesse meer algemeen is. 'n Bylae lys die belangrikste voorbeelde van sulke afleidings.

Sleutelwoorde: AANLEERDERSWOORDEBOEKE, PEDAGOGIESE LEKSIKOGRAFIE, WOORDEBOEKONTWERP, VREEMDETAALONDERWYS, MORFOLOGIE, MORFOLOGIESE ONDEURSIGTIGHEID EN DEURSIGTIGHEID, STRUKTUUR EN KARAKTER VAN DIE WOORDEBOEK.

1. Introduction¹

English, as is well-known, unites strands of vocabulary from many sources, due to the chequered political, social, cultural and therefore linguistic history

of the English-speaking peoples. One of the specific consequences of this history is the remarkably rich, versatile and flexible vocabulary of English, which often gives the user a choice of several words with subtle and useful differences in meaning. One of the *disadvantages* of this development is the frequently opaque character of the English vocabulary, which not only presents problems to the learner but occasionally to the native speaker as well.²

The mixed origins of English have led to a situation where words with obvious semantic links are frequently *formally* totally unrelated in cases where other languages such as Dutch and German have both a semantic and a formal link by means of morphological derivation and composition. Take for example the adjective to the Dutch noun *koning* "king", which is *koninklijk* "royal, regal". This adjective is semantically clearly motivated and comparatively transparent, because it is morphologically (so *formally*) largely nonopaque,³ whereas the English near-equivalents *royal* and *regal* are not, though the *-al* suffix to some extent does at least suggest that the words are adjectives. The fact that English, in addition, also has the transparently derived adjective *kingly* demonstrates on the one hand the richness of the English vocabulary, and on the other hand the choices with which the speakers of English, and especially the nonnative ones, are confronted. For though the three words are semantically related, they are certainly not synonymous.

The mainly French, Latin and Greek foreign heritage of the English language has made huge zones of its vocabulary (relatively) unmotivated, or one might also say, morphologically "opaque", i.e. morphologically and hence semantically unanalysable. As a further example, the words *opaque* and *transparent* themselves might be considered. The "derived" nouns to these adjectives are *opacity* and *transparency* respectively, somewhat less than productive formations. The adjectives themselves are totally unmotivated as well, apart maybe from the fact that the *-ent* of *transparent* may be intuited as an adjectival ending. Perhaps the "prefix" *trans-* also adds some modest amount of semantic motivation, or at least some suggestion of morphological complexity.

The Dutch equivalents are *doorzichtig* and *ondoorzichtig* respectively, meaning roughly "through-see-able" and "un-through-see-able".⁴ The derived nouns are *doorzichtigheid* and *ondoorzichtigheid*, "through-see-able-ness" and "un-through-see-able-ness", formed by means of adding the nominalising suffix *-heid*. German, of which the vocabulary in this respect is often even more systematic and hence morphologically transparent than that of Dutch, offers the example of the four words *Jahr*, *Jahrzehnt*, *Jahrhundert* and *Jahrtausend*, for which English has *year*, *decade*, *century* and *millennium*, which concisely sum up the origins of English (from Germanic, Greek, and Latin (twice) respectively).⁵

Though speakers of English may consider such words ponderous, this is certainly not the view of speakers of other languages such as Dutch and German, who find it strange that the adjective to *sun* is *solar*, as in *solar eclipse*, which to them is simply "sun-darkening" or "sun-darkness", cf. Dutch *zonsverduistering* and German *Sonnenfinsternis* respectively.⁶ The words *sun* and *solar* still have the initial consonant in common, but the situation is more difficult in

most other cases, for example in *filial devotion*, where there is phonetically no resemblance between *child/son/daughter* and *filial*. Though in such cases learner's dictionaries often offer, by means of keywords in their sense definitions, the direction from adjective to noun (*solar* > *sun*), the opposite normally does not hold. In this article, I intend to examine to what extent the users of such dictionaries can, or cannot, find their way to the nonmorphologically derived words. By "nonmorphologically derived", I somewhat loosely mean cases such as (*sun* >) *solar*, (*dog* >) *canine*, etc. and generally similar cases, where — to put it informally — the semantic description of the "derivative" includes, or refers to, the semantic description of the morphologically absent "base", without there being any formal morphological link, as there clearly is in *sun* > *sunny*. After examining the four main learner's dictionaries, I will formulate a proposal as to how the prevalent situation can be improved and remedied.

2. "Nonmorphological derivations" and the learner's dictionaries

The four best-known learner's dictionaries, the "big four" (OALD, CC, CIDE, LDOCE), are intended for both encoding and decoding purposes, though it seems that recently the latter has been gaining ground at the expense of the former (Cowie 1999: 176). In my opinion, however, one of the tasks of such dictionaries is to make learners (more) aware of the major links between vocabulary items. In this section, I intend to examine how the four dictionaries manage, or do not manage, to link "base" and "derivative" in cases such as *sun* > *solar*. In other words, how can learners speaking Dutch, German or any other language that far less often than English uses Latin and Greek "derivations", find their way from *sun* to *solar*, *moon* to *lunar*, etc. without using a bilingual dictionary? To this end I have checked the entries for a number of test cases drawn from a larger collection listed in the appendix. This list of test cases is mainly based on Van der Meer (2000) with some recent additions.

The following conclusions can be drawn. In most cases it is not too difficult to discover to which English "base" word a Latin or Greek "derived" form belongs, since the sense definitions usually mention the "base", e.g. *feline* (in all four dictionaries) mentions *cat*.⁸ It is therefore possible to find one's way back to the "base". However, the compilers of these dictionaries seem unaware that many foreign learners speak languages which show semantic links much more clearly by means of their morphology. Such learners naturally tend to expect the same in English. The consequence of this is that they find it impossible to trace adjectives such as *solar* and *lunar* when starting from the "bases" *sun* and *moon*. Hence any user wishing to form words or word combinations involving *moon* is left guessing and has to consult a bilingual dictionary.

Occasionally, however, certain links are presented, but this is not due to a real policy but entirely coincidental, and based on synonymy rather than derivational considerations. Thus, at *body* CIDE uses *physical structure* as a "guide-word", which would allow the attentive learner to find their way to *physical*. In other cases certain jumps are made possible from the Anglo-Saxon derived

form to the foreign "derived" adjective, as in the case of *eatable*, where both OALD and CIDE cross-refer the user to *edible*. Reversely, CIDE and LDOCE, but not OALD, also allow the user to go from *edible* to *eatable*.

CIDE, in a quite unpredictable and unexplainable way, occasionally offers some help: under *flower* the reference "See also FLORAL" is given. Since CIDE, under *father*, *mother*, *brother* and *moon*, for example, does not provide such a reference, this is apparently a random, though not quite unique, hit: CIDE does the same under *hell* (where there is a cross-reference to *hellfire*; *hellhole*; *infernal*) and under *night* (where there is a rather unexpected reference to *nocturnal*).⁹

There are also some apparently random cases where a cross-reference is given from the one foreign adjective to the other, as at *regal* and *royal*: from *regal* the OALD gives a reference to *royal*, and *vice versa*. Likewise, LDOCE refers *royal* to *regal*, though, rather inconsistently, not back again. For all four dictionaries the *male*, *masculine*, *female*, *feminine* group also involves a certain amount of cross-referencing (in the case of CC in the extra column, as was to be expected).¹⁰

The emerging picture is clear, however: not one of the four learner's dictionaries is really concerned with helping the user find the way from *moon* to *lunar* and similar cases, though there are some random exceptions. The possibly relevant cross-references characteristically involve synonymy instead of derivational relations. Even in OALD, which in its latest edition has introduced the novelty of "word families" that "show all words related to the headword", this usually means only families linked by strictly formal, morphological, relations. The only other family members recognised by OALD are synonyms and antonyms. Moreover, the word family boxes are comparatively rare.

One would have expected that CIDE, the only learner's dictionary left with nesting of derived forms whenever alphabetical order is not disturbed too dramatically, and hence with a certain awareness of important semantic links, would have shown the greatest awareness of the strong links between e.g. *moon* — *lunar*, yet this is not the case. The links that do occur are obviously haphazard.

3. Suggestions for improvement

Learner's dictionaries, with their focus on encoding and vocabulary building in addition to decoding, should have shown a greater awareness of the need of foreign learners in this particular respect. The character of English, with its large Latin and Greek vocabulary, is such that it often relies on foreign roots and borrowings in comparison with other languages with a different linguistic history. Learners from other European languages with a certain common classical linguistic heritage may still be able to overcome this particular hurdle in English, since not all English "nonmorphological derivatives" will be unfamiliar to them. Yet they too would be better served by a learner's dictionary which does recognise this problem. And learners from non-European backgrounds would, I suspect, need help in this respect even more.

This particular problem could therefore be solved in one of two possible ways. The first and most natural possibility would be to make liberal use of cross-references from e.g. *sun* to *solar*, possibly, though not necessarily, from *sun* to *sunny*, and certainly also from *sunny* to *solar* and *vice versa*. To the extent that the definitions and possible examples would not sufficiently discriminate between *solar* and *sunny* short usage notes might be inserted explaining when to use which. A second, though much less preferred solution would be to list the most important cases in an appendix to the dictionary. Its disadvantage would be that dictionary users usually do not seem to consult such appendices because they are often unaware of them.

4. Conclusion

In the appendix, I have listed the 166 words with "nonmorphological derivatives" I was able to find by introspection and by more direct searches. A number of them will not be directly relevant to learners but in my opinion most of them will. Since I have also included cases such as *calf* — *veal*, the notion of "derivative" should not be interpreted too strictly. The number of such cases is, however, small, and the majority of cases listed are examples of "derived" adjectives — sometimes verbs — taken from either Latin (directly or by way of French) or Greek, thus from a foreign source. Frequently, the base is an Anglo-Saxon word (e.g. *blood*) or an originally borrowed word (e.g. *bishop*, itself ultimately from Greek) which has become part of the ordinary "core" vocabulary. The "nonmorphological derivatives" are then occasionally the less adapted forms from the same source (such as *episcopal*) or forms from a totally different source (e.g. *ventral*, adjective to *belly*).

The list clearly reveals the character of the English vocabulary: derived from a number of highly diverse sources and, in addition to and alongside morphological derivation proper for its word formation, frequently using borrowings from another language. The degree of morphological transparency is therefore irregular, depending on whether there is real derivation (e.g. *breakable*) or not (e.g. *fragile*). Since *breakable* clearly connects with *break*, it is not only morphologically but also semantically dependent on *break* in a way *fragile* is not. English, in comparison to Dutch, seems to have a vocabulary with a higher percentage of what can be termed "semantically independent" words, by which I mean words such as *fragile* which have no morphological links. It also seems that English more often has a separate word for each different meaning. Thus, where most languages would say "pig meat", English has *pork*, and where many languages would have a derived adjective "feelable", English has *palpable* and *tangible*. This has, as I have argued, consequences for learners from languages with a different character. It will be interesting to see how the morphologically derived and "nonmorphologically derived" forms have semantically diverged (or not) and what particular divergence patterns can be observed.

Appendix

List of words and their "nonmorphological derivations"

1. air — aerial (assault, photography)
2. animal — bestial (cf. beast)
3. ape (cf. monkey) — simian (forehead)
4. avoid — inevitable, (un)avoidable
5. back — dorsal (fins)
6. bad — deteriorate (health), worsen (political situation)
7. ball — spherical (object)
8. beast — bestial (savagery), beastly (cf. animal)
9. beginning — initial (stages)
10. believe — credible (alternative), believable (explanation)
11. belly — ventral
12. bend — flexible (rules), bendable
13. bird — avian (malaria)
14. bishop — episcopal (duties)
15. blood — sanguinary (wars), bloody
16. body — physical (defect), corporeal (needs), corporal (punishment), somatic (cells), bodily (functions)
17. bone — ossiferous, bony; ossify
18. brain — cerebral (haemorrhage), brainy
19. break — fragile (china, peace), breakable
20. brother — fraternal (feelings), brotherly
21. calf — veal
22. carry — portable (phone), carryable
23. cat — feline (grace), catty (comment)
24. century — centennial, centenary (celebrations)¹¹
25. chalk — calcareous (shells), chalky
26. chest — pectoral (muscles), chesty
27. child — infantile (diseases, behaviour), puerile (jokes, behaviour), childish, childlike
28. church — ecclesial (teachings), ecclesiastic(al) (history), churchy
29. citizen — civil (strife, disobedience)
30. city — urban (development)
31. coast, shore — littoral (forests)
32. compel — compulsive (liar, eating), compelling (novel), compulsory (education, subject, test)
33. correct — corrigible
34. country(side) — rural (community), rustic (farmhouse, idyl)
35. cow — bovine (diseases, stupidity); beef
36. daughter — filial (duties), daughterly
37. day — diurnal (rhythms, clock), daily

38. death — lethal (weapons, blow), mortal (wound, enemy), deathly, deadly
39. deceive — deceptive (appearances)
40. do — feasible, practicable; doable
41. dog — canine (tooth, behaviour), doggy, doglike
42. donkey, ass — asinine (remark)
43. ear — aural (test)
44. earth — terrestrial (TV channels), earthly, earthy
45. east — oriental (art), easterly, eastern
46. eat — edible (snails, fungi), eatable
47. eighty — octogenarian¹²
48. empire — imperial
49. enemy — hostile (attitude), inimical (climate, conditions, influence)
50. eagle — aquiline
51. eye — ocular (defects)
52. fast — accelerate (rate of growth), speed up
53. father — fatherly, paternal (authority); fatherhood, paternity
54. feel — palpable, tangible
55. fever — febrile (activity), feverish
56. flesh — carnal (desires, knowledge), fleshy
57. flower — floral (pattern), florid (style), flowery
58. forget — oblivious; forgetful
59. fox — vulpine; foxy
60. friend — amicable (settlement), friendly
61. god — divine (wisdom, inspiration), godlike, godly
62. good — improve (health), ameliorate (working conditions)
63. hair — hirsute (animals), hairy
64. hear — audible
65. heart — cordial (smile), cardiac (arrest), hearty
66. heat — thermal (energy)
67. heaven — celestial (bodies, beauty), heavenly
68. heavy — gravity, heaviness
69. hell — infernal (heat, machine), hellish
70. horse — equine; hors(e)y
71. ice — glacial; icy
72. iron — ferrous (metals), iron (constitution, determination)
73. island — insular
74. joke — jocular (remarks), jocose, jokey
75. king — regal (splendour), royal (family), kingly (bearing) (cf. queen)
76. language — linguistic, lingual
77. laugh — risible, ridiculous, laughable
78. law — legal, juridical
79. lie — recumbent (figure), lying
80. life — vital (functions, organs), lively
81. light — luminous (paint), illuminate

82. lion — leonine (head)
83. lips — labial (sounds), lippy
84. love — amorous (looks, advances), loving, lovely
85. lung — pulmonary
86. man — male (child, voice); masculine (face, word); virile (body), manly; human, humane¹³
87. many — multitude, mass; multiply, proliferate
88. marriage — nuptial (bliss, ceremony), connubial
89. marry — nubile; marriageable
90. middle — medial; median
91. milk — lactic, lacteal; milky
92. mind — mental (image, process)
93. mistake — erroneous (beliefs, conclusions), mistaken(ly)
94. money — pecuniary (gains, advantage), financial (gains), monetary
95. monkey (cf. ape) — simian (forehead), monkeyish
96. moon — lunar (eclipse)¹⁴, moony
97. mother — maternal (grandfather, love), motherly; motherhood, maternity
98. mouth — oral (examination, sex), mouthy
99. move — mobile (phone, home), movable (goods, items)
100. neglect — negligible, neglectful
101. new — novelty, newness
102. night — nocturnal (visit, animal), nightly
103. north — boreal (forests), northerly (winds), northern
104. nose — nasal (voice, sounds), nos(e)y
105. old — senescence, senescent
106. owner — (proprietor) proprietorial (air, rights)
107. peace — pacific (community), peaceful; pacify
108. people — popular (vote, misconception, wisdom)
109. pig — porcine (aroma), piggish; pork
110. place — local (authorities, hero)
111. queen — regal (splendour), royal (family), queenly (bearing) (cf. king)
112. plough — arable
113. pope — papal; popish
114. punishment — penal; penalise
115. rain — pluvial (weather conditions), rainy
116. read — legible (signature, hand), readable
117. receive — receptive (audience, market), recipient, receiver
118. river — fluvial (deposits), riverine
119. river bank — riparian (wildlife); fluvial, fluviatile
120. rule — regular (routine, intervals)
121. salt — saline, salty (solution, taste)
122. sea — marine (biologist, pollution, habitat)
123. see — visible, seeable
124. sheep — ovine (offal), sheepish; mutton

125. shore — littoral
126. side — lateral (branch, pressure)
127. sight, see — visual (field, arts, image), sightly
128. sister — sisterly, sisterhood, sorority
129. sit — sedentary (life), sitting
130. skull — cranial (nerve, injury)
131. sleep — dormant (inflation, giant), somnolent (cat, village), sleepy, sleeping
132. slow — decelerate, slow down
133. snake — serpentine (course of a river), snaky
134. son — filial (duty, love)
135. sound — phonetic (alphabet, elements), phonic (skills), acoustic (properties, possibilities), sonic (waves, boom)
136. south — austral (winter), southern, southerly
137. space — spatial (awareness), spacious (room, kitchen), spac(e)y
138. spider — arachnoid, spidery
139. split — fissile, splittable
140. spring — vernal (equinox, grove)
141. star — stellar (constellations, performance, career), sidereal (day, zodiac), astral (bodies, navigation), starry
142. sun — solar (eclipse, temple), sunny
143. table — tabular (form)
144. three — *predet.* treble (the usual charges), *adj.* treble (line), triple (murder, jump), triplet(s)¹⁵
145. time — temporary (accommodation), temporal (perspective, distance), timely
146. tooth — dental (surgeon, treatment), toothy
147. touch — tangible, tactile, palpable; touchable, touchy
148. town — urban (development)
149. ree — arboreal (birds)
150. two — double (figures), duplicate
151. uncle — avuncular (manner)
152. understand — intelligible (English), understandable
153. voice — vocal (chords, performance)
154. wall — mural (design)
155. war — martial (arts, law), warlike
156. water — aquatic (sports, environment), aqueous (lotions), watery
157. west — occidental, westerly, western
158. whale — cetacean (anatomy)
159. winter — hibernal, wintry; hibernate, (over)winter
160. wolf — lupine (pack), wolfish
161. woman — womanly, female (animal), feminine (voice, figure), effeminate (manners), womanhood, femininity, womanliness
162. word — verbal (abuse, skills), verbose (style), wordy
163. world — global (warming, problems), mundane (life, matters), worldly
164. year — annual (salary, report), yearly

165. young — rejuvenate
166. youth — juvenile (delinquency, offender), youthful

Endnotes

1. This part of my article is an elaboration of Van der Meer (2000).
2. Cf. the frequent error of *mitigate against* instead of *militate against* in English newspapers, caused by the phonetic similarity of two otherwise formally opaque, i.e. morphologically unmotivated, words.
3. "Largely", because the (second) [k] in *koninklijk* is due to historical conditions, and no longer productive.
4. Strictly speaking, *-zicht-* is a noun meaning "vision", so that the gloss might instead be "*through-sight-ish(-ness)*", etc.
5. See Fill 1980: 136 for this example. This quite original study compares English with German word transparency. *Wortdurchsichtigkeit* in the title of Alwin Fill's book, to illustrate my point once again, translates literally as "word-through-see-able-ness".
6. Incidentally, the word *eclipse* itself, when compared with Dutch *verduistering* (consisting of the prefix *ver-*, the adjective *duister* and the nominalising suffix *-ing*) might also be called opaque and hence "unmotivated". This too is an example of how many loanwords reduce what might be called the "morphological content" of a language.
7. Aptly called "noncognate relationships" by Heuberger (2000).
8. In some cases, e.g. *aural*, the compilers themselves may not have felt any strong connection between *ear* and *aural*, since none of the four learner's dictionaries uses *ear* in the definition.
9. Also cf. *nose* where *nasal* is pointed out.
10. CC uses the extra column for synonyms and antonyms, so that this column occasionally provides the looked for "base".
11. Though strictly speaking this is Latin from Latin, I have included a couple of such cases because these derivations, too, are not so easy for foreign learners.
12. Also cf. *hexagenarian*, *septuagenarian* and of course *nonagenarian*.
13. Here man as a "human being" is meant.
14. There is even a further complication in *selenography*, where the combining form *seleno-* (from Greek) is used, where Dutch, for example, would simply have *maanbeschrijving* "moon description".
15. Cf. *quadruplet (four)*, *quintuplet (five)*, *sextuplet (six)*, etc.

Bibliography

A. Dictionaries

- CC: Sinclair, J. (Ed.). 1995². *Collins COBUILD English Language Dictionary*. London: HarperCollins.
CIDE: Procter, P. (Ed.). 1995. *Cambridge International Dictionary of English*. Cambridge: Cambridge University Press.

LDOCE: **Summers, D. (Ed.)**. 1995³. *Longman Dictionary of Contemporary English*. Harlow: Longman.
OALD: **Wehmeyer, S. (Ed.)**. 2000⁶. *Oxford Advanced Learner's Dictionary of Current English*. Oxford: Oxford University Press.

B. Other sources

- Cowie, A.P.** 1999. *English Dictionaries for Foreign Learners. A History*. Oxford: Oxford University Press.
- Fill, Alwin.** 1980. *Wortdurchsichtigkeit im Englischen. Eine nicht-generative Studie morphosemantischer Strukturen. Mit einer kontrastiven Untersuchung der Rolle durchsichtiger Wörter im Englischen und Deutschen der Gegenwart*. Innsbruck: Universität Innsbruck.
- Heuberger, Reinhard.** 2000. *Monolingual Dictionaries for Foreign Learners of English. A Constructive Evaluation of the State-of-the-Art Reference Works in Book Form and on CD-Rom*. Vienna: Wilhelm Braumüller.
- Van der Meer, Geart.** 2000. Further Ways to Improve the Active Dictionary: Collocations, Nonmorphological Derivations, Grammar. Mogensen, Jens Erik, Viggo Hjørnager Pedersen and Arne Zettersten (Eds.). *Symposium on Lexicography IX. Proceedings of the Ninth International Symposium on Lexicography, April 23-25, 1998 at the University of Copenhagen*: 125-141. Tübingen: Max Niemeyer.

The Hausa Lexicographic Tradition

Roxana Ma Newman, *International Programs and Department of Linguistics (rmnewman@indiana.edu)* and
Paul Newman, *Department of Linguistics and West African Languages Institute (pnxxpn@indiana.edu), Indiana University, Bloomington, United States of America*

Abstract: Hausa, a major language of West Africa, is one of the most widely studied languages of Sub-Saharan Africa. It has a rich lexicographic tradition dating back some two centuries. Since the first major vocabulary published in 1843 up to the present time, almost 60 lexicographic works — dictionaries, vocabularies, glossaries — have been published, in a range of metalanguages, from English to Hausa itself. This article traces the historical development of the major studies according to their type and function as general reference works, specialized works, pedagogical works, and terminological works. For each work, there is a general discussion of its size, accuracy of the phonological, lexical, and grammatical information, and the adequacy of its definitions and illustrative material. A complete list of the lexicographic works is included.

Keywords: ARABIC, BILINGUAL LEXICOGRAPHY, DIALECTAL VARIANTS, DICTIONARIES, ENGLISH, ETYMOLOGIES, FRENCH, GERMAN, GLOSSARIES, GRAMMATICAL CATEGORIES, HAUSA, LANGUAGE LEARNING, LOANWORDS, NEOLOGISMS, NIGER, NIGERIA, ORTHOGRAPHY, PHONETIC TRANSCRIPTION, PHONOLOGY, RUSSIAN, STANDARD DIALECT, STANDARDIZATION, TERMINOLOGY, VOCABULARIES, WEST AFRICA.

Opsomming: Die leksikografiese tradisie in Hausa. Hausa, 'n belangrike taal van Wes-Afrika, is een van die tale van Afrika suid van die Sahara wat die wydste bestudeer word. Dit het 'n ryk leksikografiese tradisie wat ongeveer twee eeue oud is. Van die eerste groot woordeboek wat in 1843 gepubliseer is tot die hede is ongeveer 60 leksikografiese werke — woordeboeke, naamlyste, woordelyste — gepubliseer in 'n reeks metatale van Engels tot Hausa self. Hierdie artikel gaan die historiese ontwikkeling van die groter studies aan die hand van hulle tipe en funksie as algemene naslaanwerke, gespesialiseerde werke, opvoedkundige werke, en terminologiese werke na. Vir elke werk is daar 'n algemene bespreking oor sy grootte, akkuraatheid van die fonologiese, leksikale en grammatikale inligting, en die toereikendheid van sy definisies en illustratiewe materiaal. 'n Volledige lys van die leksikografiese werke is ingesluit.

Slutelwoorde: AANLEER VAN TAAL, ARABIES, DIALEKTIESE VARIANTE, DUTS, ENGELS, ETIMOLOGIEË, FONETIESE TRANSKRIPSIE, FONOLOGIE, FRANS, GRAMMATIESE KATEGORIEË, HAUSA, LEENWOORDE, NAAMLISTE, NEOLOGISMES, NIGER, NIGERIEË, ORTOGRAFIE, RUSSIES, STANDAARDIDIALEK, STANDAARDISERING, TERMINOLOGIE, TWEETALIGE LEKSIKOGRAFIE, WES-AFRIKA, WOORDEBOEKE, WOORDELYSTE.

1. Introduction

This article is a comprehensive survey of Hausa lexicography, comprising dictionaries, vocabularies, and glossaries. Hausa, a major language of West Africa,

with an estimated 35-40 million speakers, is the largest member of the Chadic language family, which in turn belongs to the Afroasiatic phylum (P. Newman 1980a). It is the first language of ethnic Hausas and settled Fulanis in Hausaland, which stretches across the northern states of Nigeria and into southern Niger, and also of Hausa communities in the Sudan (Abu-Manga 1999). Hausa is spoken as a first language by scattered settlements throughout West Africa, and as a second language or lingua franca by millions of non-Hausas in northern Nigeria and in the northern parts of Benin, Togo, and Ghana. It is one of the three official national languages of Nigeria.

Due to its geographic spread, Hausa has a number of dialects, marked by differences in pronunciation, grammar, and vocabulary. The Hausa spoken in Kano State, Nigeria, is considered the standard dialect. This is the norm for the written language as well as the main dialect used by announcers in radio and television. Differences in modern vocabulary between the Hausa of Nigeria and that of Niger are due to their colonial histories, with heavy borrowings from English in the former and from French in the latter. Because most Hausa people are Muslim, a substantial portion of the lexicon reflects Arabic borrowings, especially with respect to religion, science, and government. There are also borrowings from neighboring African languages, especially Tuareg, Kanuri, Fulani (Fulfulde), and Yoruba.

Hausa has two writing systems. One orthography utilizes Arabic script, termed *ajami*, which has been employed at least since the beginning of the nineteenth century, and the other uses the Roman alphabet, termed *boko*, which was introduced by the British and the French at the beginning of the twentieth century. The current-day standard orthography uses Roman script with neither distinctions of vowel length nor of tone being marked.

Before examining the range of Hausa dictionaries that have been produced, it is important to provide some essential phonological and morphological information about the language itself so that one can appreciate some of the difficulties that Hausa lexicographers have had to face. This sketch, based on the reference grammar of P. Newman (2000), is limited to the standard Kano dialect.

Hausa has 25 consonants, leaving aside the labialized and palatalized (e.g., /**k**w/ and /**k**y/) consonants, namely /', **b**, **ɓ**, **c**, **d**, **ɗ**, **f**, **g**, **h**, **j**, **k**, **ƙ**, **l**, **m**, **n**, **r**, **ɾ**, **s**, **sh**, **t**, **ts**, **w**, **y**, 'y, **z**/. (Note the orthographic use of /c/ to represent the affricate /ch/ as in English *church*, and the use of the apostrophe /'/ to represent glottal stop, this latter only being indicated in word-medial position.) Most of the consonants are similar to what are found in English and other European languages, but there are some significant differences.

First, Hausa has five "glottalized" consonants. Two of these, /**ɓ**/ and /**ɗ**/, are laryngealized/implosive stops. The third is an ejective velar /**ƙ**/. There is also an apical ejective, orthographically represented as /**ts**'/, whose pronunciation varies between [s'], [ts'], and [ch'] depending on position and dialect. The

last of the glottalized consonants is a laryngealized semivowel /'y/, a historically recent phoneme derived from /d̥y/ which is found only in a small number of lexemes of high frequency, e.g. 'yā 'daughter, small', 'yā'yā 'children'.

Second, Hausa has two phonemic /R/ sounds which are not distinguished in the standard orthography (P. Newman 1980b). One, which does not occur in word-final position, is a retroflex flap represented by /r/ e.g., **barà** 'servant'. The other, commonly represented in linguistic works by /r̄/, is an apical tap or trill, e.g., **taṛà** 'nine'.

Third, most Hausa consonants can occur as geminates, indicated by double letters, e.g., **gammō** 'head pad', cf. **gāmō** 'meeting'.

Hausa has five vowels, **a, e, i, o, u**, all of which can occur long or short, e.g., **tāfi** 'palm of hand' vs. **tāfi** 'go', and two diphthongs **ai** and **au**. In scholarly works, long vowels are either indicated by means of a macron (**ā**) or by double vowels (**aa**). In closed syllables, only short, monophthongal vowels occur.

Hausa is a tone language with three distinctive tones, which are not indicated in standard orthography. These are high (H), low (L), and falling (F). (There is no rising tone.) In modern Hausa linguistic scholarship, low is marked with a grave accent, e.g., **à**, falling is marked with a circumflex, e.g., **â**, and high tone is left unmarked, e.g., **wùya** (L-H) 'trouble', **fuskà** (H-L) 'face', **yārā** (F-H) 'children'. (Most Hausa examples in this article use this tone/length notation, except when quoting examples drawn directly from works that are not phonologically marked.)

Hausa has two grammatical genders, masculine and feminine, distinguished in the singular only. Most feminine nouns end in /a/, e.g., **rānā** 'sun', but some masculine nouns do also, e.g., **watà** 'moon', and thus gender assignment has to be overtly marked. A striking and complex feature of Hausa is the multiplicity of noun and adjective plural formations involving suffixation, infixation, reduplication, and tone change. Although there are some correlations between the phonological shapes of the singulars and the choice of the plurals, this is essentially word-specific information that has to be lexically indicated. Moreover, many words allow alternative plural forms, e.g., **dōkī** 'horse', pl. **dawākī**, **dàwàkai**, and (less often) **dōkunà**.

There are two main morphogrammatical features of verbs that raise lexicographic problems. First, in certain syntactic environments, such as in a continuous tense or when following a modal verb (such as 'begin', 'repeat', 'quit', etc.), Hausa uses a "verbal noun" gerundive-type form rather than a finite verb, e.g., **yā cīzā** 'He bit (it)', but **yanā cīzō** 'He is biting', **yā dainā cīzō** 'He quit biting'. The formation of verbal nouns from corresponding verbs is not regular or predictable and therefore needs to be noted in dictionary entries.

Second, Hausa verbs occur in morphologically productive forms termed "grades" (Parsons 1960), some of which are basic and some of which are derivational classes indicating "action totally done, action done away, action done towards or for the speaker, action affecting the subject (passive)", etc. Examples are: **sàya** 'buy' (grade 2), **sayè** 'buy all' (grade 4), **sayā** 'sell' (grade 5), **sayō** 'buy

and bring' (grade 6), *sàyu* 'be bought' (grade 7). The grades present two lexicographic problems. First, not all verbs function in all grades, their occurrences being lexically specific. Moreover, although some grade forms (especially grades 2 and 3) tend to be more basic than others, this is not always the case with individual verbs, cf. *dafà* 'cook' (grade 1 basic), *kashè* 'kill' (grade 4 basic), *dāwō* 'return' (grade 6 basic), *jitu* 'get along' (grade 7 basic). Second, the meaning of a verb in some grade forms is fully predictable from the core meaning of the verb, but in others it is not. Thus the issue of how many and which verbs of this rich verbal system to include in a dictionary constitutes a vexing problem.

2. Types of Hausa lexicographic works

Hausa (and other Chadic languages) can boast of a long and distinguished tradition of linguistic scholarship going back some two centuries (see the general bibliographies of Baldi (1977) and P. Newman (1996), and the lexicographic overviews of R. M. Newman (1974) and Newman and Newman (1991)). As early as 1790, explorers and travellers in West and North Africa began to note down numerals and a few nouns in Hausa (see Hair 1967), culminating in the more extensive wordlists of Koelle (1854) and Barth (1862-1866). But if one had to pick one work from which to represent the beginning of modern Hausa lexicography, that would have to be the Hausa vocabulary of J. F. Schön (1843), a missionary who worked in West Africa for the Church Missionary Society. This productive lexicographic tradition continues to the present day, with the most recent dictionary being published in 1999. Over this period of 150+ years, 59 dictionaries, vocabularies, and glossaries have appeared, most of them bilingual. The metalanguages used in these works range from English, German, and French to Arabic, Italian, Japanese, and Russian. More recently, a few monolingual dictionaries have begun to appear, written by Hausa speakers themselves.

This survey categorizes these 59 works into four functional types, based on their primary purpose:

- *Reference works*: These are general dictionaries whose purpose and use are primarily scholarly. They are constructed on the current linguistic and analytical principles of their day. These include the major descriptive bilingual dictionaries.
- *Specialized works*: These are mostly glossaries or vocabularies devoted to restricted domains of the Hausa lexicon (crafts, fauna, idioms, music, etc.) or designed for special purposes, such as dialect or etymological studies.
- *Pedagogical works*: These are smaller bilingual (or trilingual) and monolingual dictionaries aimed at language learners, the focus being the basic vocabulary and expressions of every-day usage.

- *Terminological works*: These are glossaries created to aid the expansion and development of Hausa vocabulary to meet the needs of a modern society. Standardization of loanwords and the creation of neologisms are explicit goals.

This article does not attempt to describe all of these works — a full list appears in the bibliography — but rather to comment on those which are considered by Hausa scholars to be milestones in the long development of Hausa lexicography. Each work reflects the progressive struggle of scholars and compilers to provide a better understanding of numerous aspects of Hausa, whether of pronunciation, orthography, tone, dialectal variation, etymological origins, grammatical categorization and grammatical relations, translation vs. explanatory equivalence, semantic relations between words, and neologisms and semantic extensions.

3. Reference works

James Frederick Schön, the father of Hausa studies, never actually set foot in Hausaland. He spent some fifteen years in Freetown, Sierra Leone, where he studied Hausa with freed slaves (Schön 1843). After returning to England, he continued his interest in Hausa, working with two speakers who had been brought back to Europe in the mid-1850s by the famous traveller Heinrich Barth (see Kirk-Greene and Newman 1971). One (Dorugu) was a native speaker, the other (Abbega) was a native Margi who spoke Hausa as a second language. With the help of Dorugu, Schön compiled the first large-scale lexicographical work on Hausa that can properly be called a dictionary (Schön 1876). This work consists of a Hausa-English section of some 3 800 head entries and a slightly smaller English-Hausa section of about 3 200 entries. In spite of its age, it is surprisingly modern in its approach and grammatical analysis, describing the major grammatical morphemes, pronoun classes, and derivational suffixes. Verbs are marked for transitivity, and nouns for plurality, though not for gender. The definitions are accurate and often describe the social situations in which words are used. More unfamiliar, however, are his phonological and orthographic notations, as Schön generally did not recognize glottalized or geminate consonants, and only occasionally marked long vowels by a stress mark. He was apparently unaware of the existence of tone in Hausa. Despite these weaknesses, Schön's orthography is still quite readable, with a basically phonemic alphabet and correct division of words.

A few years after Schön, the first vocabulary with French as the metalanguage appeared, compiled by J. M. Le Roux (1886) and published in Algiers. This 300-page work, which includes a 30-page grammatical introduction, is unusual in that the published volume is an entirely handwritten manuscript. It is the first lexicographic work on Hausa to include head entries written in both Arabic and Roman scripts. Although little is known about this work or the

compiler, the fact that many entries are actually Arabic rather than Hausa indicates that it was likely based on the speech of a Hausa speaker living in North Africa who had an incomplete command of his native language. The work therefore remains more a curiosity in the history of Hausa lexicography than a work of scholarly significance.

The next period of lexicographic productivity took place around the beginning of the 20th century, when three substantial Hausa dictionaries appeared, quite independently of each other, each demonstrating significant advances in the understanding of Hausa. These dictionaries were compiled by English (Robinson 1899-1900), German (Mischlich 1906), and French scholars (Landeroin and Tilho 1909). Though inadequate from the point of view of phonological analysis, these early works contain useful grammatical information (such as gender and some noun plurals), short translation equivalents or reasonably accurate explanations, and some idiomatic expressions.

The Hausa-French dictionary of Landeroin and Tilho, published in Paris, is a substantial work, with a 172-page Hausa-French section followed by a 163-page French-Hausa section, each containing about 6 000 entries. While acknowledging the existence of Hausa written in Arabic script, the authors chose the Roman script as being more appropriate, especially for Europeans working in West Africa who wanted to learn Hausa. Their transcription recognizes the existence of the apical ejective, which they transcribed as /ts/, but not the other glottalized consonants. They were aware of some kind of distinction in vowel length for a small number of Hausa word pairs, which they attempted to mark with a "tonic accent", but failed to recognize that tonal differences were also involved. The major Hausa word classes are marked (nouns, verbs, adverbs, etc.). Grammatical gender was not recognized, possibly because they were working with a nonstandard Hausa dialect which did not make that distinction. Straightforward loanwords from Arabic were noted. Their transcriptions are surprisingly accurate in terms of word and morpheme boundaries, and their Hausa examples quite colloquial. The French-Hausa section is more properly a glossary or index, with few examples.

Charles H. Robinson was an Arabist and a member of an elite group of Hausa scholars in England who were dissatisfied with what they claimed was the overly "European" cast to Schön's works. After several trips to Hausaland, where he also collected manuscripts of Hausa written in Arabic script, Robinson published a two-volume dictionary in 1899/1900, which was revised three times in the following 25 years. Robinson's first two editions (1899/1900 and 1906) noted each Hausa headword in both Roman and Arabic scripts. Later editions dropped the *ajami* since most of his European readers did not know Arabic. He supplied Arabic roots for numerous Hausa words where he detected borrowing. Surprisingly, Robinson did not utilize vital phonemic information contained in *ajami*, where the differences between long and short vowels are clearly noted. By the time of his final and greatly expanded fourth edition (1925) of 11 000 entries, he had solved the problem of the glottalized

consonants /**ḅ**, **ḍ**, **ḗ**/, which he marked with a subscript dot under the normal consonants. However, he did not alphabetize them separately. Like Schön, Robinson acknowledged the existence of some type of tonal and/or accentual system, and used a diacritic to mark "short" syllables although this was very sporadic and unsystematic. But as he noted (1925⁴: x), "The tones of pronunciation cannot be represented satisfactorily by any system of printing ..." In fact, the Hausa tonal system would not be fully recognized until the 1920s and not systematically analyzed and marked in Hausa dictionaries until the 1930s and 1940s. In general, Robinson's translation equivalents and definitions are accurate, with many entries having illustrative phrases and explanations. Some entries also have philosophical explanations of connotative and inferential meanings, as well as information of a more encyclopedic nature. Among his major innovations were a cross-referencing system to show the grammatical and semantic relations between words, the identification of major dialectal forms, and the stylistic use of the asterisk to indicate more esoteric scholarly loanwords from Arabic that were not yet fully incorporated into Hausa.

Two small volumes, Harris (1908) and Fletcher (1912), are collections of Hausa folklore, riddles, sayings, and customs, each accompanied by 1 200-word vocabularies, excellent for their time. Harris's simple glossary contains a number of common idiomatic expressions. Fletcher's fuller work was intended as a supplement to the early editions of Robinson. It lists words missing in Robinson's first two editions, including a number of correctly analyzed compounds, as well as expanded and improved definitions. For example, whereas Robinson simply defines Hausa **garaya** as 'harp', Fletcher indicates 'a two-stringed musical instrument with a long oval body, bigger than *molo*. The word is also used to describe the *molon arna*, an oblong instrument like a zithern ...' These additional entries and fuller explanations were incorporated, sometimes verbatim, into subsequent editions of Robinson.

Adam Mischlich's important Hausa-German dictionary (1906) was the result of years spent working as a missionary in a large dialectally diverse Hausa trading community in Kete-Krachi, Togo, complete with four mosques, a number of Koranic schools and many educated scholars. He published a grammar in 1902 and a dictionary in 1906, completed before he had seen Robinson's dictionary. This almost 700-page work contains approximately 7 000 head entries written in both Roman and Arabic scripts. In contrast to Robinson, Mischlich incorporated the long/short vowel distinction noted in *ajami* into his Roman transcription, marking long vowels with a macron although this was not completely reliable. He treated the ejective /ts/ as a unit phoneme which he alphabetized separately from /t/, a practice not subsequently followed by lexicographers until quite recently. The dictionary also noted dialectal forms and it was this research that Robinson incorporated in the last two editions of his own dictionary. Mischlich's dictionary reflects a deeper analysis of Hausa grammatical structure and understanding of derivational morphology, with many plurals and other related forms being included under the head entries.

In the 1920s, the British colonial officer, F. W. Taylor, produced a series of grammars and vocabularies of Hausa and Fulani, the languages of the politically dominant elite in northern Nigeria. His trilingual Fulani-Hausa-English vocabulary of 3 800 words (Taylor 1927) is more properly a specialized glossary than a general reference work. The vocabulary is organized according to an alphabetically organized mix of grammatical and semantic categories, e.g., abstract nouns; adjectives; animals (domestic); animals (wild); farming; horses; imperatives; verbs; war and hunting. Within each category, the words are alphabetized according to the English word. This specialized glossary is included in this section because it was the first Hausa lexicographic work to explicitly include the marking of tone. Taylor utilized a curious numerical system, which he first presented in his earlier Hausa grammar (Taylor 1923). Each word was followed by a series of numbers, e.g., **nama** 35 'meat' (= H-L); **mata** 13 'wife' (= L-H); and **sabo** 23 'new' (= H-H). Although the tone marking is far from reliable in its details, this work, along with his grammar, served to establish that Hausa was a tone language rather than a stress or accentual language, and thus set the stage for all subsequent lexicographic works.

The year 1934 saw the publication of what Hausa scholars universally acknowledge to be the masterwork of Hausa lexicography, the dictionary of G.P. Bargery (1934), a missionary linguist who devoted over 20 years in Nigeria to its compilation. It is a monumental volume of more than 1 200 tightly-packed pages that has not only stood the test of time but will not likely be surpassed in the foreseeable future. It reflects a level of analytical sophistication and descriptive thoroughness that is rarely found in dictionaries or grammars of African or other "exotic" languages. It has justly been hailed as "one of the most outstanding linguistic achievements of the first half of the 20th century" (Green 1967: 59). The work is divided into two parts: the large Hausa-English section containing some 52 000 head entries, of which about 39 000 are main lexical entries and another 13 000 derived and dialectal forms; and a concise English-Hausa index of about 5 000 words.

For the first time, all Hausa head entries are written in phonologically-based, (nearly) accurate transcriptions showing the glottalized consonants, the distinction between the two /R/s, vowel length (except for final vowels with low tone), and tone. The major phonological breakthrough was his systematic analysis of the tonal structure of Hausa and the very phonetically accurate marking system that he devised, including perceptual differences that later turned out to be analyzable as intonational or subphonemic phenomena. His classification of verbs was also a major contribution, enabling him to identify derived forms and cross-reference them to their basic roots. Definitions are remarkably full and well illustrated, including numerous proverbs, epithets, and notes on the social and cultural contexts of usage. Many entries are written like paragraphs containing information more often relegated to an encyclopedia. Bargery also made efforts to group together semantically related words and synonyms. For example, under the entry for **kātō** 'huge', he lists about 330

different words (some nouns, some ideophonic adjectives, some basic, some derived) that contain any semantically related notion of 'hugeness.'

Bargery was the first lexicographer to explicitly base his dictionary on the now standard dialect of Kano, Nigeria. Nevertheless he also included many dialectal forms and his work remains an important and accurate source for dialectal information. He also included much etymological information, noting borrowings not only from Arabic, but more recent ones from English and from other Nigerian languages.

Some 50 years later, in order to update Bargery's monumental work, Neil Skinner, a prolific lexicographer in his own right, published a supplement of 1 500 entries which included neologisms, new borrowings from English and French, and semantic extensions of older terms (1985). In 1993, this supplement was incorporated in a new printing of Bargery by Ahmadu Bello University Press in Nigeria.

A second comprehensive Hausa dictionary that was modern in conception and impressive in scope appeared in 1949. It was that of R. C. Abraham, who served for two years as Bargery's assistant working primarily on Arabic origins and verb classification. Abraham, a British government anthropologist, was a truly remarkable and prolific scholar of African languages, who also compiled dictionaries and grammars of Amharic, Idoma, Somali, Tiv, and Yoruba (see Jaggar 1992). In some ways, his Hausa dictionary appears to be a revised edition of Bargery, but in other ways should be considered as an entirely new work. Abraham correctly phonemicized the tonal system and simplified its notation, marking only low and falling tones by diacritics under the vowel. While fully recognizing the distinctiveness of the two /R/s, he unfortunately chose not to mark them in his dictionary, except in the case of a few minimal pairs. He developed a more elegant classification of the verbal system, which he used to identify each head verb in the dictionary, and included a great number of syntactic and semantic collocations to illustrate these verbs. An outstanding feature of Abraham's dictionary is the complete phonemic marking of every Hausa word in the book, in the examples as well as in the headwords, so that tone and vowel length distinctions are indicated throughout in context. (In Bargery, tone and vowel length are marked only in the headwords.)

Abraham also devoted great attention to grammatical analysis and information, not surprisingly, as he had also written two grammars of the language (1934, 1959). (By comparison, Bargery was one of the few Hausa lexicographers who did not also produce a grammar.) Abraham's presentation of related grammatical and morphological forms was much more compact than Bargery's, resulting in an estimated total of 35 000 head entries. He excluded such predictable verbal derivational forms as intensives, past participles, statives, and regular verbal nouns. He also left out many nonstandard dialectal forms. In his definitions, Abraham was also more analytical, imposing a more narrowly linguistic interpretation of what properly should be included in a dictionary entry. His method was to let the various meanings and nuances of a

word manifest themselves by illustrating the word in as many varied linguistic contexts as necessary, including proverbs and epithets. Bargery, on the other hand, often failed to include illustrative examples, rather choosing to describe the cultural and social contexts and referents associated with a word (see R. M. Newman (1992) for a discussion of their methodological differences).

In the 1960s, Russian scholars began to take an interest in Hausa and, acknowledging the earlier contributions of Bargery, Abraham, Mischlich, and Robinson, produced two comprehensive dictionaries. The Hausa-Russian volume of 460 pages containing 18 000 entries was published in 1963 (Olderogge 1963), followed four years later by the Russian-Hausa 400-page volume of 22 000 entries (Laptukhin and Kano 1967). These books incorporated a few linguistic advances over previous dictionaries. In the first volume, Hausa headwords beginning with glottalized consonants, as well as the phonemes /sh/ and /ts/ represented by digraphs, are alphabetized as distinct letters in their own sections, and similarly within the word. (Although Bargery and Abraham had correctly noted the distinction between the three nonglottalized/glottalized consonant pairs (**b/ḅ**, **d/ḁ**, **k/ḅ**), they chose not to have separate alphabetic sections for them.) Phonological information such as tone and vowel length are not provided. In terms of grammatical information, this is the first dictionary to attempt to include a system for marking a few derived "grade" suffixes for common verbs, as well as a codification of common plural suffixes for most nouns. The entries themselves are not particularly deep, giving many Russian translation equivalents but few examples of how they are used. The entries that are best illustrated are high frequency grammatical morphemes, the most common verbs, and idiomatic compounds. Because there is nothing on pronunciation and little on exemplification, the dictionary is clearly designed to be used for Russian speakers translating from written Hausa sources, and not for productive use. Similarly, the Russian-Hausa volume is designed to be used for reading comprehension by Hausas studying in Russia. Because many Russian words and concepts do not have exact Hausa translation equivalents, explanatory Hausa phrases are provided though these sometimes seem stilted and unnatural. Interestingly, there is some attempt to provide Hausa "equivalents" for modern Russian technical vocabulary by literal transliteration, so that *radio-activity* is rendered by the made-up Hausa word **rediyooaktiviti**, followed by a phrasal explanation.

The most recent scholarly reference work is the Hausa-French dictionary of Abdou Mijinguini (1994), a native Hausa linguist from Niger. His 500-page dictionary contains approximately 10 000 head entries, well-illustrated with examples of colloquial usage. It is based largely on standard Niger Hausa, essentially the dialect of Damagaram rather than the standard Kano dialect that has dominated most previous lexicographic research. It represents a major step in the history of Hausa lexicography in that it is the work of an all-Hausa team of researchers, and it is printed in Niger.

A number of innovative practices are found in this dictionary. Mijinguini adheres to the phonemic principle in the alphabetization of Hausa entries beginning with the digraphs /sh/ and /ts/, like the compilers of the Hausa-Russian dictionary. (By contrast, all of the contemporary pedagogical dictionaries — see following sections — follow the graphemic principle in the alphabetization of entries, with /sh/ being alphabetized within /s/ and /ts/ within /t/.) Regarding other phonological features, each headword in Mijinguini's dictionary is written in standard orthography, followed by a phonetic transcription marking tone, vowel length, and the distinction between the two /R/s; this transcription is totally reliable.

The major departure from standard practice is in the choice of the form of the verb to use for the head entry. Since the earliest dictionaries, Hausa scholars have conventionally used the finite form of the verb occurring in phrase-final position (with transitive verbs, when no object is expressed), e.g., *yā kāmà* 'he caught (something)'; *àbín dà ya sàtā* (with L-H tone) 'the thing that he stole'; *yā tàfi* 'he went'. This finite form occurs in all tenses/aspects except the continuous. For his dictionary, Mijinguini has chosen the gerundive or verbal noun form used in the continuous, corresponding roughly to *-ing* in English, e.g., *kāmāwā* 'catching'; *sātā* 'stealing, theft' (with H-L tone); *tàhiyā* 'going'. He gives examples illustrating finite forms where appropriate, e.g., under *tàhiyā* 'going' will be found the sentence, *haɗ sun tàhi?* 'Have they gone already?' For Mijinguini, it seemed counterintuitive to give finite, syntactically contextual forms (equivalent to 'caught, stole, went') as head entries of verbs. Rather, his verbal noun forms are the closest equivalent in Hausa to nonconjugated, nonfinite verb forms that can be expressed out of context. It remains to be seen whether this innovative practice will be adopted by other Hausa lexicographers.

4. Specialized works

In addition to the general, all-purpose reference dictionaries, there have been a number of special-purpose lexicographic works, mostly dealing with specific or restricted semantic domains. An early study of this type, which focused on medical terminology, was published in 1915 by A.C. Parsons, father of the renowned British scholar F.W. Parsons, who, in the period after the Second World War, would establish himself as the world's leading expert on Hausa grammar.

A key study of botanical terminology was that of Dalziel (1916), a medical officer in the British colonial administration. This work included the scientific classification and Latin terminology for Hausa flora and has been a standard resource for all subsequent Hausa lexicographers. There are about 1 300 entries, presented in Hausa alphabetical order with extensive English explanations describing the plants and providing information about their use. For example, Hausa *garafuni* (*Momordica balsamina*) is described as "'Balsam Apple'; a twiner

of the cucumber family, with yellow flowers and orange-yellow tubercled fruits; used medicinally and in soup, and as a soap forming a viscid solution in water". The dictionary is accompanied by a ten-page index to *genera* according to the Latin terms and their common English names. Other studies providing Hausa terminology for trees and plants found in the area include Vischer (1936) for Nigeria, and, in more modern times, Lévy-Luxereau (1972) for Niger.

The Hausa-English dictionary of Ames and King (1972) provides a rich lexical study of Hausa musical terms presented in semantic and functional categories, namely, instruments (divided into types), parts of instruments, performers, patrons, musical occasions, and musical performances. There is an alphabetical Hausa index and an English-Hausa index. The book ends with ten pages of high quality photographs. It does not, however, contain pictures or drawings to illustrate the individual items being described.

The brief (18-page) study by Kafin-Hausa (1977) presents terms describing horses — these traditionally being of great importance in Hausa (upper class) culture — and associated items. Included are names for kinds and colors of horses, equine diseases, horse trappings, and different gaits. It is notable in that it is a monolingual Hausa-Hausa work directed primarily at a Hausa audience, although the information included is also of value for the international scholar.

Baldi (1989) is a presentation of almost 900 basic and cultural vocabulary items, numbered and organized into 21 semantic categories, including people, religion, education, animals, countries of the world, etc. Included are many modern concepts matched by the semantic extension of traditional Hausa vocabulary as well as by recently accepted new terminology. The headwords are given in five (!) languages, namely Italian, English, French, German, and Spanish, followed by Hausa equivalents written in standard orthography as well as accurate phonological transcription. The body of the glossary is followed by alphabetical indexes in the five European languages as well as Hausa that are cross-referenced to the numbered entries. There is also a brief index of technical Latin names for flora and fauna, a useful addition that is seldom provided in other works. This is a high-quality glossary that, unfortunately, has been generally neglected by Hausaists, perhaps because it is written in Italian.

Bross and Tela (1996) is a Hausa-English lexical study of crafts in the broadest sense of the term, i.e., it includes terms for objects, tools, verbs describing processes, as well as basic plants and materials used in preparing crafts. The volume is amply illustrated throughout with line drawings. The book is said to be a dialectal documentation — the basic research was carried out in areas other than that of standard Kano Hausa — but most of the entries are the same as one would find in standard Hausa.

Two dictionaries that truly focus on dialects other than on standard Kano Hausa are those of Matsushita (1991, 1993-1994). The first work is written in Japanese. It is a study of the dialect spoken in Sokoto in the western part of Hausaland. The entries are organized by topical and grammatical categories, e.g., body parts, animals, natural features, adjectives, verbs, etc. The headword

(which is numbered) is provided in English and Japanese followed by the Hausa translation. These headwords are then accompanied by Hausa phrases and sentences with Japanese, but not English, translations. The book concludes with indexes containing all the Hausa and English words in alphabetical order cross-referenced to the headword numbers.

Matsushita's two-volume *Hausa Dialect Vocabulary* (1993-1994) is a computerized recompilation of material found in Bargery (1934). In Bargery, distinct dialectal variants were noted for specific items, both involving different words and variant pronunciations. There are some twenty different dialect labels covering geographical regions in Nigeria such as Daura, Sokoto, Zaria, etc., as well as features of Hausa spoken in Ghana. Dialectal forms were scattered throughout Bargery, wherever they might occur in their normal alphabetical place. Matsushita extracted all these labelled words from Bargery and collated them into separate dialect sections, in Hausa-English alphabetical order. Thus items marked (D.) are found in the Daura dialect section (pp. 15-58), those marked (S.) in the Sokoto dialect section (pp. 639-757), and so forth. Although the materials themselves are identical to the original Bargery entries, Matsushita's presentation makes lexical information about each dialect immediately accessible. It is a creative example of what computerization can accomplish.

The third work by Matsushita (1995-1996), also based on Bargery, is a reverse dictionary of Hausa, although a strange one. All the head entries in Bargery are presented in graphemically reverse order and alphabetized according to this order, followed by the word in its normal spelling. For example, the headword ABAS (p. 4) = Hausa **saba** 'be accustomed to', EFAS (p. 361) = Hausa **safe** 'in the morning', NITALAT (p. 645) = Hausa **talatin** 'thirty', and UBUD (p. 736) = Hausa **dubu** 'thousand'. All the entries are a maximum of one line long, so that if the Bargery entry is longer — as it inevitably is — it will only be partially reproduced and is often truncated midsentence or midword. The book contains no introduction and thus there is no explanation regarding how to use the work nor its supposed purpose. Nevertheless, for the linguist who might be interested in knowing something about the final rather than the initial phoneme(s) of words, the work can be useful. For example, it is commonly stated that Hausa words typically end in a vowel, which is indeed the norm. (In the reverse dictionary, words beginning (i.e., ending) in /a/ occupy some 324 pages.) It is thus surprising to discover that there are more than nine full pages of Hausa words ending in /s/, an inventory that no one could have been aware of before this reverse dictionary.

McIntyre and Meyer-Bahlburg's (1991) work falls somewhere between a specialized dictionary and a general dictionary. It is specialized in the sense that it is drawn almost exclusively from a specific corpus, mostly spoken materials from Hausa language radio programs broadcast by the Deutsche Welle (German World Service). However, the choice of words included does not constitute a coherent semantic set but rather represents a range of ordinary Hausa words randomly chosen, depending on the accident of what happened to occur in the texts consulted. There is thus no coherence as to what is included and

what is not. For example, the work includes *ten*, but not *seven*, *white*, but not *black*, *elephant*, but not *lion*, and so on. The main entries are trilingual, in Hausa, English, and German. The Hausa headwords are given in normal orthography followed by phonologically accurate transcriptions indicating tone, vowel length, and the two /R/s. Where the dictionary is excellent is in the examples of contemporary natural language and idiomatic usage, especially regarding modern political matters and world affairs. Part I, the Hausa-English-German section covers some 144 pages. Part II, which is of equal length, consists of an English-Hausa index and a German-Hausa index.

Dikko and Maccido (1991) is a specialized Hausa-Hausa dictionary of idioms, compounds, and closely-bound phrases, e.g., **fuska biyu** (*lit.* two faces) 'hypocrisy'; **farin jini** (*lit.* white blood) 'popularity'; **sha kai** (*lit.* drink head) 'to pester.' This excellent work contains some 4 000 entries written in standard orthography and alphabetized by the first word in the compound expression. The literal meaning of the expressions is not provided, presumably because the book is designed for Hausa speakers who would already know the meaning of the constituent parts. But it is also a valuable resource for advanced language learners, who could benefit from a well-ordered list of idioms with their explanations.

Many of the Hausa reference dictionaries discussed earlier differ in the amount of information they offer regarding the etymologies of words. There are only two lexicographic works devoted exclusively to such a study. Baldi (1988) is a specialized etymological work comparing approximately 600 loanwords from Arabic found both in Hausa and in Swahili. The organization is by semantic category, e.g., nature, man as a physical being, man as a spiritual being, social organization, natural laws, plus the catch-all category, interjections and particles. Each Arabic word, given in Roman transliteration, is accompanied by the loanword in Hausa and Swahili, with indications of semantic divergence where appropriate. The main section is followed by alphabetical Hausa and Swahili indexes.

Skinner (1996) is the first full etymological dictionary of Hausa, the words having been selected on the basis of frequency of use. Loanwords, whether from English, French, Arabic, or some other African language, are clearly identified as such. In the case of Arabic, the exact form of the source word is usually provided. For nonloanwords, Skinner normally does not attempt to postulate a precise etymology; rather, he provides a relevant proto-Chadic reconstruction when available; otherwise he presents a large array of comparative citations. The hope is that subsequent scholars will be able to draw on these citations in order to determine the true cognate forms and the true etymological connections.

5. Pedagogical works

As early as the 19th century, wordlists and grammatical notes on Hausa were prepared as practical aids for Europeans in West Africa wanting to learn and

use the language, these being primarily missionaries, explorers, or government agents and officials. In the 20th century, the range of learners expanded to include businessmen, researchers, university language students in the United States, Europe, and elsewhere, as well as international service and voluntary organizations such as the U.S. Peace Corps.

From the start, the target audiences of the early pedagogical works were primarily English, German, and French speakers, and this remains generally true up to the present. Among early efforts were the works of Delafosse (1901), Seidel (1906), and Haywood (1907), who included vocabularies of greater or lesser extent to accompany grammatical lessons intended for language learners. In Seidel, for example, separate grammatical sketches appear in German, French, and English, followed by a combined Hausa-English-German-French vocabulary of some 150 pages organized in semantic sets. By contrast, Miller (1907) is a small stand-alone glossary with both Hausa-English and English-Hausa sections.

Practical (as well as other) dictionaries after the Second World War were able to draw on the tremendous advances in Hausa lexicographic scholarship represented by the major reference dictionaries of Bargery and Abraham, as described earlier. In this section, the discussion of pedagogical works is organized in terms of the language of the target learners, beginning with English.

The Hausa-English dictionary of Skinner (1959) is a small dictionary of 70 pages designed to be used as a handy reference tool. (The revised second edition of 1968 is truly pocket-sized in a 9 x 12-cm format.) It is written in standard Hausa orthography with many — though not all — words, especially homographs, having phonetic transcriptions indicating vowel length, but not tone. All headwords are identified for part of speech, and include essential information such as noun gender, plural formation, and verb transitivity/intransitivity.

Newman and Newman (1977) is a moderate-sized Hausa-English learners'/users' dictionary of 150 pages containing approximately 6 000 head entries. Along with basic core vocabulary are incorporated many loanwords and neologisms that had entered the language over the previous half century. Linguistically it is totally up-to-date, marking the distinction between the two /R/s, vowel length, and tones for all headwords. This is the first dictionary to mark the length of word-final vowels accurately, especially those with low tone. Both Bargery and Abraham consistently marked these low tone vowels as if they were short, as the length distinction in that environment can be phonetically quite reduced. A striking departure from all previous — and, as it has turned out, all subsequent — dictionaries was the decision to mark the short vowel (with a cedilla, e.g., **ḡ**, **ḣ**, **ḥ̇**, **ḥ̣̇**) leaving the long vowels unmarked, as compared with the Bargery/Abraham tradition of only marking long vowels (with a macron). The decision was linguistically sound in that long vowels in Hausa are statistically more frequent than short. Almost all common nouns, for example, end in long vowels. Furthermore a number of derivational processes exist that are marked by vowel shortening, e.g., **hannū** 'hand'; but **hannu**

'in/on the hand', **gàjērē** 'short', but **Gàjēre** 'Shorty' (nickname). However, the tradition in Hausa scholarship was too well established to be overcome, and this innovation was never accepted.

Published in Nigeria, this dictionary was designed primarily for non-Hausa speakers, not only foreigners but more especially Nigerians with other first languages who want to learn Hausa, either formally in public schools or in practical settings. In the absence of an appropriate monolingual dictionary, native Hausa speakers have also used the dictionary as a guide to orthographic standardization and grammatical usage. Since first being published, the book has gone through some five printings.

Awde (1987) is, as its title indicates, an English-Hausa glossary divided into 43 mostly semantic categories (e.g., animal world; emotions, vices and virtues; herbs and spices; occupations; weather and elements; etc.), each with a selection of verbs appropriate to the category. There is also a small section of useful expressions. The English headwords are translated into several Hausa equivalents, with accurate phonological transcription, grammatical information on gender, plurals, and verb classes or "grades", but virtually no examples to illustrate the different equivalents.

Awde (1996) is a handy modern Hausa-English English-Hausa dictionary of about 18 000 head entries of useful every-day vocabulary and short phrases, designed for both the non-Hausa person learning Hausa and the Hausa speaker learning English. Headwords or word compounds, fully marked for pronunciation, are matched with direct translation or short-phrase equivalents in the other language. Many entries contain numbered items and short examples, but this enumeration does not necessarily reflect semantic categorization. This practice is least helpful for the learner in the case of homophones, which are gathered together under a single entry rather than appearing as separate lexemes. For example, the Hausa homophone/homograph **sâ** is actually three separate words; it is a pronoun 'they', a noun 'bull', or a series of loosely related verbs 'cause, put (on), wear, fix, appoint'. On the English side, words like **fire** and **watch** have separate meanings as nouns or verbs, but they are not accordingly differentiated. In other respects, this dictionary is admirable for its choice of words, readability, and useful appendices.

R.M. Newman (1990) is the first large-scale dictionary (over 300 pages), going from a Western language to Hausa, that is designed to guide the learner in acquiring expressive, productive language competence. For each English entry, the aim is to provide the means for expressing the same idea in Hausa. The Hausa entries are not intended as literal translations but rather as conceptual, pragmatic equivalents. The book is rich in collocations and illustrative sentences. For example, under English 'tears', one finds the Hausa word **hawàyē** with the illustrative example, 'the tears ran down his cheeks', **hawàyē sun zubō masà** (*lit.* tears poured on him). But also, 'burst into tears', **fashè dà kūkā** (*lit.* break with crying) and 'be on the verge of tears', **yi kwàllā** (*lit.* do tearfulness). Where different loanwords have entered Hausa in Nigeria (anglo-

phone) versus Niger (francophone), both variants are presented, e.g., *bucket* **bōkiti** or **sō** (French *seau*). There are a number of important appendices, including a list of Hausa names, male and female, phonologically transcribed; the Hausa pronunciation of cities and states in Nigeria, Niger, and some international locations; the Hausa equivalents for countries outside of Nigeria and Niger; and the names for governmental organizations and international agencies. A brief grammatical sketch is also provided.

The dictionary provides essential grammatical information about Hausa equivalents that the learner should know to properly speak the language. Nouns are marked for grammatical gender and at least one common plural form. Verbs are overtly classified according to the Parsonian "grade" system (see Introduction above). Verbal nouns, namely, nominalized forms that are required when a verb is used in the progressive/continuous tense, are also included. Because of the richness of its exemplification, and the fullness of grammatical information provided, this work represents a major advance over previous English-Hausa dictionaries, whether viewed as scholarly or pedagogical.

Unlike the pedagogical works just described, which are intended primarily for non-Hausa speakers interested in learning Hausa, NORLA (1957), Skinner (1965), and Baki (1997), all published in Nigeria, are English-Hausa practical dictionaries designed to help Hausa speakers learn English. NORLA is a 335-page "manual of conversation" organized in dictionary format. The pronunciation of each English headword is provided, accompanied by a brief Hausa equivalent or equivalents written in standard orthography. The word is then amply illustrated with full English sentences followed by Hausa translations, generally going from literal to more metaphorical uses. For example, the entry for 'to die', Hausa **rasu**, **mutu**, contains the English sentences such as 'His son has died recently', **Kwanan nan d'ansa ya rasu**; 'The flowers have all died', **Duk furen sun bushe** (*lit.* All the flowers have dried up), and 'The fire died down', **Gobarar ta lafa** (*lit.* The conflagration has subsided). No information regarding part of speech or conjugational forms is given either for English or for Hausa.

Skinner (1965), a revised and illustrated edition of which appeared in 1978, is similar in many ways to the NORLA volume. It differs in that the example sentences are greatly reduced so that it appears more like a dictionary rather than a conversation book. As with NORLA, pronunciation is provided for the English headwords, but otherwise lexicogrammatical information is not provided, apart from separate headwords for irregular past-tense verb forms such as 'bought', 'drove', 'went', cross-referenced to their infinitival forms. The book does include a brief introduction written in Hausa explaining the basics of English phonology and morphology.

Baki (1997) is a trilingual English-Arabic-Hausa dictionary of 235 pages, produced in Nigeria. The English headwords are accompanied by brief Arabic and Hausa translation equivalents written in the standard orthographies of those languages. No grammatical information or illustrative examples are

included. According to the author, it is intended for Hausa beginners learning English and/or Arabic. However, in the absence of any other Arabic-Hausa dictionary, this could serve as a pocket guide for Arabic speakers learning Hausa.

Among practical dictionaries whose target users are German speakers, there are two modern ones. Herms (1987) is an attractively produced Hausa-German dictionary, closely modeled on Newman and Newman (1977), but containing more headwords (approximately 8 500 as opposed to 6 000). On the whole, the dictionaries of Herms and of Newman and Newman are of comparable quality, except that Herms, unfortunately, contains a number of small errors in the marking of tone and vowel length in the Hausa head entries.

McIntyre and Meyer-Bahlburg (1999), a more recent work, is a small German-Hausa vocabulary of 68 pages (plus appendices) aimed at German students. Although limited in size, and consisting mostly of short one- or two-word translations, the work is actually quite informative in that it includes essential grammatical information (e.g., grammatical gender, associated noun plurals, etc.), as well as common expressions and collocations. In addition, the Hausa examples are transcribed carefully and reliably.

French learners of Hausa now have a good moderate-sized dictionary with the well-produced French-Hausa work of Caron and Amfani (1997). A work similar in objective to R.M. Newman but with somewhat fewer headwords, approximately 3 700, it is intended both for French speakers learning Hausa and for Hausa speakers learning French. As such, it includes grammatical information and phonetic transcription for all the French headwords as well as the Hausa translations. Interestingly, the Hausa presented is that of the standard dialect of Nigeria, and the French loanwords used in Niger Hausa supplied by Newman are totally ignored. The book includes an extensive Hausa-French index of some 85 pages and a long 70-page introduction containing background information on Hausa classification and a grammatical sketch.

The trilingual Russian-Hausa-Yoruba dictionary (Laptukhin 1987) of 1 200 entries is said to be expressly designed for Hausa and Yoruba speakers learning spoken Russian, with an introduction written in Russian and English. Phonetic transcriptions are therefore provided for each headword, but these are indicated in Cyrillic rather than in a more universal notation such as IPA (International Phonetic Alphabet). Declensions of Russian words, as well as verb perfective/imperfective stems, are listed after the head entry for the Hausa/Yoruba learner to consult. The entries themselves contain good examples and appear quite conversational. Sometimes one finds an overly literal Hausa translation of a Russian word, e.g., the word for *blackboard* is rendered literally as **bafin allo** 'black board' when more normal usage would simply be **allo** ('writing board'). Many photographs accompany the entries, but they are mainly of government buildings and monuments in Moscow!

The final pedagogical dictionary discussed here is noteworthy in that it is the first monolingual Hausa dictionary ever published. Produced in Nigeria by an educator who trains teachers of Hausa, Garba (1990) is a modest first effort of some 2 200 headwords, transcribed in standard orthography but with handwritten tones (low tone only) marked for each headword. This small dictionary is meant for upper primary school students and is admirable in its simplicity and design. Each word is labeled with its part-of-speech classification, these metalinguistic terms being explained in simple language in the short introduction. Nouns, which constitute the largest part of the list, are identified as masculine or feminine, with the most common plural forms given. Words are either defined by a synonym or a simple explanatory definition. Polysemic words have each meaning numbered and explained, although there are no examples of usage. A number of line drawings accompany many nouns. The book is a good first dictionary for the young native Hausa speaker, but could also be very useful to the nonnative learner looking for good conversational explanations in Hausa.

6. Terminological works

As a major world language, Hausa has been undergoing rapid modernization and lexical expansion to cope with the changing world. For many years, regular international broadcasts have been produced in Hausa by the British Broadcasting Corporation, the Voice of America, Deutsche Welle, and Radio Beijing, and more recently by other nations in Africa. The need for Hausa standardized terminology to express modern concepts, sociopolitical structures, and material goods is urgent, and has been recognized for years. Thus the fourth major type of Hausa lexicographic work discussed here consists of those terminological vocabularies and glossaries that have been published in Nigeria and Niger. They result from projects promoted by language planners and educators whose goals are to modernize and expand the Hausa lexicon to meet the needs of the modern world. Some of these works were individual efforts; others have been carried out or commissioned by governmental or pseudogovernmental bodies.

In Nigeria in the 1950s and 1960s, the Hausa Language Board, which was set up by the Northern Region House of Assembly in 1955 (see Kirk-Greene 1964), was the main organization focusing on language development and standardization. Pamphlets that they produced include *Alphabetical List of Words Imported into Hausa* (ca. 1958), *Glossaries of Technical Terms (Bayanin Kalmomin da a ke Anfani da su a Ma'aikatu Iri Iri)* (Explanations of Words Used in Various Work Places) (ca. 1959), and *Vocabulary of Modern Political and Administrative Terms* (1964). A pamphlet of a similar nature focusing on governmental terms (from English to Hausa) was prepared by the then Head of the Department of Local Government at the Institute of Administration (Campbell ca. 1958).

In more recent years, the responsibility for all indigenous Nigerian language matters has fallen to the Nigerian Educational Research and Development Council (NERDC). An important product of NERDC's efforts was the

publication of *A Glossary of English-Hausa Technical Terms in Language, Literature and Methodology* (Muhammed 1990). The first half of the volume is an English to Hausa vocabulary organized by topic, e.g., linguistics, literary analysis, etc. The second half consists of a straight alphabetical Hausa to English vocabulary in which tone and vowel length are indicated for all the Hausa entries. The *Quadrilingual Glossary of Legislative Terms* (NERDC 1991) is an ambitious book aimed at expanding the lexicons of Nigeria's three official national languages, Hausa, Yoruba, and Igbo. It is a work of almost 300 pages containing over 10 000 head entries, and is semantically much broader than the title would suggest. In addition to legislative and governmental terms, it also includes many words and expressions having to do with economy, education, health and medicine, sports, and the military. Each page consists of four columns for each language with the words being arranged in English alphabetical order.

In both of the NERDC books, the attempt is not to explain the English words or expressions, but rather to provide direct translation equivalents. In some cases, the Hausa equivalents represent established usage. But in many others, they have been created by the compilers themselves, working collaboratively with other native-speaker educators and linguists. Only time will tell whether the proposed neologisms (in Hausa, Yoruba, Igbo) will or will not be adopted and become part of these functioning languages.

Another terminological work is Jinju (1990), a trilingual Hausa-English-French presentation of Hausa vocabulary in the areas of science and technology. Sakkwato (1993) is a similar work (but without French included) focusing on the vocabulary of geography.

In Niger, the main government bodies promoting lexicographic work have been the Centre d'Études Linguistiques et Historiques par Tradition Orale (CELHTO) and the Institut National de Documentation de Recherche et d'Animation Pédagogiques. Mahamane (1982) is a listing of French words relating to scientific matters with simple one-word or short-phrase Hausa counterparts. Mijinguini (1983a) gives Hausa equivalents for French words describing history and geography, including flora and fauna and agricultural produce. Although the Hausa entries are usually simple translations, alternative forms and brief explanations are sometimes provided, thereby making this work particularly useful. Mijinguini (1983b) provides Hausa equivalents of French words concerning language and linguistics, again presented in straightforward French-Hausa alphabetical order. In these three vocabularies, the Hausa entries are given in the orthography that is now standard throughout Hausaland, both in Nigeria and in Niger, but the forms of words reflect the dialectal characteristics of northwest Hausa, e.g., 'plain (land)' **hwangali**, cf. Kano Hausa **fangali**; 'telephone' **talho**, cf. Kano Hausa **tarho**; 'hoe' **kalmi**, cf. Kano Hausa **fatanya**.

7. Conclusion

A number of themes recur throughout this description of the historical development of Hausa lexicography over the past two centuries: the increasingly

accurate linguistic analysis of the language, especially its phonology and morphology; the increasing naturalness of definitions and the progressive inclusion of idiomatic usage and modern terminology; and the ever greater participation of native Hausa speakers as compilers of these works. Hausa has been one of the most extensively studied languages in Sub-Saharan Africa. This comprehensive guide to its rich lexicographic tradition attests to the vitality of Hausa and to its place as one of Africa's most important and influential languages.

Bibliography

A. Hausa Dictionaries, Vocabularies, and Glossaries

- Abraham, R.C., and Mai Kano.** 1949. *Dictionary of the Hausa Language*. London: Crown Agents for the Colonies. (2nd ed. with new introduction and second author's name omitted, London: University of London Press, 1962.)
- Ames, David W., and Anthony V. King.** 1972. *Glossary of Hausa Music and its Social Contexts*. Evanston: Northwestern University Press.
- Awde, Nicholas.** 1996. *Hausa-English English-Hausa Dictionary*. (Hippocrene Practical Dictionary.) New York: Hippocrene Books.
- Awde, Nicholas, with Sa'idu Babura Ahmad and Malam Barau.** 1987. "21st Century" Hausa: An *English-Hausa Classified Wordlist*. London: Centre for African Language Learning.
- Baki, Alhaji Issah Alhassan.** 1997. *Modern Dictionary of English, Arabic and Hausa. / Kamus na Turanci da Larabci da Hausa*. Zaria (Nigeria): Hudahuda Publishing Co.
- Baldi, Sergio.** 1988. *A First Ethnolinguistic Comparison of Arabic Loanwords Common to Hausa and Swahili*. (Supplement 57 to *AION* 48(4).) Naples: Istituto Universitario Orientale.
- Baldi, Sergio.** 1989. *Manuale propedeutico di hausa con glossario e lessico concettuale di base*. Naples: Istituto Universitario Orientale.
- Bargery, G.P.** 1934. *A Hausa-English Dictionary and English-Hausa Vocabulary*. London: Oxford University Press. (Reprint with supplement and new introduction by N. Skinner, Zaria: Ahmadu Bello University Press, 1993.)
- Bross, Michael, and Ahmed Tela, with drawings by A. T. Sati Baba.** 1996. *Dictionary of Hausa Crafts: A Dialectal Documentation*. (Westafrikanische Studien: Frankfurter Beiträge zur Sprach- und Kulturgeschichte, 10.) Cologne: Rüdiger Köppe.
- Campbell, M.J.** ca. 1958. *A Word List of Government and Local Government Terms (English-Hausa)*. Kaduna (Nigeria): Baraka Press.
- Caron, Bernard, and Ahmed H. Amfani.** 1997. *Dictionnaire français-haoussa suivi d'un index haoussa-français*. Paris: Karthala.
- Dalziel, J.M.** 1916. *A Hausa Botanical Vocabulary*. London: T. Fisher Unwin.
- Delafosse, Maurice.** 1901. *Manuel de langue haoussa ou Chrestomatie haoussa, précédé d'un abrégé de grammaire et suivi d'un vocabulaire*. Paris: J. Maisonneuve.
- Dikko, Inuwa, and Usman Maccido.** 1991. *Kamus na Adon Maganar Hausa (Dictionary of Hausa Idioms)*. Zaria (Nigeria): NNPC.
- Fletcher, Ronald S.** 1912. *Hausa Sayings and Folklore, with a Vocabulary of New Words*. London: Oxford University Press.

- Garba, Calvin Y.** 1990. *Kamus na Harshen Hausa* (Dictionary of the Hausa Language). Ibadan (Nigeria): Evans Brothers.
- Harris, Hermann G.** 1908. *Hausa Stories and Riddles with Notes on the Language, etc., and a Concise Hausa Dictionary*. Weston-super-Mare: The Mendip Press.
- Hausa Language Board.** ca. 1958. *Alphabetical List of Words Imported into Hausa*. Kaduna (Nigeria): Baraka Press.
- Hausa Language Board.** ca. 1959. *Glossaries of Technical Terms (Bayanin Kalmomin da a ke Anfani da su a Ma'aikatu Iri Iri* (Explanations of Words Used in Various Work Places)). Kaduna (Nigeria): The Hausa Language Board, Ministry of Education, Northern Region.
- Hausa Language Board.** 1964. *Vocabulary of Modern Political and Administrative Terms*. Kaduna (Nigeria): Baraka Press.
- Haywood, A.H.W.** 1907. *English-Hausa Vocabulary of Words in Everyday Use*. London: Kegan Paul. (2nd ed. 1914.)
- Herms, Irmtraud.** 1987. *Wörterbuch Hausa-Deutsch*. Leipzig: VEB Verlag Enzyklopädie.
- Jinju, Muhammadu Hambali.** 1990. *Garkuwar Hausa da Tafarkin Ci Gaba: Kalmomin Kimiyya da Ilimin Fasaha na Hausa* (Hausa Scientific and Technological Words). Kaduna: Fisbas Media Services Publications.
- Kafin-Hausa, Abdullahi Umar.** 1977. *Sunayen Dawaki da Kayan Hawa* (Terms for Horses and Associated Accoutrements). Kano (Nigeria): CSNL, Bayero University.
- Landeroin, M., and J. Tilho.** 1909. *Dictionnaire haoussa, comprenant haoussa-français et français-haoussa*. Paris: Émile Larose.
- Laptukhin, Viktor V., and Ado Gwadabe Kano.** 1967. *Russko-xausa slovar'*. *Kamus na Rashanci-Hausa* (Russian-Hausa Dictionary). Moscow: Sovetskaja Enciklopedija.
- Laptukhin, Viktor V. et al.** 1987. *Učebnyj russko-xausa-joruba slovar'* (Student's Russian-Hausa-Yoruba Dictionary). Moscow: Russkij Jazyk.
- Le Roux, J.M.** 1886. *Essai de dictionnaire français-haoussa et haoussa-français*. Algiers: Adolphe Jourdan.
- Lévy-Luxereau, Anne.** 1972. *Étude ethno-zoologique du pays hausa en République du Niger*. Paris: Société d'études ethno-zoologiques et ethno-botaniques.
- Mahamane, Issoufou.** 1982. *Lexique scientifique français-hausa*. Niamey: Centre d'Études Linguistiques et Historiques par Tradition Orale.
- Matsushita, Shuji.** 1991. *A Basic Vocabulary of the Sokoto Hausa (Northern Nigeria)*. (Asian and African Lexicons, 22.) Tokyo: Tokyo University of Foreign Studies, Institute for the Study of Languages and Cultures of Asia and Africa.
- Matsushita, Shuji.** 1993-1994. *Bargery Toolbox 1-2 (Based on Rev. G.P. Bargery's A Hausa-English Dictionary): Hausa Dialect Vocabulary*. 2 vols. Tokyo: Tokyo University of Foreign Studies, Institute for the Study of Languages and Cultures of Asia and Africa.
- Matsushita, Shuji.** 1995-1996. *Bargery Toolbox 3-4 (Based on Rev. G.P. Bargery's A Hausa-English Dictionary): Hausa / asuaH: Back to Front, Volume 1 (A-E); Volume 2. (F-Z)*. Tokyo: Tokyo University of Foreign Studies, Institute for the Study of Languages and Cultures of Asia and Africa.
- McIntyre, J.A., and Hilke Meyer-Bahlburg, assisted by Ahmed Tijani Lawal.** 1991. *Hausa in the Media: A Lexical Guide*. Hamburg: Buske.
- McIntyre, J.A., and Hilke Meyer-Bahlburg.** 1999. *Arbeitsvokabular Deutsch-Hausa*. (Arbeitsmaterialien zur Afrikanistik, 1.) Hamburg: LIT.

- Mijinguini, Abdou.** 1983a. *Petit lexique français-hausa d'histoire et de géographie*. (Études et Documents, 165.) Niamey (Niger): Institut National de Documentation de Recherche et d'Animation Pédagogiques.
- Mijinguini, Abdou.** 1983b. *Vocabulaire technique des sciences du langage (français-hausa)*. Niamey: CELHTO.
- Mijinguini, Abdou.** 1994. *Karamin Kamus na Hausa zuwa Faransanci* (Concise Dictionary of Hausa and French). Niamey (Niger): SP-CNRE/PS-UNESCO.
- Miller, Walter R.** 1907. *Hausa and English Vocabulary*. Lagos: C.M.S. Bookshop. (2nd ed. 1930.)
- Mischlich, Adam.** 1906. *Wörterbuch der Hausasprache*. (Lehrbücher des Seminars für Orientalische Sprachen, 20.) Berlin: G. Dietrich Reimer.
- Muhammed, D. (Ed.)** 1990. *Hausa Metalanguage. Kamus na Keɓaɓɓun Kalmomi*. Vol 1: *A Glossary of English-Hausa Technical Terms in Language, Literature and Methodology*. Ibadan (Nigeria): University Press.
- NERDC.** 1991. *Quadrilingual Glossary of Legislative Terms (English, Hausa, Igbo, Yoruba)*. Lagos: Nigerian Educational Research and Development Council.
- Newman, Paul, and Roxana Ma Newman.** 1977. *Modern Hausa-English Dictionary (Sabon Kamus na Hausa zuwa Turanci)*. Ibadan and Zaria (Nigeria): University Press.
- Newman, Roxana Ma.** 1990. *An English-Hausa Dictionary*. New Haven: Yale University Press. (2nd ed., Ikeja (Nigeria): Longman (1997).)
- NORLA.** 1957. *Hanyar Tafi da Turanci: A Dictionary of English Conversation for Hausa Students*. Zaria (Nigeria): NORLA, London: Longmans, Green and Co.
- Olderogge, Dmitrii A.** 1963. *Xausa-russkij slovar'* (Hausa-Russian Dictionary). Moscow: GIS.
- Parsons, Allan C.** 1915. *A Hausa Phrase Book with Medical and Scientific Vocabularies*. Oxford: Humphrey Milford, OUP.
- Robinson, Charles H.** 1899/1900. *Dictionary of the Hausa Language*. Cambridge: Cambridge University Press. Vol. 1 Hausa-English (1899); Vol. 2 English-Hausa (1900). (2nd ed. 1906, with additions on interleaved pages, numbered 1a, 2a, etc., plus list of plants with Hausa names provided by J. M. Dalziel (added at end, numbered 1-8); 3rd ed., 2 vols., 1913; 4th ed., 2 vols., 1925 (reprinted by Gregg International Publishers, 1969).)
- Sakkwato, Bello Adamu.** 1993. *Kamus na Jugorafiyya* (Geography Dictionary). Sokoto: The Author.
- Schön, James Frederick.** 1843. *Vocabulary of the Hausa Language. Phrases, and Specimens of Translations, to which are prefixed, the Grammatical Elements of the Hausa Language*. London: Church Missionary Society.
- Schön, James Frederick.** 1876. *Dictionary of the Hausa Language, with appendixes of Hausa Literature*. London: Church Missionary House.
- Seidel, August.** 1906. *Die Haussasprache. / La langue haoussa. / The Hausa Language*. Heidelberg: Julius Groos.
- Skinner, Neil.** 1959. *Hausa-English Pocket Dictionary*. Zaria: Longmans, Green and Co. (2nd ed. 1968).
- Skinner, Neil.** 1965. *Kamus na Turanci da Hausa. Babban Ja-Gora ga Turanci* (English-Hausa Dictionary. A Major Guide to English). Zaria (Nigeria): Gaskiya Corporation for the Hausa Language Board. (2nd ed. 1970, Zaria: NNPC; illustrated ed. 1978.)
- Skinner, Neil.** 1985. *Hausa Lexical Expansion since 1930: A Supplement to Bargery's Dictionary*. Madison: African Studies Program, University of Wisconsin.

- Skinner, Neil.** 1996. *Hausa Comparative Dictionary*. (Westafrikanische Studien: Frankfurter Beiträge zur Sprach- und Kulturgeschichte, 11.) Cologne: Rüdiger Köppe.
- Taylor, F.W.** 1927. *A Fulani-Hausa Vocabulary*. Oxford: Clarendon Press.
- Vischer, Hanns.** 1936. *Vocabulary of Nigerian Names of Trees, Shrubs and Herbs*. Lagos: Government Printer.

B. Other works cited

- Abraham, R.C.** 1934. *The Principles of Hausa*. Kaduna (Nigeria): Government Printer.
- Abraham, R.C.** 1959. *The Language of the Hausa People*. London: University of London Press.
- Abu-Manga, Al-Amin.** 1999. *Hausa in the Sudan: Process of Adaptation to Arabic*. Cologne: Rüdiger Köppe.
- Baldi, Sergio.** 1977. *Systematic Hausa Bibliography*. (Istituto Italo-Africano. Collana di Studi Africani 3.) Rome: Tip. Pioda.
- Barth, Heinrich.** 1862-1866. *Collection of Vocabularies of Central-African Languages. / Sammlung und Bearbeitung zentralafrikanischer Vokabularien*. Gotha: Justus Perthes.
- Green, M.M.** 1967. G.P. Bargery (1876-1966). *Journal of West African Languages* 4(2): 59-60.
- Hair, P.E.H.** 1967. *The Early Study of Nigerian Languages: Essays and Bibliographies*. (West African Language Monographs, 7.) Cambridge: Cambridge University Press. (Reprint with a new introduction by Paul Newman, Aldershot: Gregg, 1994.)
- Jaggard, Philip J.** 1992. Roy Clive Abraham: A Biographical Profile and List of Writings. Jaggard, P.J. (Ed.). 1992. *Papers in Honour of R. C. Abraham (1890-1963)*: 1-4. (African Languages and Cultures, Supplement 1.) London: SOAS.
- Kirk-Greene, A.H.M.** 1964. The Hausa Language Board. *Afrika und Übersee* 47: 187-203.
- Kirk-Greene, Anthony, and Paul Newman.** 1971. *West African Travels and Adventures: Two Autobiographical Narratives from Northern Nigeria*. New Haven/London: Yale University Press.
- Koelle, Sigismund W.** 1854. *Polyglotta Africana*. London: Church Missionary Society. (Reprint with an historical introduction by P. E. H. Hair, Freetown: Fourah Bay College, 1963.)
- Mischlich, Adam.** 1902. *Lehrbuch der hausanischen Sprache*. (Archiv für das Studium der deutscher Kolonialsprachen, 1.) Berlin: Dietrich Reimer.
- Newman, Paul.** 1980a. *The Classification of Chadic within Afroasiatic*. Leiden: Universitaire Pers.
- Newman, Paul.** 1980b. The Two R's in Hausa. *African Language Studies* 17: 77-87.
- Newman, Paul.** 1996. *Hausa and the Chadic Language Family: A Bibliography*. (African Linguistic Bibliographies, 6.) Cologne: Rüdiger Köppe.
- Newman, Paul.** 2000. *The Hausa Language: An Encyclopedic Reference Grammar*. New Haven: Yale University Press.
- Newman, Paul, and Roxana Ma Newman.** 1991. Lexicography of the Chadic Languages. Hausmann, F. J. et al. (Eds.). 1991. *Wörterbücher/Dictionaries/Dictionnaires, vol. 3*: 2457-2460. Berlin: Walter de Gruyter.
- Newman, Roxana Ma.** 1974. Dictionaries of the Hausa Language. *Harsunan Nijeriya* 4: 1-25.
- Newman, Roxana Ma.** 1992. Lexicographic Method in R. C. Abraham's *Hausa Dictionary*. Jaggard, P.J. (Ed.). 1992. *Papers in Honour of R. C. Abraham (1890-1963)*: 79-97. (African Languages and Cultures, Supplement 1.) London: SOAS.
- Parsons, F.W.** 1960. The Verbal System in Hausa. *Afrika und Übersee* 44: 1-36.
- Taylor, F.W.** 1923. *A Practical Hausa Grammar*. Oxford: Clarendon Press. (2nd ed. 1959.)

Tweetalige Frasewoordeboek/Bilingual Phrase Dictionary: 'n Fokus op funksies en enkele probleemareas

Anna Nel Otto, *Departement Afrikaans, Universiteit Vista,
Port Elizabeth, Republiek van Suid-Afrika (OTTO-AN@pelican.vista.ac.za)*

Opsomming: Hierdie woordeboek deur P.A. Joubert sal bespreek word aan die hand van die volgende aspekte: die volledigheid van die woordeboek, met inagneming van die teikengebruiker en die beginsels van outentisiteit en tipiesheid, die verwerking van woordeboekinligting om gebruikers se potensiële vrae te beantwoord, die mate waarin inligting vinnig opgespoor en maklik ontsluit kan word, en die formaat en redigering van die woordeboek.

Daar word geargumenteer dat dié woordeboek redelik volledig is, maar weens die gebrek aan 'n sistematiese teoretiese onderbou tog enkele leemtes het. Tipiesheid word soms ingeboet deur die opname van baie formele, weinig gebruiklike of verouderde frases, wat ook nie as sodanig geëtiketteer is nie. Inligting kan oor die algemeen maklik opgespoor word, maar vollediger aanduidings in die gids tot die woordeboek oor die plasing van multileksikale lemmas sal opspoorbaarheid verder verhoog. Die formaat en redigering van die woordeboek is goed.

Frasewoordeboeke vervul 'n belangrike funksie deurdat daar nie weens 'n hoë inligtingsdigtheid, soos in algemene tweetalige woordeboeke, na inligting oor frases gesoek hoef te word nie. Van 'n aktiewe frasewoordeboek sou 'n mens egter soms meer geleidingsverwag in die vorm van veral voorbeeldsinne en die konsekwente gebruik van etikette.

Sleutelwoorde: TWEETALIGE FRASEWOORDEBOEK, VOLLEDIGHEID, OUTENTISITEIT, TIPIESHEID, OPSPOORBAARHEID, GEBRUIKERSVRIENDELIKHEID, VOORBEELDSINNE, ETIKETTE, REDIGERING, MULTILEKSIKALE LEMMA, KOLLOKASIE, FRASE

Abstract: *Tweetalige Frasewoordeboek/Bilingual Phrase Dictionary: A Focus on Functions and a Few Problem Areas.* This dictionary by P.A. Joubert will be discussed with reference to the following aspects: the completeness of the dictionary, with due allowance for the target users and the principles of authenticity and representativeness, the processing of dictionary information in order to answer the potential questions of users, the degree to which information can be accessed quickly and easily, and the format and editing of the dictionary.

It will be argued that though this dictionary is fairly complete, it has shortcomings because of a lack of systematic-theoretical grounding. Representativeness is at times sacrificed by including phrases which are very formal, used infrequently or being obsolete, and which are also not labelled as such. In general information can easily be retrieved but in instances where multilexical lemmata are involved, retrievability could be enhanced by more detailed indications of placement in the guide to the dictionary. The format and editing of the dictionary is good.

Other than general bilingual dictionaries with their high information density, phrase dictionaries play an important role in that information about phrases does not need to be sought. One

would however sometimes expect more usage guidance from an active phrase dictionary in the form of especially example sentences and a consistent use of labels.

Keywords: BILINGUAL PHRASE DICTIONARY, COMPLETENESS, AUTHENTICITY, REPRESENTATIVENESS, RETRIEVABILITY, USER-FRIENDLINESS, EXAMPLE SENTENCES, LABELS, EDITING, MULTILEXICAL LEMMA, COLLOCATION, PHRASE

1. Inleiding

By die beoordeling van enige woordeboek kan daar na minstens vier aspekte gekyk word:

- die volledigheid van die woordeboek, met inagneming van die teiken-gebruiker en die beginsels van outentisiteit en tipiesheid,
- die verwerking van die woordeboekinligting om gebruikers se potensiele vrae te beantwoord,
- die mate waarin inligting vinnig opgespoor en maklik ontsluit kan word, en
- die formaat en finale redigering van die woordeboek.

2.1 Die volledigheid van die woordeboek

Die volledigheid al dan nie van enige woordeboek moet gemeet word aan die tipe woordeboek en die teikengroep vir wie die woordeboek opgestel word. Hierbenewens geld twee algemene beginsels wat vir alle woordeboeke geld, naamlik outentisiteit en tipiesheid. Svensén (1993: 40-42) verduidelik hierdie twee terme soos volg.

Outentisiteit

Baie woordeboeke word deur leksikograwe opgestel wat hul seleksie van die woordeskat slegs baseer op hul eie taalvermoëns. Hierdie metode word as gevaarlik bestempel. Die seleksie van woordeboekmateriaal moet outentiek wees, d.w.s. dit moet slegs linguistiese materiaal bevat wat werklik bestaan. Onafhanklike bronne moet gebruik word om die outentisiteit van linguistiese materiaal te bevestig. Getuienis word ongelukkig meestal uit geskrewe bronne verkry. In hierdie verband sê Svensén die volgende:

"Evidence" normally means "evidence from written sources". Unfortunately, the spoken language makes up by far the larger part of all linguistic production, whereas by far the larger part of all evidenced linguistic production is in writing. The requirement of authenticity is therefore unfavourable to several of those forms of the language which the user in many cases might want information about, such as everyday colloquial speech, slang, dialect, and technical jargon.

Verder waarsku Svensén (1993: 41) teen die opname van "spookwoorde", d.i. woorde wat in woordeboeke bestaan, maar nie in die werklikheid nie. Dit word nog moeiliker om lewenseg te wees wanneer grammatiese en semantiese inligting geïllustreer moet word. Wanneer die leksikograaf 'n bepaalde insig by die gebruiker wil tuisbring, is dit egter volgens Svensén (1993: 41) geregverdig om byvoorbeeld poëme te gebruik wat afgelei is van outentieke voorbeelde.

Tipiesheid

Dit is nie net noodsaaklik om outentieke woorde en uitdrukkings in die woordeboek op te neem nie. 'n Verdere vereiste is dat elke woord of uitdrukking gereeld genoeg gebruik moet word in die taal. Svensén (1993: 42) verduidelik tipiesheid soos volg:

Representativeness ... means that the word or expression occurs with a certain frequency in the general use of the language. This condition of representativeness primarily concerns active dictionaries.

In die bespreking oor die tipe woordeboek, sy doel en teikengebruikers, en die gedeelte oor die verwerking van die inligting hieronder, sal aangetoon word dat die *Tweetalige Frasewoordeboek/Bilingual Phrase Dictionary* (voortaan *TFW/BPD*) verbeter sou kon word deur meer outentieke en tipiese frases in te sluit, aangesien hierdie woordeboek ten doel het om as aktiewe woordeboek praktiese geleidings te verskaf.

2.2 Die tipe woordeboek, sy doel en teikengebruikers

Die eerste vraag wat 'n mens in hierdie verband kan vra, is: wat is 'n frasewoordeboek en aan watter kriteria behoort 'n tweetalige frasewoordeboek gemeet te word? Schemann (1989: 1019 e.v.) wys soos volg op die gebrek aan 'n algemeen geldige antwoord op hierdie vraag:

Mir ist kein einziges phraseologisches Wörterbuch bekannt, das (a) sich auf ein Material beschränkte, welches von irgendeiner Definition des phraseologischen Ausdrucks in seinem Umfang exakt abgesteckt werden könnte; und das (b) in den Definitionen, die (in den Vor- oder auch Nachworten) gegeben werden, systematisch-kohärent wäre.

Volgens Schemann (1989: 1021) is die enigste sinvolle vraag wat 'n mens oor hierdie woordeboek tipe kan stel: oor watter frase gaan dit in hoofsaak? En die antwoord sal ten nouste saamhang met die doel van die woordeboek.

Verder wys Schemann (1989: 1021-1022) op die gevaar van 'n frasewoordeboek wat nie hou by die algemeen geldende kriterium dat 'n frase 'n kontekstueel gebonde item moet wees en nie 'n vrye verbinding nie.

In die *TFW/BPD* word bloot in die voorwoord genoem dat die woord "frase" in die titel van die woordeboek in sy ruimste betekenis gebruik word. In die *HAT* word die woord "frase" gedefinieer as: "groep samehangende woorde wat 'n deel van 'n sin vorm, maar geen volsin nie; sinsnede". In die *TFW/BPD* word verder genoem dat die begrip "frase" 'n verskeidenheid soorte woord-groepe omvat, wat soms duidelikheidshalwe tot volsinne uitgebrei word. Hierdie woorde toon hoe woorde in Afrikaans en in Engels saam ingespan word. Boonop is 'n aantal spreekwoorde opgeneem. Die moontlike gebruike van die woordeboek word verder soos volg aangedui: "Antwoorde word verskaf op vrae soos wat die gepaste werkwoorde, voorsetsels, byvoeglike naamwoorde of bywoorde in bepaalde gevalle is, wat die woordorde is en of 'n lidwoord gebruik word of nie."

En in 'n volgende paragraaf word gesê: "Die woordeboek is bedoel as 'n praktiese gebruiksgids waarin naslaners maklik die woordkombinasies kan vind waarmee hulle hulle in die twee tale kan uitdruk; dit kan hulle ook help om die twee tale se onderskeie maniere om iets te sê nie met mekaar te verwar nie."

Met so 'n pragmatiese uitgangspunt is daar nie fout te vind nie. Hanks (1979: 33) het immers gesê: "When theory comes into lexicography, all too often common sense goes out." Die gevaar bestaan egter dat as daar nie 'n teoretiese onderbou is nie, daar aan konsekwentheid ingeboet kan word, wat daartoe kan lei dat belangrike inligting per ongeluk weggelaat word of dat die gebruiker verwar kan word deur 'n inkonsekwente aanbod.

Hoewel uitgewers gewoonlik beweer dat die toeligtingsgedeelte so kort moontlik moet wees, aangesien die meeste gebruikers dit in elk geval nie lees nie, is dit noodsaaklik dat daar in die toeligtingsgedeelte baie duidelik uitgespel moet word watter tipes frases opgeneem word en wat die kriteria vir opname was, anders sal die gebruikers nie weet of die woordeboek aan hulle behoeftes behoort te voldoen of nie. Dit is nie nodig om die gebruiker met teoretiese kennis te oorlaai nie: 'n voorbeeld of twee van elke tipe sal dadelik aan hom/haar 'n beeld gee van die omvang van die woordeboek.

Bogenoemde probleme sal vervolgens aan die hand van enkele voorbeelde bespreek word.

In die *BBI Dictionary* word kollokasies konsekwent op grond van geïdentifiseerde patrone opgeneem. Dit is ook 'n tipe "frase" wat die *TFW/BPD* insluit, maar anders as in die *BBI* word kollokasies nie op 'n konsekvente wyse opgeneem nie.

In die *BBI* word die positiewe en die negatiewe konsekwent by werkwoorde aangedui. Die woord *aanbeveling* kan byvoorbeeld met *bekragtig* en *verwerp* verbind. Hierdie inligting ontbreek by *aanbeveling* in die *TFW/BPD*. Op 'n soortgelyke wyse ontbreek *toestaan* en *afkeur / afwys / teenstaan* byvoorbeeld by die woord *aansoek*.

In die geheel beskou kan nogtans gesê word dat die woordeboek se makrostruktuur, wat volledigheid betref, aan die meeste gebruikers se verwagtinge behoort te voldoen.

3. Die verwerking van die inligting om gebruikers se potensiële vrae te beantwoord

Voorbeeldsinne en etikette

In die voorwoord word aangedui dat die woordeboek 'n praktiese gebruiksgids is en dit veronderstel dat dit bedoel is om 'n aktiewe woordeboek te wees. In 'n aktiewe woordeboek moet die gebruiksliding so volledig wees dat die gebruiker sal weet hoe om elke frase op 'n gepaste wyse te gebruik. Dit is te betwyfel of die gebruiksliding wat wel verskaf word, in alle gevalle aan gebruikers se behoeftes sal voldoen. Enkele voorbeelde ter illustrasie:

Onder die woord *baadjie* verskyn o.a. die frase *baadjie uittrek** met as vertaalekwivalente *buckle to§, get cracking§; have a scrap§*. In die afwesigheid van voorbeeldsinne kan gebruikers wat nie goed vertrou is met Afrikaans en/of Engels nie, foutiewe aannames maak oor die gebruik van hierdie uitdrukkings. Die aanwending van minstens die etiket "figuurlik (fig.)" sou hier noodsaaklike gebruiksliding kon verskaf. Daar is ook enkele ander gevalle waar 'n mens graag 'n voorbeeldsin sou wou sien. By die frase *he has his work cut out to ...* sou 'n mens graag 'n voorbeeldsin wou hê, wat kan aandui hoe hierdie frase voltooi kan word.

Die grootste probleem met die verwerking van inligting in hierdie woordeboek lê op die terrein van etiketgebruik. Daar is baie gevalle waar 'n uitdrukking aangedui word as "informeel", maar waar "minder gebruiklik"/"verouderend" meer gepas sou wees. (Die tekens * en § dui in die *TFW/BPD* aan dat 'n frase "informeel" is in onderskeidelik Afrikaans en Engels.) Vergelyk enkele voorbeelde:

baadjie HOM op sy ~ **gee***
HY kry op sy ~ *
baan 'n ~ **opskop***
baard bietjie ~ maar klipsteenhard*
baken HY steek ~ *
bang-bang ~ speel*

In ander gevalle, waar 'n mens die etiket "formeel", "minder gebruiklik" of "weinig gebruiklik" sou verwag, word geen etiket verskaf nie. Die gebruiker sal graag verstaan wil word en moet gevolglik gebruiklike vertaalekwivalente kry wat vir iemand in die jaar 2001 in Suid-Afrika duidelik sal wees. Dit is te betwyfel of selfs moedertaalsprekers van Afrikaans mekaar sal verstaan indien die volgende vertaalekwivalent gebruik sal word vir "make s.o. a present of

s.t.: *iem. iets kado gee*. Die nuutste uitgawe van die *HAT* dui tereg aan dat die woord *kado* weinig gebruiklik is. Vergelyk ook die volgende voorbeelde:

baan breek
iem. moet die **bed hou**
iem. iets in **bedenking gee**
die **bedrieër** bedroë
iem. laat **beesvel** ry
befaamd wees om/vir
goed **beklant** wees
HY het SY **bekoms** van iets
HY **bekla** HOM oor iets
HY staan onder die **besemstok**
beslaan raak (*bv. 'n bril*)
'n **beswaar billik/handhaaf**
'n **betaling voldoen**
dit **behoef** geen **betoo**g nie dat –
na **bevind van** sake
iem. se duur **bevogte** –

Grammatiese inligting

Taal is nie staties nie, gevolglik behoort die leksikograaf ook erkenning te gee aan grammatiese veranderinge wat reeds ingeburger is. Al-Kasimi (1983: 109) noem o.a. die volgende kriterium wat in tweetalige woordeboeke behoort te geld: "Are the recent findings of modern linguistics in phonology, grammar, and semantics applied in the dictionary?" 'n Mens sou hierdie kriterium nie net op taalteorie nie, maar ook op die taalpraktyk van toepassing kon maak. In hierdie verband sou die leksikograaf byvoorbeeld erkenning moes gee aan die verlies aan wederkerendheid by sommige werkwoorde in Afrikaans. Ponelis (1979: 228) het reeds in 1979 daarop gewys dat wederkerendheid by 'n klompie werkwoorde, veral in die omgangstaal, verlore gaan. Dit is gevolglik nie duidelik waarom wederkerendheid in die volgende gevalle voorkom nie:

waag HY ~ HOM **buite(kant)**
HY ~ HOM **in** die water
waardig HY toon HOM die geleentheid ~.

In die voorwoord word vermeld dat die korrekte woordorde aangedui word. Oor die algemeen verskaf die volgorde waarin die woorde van die frases verskyn wel 'n korrekte aanduiding van die woordorde, maar dit kan nie konsekwent as 'n riglyn gebruik word nie. Dit is in elk geval baie moeilik om die korrekte woordorde aan te dui sonder volsinne. Hiermee enkele voorbeelde:

wag die **wag aflos**
die **wag betrek**

oor – wag hou/staan
'n wag opstel/uitsit
(op) wag staan
wagte uitsit
waardigheid aan – verleen

4. Die mate waarin inligting vinnig opgespoor en maklik ontsluit kan word

Hierdie kriterium hou verband met die konsep "gebruikersvriendelikheid" waarop toenemend klem gelê word deur leksikograwe. In hierdie verband sê Kromann, et al. (1991: 2713): "The user aspect ... involves consideration on the part of the lexicographer for the dictionary's target group, its needs and competence, and the types of user situations that occur."

In die voorwoord van die *TFW/BPD* word die teikengebruiker nie duidelik gespesifiseer nie, maar daar kan waarskynlik aanvaar word dat die woordeboek veral op taalpraktisyns gerig is. Indien dit die geval is, behoort die gebruikers oor die algemeen nie probleme te hê met die opspoor en ontsluiting van inligting nie. Kruisverwysings in die woordeboek dra by tot 'n gebruikersvriendelike aanbod. By die woord *waarheid* is daar byvoorbeeld 'n verwysing na *suigeling* en by *mond* is daar ook 'n verwysing na *suigeling*.

Die gebruikersvriendelikheid van die woordeboek sou verder verhoog kon word deur in die toeligtingsgedeelte aan te dui onder watter woord(e) multileksikale lemmas opgeneem word. Dit is veral belangrik, aangesien die plek van aanduiding omstrede is. (Vergelyk Gouws (1989: 97-108) vir 'n bespreking van die voor- en nadele van verskillende plasinge.) Enkele voorbeelde sal ter illustrasie bespreek word.

Die uitdrukking *uit die mond van die suigeling sal jy die waarheid hoor* is hierbo by implikasie genoem. Hoewel dit die gebruiker help dat daar by verskeie lemmas verwysings is na *suigeling*, waar die vertaalekwivalent verskyn, sou dit die gebruiker nog meer gehelp het as daar in die toelighting 'n aanduiding was van die woord waaronder 'n vaste uitdrukking opgeneem word. Die meeste gebruikers sou waarskynlik onder die eerste selfstandige naamwoord in die uitdrukking, naamlik *mond*, soek vir die vertaalekwivalent, maar dan verder moet blaai na *suigeling* toe.

By *walgens* verskyn die frase *tot walgens toe* met die vertaalekwivalente *to satiety*; *ad nauseam*. Aan die Engelse kant is *ad nauseam* nie opgeneem onder *ad nauseam* by die letter A nie, maar wel by die letter N onder *nauseam*. Gebruikers van die *HAT* is byvoorbeeld gewoond daaraan om leenwoordgroepe alfabeties na te slaan. Interessant genoeg verskyn *ad nauseam* glad nie aan die Afrikaanse kant van die woordeboek nie.

Die omkeerbaarheidsbeginsel (vgl. Gouws 1989: 162) word ook byna deurgaans toegepas. 'n Uitsondering is teëgekome by die uitdrukking *hy staan onder*

die besemstok wat aan die Afrikaanse kant vertaal word met *he is henpecked*, maar laasgenoemde uitdrukking verskyn nie aan die Engelse kant nie.

5. Die formaat en finale redigering van die woordeboek

Al-Kasimi (1983: 112) noem drie kriteria wat verband hou met die voorkoms van 'n woordeboek, naamlik:

- 'n aangename en aantreklike formaat,
- goeie drukwerk en maklik vindbare inskrywings, en
- min tipografiese foute.

Hierby sou 'n mens bekostigbaarheid en hanteerbaarheid as bonusse kon byvoeg.

Die *TFW/BPD* het beslis 'n aantreklike voorkoms en die drukwerk en uiteensetting is besonder netjies. Die woordeboek verskaf ook goeie waarde vir geld en kan boonop maklik rondgedra of ingepak word. Tipografiese foute is ook tot die minimum beperk en die gebruiker sal dit dadelik as sodanig herken, byvoorbeeld die woord *be* wat herhaal word by die *Guide: English-Afrikaans*, die woord *to* wat herhaal word by die Engelse vertaalekwivalent onder die lemma *berus* en die woord *idea* wat weggelaat is in die frase *not have the faintest / foggiest idea* onder die lemma *benul*.

6. Gevolgtrekking

Die waarde van die *TFW/BPD* is geleë in die groot versameling gebruiklike frases wat in een woordeboek saamgevat is, sonder dat die gebruiker deur 'n magdom inligting moet worstel om 'n bepaalde frase op te spoor, soos wat die geval dikwels in algemene tweetalige woordeboeke is. Hierdie woordeboek sal verder verbeter kan word deur 'n vollediger toeligtingsgedeelte, deur slegs te konsentreer op gebruiklike frases, deur etikette op 'n sistematiese wyse te gebruik en deur meer volsinvoorbeelde naas die frases in te sluit.

Verwysings

- Al-Kasimi, Ali M. 1983. *Linguistics and Bilingual Dictionaries*. Leiden: E.J. Brill.
- Benson, M., E. Benson en R. Ilson. 1986. *The BBI Combinatory Dictionary of English. A Guide to Word Combinations*. Amsterdam/Philadelphia: John Benjamin.
- Gouws, R.H. 1989. *Leksikografie*. Pretoria/Kaapstad: Academica.
- Hanks, P. 1979. To What Extent Does a Dictionary Definition Define? Hartmann, R.K.K. (Red.). 1979: 32-38.
- Hartmann, R.K.K. (Red.). 1979. *Dictionaries and their Users: Proceedings of the 1978 BAAL Seminar on Lexicography*. Exeter: University of Exeter.

- Hausmann, F.J., O. Reichmann, H.E. Wiegand en L. Zgusta** (Reds.). 1989-1991. *Wörterbücher: Ein internationales Handbuch zur Lexikographie. / Dictionaries: An International Encyclopedia of Lexicography. / Dictionnaires: Encyclopédie internationale de lexicographie*. Berlyn/New York: Walter de Gruyter.
- Joubert, P.A.** 1997. *Tweetalige Frasewoordeboek Afrikaans-Engels: 'n Praktiese Gebruiksgids./ Bilingual Phrase Dictionary English-Afrikaans: A Practical Usage Guide*. Kaapstad: Pharos.
- Kromann, Hans-Peder, T. Riiber en P. Rosbach.** 1991. Principles of Bilingual Lexicography. Hausmann, F.J. et al. (Reds.). 1989-1991: 2711-2728.
- Odendal, F.F. en R.H. Gouws.** 2000. *HAT Verklarende Handwoordeboek van die Afrikaanse Taal*. Midrand: Perskor.
- Ponelis, F.A.** 1979. *Afrikaanse Sintaksis*. Pretoria: Van Schaik.
- Schemann, H.** 1989. Das phraseologische Wörterbuch. Hausmann, F.J. et al. (Reds.). 1989-1991: 1019-1032.
- Svensén, B.** 1993. *Practical Lexicography. Principles and Methods of Dictionary-Making*. Oxford/New York: Oxford University Press.

Herbert Wiegand with a "Metalex-icographical Panga" in the Jungle: An Unlocking of *Wörterbuch-* *forschung*

Maria Smit, *Department of Music, University of Stellenbosch, Stellenbosch, Republic of South Africa (msmit@maties.sun.ac.za)*

Abstract: This article discusses volume 1 of Wiegand's *Wörterbuchforschung*, which contains a section on the structure of the research field of lexicography, and a section on research on dictionary use. Wiegand firstly discusses different aspects of language lexicography and dictionary research to determine whether this type of research constitutes a scientific discipline. He concludes that it is a scientific research field on its way to become a scientific discipline. In the second section, Wiegand puts forward a theoretical framework for research on dictionary use, based on action theory, and a detailed methodology which includes empirical research. Such a theoretical and methodological foundation, which also takes into account the systematic research on dictionary structure, can be informative to lexicographers and enable them to plan the microstructures of future dictionaries in a more systematic way. This might also lead to a more scientific concept of user-friendliness in dictionaries.

Keywords: DICTIONARY RESEARCH, METALEXICOGRAPHY, DICTIONARY USE, LANGUAGE LEXICOGRAPHY, COMPUTER LEXICOGRAPHY, ACTION THEORY, METHODOLOGY, EMPIRICAL RESEARCH, ACCESS ACTIONS, USER ACTIONS, QUESTIONNAIRES, USER TESTS, PROTOCOLS, EXPERIMENTS.

Opsomming: Herbert Wiegand met 'n "metaleksikografiese panga" in die oerwoud: 'n Ontsluiting van *Wörterbuchforschung*. Hierdie artikel bespreek deel 1 van Wiegand se *Wörterbuchforschung*, wat 'n afdeling oor die struktuur van die navorsingsveld binne leksikografie bevat, en 'n afdeling oor navorsing oor woordeboekgebruik. Eerstens bespreek Wiegand verskillende aspekte van taalleksikografie en woordeboeknavorsing om vas te stel of hierdie soort navorsing 'n wetenskaplike dissipline vorm. Hy kom tot die gevolgtrekking dat dit 'n wetenskaplike navorsingsveld is wat op pad is om 'n wetenskaplike dissipline te word. In die tweede afdeling stel Wiegand 'n teoretiese raamwerk voor vir die navorsing oor woordeboekgebruik wat gebaseer is op aksieteorie, en 'n gedetailleerde metodologie wat empiriese navorsing insluit. So 'n teoretiese en metodologiese basis, wat ook die sistematiese navorsing oor woordeboekstruktuur in aanmerking neem, kan insiggewend wees vir leksikograwe en hulle in staat stel om toekomstige woordeboeke se mikrostrukture op 'n meer sistematiese wyse te beplan. Dit kan ook lei tot 'n meer wetenskaplike siening van gebruikersvriendelikheid in woordeboeke.

Slutelwoorde: WOORDEBOEKNAVORSING, METALEKSIKOGRAFIE, WOORDEBOEK-
GEBRUIK, TAALLEKSIKOGRAFIE, REKENAARLEKSIKOGRAFIE, AKSIETEOORIE, METODO-

LOGIE, EMPIRIESE NAVORSING, TOEGANGSAKSIES, GEBRUIKERSAKSIES, VRAELYSTE, GEBRUIKERSTOETSE, PROTOKOLLE, EKSPERIMENTE.

1. Introduction

In the first section of his article "Was eigentlich ist Fachlexikographie?" (1988), the well-known German metalexicographer Herbert E. Wiegand uses the powerful metaphor of a helicopter flying over a jungle. He compares the superficial observation of the jungle by the passengers with observations that have traditionally been made in the field of subject lexicography: they do not even get out of the helicopter, but try to form an idea of what the jungle looks like by taking some aerial photographs. Although some researchers in subject lexicography have occasionally made a couple of intermediate or forced landings, they mostly did so in order to obtain information about other matters. Hardly ever has anybody ventured out of the helicopter to tackle the jungle with a "metalexicographical panga" in order to get a more profound theoretical grip on this area.

What Wiegand said about subject lexicography, can also in a sense be said of his own unremitting efforts within metalexicography as a whole. In contrast to many others who have a more cautious approach, Wiegand has never been afraid of braving into the unexploited and uncultivated areas of metalexicography. After working for many years on his "book", he finalized the first volume in 1998. The title is *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie*, published in Berlin and New York by De Gruyter. This volume deals with Wiegand's structuring of the first field of lexicographical research, namely, research on dictionary use ("Wörterbuchbenutzungsforschung"). Volume 2, when it is published, will treat the other three fields of lexicographical research, namely, research on dictionary structure ("systematische Wörterbuchforschung"), research on dictionary history ("historische Wörterbuchforschung") and research on dictionary criticism ("kritische Wörterbuchforschung").¹ It is a pity that only the first volume is available at present, because Wiegand often refers to the second volume in which important aspects of his theory are discussed. This means that, for the time being, the reader can only see half the picture.

Volume 1 ends with a comprehensive bibliography, as well as registers of (i) the definitions formulated in the book; (ii) conventions regarding terminology coined by Wiegand ("Namengebungskonventionen"); (iii) figures and graphs used; (iv) illustrative examples to explain lexicographical procedures and instances of dictionary use; (v) lists of citations from dictionaries; (vi) the pages on which authors are mentioned or cited; and (vii) a general index.

The same high quality of printing which characterizes Wiegand's other works can also be found in this publication in which the multitude of intricate

tables, diagrams and rule formulations have been executed with almost faultless precision.

2. The argument in the first part of *Wörterbuchforschung*

As is characteristic of this metalexigrapher, this volume is a formidable reworking of his many multifaceted and lengthy articles which regularly appeared since the 1970s. He once again follows the path from beginning to end, writing in great detail and with typical German precision about his passion in life: *Wörterbuchforschung* (dictionary research). In a most systematic way he revisits the metalexigraphical beacons he has erected over the years, only to show even more clearly what he regards as the academic and scientific status of dictionary research.² A most frustrating fact for readers who do not have a command of German has always been that so few of Wiegand's works have been published in English. Only one volume containing a selection of articles is available in English translation.³ It is therefore very difficult, if not impossible, for anyone who does not know German well, to get an in-depth insight into the extensiveness and complexity of Wiegand's lexicographical theory. Even those with a good command of German know that Wiegand's works are not easily assimilable, and this first volume certainly is not.

Wiegand has written many articles in journals, monographs, and textbooks, as well as many conference papers, reviews, review articles, reports, bibliographies, discussions, et cetera.⁴ Many of his articles are of substantial length and might just as well have appeared as books in their own right. He is editor and coeditor of numerous publications which serve as the canon for metalexigraphy, such as the journal *Lexicographica*; the three-volume *Wörterbücher, Dictionaries, Dictionnaires (Handbücher zur Sprach- und Kommunikationswissenschaft, Vol. 5.1-5.3)* (Hausmann et al. 1989-1991); the *Studien zur neuhochdeutschen Lexikographie, Vols. I-VI.1/2 (Germanistische Linguistik)*; and the *Lexicographica Series Maior* of which over 90 volumes have been published so far.

The first volume of *Wörterbuchforschung* starts out with Wiegand's classical explanation of what it is all about — what language lexicography is, and what dictionary research is. His main aim is to develop a general theory of lexicography. He claims (1998: VII), and rightly so, that there were no previous examples that he could use in drawing up this theory, and that it took him two decades to do so. The reason for his work has been to establish dictionary research as a scientific discipline, so that it would obtain the status of an official academic discipline (1998: VII). The first part of the first volume deals with "Studies towards the structuring of a research field"⁵, which has as subheading "Language lexicography and dictionary research". This part contains the following sections: (1) What it is all about; (2) What is language lexicography?; (3) Dictionary research: its object of research, its tasks and aims; (4) Is dictionary research a scientific discipline?; (5) Computers, scientific lexicography and dictionary research; (6) The object of dictionary research: a concluding charac-

terisation; (7) Scientific lexicography and other academic disciplines; and (8) A resumé on the status of language lexicography and dictionary research in the mid-1990s.

A lengthy discussion of the question "What is language lexicography?" leads to several subquestions. Wiegand (1998: 15) uses his customary way of building his arguments on citations from other works on lexicography (in many cases in order to refute or correct the arguments of those authors).⁶ He (1998: 15) warns the reader that he will use relatively many citations, and it becomes clear that he has indeed, in his reworking of his theory, added a still greater number of bibliographical sources to strengthen his previous arguments.⁷ It appears that the citations mainly help him to build his arguments, and by interpreting them (and refuting them) he can refine them, thus forming the clear distinctions needed for his comprehensive theory.⁸

The subquestions (i) whether language lexicography is a science or not, (ii) whether it is applied linguistics, and (iii) whether it is a branch of lexicology, remind the reader of the discussions in earlier articles (Wiegand 1983; 1983a; 1984; 1984a; and 1989b), although he (1998: 4) states that these earlier versions are only partially included in the present publication. This certainly is the case, especially because Wiegand adds a wealth of recent literature to support his arguments. As usual, his excessiveness for the sake of completeness overwhelms the reader (for example, the last paragraph on page 18, as well as pages 20, 27 and 43).

Wiegand (1998: 46) makes the important distinction between language lexicography and dictionary research, where the latter forms the scientific metadomain for language lexicography. But before he gives a final verdict on what language lexicography is, he (1998: 47-58) discusses the difference between language lexicography ("Sprachlexikographie"), special-field lexicography ("Sachlexikographie") and the lexicography of encyclopaedic dictionaries ("Allbuchlexikographie"). According to the definition on page 62, language lexicography can either be a self-taught, nonscientific cultural practice, or an independent cultural and scientific practice, consisting of a more or less clearly demarcated self-reflecting component, which can only be mastered, because of its multilayered comprehensiveness, by academically trained staff. Language lexicography is a calculable, analysable, checkable, manageable, testable and teachable practical process, which draws on various results and methods from different disciplines. It forms the central part of the pretheoretical empirical subject domain out of which the various scientific objects for the different research fields of dictionary research can be extrapolated.⁹

Wiegand proceeds by distinguishing the differences between various denotations for lexicographical works (1998: 64-71),¹⁰ and the difference between the terms "metalexicography" and "dictionary research" (he uses "metalexicography" to denote the total metadomain whose object of study consists of (i) dictionary research, (ii) lexicon research and (iii) the research of encyclopaedic dictionaries) (1998: 72-76). This is followed by a characterisation of dictionary

research (1998: 76-88). The question whether dictionary research is a scientific discipline can only be answered after determining what the criteria are for any field of study to be called a "scientific discipline" (Wiegand 1998: 89-97). Wiegand presents numerous citations and reflections on the nature of the notion "discipline". Finally, after giving an exposition of the present status of dictionary research, Wiegand (1998: 102-103, also 256) concludes that, in the mid-1990s, dictionary research is not a scientific discipline yet, but rather on its way to become one. At present, one should rather use the designation "scientific research field" to refer to dictionary research. It is possible for various subdisciplines to emerge, for example (i) dictionary research of the German language, (ii) bilingual dictionary research, or (iii) the use of English monolingual dictionaries (Wiegand 1998: 104-118). Wiegand (1998: 133) ends this section by expressing the wish that dictionary research will in the future become an academic discipline with official status.¹¹

The next section (Wiegand 1998: 133-246) deals with computer lexicography and dictionary research. Firstly, Wiegand (1998: 134-153) discusses crucial aspects of lexicographical processes without the use of computers, and then proceeds to aspects where computers are partially involved (1998: 153-160), as well as processes which are from the very beginning completely computerised (1998: 171-248). Many of these issues have previously been dealt with in Zai-ping and Wiegand (1987), where the organisation of *Das Grosse Deutsch-Chinesische Wörterbuch* was discussed, and in Wiegand (1986), which specifically focused on the computerisation of dictionary projects. Now, Wiegand's arguments are more detailed and more universal, also referring to works by other scholars. The section on computer lexicography especially is now more up-to-date and greatly expanded, because most of the greater lexicographical projects undertaken today make use of corpora stored and sorted in electronic databases. The lexicographical process consists basically of (i) the preparation phase, which can be divided into several steps (conceptualising a dictionary programme, and drawing up a dictionary plan); (ii) the phase of collecting the material; (iii) the phase of preparing the material; (iv) the phase of exploiting the material (the actual writing of dictionary articles); and (v) the phase of preparing the manuscript for publication. Most of these phases are also used when a dictionary is computerised. As earlier in Wiegand (1987a), he once again makes sure that the reader understands the distinction between the notions "lexicographical data" and "lexicographical information" (Wiegand 1998: 160-171), also by referring to many other authors who have dealt with these concepts.¹² Wiegand argues that users have to extract or reconstruct lexicographical information presented in dictionary articles — and different users can extract different information from the same data (1998: 170).

He (1998: 248-256) ends the first part of volume 1 by discussing scientific lexicography and its relationship to other academic disciplines. It is clear that greater lexicographical projects can nowadays only be undertaken in teams, in most cases also involving experts from other disciplines.

3. The argument in the second part of *Wörterbuchforschung*

The first part of volume 1 is the easier part. In the second part of *Wörterbuchforschung* which spans pages 259 to 1031, Wiegand takes research on dictionary use under close scrutiny. Chapter 2 (Wiegand 1998: 259-267) gives an overview of research on dictionary use in the mid-1990s, whereas chapter 3 deals with the theoretical foundation of dictionary use based on action theory (Wiegand 1998: 268-567). The next chapter, chapter 4, deals with the methodology of research on dictionary use (Wiegand 1998: 568-1025), followed by a short resumé in chapter 5 on the position and future tasks of research on dictionary use in the mid-1990s (Wiegand 1998: 1026-1031).

The overview of research on dictionary use in chapter 2 contains many references to publications that have until now dealt with this aspect. In chapter 3, Wiegand (1998: 268-567) gives a lengthy discussion of how research on dictionary use can and should be based on action theories ("Handlungstheoretische Grundlegung der Wörterbuchbenutzungsforschung") as developed in social science research. This section is a much more comprehensive exposition than in any of his previous publications (for example, Wiegand 1987a). He draws on different "action" categories and gradually builds up his own analytical model of different types of dictionary use. He gives examples of dictionary use by "constructed" dictionary users, where these users pose relevant search questions. By means of these examples, Wiegand aims to describe certain central features of user actions, in order to classify them systematically. The examples are representative of search questions that arose during earlier pilot studies, and Wiegand makes general statements and formulates numerous definitions (in fact, 175 of them!) with regard to dictionary use and terminology (either well-known or newly-coined terms).¹³ Some of his examples have been used in earlier work, but many of them are new, or expanded. According to Wiegand (1998: 283-284), it is very important to determine the patterns of action ("Handlungsmuster") that are actually performed when users consult dictionaries. One cannot determine these without some kind of pretheoretical framework, because otherwise it will not be possible to interpret the empirical data. On the other hand, an adequate theory of dictionary user actions cannot be developed without empirical pilot studies.¹⁴ Later in the volume it becomes clear that Wiegand has ample experience of empirical research, when he admits (1998: 1001, note 90) that, because of the empirical projects that he undertook, he received many stimuli for the formulation of hypotheses about user actions.

Hartmann (1998: 160) in his review of Wiegand's *Wörterbuchforschung*, criticises the fact that "the problem situations are invented", meaning that Wiegand devised many of the search questions, the search actions and the features of the constructed users he describes in his examples. This may, however, not be problematic — linguists in the generativist tradition were very positive about the "values of introspective data" as opposed to data obtained from experiments.¹⁵

Wiegand (1998: 296) distinguishes between usual, unusual and didactical uses of dictionaries, before he (1998: 298-301) explains to the reader what the genuine purpose of a dictionary is. Usual search actions include (i) using a dictionary as a reference work (Wiegand 1998: 304-350); and (ii) using a dictionary as reading matter on language (Wiegand 1998: 350-355). Using a dictionary in an unusual way includes (i) using it in order to learn something about it, for example, when writing a review of the dictionary (Wiegand 1998: 355-359); and (ii) using it not as a dictionary but, for example, for drying leaves in it (Wiegand 1998: 359). The latter type of user action is noncommunicative, not related to the purpose of the dictionary, and if successful, without errors. The didactical use of dictionaries is also discussed (Wiegand 1998: 363-366).¹⁶

In a next section, Wiegand (1998: 370-480) adds another dimension when he distinguishes between general, special and dictionary-specific user practice. The first of these practices, as the designation suggests, occurs when a user masters the general system of dictionary consultation, without having to supplement the user act with additional knowledge (for example, the alphabetical order in which dictionaries are usually presented, can be considered as a general user practice in languages using the alphabetical system). Dictionary-specific user practices occur when users are able to perform action types for specific types of dictionaries (not a specific dictionary). As an example, Wiegand discusses user practices specific to retrograde dictionaries (1998: 380-384) and to orthographic dictionaries (1998: 384-390). Special user practice occurs when a user has enough experience of using dictionaries that he/she masters more than one specific user practice.

Many terms are used to determine the exact nature of different user actions, for example, "external and internal access actions" ("externe und interne Zugriffshandlungen"). External access actions are performed when a user, for example, consults the outer texts and/or register(s) of the dictionary (Wiegand 1998: 394), whereas internal access actions occur when a user consults dictionary articles in order to obtain the needed lexicographical data (Wiegand 1998: 404). These external and internal access actions are also subdivided into smaller subtypes, such as "actions following cross-references" ("Verweisbefolgungshandlungen") and "consultation actions motivated by cross-references" ("verweismotivierte Konsultationshandlungen") in the case of internal access actions (Wiegand 1998: 408-423). Of course, cross-references can differ in format and content and they can also cross-refer the user to different parts of the dictionary text or even to other dictionaries, which once again leads to very specific actions. It is clear that Wiegand has considered these many instances of dictionary use very carefully, as can be illustrated by his remarks on the "action following a cross-reference": The "dictionary-incurrent action following a cross-reference" ("wörterbuch-inkurrente Verweisbefolgungshandlung") means that a user is cross-referred to another dictionary by means of a bibliographical index. This does not imply, however, that the user gives up the first search question in favour of a new one. If this should happen, it constitutes another

type of action, namely, a "consultation action motivated by a cross-reference" (Wiegand 1998: 421).

Wiegand, in his discussion of these different action types, resorts to many devices in order to build his theory. Firstly, when using examples, he often refers to dictionary articles in existing dictionaries to illustrate his arguments on types of access actions. By doing so, he also criticises elements in these dictionary articles, which belong under the fifth component of his lexicographical theory, namely, the research on dictionary criticism. Examples of such dictionary criticism can be found in Wiegand (1998: 440 and 446). Hopefully, this aspect will eventually also be mentioned in part V, volume 2, as a specific type of dictionary criticism.¹⁷

Secondly, Wiegand coins new terms and uses already known terms in order to designate all the instances of action types of which he can think. For each term, a definition is given — there are 29 definitions in this section alone. User actions are subdivided into various "families" (18 in total) and "subfamilies" (11 in all), and in each case various designations are given for different actions (1998: 480-494).

Thirdly, as in most of his other publications, Wiegand makes extensive use of graphic illustrations in order to help the reader to get a better overview of the theory.

In the fourth place, in his distinction of "action types", Wiegand represents these actions in the following format:

AUF DAS LEMMA *sich etwas vom Herzen reden* IM WÖRTERVERZEICHNIS DES IDIOMATISCHEN WÖRTERBUCHES (=WB8) "Friedrich (1976)" ZUGREIFEN.¹⁸

This indicates the type of action, the problematic expression which instigated the user action, and the dictionary used during the search. Not all rules are as specific as this one.

Fifthly, Wiegand gives formulas which represent rule systems. These can be simple or very intricate, dazzling the mind of the reader, and taking some time to decipher. An example of a simple rule (Wiegand 1998: 390) is:

{HT-4A-3.3} → {HT-5A-3.3} → {HT-5aA-3.3}

Lastly, Wiegand presents explanations of his rule formulation by means of intricate descriptions in separately marked sections of the conventions regarding his terminology. In coining his terminology, he follows the conventions of formal logical languages (Wiegand 1998: 497). This means that if the reader is not trained in these conventions, Wiegand's works, and especially this part of volume 1, will be very hard to follow.

Wiegand (1998: 500-510) devotes a section (paragraph 3.3.7.) to types of users. He discusses user types, user roles, potential users and addressees. There are learners of dictionary use ("Benutzungslerner"), experienced dictionary users, and trained and initiated users. Dictionary users can be successful or

unsuccessful; they can either be laypersons or scientific users. They can be mother-tongue users or foreigners.

There are also, Wiegand (1998: 510) contends in paragraph 3.3.8., different user modalities and user skills. Predicates which evaluate user skills have not been dealt with extensively in dictionary research, and it falls under the didactical component. Wiegand (1998: 510-519) therefore discusses some predicates which are of importance for the research of dictionary use. The following examples are predicates which can describe user actions: "experienced/inexperienced"; "appropriate/inappropriate"; and "without errors/correcting the errors/with errors".

In addition, there are contexts ("Benutzungskontexte") in which consultation actions can take place (Wiegand 1998: 519-562). External contexts take into account, for example, the place, time, and duration of user action, and different social and communicative aspects. Users have different interests, and different reasons for performing a user action, and the results of their user actions can be different.

Wiegand (1998: 563-567) concludes chapter 3 on the theoretical foundation of dictionary use by an overview of his proposed structure in the format of a table (1998: 564-566). Examples of types of search questions listed in this table are: (i) WHO? (uses, can use); (ii) HOW? (how often, under which social circumstances, under which conditions); (iii) WHERE?; (iv) WHEN?; (v) HOW LONG?; (vi) WHY?; (vii) WITH WHICH CONSEQUENCES? He claims that his proposals in this chapter can lead to a theory-based orientation towards empirical research, and adds that his contemplations are, in fact, based on empirical work previously done on a small scale (1998: 567).¹⁹

Chapter 4 deals with the methodology of dictionary use (Wiegand 1998: 568-1025), and Wiegand points out that one should ask oneself whether methods from empirical social research can successfully be used in the research of dictionary use, and whether the questions asked in research of dictionary use can appropriately be investigated by means of these methods (1998: 568). Some of the methods that Wiegand (1998: 568-823) discusses in great detail are observation, questionnaires, interviews, tests, experiments (only a short section), content analysis (also a very short section), and the use of statistical methods. His discussion includes excerpts from previous studies conducted at the University of Heidelberg, as well as newly devised attempts to structure his theoretical framework. He also refers to the work of other scholars, and weighs the advantages of the methods against the disadvantages.²⁰

Once again, Wiegand breaks down into subtypes the different methods that can be used. In the section on questionnaires, for example, he distinguishes among others between oral and written questionnaires; personal and group interviews; interviews with laypersons and with experts; and standardised, partially standardised and nonstandardised interviews, to name a few.

Wiegand (1998: 587-588) gives four reasons why certain questions have not yet been posed in the research of dictionary use. Firstly, some publishers

have only been interested to market the product, but did not systematically take into account user needs. Secondly, various diverging models have so far been used to study dictionary use, some of which were not specific or theoretical enough for empirical research. Thirdly, Wiegand contends that only very recently has research on dictionary structure presented useful formal and structural devices with which empirical research can be done.²¹ Finally, as fourth reason, Wiegand puts forward that, apart from the "EURALEX/AILA research project into dictionary use", no other extensive research projects have been undertaken in the area of dictionary use.

Wiegand (1998: 589-664) then engages in lengthy discussions on the nature of questionnaires and observations, user profiles, user experience, user habits, the types of questions to be included, techniques to be used in order to obtain useful information, rating scales for grading answers, the layout of the questionnaire, and other issues. These discussions are accompanied by detailed examples of two attempts to devise a questionnaire (Wiegand 1998: 591-597 and 646-656), leading to evaluations of the advantages and disadvantages of the questionnaires and their results when they were first used, and followed by a revised questionnaire (Wiegand 1998: 664-677).

According to Wiegand, one will have to work out in more detail what the methods and contents should be if one wants to use the test as measurement of dictionary use (1998: 677). Three types of tests are possible, namely, (i) the user test, (ii) a test to determine metalexicographical knowledge, and (iii) a dictionary test. The latter has been developed within the framework of the second component of Wiegand's lexicographical theory, i.e. the research on dictionary criticism, and Wiegand refers the reader to part V of volume 2 (1998: 677). A user test can test user skills, language competence, user experience, knowledge of dictionaries, knowledge of patterns of user actions, and the skills to perform specific user actions of a specific type, thereby demonstrating the mastering of a user practice. A user test can be monothematic if it concerns the use of one dictionary only, or polythematic if more than one dictionary is involved. It can test one user in comparison to a group (making a cross-section), or it can test the developing process of one user over a period of time (making a longitudinal section) (Wiegand 1998: 678-679). Tests of this nature can inform the didactical component of dictionary use. Furthermore, tests can be divided into classes of (i) tests for users-in-action (i.e. while they are consulting a dictionary); (ii) tests for users-not-in-action (i.e. usually comprising tests about users' knowledge of dictionaries); and (iii) a mixture of (i) and (ii). It is important to base the development of user tests on theoretical foundations, which means that the features of the users to be tested have to be determined theoretically, and not on an ad hoc basis (Wiegand 1998: 683-684). One should also draw on studies within the field of constructing tests, as done in pedagogical diagnostics, psychology and empirical social research (Wiegand 1998: 684).²² These user tests can, once more, be divided into three categories: (i) subjects have to carry out specific instructions; (ii) the choice of dictionaries can be the focus; or (iii)

questions can focus on determining the subjects' knowledge of dictionaries (or a specific dictionary). Wiegand engages in a detailed discussion of the possibilities and problems of these types of tests, accompanied by numerous examples from his own empirical research (1998: 685-818). Aspects such as objectivity (1998: 760-762), reliability (1998: 762-766), and validity (1998: 766-767; 804-812) are raised.

Paragraph 4.2. is devoted to a detailed exposition of different "genuine metalexicographical concepts for the methodology" in research on dictionary use (Wiegand 1998: 823-1026). Here he categorises situations of dictionary consultation, using amongst other devices, tree diagrams and symbols to illustrate relationships between categories. Once again, this part is richly provided with examples of dictionary use to enable the reader to form a clear picture of what Wiegand aims at. This is a very important part of the volume, and filled with information that has not yet been published in such a comprehensive way. Wiegand (1998: 978, note 87) also points out to the reader that one of the purposes of building typologies is to develop the terminology of a research field, as Wiegand himself undoubtedly does in this part of the book.

The following section discusses protocols of dictionary use (Wiegand 1998: 974-1025). Types of protocols and devices used for protocols are discussed, and examples from Wiegand's own research are used as illustrations.

With regard to experiments, Wiegand (1998: 1023-1024) contends that one cannot successfully describe metalexicographical experiments about the relationship between dictionary structure and dictionary use without a detailed theory of dictionary structure. This aspect is only dealt with in part III of volume 2. But, if both a theory of user actions and a theory of dictionary structure are available, one can empirically prove by means of experiments (in which the independent variable features can be controlled) how the features of the dictionary structure affect the use of a dictionary. This will enable a scientific concept of user-friendliness, which can also make it possible to compare dictionaries on empirical grounds, and not just by approximate judgements.

Wiegand (1998: 1026-1031) ends this comprehensive volume with a résumé on the status and tasks of research on dictionary use in the mid-1990s. He (1998: 1026) claims that it is the youngest of all the research fields, and that there are only about 100 publications dealing with this field at the time of writing. Studies so far have concentrated mostly on dictionary use by laypersons, especially learners of foreign languages. Wiegand (1998: 1026-1027) notes that a fully developed research on dictionary use also has a historical dimension, in which one could, for example, study the dependence of one historical dictionary on preceding ones.²³ He (1998: 1027) states that even though some previous studies pose detailed questions and present worthwhile ideas and interesting results, most of the results cannot be generalised, particularly because of the lack of a common theoretical framework. His own research within the framework of the Heidelberg Project "Empirische Wörterbuchbenutzungsforschung" did, from the start, focus on working out a theoretical basis for re-

search on dictionary use. This effort now forms the backbone of the second part of this volume. Apart from the theoretical basis for research on dictionary use, it is important to develop a methodology. This is what Wiegand did in this volume. He (1998: 1030) expresses the hope that these chapters will help the research on dictionary use to establish a sound theoretical and methodological framework.

According to Wiegand (1998: 1030-1031), tasks for the future will be: (i) to undertake an extensive interinstitutional team research project on the use of dictionaries, where not only students learning German as a foreign language will be involved; (ii) to develop, by taking into account the systematic research on dictionary structure, a standardised concept of user-friendliness; (iii) to develop usable user achievement tests ("Benutzungsleistungstests") in order to be able to use different dictionaries with the same effect; (iv) to develop, together with the systematic study of dictionaries, concepts which enable more systematic planning of microstructures in dictionaries.

4. Conclusion

The preceding discussion demonstrates how successful Wiegand has been in devising a closely-knit theory on dictionary research. He has embedded this theory within the framework of well-established German schools of thought as is shown by his reference to many German scholars in other disciplines, who have done worthwhile research. For example, his arguments, conventions for coining terminology, ways of formulating definitions, and ways of writing down rule systems, are based on those found in social science research, formal logic, mathematics, the natural sciences, et cetera. On page 767 he suggests that, in order to form a theory, one should consult the "canon of scientific theory". Obviously he refers in the first place to a vast collection of German scholarly work (as can also be seen from the bibliography).²⁴ Throughout the years, disciplines have often grown and expanded rather independently in different countries. Sometimes there were influences from German scholars on English scholars, and vice versa. But it is clear that this "canon" that Wiegand refers to, is a closed book to many who have not been trained in the German tradition. So is the work of Wiegand — if you do not read German.

In conclusion, Wiegand (1998: 1025) advocates that lexicographers who are working within a theoretical framework, should take dictionary structure far more seriously than is the case at present. We are looking forward to reading volume 2 which contains the theory of dictionary structure.

Notes

1. Cf. Hartmann (1999: 156-157) for the English equivalents used in this article.
2. Cf., for example, Wiegand (1998: 435), where he explains that, in this work, he has to elaborate on his previous work regarding the instruments he uses for developing his theory in detail.

3. Cf. Immken and Wolski (1999).
4. According to the website which lists Wiegand's publications, he has now published more or less 378 titles.
5. "Studien zur Strukturierung eines Forschungsfeldes" (Wiegand 1998: XI).
6. Once, towards the end of the volume, Wiegand (1998: 930, note 78) admits that he changed his mind because of the work done by another scholar, when he speaks about the concept of congruence.
7. Cf., for example, Wiegand (1998: 15, note 5), where he demonstrates that he did not only consult German, English and French works, but also Italian, Spanish, Portuguese and Russian works on lexicography!
8. Cf., for example, Wiegand (1998: 44), where he interprets Schaefer, and even explains to the reader what Schaefer actually wanted to say. Schaefer (1998: 294), on his part, humorously reacted in his review of Wiegand's book by "quoting" the following part of Wiegand's sentence "Was SCHAEDER ... wahrscheinlich ausdrücken will, ist dies" before he made his final statement about Wiegand's work.
9. This is a very oversimplified rendition of what is actually included in Wiegand's (1998: 62-63) "preliminary characterisation" of language lexicography.
10. This part is a revised and expanded version of the introductory paragraphs of the article "Shangai bei Nacht" (Wiegand 1988a).
11. According to the impressive list of academic institutions in German-speaking countries which already offer courses in lexicography (Wiegand 1998: 121-127), great progress has been made at least in those countries towards establishing dictionary research as an official academic discipline.
12. It is important to note that Wiegand (1998: 252) refines his own terminology with regard to *other disciplines* which can supply data to lexicography. He used to call these disciplines *informationsspendend* ("information-supplying") (Wiegand 1989b: 265), but brings this in congruence with his explanation in section 1.5.2.2. of this volume, changing the expression to *datenliefernd* ("data-supplying").
13. Cf., for example, Wiegand (1998: 292) for an overview of typical search questions, mapped to types of user actions and terminology that can be used in research on dictionary use.
14. Cf. also Wiegand (1987a: 181) where he explains that before one can do any empirical research, one will have to acquire rudimentary knowledge. This may include (i) general, pre-scientific knowledge about dictionary use; (ii) accumulated experience in the use of dictionaries, and reflecting on them; and (iii) theoretically based hypotheses on dictionary use.
15. Cf., for example, Newmeyer (1983: 48-50).
16. Cf. Wiegand (1998: 367) for a figure depicting the types of dictionary use.
17. It is also interesting to note that in the second part of volume 1, Wiegand often refers to many dictionary articles in existing dictionaries, but in some sections he seldom refers to work done by other scholars within the field of research in dictionary use. This is probably because many of the distinctions that he makes within this field, were introduced by himself, and were yet not extensively treated by other scholars.
18. This example is found in Wiegand (1998: 454).
19. Cf., for example, Wiegand (1985; 1987), Ripfel (1989; 1990; 1991) and Ripfel and Wiegand (1988).

20. Cf., for example, Wiegand (1998: 583, 588-589) where he refers extensively to previous work done by other scholars within the area of questionnaires on dictionary use.
21. These formal and structural devices have, of course, mainly been worked out by Wiegand himself, as can be seen in the references on page 589, which include, amongst others Wiegand (1989; 1989a; 1990; 1995).
22. It is interesting to note that Wiegand (1998: 720) observes that psychologists and pedagogues have up to now not realised that a test is a *text* which belongs to a certain text type and that an instruction in a text is a partial text. Consequently they frequently move to and fro between cognitive and textual levels. Wiegand urges for a linguistic study to be done on the test as a text type.
23. Cf. also Wiegand (1998a: 647), where he discusses the practice of "compiling dictionaries from other dictionaries" ("Wörterbücher aus Wörterbücher zu erarbeiten").
24. Cf. also Wiegand (1998: 797), where he draws on concepts and methods used in mathematics and the natural sciences.

Bibliography

- Hartmann, R.** 1999. What is "Dictionary Research"? *International Journal of Lexicography* 12(2): 155-161.
- Hausmann, F.J. et al. (Eds.)**. 1989-1991. *Wörterbücher. Ein Internationales Handbuch zur Lexikographie. / Dictionaries. An International Encyclopedia of Lexicography. / Dictionnaires. Encyclopédie Internationale de lexicographie.* Handbücher zur Sprach- und Kommunikationswissenschaft, vol. 5.1.-5.3. Berlin: De Gruyter.
- Immken, A. and W. Wolski.** 1999. *Semantics and Lexicography. Selected Studies (1976-1996)*. Tübingen: Niemeyer.
- Newmeyer, F.J.** 1983. *Grammatical Theory. Its Limits and its Possibilities*. Chicago: University of Chicago Press.
- Ripfel, M.** 1989. Ergebnisse einer Befragung zur Benutzung ein- und zweisprachiger Wörterbücher. *Lexicographica* 5: 178-201.
- Ripfel, M.** 1990. Probleme der Erhebung metalexikographischer Daten. Hausmann, F.J. et al. (Eds.). 1989-1991: 1631-1638.
- Ripfel, M.** 1991. Wörterbuchbenutzung bei Muttersprachlern. Untersuchungsbericht über eine Befragung erwachsener muttersprachlicher Sprecher zur Wörterbuchbenutzung. *Lexicographica* 6: 237-251.
- Ripfel, M. and H.E. Wiegand.** 1988. Wörterbuchbenutzungsforschung. Ein kritischer Bericht. *Germanistische Linguistik: Studien zur neuhochdeutschen Lexikographie* 6(2): 490-520.
- Schaeder, B.** 1998. Review: Herbert Ernst Wiegand: Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie. *Lexicographica* 14: 284-294.
- Wiegand, H.E.** 1983. Ansätze zu einer allgemeinen Theorie der Lexikographie. Schildt, J. and D. Viehweger. 1983. *Die Lexikographie von heute und das Wörterbuch von morgen: Analysen — Probleme — Vorschläge*: 92-127. Linguistische Studien, Reihe A. Arbeitsberichte 109. Berlin: Akademie der Wissenschaften der DDR, Zentralinstitut für Sprachwissenschaft.
- Wiegand, H.E.** 1983a. Überlegungen zu einer Theorie der lexikographischen Sprachbeschreibung. *Germanistische Linguistik* 5-6: 35-72.

- Wiegand, H.E.** 1984. On the Structure and Contents of a General Theory of Lexicography. Hartmann, R.R.K. (Ed.). *LEXeter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9-12 September 1983*: 13-30. Lexicographica Series Maior 1. Tübingen: Niemeyer.
- Wiegand, H.E.** 1984a. Prinzipien und Methoden historischer Lexikographie. Besch, W. et al. (Eds.). 1994. *Sprachgeschichte. Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung*: 557-620. Handbücher zur Sprach- und Kommunikationswissenschaft 2.1. Berlin: De Gruyter.
- Wiegand, H.E.** 1985. Fragen zur Grammatik in Wörterbuchbenutzungsprotokollen. Ein Beitrag zur empirischen Erforschung der Benutzung einsprachiger Wörterbücher. Bergenholtz, H. and J. Mugdan (Eds.). *Lexikographie und Grammatik. Akten des Essener Kolloquiums zur Grammatik im Wörterbuch, 28-30.6.1984*: 20-98. Lexicographica Series Maior 3.
- Wiegand, H.E.** 1986. Metalexigraphy: A Data Bank for Contemporary German. *Interdisciplinary Science Reviews* 11(2): 122-131.
- Wiegand, H.E.** 1987. Empirische Wörterbuchbenutzungsforschung. Wiegand, H.E. (Ed.). 1987. *Theorie und Praxis des lexikographischen Prozesses bei historischen Wörterbüchern*: 255-257. Lexicographica Series Maior 23. Tübingen: Niemeyer.
- Wiegand, H.E.** 1987a. Zur handlungstheoretischen Grundlegung der Wörterbuchbenutzungsforschung. *Lexicographica* 3: 178-227.
- Wiegand, H.E.** 1988. Was eigentlich ist Fachlexikographie? Munske, H.H. et al. (Eds.). 1988. *Deutsche Wortschatz: Lexikologische Studien*: 729-790. Berlin: De Gruyter.
- Wiegand, H.E.** 1988a. Shanghai bei Nacht. *Germanistische Linguistik: Studien zur neuhochdeutschen Lexikographie* 6(2): 521-626.
- Wiegand, H.E.** 1989. Arten von Mikrostrukturen im allgemeinen einsprachigen Wörterbuch. Hausmann, F.J. et al. (Eds.). 1989-1991: 462-501.
- Wiegand, H.E.** 1989a. Der Begriff der Mikrostruktur: Geschichte, Probleme, Perspektiven. Hausmann, F.J. et al. (Eds.). 1989-1991: 409-462.
- Wiegand, H.E.** 1989b. Der gegenwärtige Status der Lexikographie und ihre Verhältnis zu anderen Disziplinen. Hausmann, F.J. et al. (Eds.). 1989-1991: 246-280.
- Wiegand, H.E.** 1990. Printed Dictionaries and their Parts as Texts. An Overview of More Recent Research as an Introduction. *Lexicographica* 6: 1-126.
- Wiegand, H.E.** 1995. Lexikografische Texte in einsprachigen Lernerwörterbüchern. Popp, H. (Ed.). 1995. *Deutsch als Fremdsprache. An den Quellen eines Faches*: 463-499. München: Iudicium.
- Wiegand, H.E.** 1998. *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie. Band 1*. Berlin: De Gruyter.
- Wiegand, H.E.** 1998a. Historische Lexikographie. Besch, W. et al. (Ed.). 1998. *Sprachgeschichte. Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung*: 643-715. Berlin: De Gruyter.
- Zaiping, P. and H.E. Wiegand.** 1987. Konzeption für das grosse deutsch-chinesische Wörterbuch (zweiter Entwurf). *Lexicographica* 3: 228-241.

Dorothea Mantzel and Bernd Schulz. *Francolin Illustrated School Dictionary for Southern Africa*. Developed by The Dictionary Unit for South African English in consultation with The Molteno Project and READ Education Trust. 1997. xvi + 336 pp. ISBN 1-86859-015-1. Cape Town: Francolin Publishers.

It was a phone-in programme on a local radio station, dealing with the overcharging of tourists who use taxis. The deejay was completely upset, interrupting the CD she was playing. How could a trip which cost her R15 cost a tourist R300? Eventually a taxi operator called in to cut through all the nonsense. It was about misunderstanding the difference between a taxi and a minibus taxi. Using the former means that you pay per kilometre, irrespective of the number of people who get into the taxi, to a destination you decide. Using the latter implies that you pay a fixed cost per person and the drop-off points are predetermined for you along a specific route. And there is little difference in cost for the trip if the minibus taxi were carrying 20 people.

A learner conveys this discussion to me, demonstrating the power of radio stations in shaping and influencing the vocabulary of our youth, irrespective of whether they are first, second or third language users. I am writing this review, and wonder whether the **Francolin Illustrated School Dictionary for Southern Africa** makes this distinction. I wonder further whether it includes the abbreviations / words "CD" or "compact disc" and "DJ", "deejay," or "disc jockey", terms that fall within the experience of the target audience of this dictionary.

Target Audience

The dictionary, the editors claim, is aimed at senior primary school learners who do not speak English as their first language, but who have English as a subject or as the language of learning. In my opinion, they have succeeded in this objective, despite a few minor complaints about the dictionary.

They have tried to include vocabulary that the primary school learner is likely to encounter, such as

- common everyday English words,
- subject jargon from their school programme, and
- some common South African English (SAE) words.

The words used in definitions are all explained in the dictionary, with the exception of "Southern Africa" and "consultation" which appear on the front of the dictionary. "South Africa" is explained, but which areas are included in the term "Southern Africa?" The inquisitive and observant learner would have to go to some other reference book to clear that up.

The question raised in my introduction is partially answered. Words such as "minibus" and "taxi" and "combi" are listed, but while sharing similar mean-

ings in the South African context, are not cross-referenced. The distinction referred to in the opening paragraph is not made. Also, everyday words such as "CD" and "deejay" are not included.

The inclusion of the curriculum words add to the usefulness of this dictionary, making it a tool to be used not only in English but also in other subjects. Words such as "cotyledon" (p. 61), "rhombus" (p. 235) or "quotient" (p. 228) are examples of subject jargon explained in simple terms, although "rhombus" could have carried a label, indicating it is used in mathematics.

Organisation of the Dictionary

The dictionary has the following sections:

- an explanation of the entries,
- a note to the teacher,
- some examples of how to use the dictionary in the classroom, and
- the entries, with the occasional illustration.

The explanation of the entries is a valuable source of learning material and even teachers of grades 8-9 (the first two years of high school) should find the dictionary useful.

The note to the teacher is clear and easy to follow, covering matters such as the guide word, entry, headword, pronunciation, part of speech, definitions and so on.

The sample exercises on how to use the dictionary in the classroom are a good starting-point and provide the teacher with many opportunities for advancing the learners' vocabulary and understanding.

The illustrations used are generally apt. They provide opportunity for further vocabulary extension. However, there are times when the illustrations could have been less confusing, e.g. what is the link between the example sentences, the headword "aim" and the illustration on page 5? The sentences read "Luvuyo threw the ball and aimed it at the player on the other side of the field", and "The man aimed the gun at the buck and then he shot it." Then follows the illustration of someone shooting with a bow and arrow at a target.

When one turns to the headword "bow" (p. 31) there is neither an illustration nor a cross-reference to page 5, nor any illustration which has bearing on the word "bow".

On page 95, there is an example sentence dealing with the flags of Namibia, Botswana and Mozambique, but the South African rather than the Mozambican flag is included in the illustration. If this is to be a Southern African dictionary, then such symbols ought to be depicted consistently.

I have used this dictionary beyond its target group and have found that both my learners whose home language is not English and the learners who have English as medium of instruction have found its uncluttered look accessi-

ble. A short explanation on how to use the dictionary was all that was required before they used it independently. In language teaching the dictionary has offered opportunities for vocabulary extension for the "weaker" learner. The explanation of the rules of how the dictionary has been written has provided the older learner with some insight into lexicography and made the use of other dictionaries easier.

This dictionary deserves a place in the classroom.

Michael Benjamin
Teacher: English (First and Second Language)
Belhar High School
Bellville
Republic of South Africa
(*bhsm1@mweb.co.za*)

Hanny Demeersseman (leiding en eindredactie), **Piet Vermeer** (eindredactie), **Roy Pheiffer**, **Fred Pheiffer**, **Anneli Terre Blanche** en **Sief Veltkamp-Visser** (redactie). *Kramers woordenboek Zuid-Afrikaans–Nederlands Nederlands–Zuid-Afrikaans*. 2000. 414 pp. ISBN 90-6882-378-7. Amsterdam: Elsevier. Prijs HFL16,90.¹

Woordenboeken die van het Afrikaans naar het Nederlands vertalen of omgekeerd, zijn schaars en beperkt van omvang. Verder richten zij zich bij voorkeur op die woorden waar een vertaling nodig is. De "woordenboeken" waar ik op doel, hebben daarom vaak iets weg van idiotica. Hiermee sluiten ze aan bij de verklarende woordenlijsten bij edities van Afrikaanse en Nederlandse literatuur in Nederland–Vlaanderen resp. Zuid-Afrika.

Het wordt dus langzamerhand tijd voor een echt woordenboek Afrikaans–Nederlands Nederlands–Afrikaans (van welke omvang dan ook) — te meer omdat er voor Afrikaans–Nederlands alleen nog maar antiquarisch gezocht kan worden: de twee edities van Bloem et al. dateren van 1947 en 1960, terwijl Asbeek Brusse alweer van 1983 is. Ook de *Lekkerleeslys* van Nothnagel (1990) is moeilijk te vinden. De beschikbaarheid van het Nederlands–Afrikaanse woordenboek van Dekker (1990) en van het taalgidsje van Veltkamp-Visser (1992) met de daarin bevatte thematische woordenlijstjes steekt hier gunstig bij af.

Nu is dan afgelopen jaar in de reeks kleine tweetalige Kramers-woordenboeken van Elsevier een deeltje Afrikaans verschenen: ruim 30 000 ingangen groot (voor beide afdelingen samen), met bij sommige ingangen wat uitdrukkingen en gezegdes en met een grammaticaatje van het Afrikaans. Het is een product voor de Nederlandstalige markt. De gebruikte metataal is dus Nederlands.

Hoewel dit laatste verkoop in Zuid-Afrika niet in de weg hoeft te staan, zou dat wel kunnen gelden voor de discutabele beslissing om het Afrikaans in dit boek consequent met *Zuid-Afrikaans* aan te duiden, wat in Zuid-Afrika voor sommigen aanstootgevend zou kunnen zijn en wat daar in ieder geval bevreemding zal wekken.

In Nederland en Vlaanderen ligt dit anders: daar zeggen nog steeds velen (ook jongeren) *Zuid-Afrikaans* — soms in vrije variatie met *Afrikaans*. En dat hangt weer samen met het feit dat het Nederlandse adjectief *Afrikaans* nooit zijn oorspronkelijke betekenis, 'van/uit Afrika', is kwijtgeraakt, zodat het ietwat bizarre Afrikaanse prefix *Afrika-* 'van/uit Afrika' geen plaatsje in het Nederlands heeft kunnen verwerven (afgezien van wat recente na-aperij onder in het Afrikaans geïnteresseerden). De uitgever kon dus kiezen uit twee homonieme adjectieven–taalnamen: *Afrikaans* en *Zuid-Afrikaans*. Door de conservatieve, tegen de trend ingaande keus voor de oudere, volksere benaming blijven we internationaal nog wat langer uit de pas lopen. Het woordenboekje zelf doet hier natuurlijk niet aan mee, al valt op dat men ndl. *Afrikaans* wel, maar ndl. *Zuid-Afrikaans* niet als homoniem behandelt. Dat is jammer, want het bij Neder-

landstaligen levende lekenidee dat het Afrikaans overall elders ter wereld natuurlijk ook naar Zuid-Afrika vernoemd is, kan niet vaak genoeg worden tegengesproken.

De redactie van dit woordenboek lijkt te bestaan uit drie Zuid-Afrikanen (Roy Pheiffer, Fred Pheiffer, Anneli Terre Blanche) en één Nederlandse (Sief Veltkamp-Visser). Aangezien de laatste, naar zij mij verteld heeft, alleen verantwoordelijk was voor het Afrikaanse grammaticaatje (pp. 409-414), lijken we te maken te hebben met een geheel Zuid-Afrikaanse redactie, met een Nederlandstalige eindredactie boven c.q. naast zich. Dit roept de vraag op welke onderdelen de redactie voor haar rekening heeft genomen: het deel Nederlands-Afrikaans, het deel Afrikaans-Nederlands of allebei?

Normaal gesproken is de redactie bij een bidirectioneel woordenboek (soms via subredacties) voor beide delen verantwoordelijk, maar een en ander geeft mij aanleiding om te denken dat de Zuid-Afrikaanse redactie het deel Nederlands-Afrikaans heeft aangeleverd, op grond waarvan door semi-automatische omkering het Afrikaans-Nederlandse deel gegenereerd is, dat — zoals dat gaat — vervolgens handmatig bewerkt moest worden. Bij dit laatste is de redactie vermoedelijk maar sporadisch betrokken geweest, al moet ze wel nog een aantal extra Afrikaanse woorden hebben aangeleverd, die echter niet (allemaal) door omkering weer in het Nederlands-Afrikaanse deel terecht zijn gekomen.

Een woordenboek Afrikaans voor de Nederlandstalige markt dient m.i. — of het nu uni- of bidirectioneel is — aan het volgende verlanglijstje te voldoen: (a) Een goede behandeling van de Afrikaanse voornaamwoorden en andere functiewoorden is gewenst, (b) Idem voor belangrijke "on-Nederlandse" Afrikaanse inhoudswoorden, waaronder (c) plant- en diernamen en (d) landschapnamen.

Aan (a) en (b) kan makkelijk voldaan worden, al is, afgezien van het geval van de voornaamwoorden, geen volledigheid of ideale keus te verwachten — en zeker niet in een klein woordenboek. Dit geldt eens te meer voor de deelverzamelingen (c) en (d).

Daarnaast dienen in een woordenboek Afrikaans voor Nederlandstaligen m.i. ook de volgende vorm-gerelateerde kwesties aan de orde te komen: (e) Afrikaanse variabele werkwoord-stammen zoals *sterf* / *sterwe* (voor het modernste Afrikaans misschien minder van belang), (f) Afrikaanse adjectieven met "on-Nederlandse" flexionele vormvariatie: *oud*, *lank*, *jonk*, *nuut*, (g) een indicatie per Afrikaans adjectief of en hoe het flechteert, en (h) een duidelijke markering van die paar Afrikaanse werkwoorden die qua scheidbaarheid niet met het Nederlands overeenkomen (m.n. *óorkom* 'overkómen' en het niet-scheidbare *hardloop* 'rennen'). Voor het overige zou er geminimaliseerd kunnen worden op de voor Nederlandstaligen toch overbodige informatie over het gebruik van het prefix *ge-*, al zou bij variatie zoals in (*ge*)*probeer* 'geprobeerd' informatie gewenst zijn. En tenslotte zou gewenst zijn: (i) een indicatie per Afrikaans tel-

baar naamwoord over de gebruikte meervoudsvorm, al kan ook hier geminimaliseerd worden.

Nu moeten we van een zakwoordenboekje, dat eigenlijk uit twee woordenboekjes van elk minder dan 200 bladzijden bestaat (Afrikaans–Nederlands: pp. 11-209, Nederlands–Afrikaans: pp. 213-408), niet al te veel verwachten, maar (f), (g), (h) en (i) en de voornaamwoorden horen er toch goed uit te komen. De rest is afhankelijk van lexicografische keuzes.

Hoe komt nu het Kramers-woordenboek (Zuid-)Afrikaans–Nederlands Nederlands–(Zuid-)Afrikaans (voortaan KWANNA) bij eerste inspectie op ons over? Hier doet zich onmiddellijk een verschil voor tussen het Afrikaans–Nederlandse en het Nederlands–Afrikaanse deel van KWANNA: het Nederlands–Afrikaanse deel lijkt verzorgd; je vindt meer annotatie, meer vertalingen per woord, contexten, betekenisnuances, enz. dan in het Afrikaans–Nederlandse deel. Bij een sterkte-zwakke-analyse aan de hand van (a) t.m. (i) in combinatie met dom bladerwerk treden er echter nog andere verschillen aan het licht.

Wat betreft de pronomina doet het Nederlands–Afrikaanse deel het redelijk tot goed, waarbij voor het Nederlands zelfs onderscheid wordt gemaakt tussen *zich* met een derde-persoons-antecedent en *zich* met *u* als antecedent, aangezien dat in het Afrikaans verschil maakt. Maar het valt op dat de vrije possessieven (*myne, joune*, enz.) niet behandeld worden en dat de korte reflexief *zich* van de derde persoon alleen maar met afr. *hom/haar/hulle-self* vertaald wordt en niet ook met *hom/haar/hulle* — terwijl er in het boekje voldoende inherent reflexieve werkwoorden genoemd worden. Verder staat er iets onduidelijks in het lemma voor ndl. *wie* (waarvoor zie onder). Een bevreemdende fout tenslotte is "**mijn** [...] *bez vnw myne*". Gelukkig wordt geen enkel ander bezittelijk voornaamwoord vertaald met de corresponderende vrije possessief. — Het Afrikaans–Nederlandse deel daarentegen doet het duidelijk slechter dan het Nederlands–Afrikaanse deel, met een aantal onnodige fouten, waarover later meer.

Wat betreft specifiek Afrikaanse inhoudswoorden (inclusief adverbia) scoort het Nederlands–Afrikaanse deel m.i. gemiddeld, terwijl ik bij de vertaling van Nederlandse inhoudswoorden betrekkelijk weinig fouten gevonden heb. Zo is de steensoort *mergel* vertaald met het Afrikaanse ww. *uitmergel*. Dit lijkt een toevallige vergissing, terwijl "**val** *zn* [...] sondeval [...]" op een Engels interface zou kunnen duiden. Maar ik heb daarvoor geen bevestiging kunnen vinden, zodat ook dit toeval kan zijn. Idem voor "**heen** *heen*; daarnatoe, soontoe; [...]", waar had moeten staan "[...] *er~* daarnatoe, soontoe; [...]". Alles bij elkaar maakt het Nederlands–Afrikaanse deel ook op dit punt een zorgvuldige indruk: zorgvuldig, dus weinig fouten. — Het Afrikaans–Nederlandse deel daarentegen krijgt op het punt van de inhoudswoorden bij mij een gemengde ontvangst. Het goede van het Afrikaanse deel is dat er allerlei woorden zijn opgenomen (feitelijk denk ik: achteraf toegevoegd) die ik in het Nederlands–Afrikaanse deel niet hebben kunnen traceren, zoals: "**piekel** *sjouwen*", "**pamper-**

lang vertroetelen, vleien, flikflooiën", "**tjaila** na het werk naar huis toe gaan", "**mamparra** domkop", "**maltrap** wildebras", enz. Het slechte van het Afrikaanse deel is echter dat het vol met fouten staat, waarover hieronder meer.

Overigens moet hier nog bij opgemerkt worden dat KWANNA als geheel op het punt van plant- en diernamen niet hoog scoort, wat jammer is voor de toerist die dit boekje mee naar Zuid-Afrika wil nemen. Qua landschapsworden doet KWANNA het beter, al ontbreken van deze toch betrekkelijk kleine set nog woorden als *krans* en *spruit*. Verder blijkt dat de kwestie van de Afrikaanse variabele werkwoord-stammen KWANNA niet echt interesseert, maar dat is niet vreemd gezien de teruggang van de sjwa-varianten die bezig lijkt te zijn.

Wat betreft flexionele en andere vormelijke informatie kan ik kort zijn. Dit hoort in het Afrikaans-Nederlandse deel van KWANNA te staan maar het staat er niet: geen informatie over adjectivale verbuiging, meervoudsvorming of (minder belangrijk) participia; geen indicatie dat "**oorkom** overkómen; óverkomen; [...]" in beide betekenissen een scheidbaar werkwoord is; geen indicatie dat "**hardloop** hollen, rennen" een niet-scheidbare samenstelling is. Op het punt van de oppositie attributief *ou*, *jong*, *lang* en *nuwe* vs. predicatief *oud*, *jonk*, *lank* en *nuut* daarentegen zijn beide delen van KWANNA opmerkelijk eensgezind: het adjectief *ou* ontbreekt, de overige vormen worden genoemd. Maar alleen in het Nederlands-Afrikaanse deel worden de vormen als attributief resp. predicatief geïdentificeerd. Dit is een van de redenen waarom het Nederlands-Afrikaanse deel zorgvuldiger overkomt dan het Afrikaans-Nederlandse deel.

Er is dus een kwalitatief verschil tussen de beide delen van KWANNA, in het voordeel van het Nederlands-Afrikaanse deel. Wat voor fouten vinden we echter in het Afrikaans-Nederlandse deel en waar komen die vandaan? Laat ik met de tweede vraag beginnen.

M.i. is de ruwe basis voor het Afrikaans-Nederlandse deel gegenereerd door semi-automatische omkering van het door de Zuid-Afrikaanse redactie aangeleverde Nederlands-Afrikaanse deel. Deze lexicografische basis moest vervolgens handmatig bewerkt worden en daarbij is nogal wat misgegaan. Gezien de aard van veel fouten lijkt het me verder onwaarschijnlijk dat de redactie daar veel bemoeienis mee heeft gehad. Wel moet de redactie nog extra woorden hebben aangeleverd, zoals ik boven al betoogd heb. En mogelijk zijn sommige toevoegsels weer in het Nederlands-Afrikaanse deel geïncorporeerd.

Evidentie voor mijn hypothese haal ik uit de vele fouten en eigenaardigheden die ik in het Afrikaans-Nederlandse deel heb aangetroffen, al is niet elke fout daarvoor bruikbaar. Zo is de weerspiegeling van de foute vertaling van ndl. *val* (zie boven) in het foute lemma "**sondeval** val" zonder bewijskracht. Er moet meer aan de hand zijn om van een omkeerfout te kunnen spreken. Mogelijk betere evidentie is het foute lemma "**myne** bez *vnw* mijn" in combinatie met het lemma "**my** me, mij", waarin de vertaling *mijn* (bez. vnw.)

ontbreekt. Deze combinatie van fouten is bij een Afrikaanstalige redacteur dermate onwaarschijnlijk, dat eerder aan het resultaat van semi-automatische omkering te denken valt. Alleen: ook de fout in het Nederlands–Afrikaanse lemma "**mijn** [...] *bez vnuw* myne" is dermate vreemd dat er misschien aan hele andere scenario's gedacht moet worden.²

Het zal bij dit alles niet verbazen dat ook de overige genoemde fouten van het Nederlands–Afrikaanse deel in het Afrikaans–Nederlandse deel terugkomen, en wel op zo'n manier dat aan (semi-)automatische omkering gedacht moet worden. Ten eerste: het bij eerste blik onmiddellijk dubieuze lemma "**uitmergel** *zn* mergel; *ww* slopen (m.b.t. ziekte)", dat uit de omkering van twee Nederlands–Afrikaanse lemmata voortkomt, is keurig geannoteerd maar de vertaling is blijkbaar niet in twijfel getrokken. Ten tweede, bij ontstentenis van lemmata voor ndl. *ernaartoe* en *daarnaartoe* kon het kleine foutje in het lemma *heen* leiden tot de foutieve Afrikaans–Nederlandse lemmata "**daarnatoe** heen" en "**soontoe** heen". De fout is zo evident dat een Afrikaanstalige redacteur dit bij controle zou hebben opgemerkt.

Maar er is meer aan de hand. Door de omkering zijn ook geheel nieuwe fouten ontstaan. Een verontrustend verschijnsel is dat allerlei Afrikaanse woorden van categorie en betekenis veranderd zijn. Zo is het naamwoord *tye* een voorzetsel geworden blijkens "**tye** onder", terwijl het "adjectief" *dikbek* blijkens "**dikbek** pruilen, mokken" een werkwoord geworden is. De corresponderende Nederlands–Afrikaanse lemmata zijn: "**onder** onder; tydens, ten *tye* van", "**mokken** dikbek/nukkerig/nors wees" en "**pruilen** pruil, dikbek wees". Het gebruikte omkeerprogramma focust blijkbaar op hoofdwoorden, en dan vooral op naamwoorden en adjectieven, wat zonder goede handmatige correctie tot rampen leidt, zoals ook waargenomen kan worden bij afr. "**kontant** *ww* verzilveren; [...]", dat is afgeleid uit ndl. "**verzilveren** [...]" in kontant omsit".

Gelukig zijn zulk soort fouten niet altijd aan de aandacht ontsnapt blijkens de Afrikaanse lemmata *nukkerig* en *nors*. En ook bij naamwoorden, die als onderdeel van werkwoordelijke uitdrukkingen door het omkeerprogramma makkelijk tot werkwoorden kunnen worden, is wel ingegrepen. Maar toch zijn er een paar onwaarschijnlijke fouten blijven zitten. Zo hebben de Nederlands–Afrikaanse lemma's "**informereren** *onoverg ww* inligting vra, navraag doen; [...]" en "**vragen** vra [...] navraag doen [...]" geleid tot Afrikaanse lemma's als de volgende: "**inligting** *zn* inlichtingen, informatie; *ww* informeren naar" en "**navraag** *zn* navraag; *ww* vragen, informeren". Vergelijk ook het Afrikaanse lemma "**hand** *zn* hand; *ww* verhandelen; afslaan; afzetten". De lezer moge zelf bedenken waar die drie werkwoorden *hand* vandaan zouden kunnen komen.

Overigens zijn er ook zwakheden van een minder erge soort die zich uit het gebruik van een omkeerprogramma laten verklaren. Zo wordt: afr. *hiernatoe* archaisch vertaald met 'herwaarts'. Dit komt doordat *herwaarts* wel een ingang heeft in het Nederlandse deel maar *hiernaartoe* niet. (Wat overigens een zwakte van het Nederlands–Afrikaanse deel is.) Verder worden gewone Afrikaanse woorden soms niet met "hetzelfde" Nederlandse woord vertaald maar met een

semi-equivalent. Vgl. "**honend** sarcastisch" en "**smalend** sarcastisch". Dit is door omkering afgeleid uit het Nederlands–Afrikaanse lemma "**sarcastisch** sarkasties; smalend, honend", dat niet gesecondeerd wordt door ingangen voor ndl. *honend* of *smalend*.

Hiermee heb ik een groot aantal typen fouten uit het Afrikaans–Nederlandse deel van KWANNA behandeld maar één type is nog niet aan bod geweest: de fouten bij de voornaamwoorden (vgl. boven). Ook hier heeft het omkeerprogramma een rol gespeeld.

Om te beginnen zijn er een drietal lemmata die zelfs een Nederlandstalige zouden moeten bevreemden: "**jou** *pers vnw* zich; jou; [...]", "**jouself** zich" en "**uself** zich". De laatste twee lemmata gaan via omkering terug op het volgende Nederlands–Afrikaanse lemma: "zich (*wederk vnw 3^e pers*) homself, haarself, hulleself; (2^e pers beleefdheidsvorm) sig *vero*, jouself, uself". Dit lemma geeft keurig aan dat het Nederlandse gecoïndexeerde paar *u ~ zich* in het Afrikaans vertaald moet worden met *ju ~ jou(self)* of *u ~ u(self)*. (M.i. is *-self* hier niet verplicht.) Deze grammatische informatie is bij de omkering blijkbaar verloren gegaan, zodat we nu twee nogal onzinnige lemmata hebben. Dit verklaart echter nog niet "**jou** *pers vnw* zich; jou [...]". Vanwaar "zich"? Dit heeft het omkeerprogramma waarschijnlijk afgeleid uit de vertaling van Nederlandse inherent reflexieve werkwoorden, zoals bijv. "**bevinden**: *zich ~ jou bevind*". Maar zonder grammatische aantekening is de vertaaloptie *zich* voor afr. *jou* hoogst onprecies.

Dit probleempje haalt het echter niet bij de problemen met het lemma voor het Afrikaanse pronominale element *wat*, dat op een aantal lemmata in het Nederlands–Afrikaanse deel teruggaat. Hier moest er wel handmatig aangevuld worden en daarbij is het een en ander misgegaan, mede omdat de redacteuren in het Nederlands–Afrikaanse deel hun explicaties soms wat kort hebben gehouden en omdat bepaalde informatie niet is meegenomen. Het bedoelde lemma nu ziet er als volgt uit: "**wat** *vr vnw* wat, waar; *aanw vnw* datgene; *betr vnw* wat; wie (met antecedent); welk".

Hier zitten allerlei fouten in. De bewering dat Afrikaans vragend *wat* ook *waar* zou kunnen betekenen komt waarschijnlijk voort uit slecht lezen van het volgende Nederlands–Afrikaanse lemma: "**waar** [...] *vrag vnw* waar? [...] *betr vnw* wat, waar". Dit verklaart tevens waarom in het lemma voor afr. *wat* sub "betr vnw" de optie *waar* ontbreekt. Dat ook de opties *die* en *dat* ontbreken moet onachtzaamheid van de bewerk(s) zijn geweest of een bug in het omkeerprogramma, want het Afrikaans–Nederlandse deel geeft deze voornaamwoorden met hun Afrikaanse vertaling *wat*.

De geheimzinnige optie "wie (met antecedent)" danken we echter aan nauwkeurige lezing van het volgende Nederlands–Afrikaanse lemma: "**wie** *vrag vnw* wie; *betr vnw* (met antecedent) wat, (sonder [sic] antecedent) wie; [...]". Aangezien de redactie op de plek van de puntjes met voorbeelden van afr. *wie se* en *aan wie* (met antecedent) aan komt zetten, bedoelt ze dus niet ndl.

wiens en *aan wie*. Ik houd het maar op het moderne gebruik van *wie* bij menselijk antecedent als in *de man wie ...* — m.i. onnodig, maar ook niet fout. Wel fout daarentegen is "*aanw vnw datgene*". Dit is te wijten aan een ommissie in het Nederlands–Afrikaanse lemma "**datgene** (dit) wat". Hoewel ook voor een Nederlandstalige duidelijk hoort te zijn dat hier eigenlijk "**datgene** ~ *dat, wat* (dit) wat" moet staan, is er klakkeloos overgenomen, terwijl het lemma "**het-geen** dit wat, wat" (dat elders wel verwerkt is) de bewerker(s) toch had kunnen waarschuwen.

Hoe had dit lemma er dan wel uit moeten zien? Wel ongeveer zo: "**wat** *vrag vnw wat; betr vnw die, dat; waar* (met afgesplitst vz); *welk(e); wat* (met ingesloten antecedent)", en dan kan nog worden toegevoegd: "*waar* (met ingesloten antecedent en afgesplitst vz); *voegw dat* (in bepaalde constructies)".

Tot zover de analyse van de fouten. De vraag is nu wat we hiermee aanmoeten.

Ik heb eerder gesteld dat we van een zakwoordenboekje als KWANNA niet al te veel moeten verwachten, maar mijn opsomming van fouten en zwakheden kan de indruk gewekt hebben dat ik van het tegendeel uitga.

Dat is echter niet het geval. Ik heb het boekje redelijk welwillend aan de hand van mijn lexicografische verlanglijstje geanalyseerd en daarbij heb ik lichte verschillen geconstateerd tussen het Afrikaans–Nederlandse en het Nederlands–Afrikaanse deel. De daarop volgende analyse van de fouten en zwakheden in het Afrikaans–Nederlandse deel heeft de bedoeling om te laten zien dat KWANNA eigenlijk uit twee producten bestaat. Ten eerste het Nederlands–Afrikaanse deel, dat redelijk "af" is, met wat foutjes en zwakheden. Dit kan op voor een verbeterde tweede editie, waarin dan tevens de aan het Afrikaans–Nederlandse deel toegevoegde woorden verwerkt kunnen worden. Daarnaast staat het Afrikaans–Nederlandse deel. Veel lemmata daarvan zijn natuurlijk in orde en een positief punt vormen de toegevoegde woorden. Maar naar mijn mening is dit te mechanisch samengesteld en zijn er teveel fouten blijven zitten, terwijl er bij de pronomina sprake is van onvoldoende deskundigheid. Dit deel had nooit zo de markt op gemogen en het zal voor een volgende editie volledig door de redactie of een andere groep deskundigen gereviseerd moeten worden, waarbij dan tevens enige flexionele informatie e.d. zou kunnen worden toegevoegd. Daarna is het wachten op een groter woordenboek.

Voetnoten

1. Voor een eerdere bespreking, geschreven voor algemeen publiek, zie Den Besten (2000).
2. Misschien is *myne* net als *joune* 'die/dat van jou, (predicatief) van jou' achteraf toegevoegd. Een Nederlandstalige kan dan de vertaling 'van mij' (vgl. het lemma van "**joune** van jou") "verbeterd" hebben, waarna *myne* als vertaling is toegevoegd aan het lemma van ndl. *mijn*. Dit zou betekenen dat de correcte vertaling *my* als enige van de bezittelijke voornaamwoor-

den in het Nederlands–Afrikaanse deel ontbrak en dus de invoeging van *myn*e mogelijk maakte. Maar zo'n omissiefoutje is denkbaar.

Verwijzingen

- Asbeek Brusse, C.E.** 1983. *Afrikaans–Nederlandse woordenlijst*. Amsterdam: Nederlands-Zuidafrikaanse Vereniging.
- Besten, H. den.** 2000. Een zakwoordenboekje Afrikaans. (Bespreking van Hanny Demeersseman et al. *Kramers woordenboek Zuid-Afrikaans–Nederlands Nederlands–Zuid-Afrikaans*. 2000. Amsterdam: Elsevier.) *Zuid-Afrika* 77: 98.
- Bloem, H., J.H. Broekman, F.J. Snijman en F.J. du Toit Spies.** 1947. *Beknopt Afrikaans–Nederlands woordenboek*. Samengesteld in opdracht van de Nederlandsch-Zuidafrikaanse Vereeniging te Amsterdam. Amsterdam: De Bussy.
- Bloem, H., J.H. Broekman, F.J. Snijman en F.J. du Toit Spies.** 1960². *Beknopte Afrikaans–Nederlandse woordenlijst*. Tweede herziene en uitgebreide druk door J.M. Tillema-de Vries. Uitgegeven onder auspiciën van de Nederlands Zuidafrikaanse Vereniging. Amsterdam: De Bussy/'s-Gravenhage: Van Stockum.
- Dekker, L.** 1990. *Nederlands–Afrikaanse woerdeboek*. Saamgestel deur ...; met medewerking van P.C. Paardekooper. Pretoria: Van Schaik.
- Nothnagel, A.** 1990. [*Lekkerleeslys*] *Afrikaans–Nederlandse woordenlijst*. Den Haag: in eigen beheer. (Latere drukken: Hoogezand: Stubeg.)
- Veltkamp-Visser, Sief.** 1992. *Afrikaans op reis: taalgids voor de Nederlandssprekende toerist in Zuid-Afrika*. Amsterdam: NZAV; Amsterdam: SAI (en latere drukken).

Hans den Besten
Algemene Taalwetenschap
Universiteit van Amsterdam
Amsterdam
Nederland
(*h.den.besten@hum.uva.nl*)

K.H. van Dalen-Oskam, K.A.C. Depuydt, W.J.J. Pijnenburg and T.H. Schoonheim (Editors). *Dictionaries of Medieval Germanic Languages: A Survey of Current Lexicographical Projects. Selected Proceedings of the International Medieval Congress, University of Leeds, 4-7 July 1994, 1997*, viii + 197 pp. ISBN 2-503-50601-1. Turnhout: Brepols. E-mail: publishers@brepols.com. Price: Belgian f1613.

From 4 to 7 July 1994 the University of Leeds hosted the first International Medieval Congress. Three sessions were dedicated to "Dictionaries of Medieval Languages". They dealt with the following themes: "Projects", in which several dictionary projects were presented, "Historical Background", which focuses on the importance of extensive knowledge of the historical background when working on a dictionary of a medieval language, and "Scribes and Scholars", elaborating on the philological problems a lexicographer encounters when working with old handwritten source material and its editions.

When preparing the papers presented at this congress for publication the editors tried, according to the introduction, to give a complete survey of the current lexicographic projects on Medieval Germanic Languages by urging the editors of the different dictionaries to write an overview of their projects for this volume. The first part therefore contains a description of the twelve projects: three on the Middle Dutch language (one of which also covers Early Modern Dutch), five on Old or Middle English, one on Middle High German and three on Old Norse or Old Norwegian.

Each overview gives information on the dictionary (the title, editors, institute and address), the contents of the dictionary (the type, subject/corpus), described period and area, and an example of an entry), the project and planning (a short history, the (planned) year of publication and the form of publication), and details on lexicographic tools and methods (the hard- and software). The information makes a comparison of the projects and their working procedures, and the types of dictionary and their social and cultural backgrounds possible. At the same time it suggests special aspects of historical lexicography such as the preparation of a corpus and delimiting the period and area.

The second part comprises seven papers which can be classified into three groups. With the exception of Tanneke Schoonheim's paper "The *Vroegmiddelnederlands Woordenboek* and the Computer" which describes how the computer has been applied in the compilation of the Dictionary for Early Middle Dutch, the other six papers deal with specific problems and problem areas encountered in historical lexicography regarding the earlier forms of a language: (a) aspects of determining the language period and area to be treated, (b) acquiring knowledge of the historical, cultural and social background of the treated period, and (c) preparing and interpreting texts used for source material. The only paper discussing (a) is "On the Preparation of a Dictionary of Middle Norwegian" by Erik Simensen. "The Importance of Legal History for Old Fri-

sian Lexicography" by Oebele Vries, "Encyclopaedic Aspects of the Historical Dictionary" by Jeffrey L. Singman, and "Identification of the Veins and Arteries in Middle English Literature" by James M. Girsch, all deal with (b). The two papers treating (c) are "The Lexicographer's Dilemma: Handling Mistranslations, Misconstruings and Paronomasia in Source Texts" by Elizabeth S. Girsch and "Lexicography and Philology" by Karina van Dalen-Oskam and Katrien Depuydt.

The most informative part of the work is the first, dealing with the different lexicographic projects. Unfortunately not all the projects approached for a contribution, seem to have responded so that it does not contain a complete overview of all the projects in progress. The selected papers in the second part of the book, although interesting, mostly deal with such specific problems and problem areas that, when reading them, only a few general outlines can be constructed for solving these problems and handling these problem areas. Nevertheless, they do make one aware of the directions in which to search for solutions when these problems and problem areas are encountered in historical lexicography.

J.C.M.D. du Plessis
Bureau of the Woordeboek van die Afrikaanse Taal
Stellenbosch
Republic of South Africa
(*wat@maties.sun.ac.za*)

Pharos. *Pharos Woordeboeke / Dictionaries 5 in 1: 5 ten volle geïntegreerde woordeboeke / Fully Integrated Dictionaries. Afrikaans-Engels / English-Afrikaans 1. Groot Woordeboek / Major Dictionary 2. Tweetalige Frasewoordeboek / Bilingual Phrase Dictionary 3. Nuwe Woorde / New Words Afrikaans 4. Verklarende Afrikaanse Woordeboek 5. Groot Tesourus van Afrikaans.* 2000. Pharos and Logos Information Systems. Prys R299.

In 2000 het 'n produk van Pharos verskyn wat die landskap van veral die Afrikaanse leksikografie ingrypend verander het. *Pharos Woordeboeke/Dictionaries 5 in 1* (voortaan Pharos 5 in 1) bied vir die eerste maal 'n geïntegreerde digitale biblioteek van woordeboeke met Afrikaans as behandelde taal wat ook kan inskakel by 'n groter digitale biblioteek op die Logos-platform. Vir die eerste keer word daar vir die gebruiker 'n kommersieel suksesvolle vertalende woordeboek met Afrikaans en Engels as behandelde taalpaar op CD-ROM gebied, nl. *Groot Woordeboek/Major Dictionary* met Pharos se *Nuwe Woorde/New Words* en *Tweetalige Frasewoordeboek/Bilingual Phrase Dictionary* ter ondersteuning. Die Afrikaanse verklarende leksikografie baat ook met aanvulling vir die ELHAT wat kom in die vorm van *Verklarende Afrikaanse Woordeboek*. Laastens word die *Groot Tesourus van Afrikaans* ter afronding van die biblioteek beskikbaar gestel.

In hierdie resensie sal daar nie gefokus word op die meriete van die individuele woordeboeke nie, aangesien hulle bloot rekenaarmatige aanpassings van reeds bestaande gedrukte woordeboeke is. Die klem sal dus eerder val op die elemente van die toegangstrukture wat by die soektogte op die interne en eksterne soekroetes hulp bied en die aktiewe kruisverwysings wat as deel van die mediostruktuur die woordeboeke se toeganklikheid en dus die gebruikersvriendelikheid daarvan verbeter.

Soektogte

Soektogte in individuele woordeboeke of oorkoepelende soektogte is moontlik by hierdie produk en word gefasiliteer deur 'n kragtige soekenjin wat deel is van die gekose platform, *Logos Library System*. Hierdie soekenjin kan veral op twee punte beoordeel word, nl. die toeganklikheid daarvan en die effektiwiteit van die soektogte wat dit toelaat.

Toeganklikheid van die soekenjin

Die *Logos Library System* bied 'n goed afgeronde en aanpasbare soekenjin wat vir eenvoudige tot komplekse soektogte voorsiening maak. Dit is egter juis vanweë hierdie kompleksiteit wat beskikbaar is, dat die soekenjin nie so toeganklik is as byvoorbeeld die aangepaste soekvorms wat beskikbaar is by die ELHAT nie. Die suksesvolle ontsluiting en benutting van Pharos 5 in 1 se soekvorms verg 'n gedetailleerde studie van die gebruikersgids.

Die gebruikersgids is egter op sigself vir die meeste gebruikers 'n intimiderende dokument wat byvoorbeeld moeilik deursoekbaar is en slegs op die CD-ROM beskikbaar is en nie ook as 'n gedrukte handleiding nie. Verder is die gebruikersadressering daarvan ook vreemd. Aangesien Afrikaanssprekendes een van die primêre teikengebruikersgroepe van Pharos 5 in 1 is, sou 'n mens met reg kon verwag dat die gebruikersgids ook in Afrikaans beskikbaar sou wees. Dit is egter nie die geval nie. Daar word tog voorsiening gemaak vir die behoeftes van Afrikaanssprekendes deur middel van 'n verkorte gebruikersgids getiteld *Quickstart/Wegspring 1-2-3!* wat die basiese soekfunksies vir die Afrikaanssprekende verduidelik, vrae antisipeer en beantwoord en die ikone verklaar, maar vir die meer gespesialiseerde gebruiker is die dokument nie gedetailleerd genoeg nie. Verder is geen van die ikoontitels, kieslystems of ander bevels op die soekvorms of kieslyste in Afrikaans beskikbaar nie. Hierdie gebruikersadressering benadeel die toeganklikheid van die produk.

Hierdie stand van sake is egter toe te skryf aan die gebruikersadressering van die *Logos Library System* en nie aan die samestellers nie. Met die volgende weergawe van Pharos 5 in 1 sal hierdie probleem hopelik opgelos word, aangesien 'n nuwe weergawe van *Logos Library System* dan dalk beskikbaar sal wees.

Een groot voordeel van *Logos Library System* is dat toegang tot soekvorms deur middel van 'n regsklik verkry kan word. 'n Kieslys word dan opgeroep waarop elemente van die eksterne sneltoegangstruktuur toegang bied tot 'n snelsoektog vir 'n ingekleurde woord of frase ("speed search"), of 'n soektog na die betrokke woord as lemma ("topic search"). Hierdie opsie versnel die soektogte van die gebruiker en verhoog dus die toeganklikheid van die produk aansienlik.

Effektiwiteit van die soekenjin

Hoewel die toeganklikheid van die soekenjin aandag verg, is dit tog 'n uiters effektiewe medium om veral komplekse soektogte, insluitende Boolese en "wildcard"-soektogte, uit te voer. Buiten die snelsoektogte wat deur middel van 'n regsklik geaktiveer word, bied Pharos 5 in 1 drie hoofsoekroetes. Eerstens kan elke "boek" afsonderlik oopgemaak word. Daar kan dan op- en afgerol word met behulp van die ingeboude rolstaaf. Hierdie metode is die naaste wat die produk kom aan 'n tradisionele naslaanervaring.

Tweedens het gebruikers die keuse om aan die linkerkant van die skerm 'n venster bekend as die "Library Browser" op te stel waarin hulle deur die afsonderlike woordeboeke se lemmalyste kan snuffel deur gebruik te maak van oopvoubare diagramme met die afsonderlike lemmas as eindkomponente. 'n Klik op die gekose lemma sal die gebruiker neem na die relevante stuk teks in die regtervenster. In Pharos 5 in 1 is verkies om soekresultate te skakel met die sentrale teks van die aktiewe woordeboek eerder as om slegs 'n lemma en die meegaande mikrostrukturele bewerking in die regtervenster op te roep soos byvoorbeeld die geval is in die *Microsoft Encarta World English Dictionary*. 'n

Volteksaanbieding is makliker om te implementeer en bied meesal 'n bevredigende naslaanervaring, maar dit kan die gebruiker se taak bemoeilik indien individuele lemmas uitgedruk of gekopieer moet word. Deeglike empiriese navorsing is nog in Suid-Afrika nodig om vas te stel wat verskillende gebruikersgroepe se voorkeure in dié verband is.

Die eerste twee soekroetes bied die gebruiker die meer tradisionele, stadiger, dog eenvoudiger soekroetes, maar die werklike innoverende waarde van 'n CD-ROM-woordeboek lê in die spesiaal ontwerpde soekvorms. By Pharos 5 in 1 het die gebruiker 'n keuse tussen 'n basiese soekvorm en 'n gevorderde soekvorm. Op die basiese soekvorm kan 'n leksikale item of frase as soekterm ingetik word, waarna die opsies uitgeoefen word om die optrede daarvan as lemma of bloot in die hele teks na te spoor en of al die woordeboeke of slegs een deurgesoek moet word. Hierdie vorm is op standaard met die meeste elektroniese woordeboeke en bied die gebruiker vinnige toegang tot die soekresultate wat dan diagrammaties aangebied word in 'n verdere venster. Van hier af kan individuele resultate vinnig en sonder moeite ontsluit word met 'n linksklik. Die leksikale item of frase waarvoor die soektog geloods is, sal dan in blou teks óf as lemma óf binne die mikrostruktuur (afhangende van die parameters van die soektog) in die sentrale teks van die aktiewe woordeboek verskyn.

Met 'n linksklik op die "more"-ikon kan die gebruiker 'n deurgang kry tot 'n gevorderde soekvorm. Aanvanklik bied hierdie soekvorm 'n aantal ekstra soekopsies (bv. "show all wildcard matches", "search across languages" en "show books' hits in context") en sekere operatore vir "wildcard"-soektogte. Binne in die soekvorm is daar egter ook twee kleiner, alternatiewe soekvorms wat met 'n linksklik opgeroep kan word. In die eerste word 'n volledige woordelys gegee in 'n venstertjie links en vormlik soortgelyke woorde regs. In die tweede word die komplekse Boolese operatore in een venster en 'n inleiding oor Boolese soektogte tesame met 'n verklaring van die funksies van elke operator in die ander venster gebied.

Die gevorderde soekvorm is 'n goed ontwerpde, funksionele werktuig, maar daar is tog 'n tekortkoming (bo en behalwe die feit dat dit slegs in Engels verskyn) wat uitgewys kan word. In vergelyking met ELHAT bied die soekvorms van Pharos 5 in 1 geen soektogte binne vooraf vasgestelde soekvelde in spesifieke woordeboeke nie. Hier moet in gedagte gehou word dat ELHAT se soekvorms spesifiek vir daardie enkele woordeboek ontwerp is, terwyl Pharos 5 in 1 se soekvorms deel is van die *Logos Library Browser*. Die produk se gebruikersvriendelikheid sal heelwat verbeter word indien sekere soekvorms vir spesifieke woordeboeke, veral vir die *Verklarende Afrikaanse Woordeboek*, ontwerp kan word, aangesien daar tans 'n komplekse prosedure met "wildcard"-operatore gevolg moet word om soektogte binne soekvelde uit te voer.

Aktiewe kruisverwysings

Pharos 5 in 1 maak deurgaans gebruik van aktiewe struktuurmerkers wat as kruisverwysings optree. Die prominentste van hierdie merkers is die teksinter-

ne, artikeleksterne aktiewe kruisverwysings wat in rooi letters onderstreep aangebied word en meesal as wisselvorm- of sinoniemverwysings dien. Hierdie verwysings bied uitstekend onmiddellike toegang tot verwyderde informasie. Ook die tekseksterne kruisverwysings word baie goed aangebied. Sommige afkortings, veral die redaksionele afkortings, word in die sentrale teks in groen aangebied. 'n Linksklik roep dan 'n klein ingevoegde teksie op waarin die leksikale item waarvoor die afkorting staan, vermeld word. Hierdie vernuwing verbeter die gebruikersvriendelikheid van Pharos 5 in 1 aansienlik.

Slot

Pharos 5 in 1 is 'n bevredigende produk wat toegevoegde waarde vir die gebruiker bied, vernaamlik met behulp van 'n goeie platform met 'n kragtige, effektiewe soekenjin. Met so 'n stewige basis om op te bou, kan toekomstige uitgawes van Pharos 5 in 1 met kleinere aanpassings hierdie produk vestig op die voorfront van die elektroniese leksikografie.

Phillip Louw
Buro van die Woordeboek van die Afrikaanse Taal
Stellenbosch
Republiek van Suid-Afrika
(*pal@maties.sun.ac.za*)

Pharos. *Pharos Speller: Speltoetsers en Woordafbrekers vir Afrikaans*. 2000. Pharos and Polderland Language and Speech Technology. Prys R299.

Van dieselfde uitgewer as *Pharos Woordeboeke/Dictionaries 5 in 1* kom die *Pharos Speller: Speltoetsers en Woordafbrekers vir Afrikaans* (voortaan *Pharos Speller*). Ook met hierdie produk bied Pharos innoverende taalleiding vir die Afrikaanse taalgebruiker. *Pharos Speller* is vinniger en omvattender as enige voorganger en skakel sonder moeite met vier produkte binne Microsoft Office. Dit neem dieselfde vorm aan as die verstekspeltoetsers in verskeie tale vir Microsoft Office, en het duidelik gebaat by die samewerking tussen Pharos en Polderland Language and Speech Technology wat die lewerransier van die Nederlandse verstekspeltoetsers is.

Daar is min om op kritiek te lewer by hierdie produk. Dit installeer maklik, beskik oor 'n deeglike, maklik verstaanbare handleiding en bied onmiddellike toegang tot speltoetsing en korrekte woordafbreking. Die woordafbreker werk op dieselfde wyse as dié vir ander tale en is 'n goed geïntegreerde produk wat maklik deur 'n gebruiker vertrouwd met Microsoft Office gebruik kan word. Die werking word ook goed in die handleiding verantwoord.

Die speltoetsers bevat heelwat van die kenmerke wat die verstekspeltoetsers van Microsoft Office so suksesvol gemaak het. Hier kan veral klem gelê word op die regs-klik-funksie wat vinnige toegang bied tot alternatiewe vir die verkeerd gespelde woord. Dit bied ook die opsie om die woord wat nie herken is nie, toe te voeg tot die verstekwoordeboek, wat baie nuttig is vir die herkenning en kontrolering van vaktaterme, samestellings of dergelike onbekende leksikale items wat gereeld deur 'n spesifieke gebruiker benut word. 'n Voorheen onherkenbare rekenaarterm soos *regs-klik* en samestellings soos *vaktaterme* en *taalleiding* kon byvoorbeeld tydens die skryf van hierdie resensie toegevoeg word.

Buiten die regs-klik-funksie is 'n ander sterk punt van die integrasie met veral Microsoft Word die keuse wat aan die gebruiker gebied word om die spelfoute in rooi onderstreep te sien of om 'n meer tradisionele soekroete te volg waartydens die spelfoute een vir een vertoon word en dan reggemaak, geïgnoreer of tot die woordeboek toegevoeg kan word. Meer vormlik soortgelyke spelvorme is ook hier teenwoordig as by die spyskaart wat deur 'n regs-klik opgeroep word. Outomatiese regstelling ("Autocorrect") word ook hier as 'n opsie gelys, wat dan help om te verseker dat leksikale items wat uit gewoonte verkeerd gespeld word, onmiddellik outomaties gekorrigeer sal word. Verdere opsies kan ook opgeroep word waar onder meer afgebaken kan word watter spelvorme nagegaan moet word, of die spelling by afhandeling van elke getikte woord nagegaan moet word en in watter woordeboek toegevoegde spelvorme geberg moet word.

Daar is egter een haakplek met hierdie produk. Komposisionaaliteit is nog nie ingebou in hierdie weergawe van *Pharos Speller* nie. Met die verskyning van die volgende weergawe behoort hierdie probleem egter uit die weg geruim

te wees. Intussen kan gebruikers dié samestellings, afleidings, of dergelike wat frekvent in hulle tekste voorkom, met behulp van die bevel "Add" voeg by die bestaande woorde wat die Pharos Speller herken.

Pharos Speller het 'n nuwe standaard gevestig vir Suid-Afrikaanse elektroniese produkte van hierdie aard in 'n mark wat vinnige groei toon. Indien die voorgestelde aanpassings gemaak word in die volgende weergawe sal dit ook as 'n produk van internasionale statuus erken kan word, wat op standaard is met Microsoft Office se amptelike speltoetsers vir ander tale.

Phillip Louw
Buro van die Woordeboek van die Afrikaanse Taal
Stellenbosch
Republiek van Suid-Afrika
(*pal@maties.sun.ac.za*)

F. Heyvaert, A. Moerdijk, M. Mooijaart, M. Smits en R. Tempelaars (redactie). *Het grootste woordenboek ter wereld: Een kijkje achter de kolommen van het Woordenboek der Nederlandsche Taal (WNT)*, s.j., 434 pp. ISBN 90 75566 96 4. URL: <http://www.sdu.nl>. Den Haag: SDU Uitgevers / Antwerpen: Standaard Uitgeverij. Prys f49.90.

Omvang

In die eerste gedeelte van hierdie resensie gaan ek aan die hand van bepaalde artikels probeer aantoon in watter opsig die *WNT* as die grootste woordeboek ter wêreld beskou kan word.

In sy artikel "Het WNT in cijfers" bied A. Moerdijk syfers aan om die omvang van die *WNT* aan te toon. Hy noem onder andere dat die *WNT* 30 dele beslaan wat in 40 volumes ("banden") ingedeel is. Dit sluit die een supplement-deel in. (Daar is meer volumes as dele, aangesien sommige dele om bepaalde redes in kleiner dele of "deelstukke" opgedeel moes word. Elke deel en "deelstuk" kry 'n eie volume.) Verder beslaan die *WNT* 45 805 bladsye (of, volgens 'n spesiale berekening van die drukker van die *WNT*, meer as 117 000 gewone "boekbladsye"). Moerdijk beweer dat die omvang van 'n woordeboek deur die hoeveelheid karakters bepaal word. Volgens 'n berekening van die drukker bevat die *WNT* meer as 309 miljoen karakters.

'n Verdere syfer wat Moerdijk aangee, is die aantal sitate wat die redaksie gebruik het vir die bewerking van die *WNT*, die sogenaamde "bruto citaten". Op versoek van Moerdijk is daar bereken dat daar vir die bewerking van een kolom in die *WNT* 29,7 bruto sitate gebruik is, wat neerkom op 'n totaal van meer as 2,7 miljoen sulke sitate. Aangesien daar verskillende metodes is waarvolgens hierdie berekening gedoen kan word en die resultaat telkens kan varieer, beraam Moerdijk die aantal bruto sitate op tussen 2,5 miljoen en 3 miljoen. Verder is daar bereken hoeveel van hierdie sitate in die woordeboek opgeneem is, die sogenaamde "netto citaten". Aangesien die gemiddelde aantal netto sitate per kolom wissel van redakteur tot redakteur en tussen die onderskeie woordeboekgeslagte, varieer die resultaat van hierdie berekening ook. Na 'n verskeidenheid van redelik ingewikkelde bewerkings stel Moerdijk hierdie sitate op ongeveer 1,6 miljoen vas.

Naas bogenoemde syfers bepaal Moerdijk op grond van steekproewe dat daar tussen 350 000 en 400 000 behandelde woorde in die *WNT* voorkom. Hy verskaf ook interessante syfers oor onder andere watter persentasie van die geheel elke afsonderlike letter uitmaak, die aantal bronne waaroor die *WNT* beskik en die getal intekenare van die *WNT*.

Met gegewens soos dié hierbo is dit seker moontlik om met redelike sekerheid die aanname te kan maak dat die *WNT* die grootste woordeboek ter wêreld is!

Die omvang van die *WNT* word egter nie net met syfers uitgebeeld nie. In sy artikel "WNT-kwalifikasies" toon Moerdijk watter enorme belangstelling

daar in en waardering daar vir die woordeboek bestaan. Volgens hom het daar vanaf die beginfase al meer as 1 000 publikasies oor die *WNT* verskyn.

In hierdie publikasies word daar nie geskroom om die omvangrykheid van die woordeboek met woorde soos "reusachtig", "gigantisch" en "enorm", met woorddele soos "reuzen-", "mammoet-" en "mega-", of met 'n frase soos "het grootste woordenboek ter wereld" te beskryf nie. Verder word die *WNT* dikwels gesien as 'n skatkamer, museum of monument in frases soos "deze enorme schatkamer van onze taal", "dat grootse taalmuseum" en "a cultural monument of inestimable value". Met woorde soos "groots", "legendarisch" en "respectabel" word die kwaliteit of waarde van die *WNT* weer uitgedruk, terwyl frases soos "de bron van alle hedendaagse woordenboeken" en "de moeder aller handwoordenboeken" in publikasies gebruik is om die besondere posisie van die *WNT* ten opsigte van ander woordeboeke aan te toon.

Moerdijk toon met emosioneel en ideologiese gelaaiete frases soos "een zoo grootsch en echt vaderlandsch werk" (1852) en "eene aangelegenheid van nationaal belang" (1864) dat daar aanvanklik 'n nasionale benadering tot die *WNT* gevolg is. In die Inleiding tot Deel I van die *WNT* word onder andere geskryf dat die *WNT* "eene edele roeping" het, dat dit 'n middel is om die nasionale bewussyn te voed en te versterk, en dat die *WNT* gebruik moet word om "de liefde voor het vaderland" (dit is, Nederland) te verhoog.

Hierdie nasionale benadering, waarin Nederland as "vaderland" beklemtoon is, het later plek gemaak vir 'n benadering waarin Nederlands as taal beklemtoon word. Dick Wortel skryf in die artikel "De taal. Niet het land" dat die *WNT* 'n bepaalde betekenis vir die Nederlandse taalgebied gehad het. So het die *WNT* byvoorbeeld 'n invloed gehad op die posisionering van Nederlands in Vlaandere nadat die Belgiese staat in 1830 uitgeroep is. In hul verset teen Franstalige oorheersing het die Vlaminge die *WNT* as 't ware as wapen gebruik om aan te toon dat Nederlands "kwalitatief gelykwaardig" is aan Frans. Hier word daar dus op die taalgebied van Nederland en Vlaandere, waar Nederlands gemeenskaplike besit is, gefokus, en nie op 'n spesifieke nasie nie.

Organisasie

Wanneer 'n mens die artikel van P.G.J. van Sterkenburg oor die hoofredakteurs van die *WNT* lees, kry jy reeds 'n aanduiding van die organisasie van die *WNT*. Matthias de Vries wat in 1851 met 'n preskriptiewe woordeboek van 19de-eeuse Nederlands begin het, kan seker as die voorloper van die *WNT* as woordeboek beskou word. Met sy *Ontwerp* en *Inleiding* het hy die basis vir die *WNT* gelê. Hy het ook begin met materiaalversameling, die selektering van woorde wat opgeneem moet word deur middel van ekserpering, en die redigering van artikels van ander redakteurs.

Dit is egter onder Jacoba van Lessen wat in 1946 as hoofredakteur aangestel is, dat die *WNT* as organisasie begin vorm aanneem het. Onder haar krag-

tige leiding het daar 'n samehorigheidsgevoel onder alle medewerkers ontstaan, is samewerking tussen redaksieledede tot "skitterende" spanwerk uitgebou en is die WNT tot 'n homogene geheel omvorm. Jare daarna, in die periode wat Van Sterkenburg as hoofredakteur waargeneem het, is die WNT verder as organisasie uitgebou deurdat daar met beplanning begin is. Hierdie beplanningsfase word veral deur W.E.M. de Clerck wat in 1978 as hoofredakteur aangestel is, verteenwoordig. Hy het as bestuurder, eerder as wetenskaplike, organisatories 'n belangrike rol gespeel. Onder sy hoofredakteurskap is die eerste langtermynbeplanning gedoen. Nadat die termyn waarbinne die WNT voltooi moes word, vasgestel is, het hy, in samewerking met A. Verschelling, begin om die oorblywende materiaal op bepaalde wyses te reduseer sodat dit binne die gegewe tydperk bewerk sou kon word. Soos Moerdijk dit stel in sy artikel "Het WNT als organisatie", het De Clerck en Verschelling die organisatoriese orde en die administratiewe logistiek gevestig op basis waarvan die voltooiing van die WNT binne die gestelde termyn moontlik sou word.

Daar was egter 'n tydperk in die geskiedenis van die WNT dat daar nie sprake was van enige organisatoriese orde nie. Veral die tydperk van die vierde geslag redakteurs van die WNT is gekenmerk deur 'n gebrek aan eensgesindheid. Daar was nie 'n hoofredakteur nie, en werksaamhede het kortsigtig en onbepland verloop. Daar het 'n onsekerheid oor die toekoms geheers. Almal was daarvan oortuig dat die WNT nooit voltooi sou word nie, tensy daar iets ingrypends plaasvind.

Nadat die WNT in 1967 as afdeling by die Instituut voor Nederlandse Lexicologie (INL) ingelyf is, is hierdie "lossere beeld", soos Moerdijk dit stel, vervang met "een strakke, zakelijker aanpak". Die INL het in die middel van die sewentigerjare 'n aantal besluite geneem om die voltooiing van die WNT voor die einde van die eeu moontlik te maak. Daar is besluit om 'n moratorium op die materiaalversameling te plaas, om 1921 as afsnypunt vir voorbeeldmateriaal te stel ten einde die omvang van die woordeboek te beperk, om 'n hoofredakteur aan te stel wat die redaksie moes rig en saambind, en om die redaksie uit te brei. Bogenoemde organisatoriese orde wat onder Van Sterkenburg en De Clerck begin vorm aanneem het, kan as 'n direkte uitvloeisel van hierdie besluite beskou word.

Onder die leierskap van De Clerck en die huidige hoofredakteur, A. Moerdijk, wat De Clerck in 1989 opgevolg het, is die WNT verder uitgebou deur die instelling van vier duidelike komponente binne die organisasie. Hulle is materiaalversameling, die redaksie, korreksie en manuskripvoorbereiding, elkeen met sy eie hoof of koördineerder. Die hoofredakteur moes nie net toesien dat die redaksionele bewerking vlot verloop, byvoorbeeld deur produksiekontrole toe te pas, nie, maar moes ook toesien dat die werksaamhede binne elkeen van hierdie komponente volgens plan verloop en dat elkeen mekaar onderling goed aanvul.

In sy artikel oor materiaalversameling gee Adriaan Verschelling 'n deeglike uiteensetting van die funksie en werkswyse van die komponent wat hier-

mee gemoeid is. Volgens Verschelling, wat eers as taalkundige assistent en later as hoof van hierdie komponent werksaam was, het 'n redakteur sitate nodig vir die betekenisomskrywing van 'n woord. Die funksie van die materiaalversamelingskomponent is die versameling en gereedmaak van die sitate vir redaksionele bewerking. Vroeër, vanaf 1966 tot die helfte van 1976, is die sitate verkry deur hulle uit te skryf, maar vanaf 1976 tot 1994, is die sitate geskep deur hulle te fotokopieer (die sogenaamde "xeroxcitaten"). Hierdie besluit is geneem om die materiaalversameling vinniger af te handel.

Verschelling gee 'n breedvoerige oorsig van die werksaamhede van die taalkundige assistente en assistente-materiaalversameling binne hierdie komponent, veral van die ontwikkeling wat die werksaamhede ondergaan het. Daar word onder meer gekyk na maniere van ekserpering, die optekening van woorde uit die registers van boeke, die skep van sitate (hetsy deur dit uit te skryf of te kopieer) en die ordening daarvan. Die metodes van materiaalversameling wat deur hierdie komponent toegepas is, sou as riglyne vir enige woordeboek kon geld.

Die belangrikheid van beplanning en onderlinge samewerking tussen die onderskeie komponente binne die WNT kom duidelik na vore wanneer 'n mens De Jong en Smits se artikel oor manuskripvoorbereiding en korreksie lees. Hieruit blyk dat die werksaamhede van 'n redakteur, die korreksie- en manuskripvoorbereidingskomponente en die hoofredakteur nou met mekaar geskakel is. Wanneer 'n redakteur 'n artikel of reeks artikels klaar bewerk het, gaan dit na die korreksiekomponent, waar dit gelees word om moontlike foute op te spoor en verskeie aspekte daarvan gekontroleer word. Van hier af gaan die werk na die manuskripvoorbereidingskomponent, waar dit drukklaar gemaak word deur dit in 'n teksverwerker in te tik. Die resultaat gaan as eerste weergawe terug na die korreksiekomponent, wat dit veral vir tikfoute kontroleer. Nadat korreksies aangebring is, word 'n tweede weergawe in viervoud uitgedruk, waarvan onder andere die hoofredakteur, die redakteur wat die materiaal bewerk het en 'n kollega 'n kopie ontvang. Die opmerkings van die hoofredakteur en kollega word deur die redakteur op sy kopie bygewerk, en word dan vir verdere aanpassings by die korreksiekomponent ingelewer. Hierna word 'n sogenaamde "definitiewe weergawe" gemaak, wat weer eens krities deurgelees word vir foute. Hierdie weergawe gaan dan na die manuskripvoorbereidingskomponent, waar die dokument gereed gemaak word vir die drukker. Die drukproewe wat van die drukker ontvang word, word nog 'n keer deurgelees. Hierdie ingewikkelde proses van samewerking sou nie kon vlot sonder behoorlike beplanning en 'n goeie organisatoriese orde nie.

Die waarde van 'n goeie organisatoriese struktuur kom ook duidelik na vore in Colman en Mooijaart se artikel oor die opleiding van redakteurs by die WNT. Alhoewel daar in die beginfase onder De Vries wel van 'n opleidingsituasie gepraat kan word, was daar geen presiese aanduiding van wat die inhoud van sodanige opleiding behels het nie. Daar kan ook slegs afleidings gemaak word oor die tydperk van opleiding, naamlik dat dit tussen een en vier jaar

geduur het. Ten spyte van die redaksionele handleidings en dokumente wat tussen 1940 en 1970 ontstaan het, was daar nog steeds nie juis sprake van 'n amptelike opleidingsituasie nie.

Eers nadat daar ingrypende organisatoriese veranderinge plaasgevind het met die WNT se inlywing by die INL, is daar in 1976 'n nuwe opleidingstruktuur ontwerp. Hierdie vernuwing moes noodwendig volg, gegee al die ander moderniserings wat gedurende hierdie tydperk by die WNT plaasgevind het. Met die nuwe opleidingsprogramme is daar 'n vaste inhoud aan die opleiding van redakteurs gegee. Daar is ook 'n vasgestelde periode daaraan gekoppel, naamlik twee jaar. Gedurende die eerste twee maande moet 'n aspirantredakteur 'n grondige kennis opbou van die geskiedenis van die WNT, van die materiaal waaroor die WNT beskik, en van die skep van sitate. In die oorblywende gedeelte van die eerste jaar word die aspirantredakteur opgelei in redaksionele vaardighede. Indien so 'n redakteur ná die eerste jaar van opleiding gunstig beoordeel word, kan hy/sy met die tweede jaar van opleiding begin. Tydens hierdie periode werk die aspirantredakteur min of meer selfstandig.

So 'n deeglike en goed omskryfde opleidingsprogram is, soos die skrywers van die artikel tereg opmerk, "onmisbaar voor een modern woordenboekproject".

In die laaste artikel in die afdeling oor die WNT as organisasie, bied Rob Tempelaars 'n uiters breedvoerige en interessante kykie op die korrespondensie wat oor woorde tussen die redakteurs, die medewerkers en die gebruikers van die WNT gevoer is. Hy fokus veral op die uitgaande korrespondensie (die korrespondensie "van binnen naar buiten"), waarmee redakteurs ander persone of instellings om inligting aangaande 'n bepaalde woord of woorde vra, en op die inkomende korrespondensie van hierdie persone of instellings (die korrespondensie "van buiten naar binnen"), waarmee daar op die vrae van die redakteurs gereageer word. Daar is reeds in die tyd van De Vries met uitgaande korrespondensie begin toe hy proefartikels vir beoordeling aan sogenaamde "deskundige vrienden" gestuur het. Daar is later selfs spesiaal gedrukte vorms gebruik waarop vrae oor 'n bepaalde woord of woorde aan medewerkers gestel is.

Aangesien die vyfde geslag redakteurs van die WNT die inwinning van inligting tot die minimum beperk het, het die rigting van korrespondensie gedurende die mees onlangse periode hoofsaaklik inkomend van aard geword. Volgens Tempelaars nader meer mense as ooit tevore gedurende hierdie periode die WNT skriftelik, maar dikwels ook telefonies, om hulp oor bepaalde taalkwessies.

Volgens Tempelaars getuig die wedersydse korrespondensie oor woorde van 'n vrugbare samewerking tussen die redakteurs, die medewerkers en die gebruikers van die WNT. Vir my getuig dit ook van 'n behoorlike organisatoriese struktuur en deeglike beplanning, aangesien die ekstra werkslading wat die inkomende korrespondensie op die redakteurs geplaas het, hulle nie daarvan weerhou het om die WNT binne die gestelde termyn af te handel nie.

Werkzaamhede en probleme

In die tweede afdeling van die boek, naamlik daardie deel wat handel oor leksikografiese aspekte, kry die leser 'n gevarieerde beeld van die werkzaamhede van die *WNT*-redakteurs.

In die eerste artikel in hierdie afdeling gee Moerdijk 'n deeglike uiteensetting van die werkswyse wat gevolg is om 'n artikel in die *WNT* tot stand te bring. Aan die hand van die woord *voorloop* illustreer hy hoe die redaksionele proses verloop. Dit begin wanneer die redakteur die sitaatmateriaal van die materiaalversamelingskomponent ontvang. Hierna analiseer die redakteur die sitate deur daardie sitate wat sekere kenmerke deel, by mekaar te plaas en hulle van ander sitate wat weer as 'n groep ander kenmerke deel, te onderskei. Sodoende word die sitate in verskillende betekenisonderskeidings gegropeer. Hierna word die gegropeerde betekenisonderskeidings (in die geval van die *WNT*, chronologies) georden. Wanneer die struktuur vasgelê is, word hierdie voorlopige weergawe van die artikel omgesit in 'n sogenaamde "definitiewe teks" deur die definisies af te rond en te verfyn. Hierna word die kop, wat nie-semantiese inligting soos grammatikale kenmerke, vormvariante en etimologiese gegewens insluit, boaan die betekenis- en sitaatgedeelte geplaas. Laasgenoemde gedeelte kan weer gevolg word deur afleidings en samestellings wat nie hul eie bewerking in 'n afsonderlike artikel vereis nie, en bloot onverklaard onder die basisvorm aangebied kan word.

In die oorblywende artikels word 'n seleksie van probleme waarmee die leksikograaf tydens redaksionele bewerking gekonfronteer word, aangebied. In die eerste van hierdie artikels wys Beijck en Westgeest op een van die vele taalkundige probleme waarmee 'n *WNT*-redakteur te doen kry, naamlik die problematiek om vas te stel tot watter woordsoortlike kategorie die eerste lid van 'n samestelling behoort. Aan die hand van verskeie voorbeelde van samestellings met *werk* as eerste lid illustreer hulle dat die grammatikale kategorie van so 'n eerste lid bepaal kan word deur na te gaan hoe die samestelling as 'n geheel gevorm is. Deur byvoorbeeld vas te stel wat die semantiese verhouding tussen die lede van die samestelling is, of deur te kyk na ander samestellings wat op 'n soortgelyke manier gevorm is, kan afgelei word of die samestelling met 'n selfstandige naamwoord of met 'n werkwoord as eerste lid gevorm is.

'n Tweede probleem word deur Hans Heestermans aan die hand van die besitlike voornaamwoorde *uw* en *zijn* verduidelik. Volgens hom is hierdie besitlike voornaamwoorde, wat saam met ander soos *hun*, *jouw*, *mijn* en *ons* 'n groep van soortgelyke woorde vorm, nie op dieselfde manier in die *WNT* hanteer nie aangesien hulle in alfabetiese volgorde bewerk is. In die geval van 'n alfabetiese bewerking is dit moeilik om die verskil in die aantal sitate wat daar vir verwante woorde bestaan, raak te sien. Dat daar ongeveer 900 sitate met *zijn* bestaan, maar slegs ongeveer 500 met *uw*, is nie deur Heestermans, wat albei woorde bewerk het, opgemerk nie. Die verskil in die aantal sitate het noodwendig tot 'n verskil in bewerking aanleiding gegee, spesifiek wat betref

die aantal betekenisonderskeidings wat daar vir elk van die woorde aangebied is. Indien die besittlike voornaamwoorde egter tematies, met ander woorde as 'n groep, hanteer is, sou hierdie verskille opgemerk kon word, en sou die onderskeie bewerkings meer met mekaar ooreengestem het.

In sy artikel "Rafels aan de grammatica" stel Frans Heyvaert dit dat die probleme wat hy bespreek, nie met definiëring verband hou nie, maar eerder met die werking van grammatikale kategorieë. Die eerste probleem wat hy demonstreer, hou verband met die onderskeid tussen, wat hier genoem word, verbindings en koppelings, spesifiek wanneer 'n verbinding as koppeling (dit is, as een woord) opgevat word. Hy toon ook aan dat alle koppelings nie noodwendig woordstatus geniet nie, en dat die begrip "woord" nie presies begrens kan word nie. Verder toon Heyvaert hoe moeilik dit is om voegwoorde van betreklike voornaamwoorde te onderskei, asook om 'n onderskeid tussen bepalinge van gesteldheid en bywoordelike bepalinge te maak. Aangesien Heyvaert hierdie probleme self as "marginale probleemstukke" beskou, gaan ek nie verder hierop in nie.

In twee van die oorblywende artikels word probleme rakende sitate aangespreek. In die eerste hiervan gee Dirk Kinable 'n goeie oorsig van die probleemsituasies waarmee 'n *WNT*-redakteur te doen kry wanneer hy/sy met moeilik interpreteerbare sitate gekonfronteer word. Hy stel dit dat sodanige sitate dikwels "een gevoel van vervreemding en onbegrip" by die leksikograaf laat ontstaan, en bied spesiale ondersoekstrategieë aan wat gebruik kan word om sulke interpretasieprobleme te kan hanteer.

Die tweede van hierdie artikels handel oor die verskynsel van identiese of byna identiese sitate in die taalmateriaal van die *WNT*. Volgens Lambermont en Westgeest gebeur dit soms dat 'n redakteur 'n bepaalde sitaat teëkom wat (min of meer) ooreenstem met een wat hy/sy reeds vantevore hanteer het. Hierdie sogenaamde "dubbelcitaten" kan volgens hulle onder meer die gevolg wees van oorskrywing ("plagiat") of nabootsing ("imitatio") deur 'n ander skrywer. Dit kan egter ook die gevolg wees van die manier waarop die materiaalversameling van die *WNT* opgebou is, deurdat daarin verskillende uitgawes van dieselfde werk kan voorkom. Volgens die skrywers is dit baie moeilik om sulke ooreenstemmende sitate te herken, veral wanneer die woorde waarvoor die betrokke sitate geskep is, alfabeties ver uitmekaar geplaas is. In gevalle waar die ooreenstemmende sitate wel naby aan mekaar voorkom en as "dubbelcitaten" herken word, moet die redakteur verkieslik slegs een opneem. Indien een van hierdie sitate egter afkomstig is uit 'n periode wat nie deur die ander sitate gedek word nie, kan die redakteur besluit of die sitaat nie wel opgeneem behoort te word nie.

In die laaste artikel oor leksikografiese aspekte word daar gekyk na hoe ander woordeboeke as sekondêre bron in die *WNT* gebruik is. Marijke Mooijart gee 'n deeglike uiteensetting van die voor- en nadele van die gebruik van woordeboeke as bron. Op grond van spesifieke gedeeltes uit die *WNT* bevind sy onder meer dat sitate uit woordeboeke, in vergelyking met literêre sitate, 'n

baie klein gedeelte daarvan uitmaak, en dat sodanige sitate hoofsaaklik in die kopgedeelte van 'n artikel aangebied word, 'n posisie waar sekondêre materiaal verwag word. Sy sluit af deur bepaalde beginsels vir die gebruik van woordeboeke as bron aan die hand te doen, byvoorbeeld dat inligting uit woordeboeke, gegee die andersoortigheid daarvan, van die primêre materiaal geskei en apart daarvan aangebied moet word, dat woordeboeke slegs as bron gebruik moet word wanneer daar nie voldoende primêre bronne bestaan nie, en dat 'n lys van gesaghebbende woordeboeke wat geraadpleeg mag word, vooraf vasgestel moet word.

Die artikels in hierdie boek is insiggewend omdat hulle deur praktiserende leksikograwe geskryf is. Die artikels wat ek in die eerste gedeelte van die resensie bespreek het, gee 'n duidelike beeld van die omvang van die *WNT*, terwyl die artikels in die tweede gedeelte weer illustreer dat beplanning en 'n goeie organisatoriese struktuur noodsaaklik is vir 'n woordeboekprojek om suksesvol te wees. Aangesien die artikels wat in die laaste gedeelte aan bod gekom het deur leksikograwe met heelwat praktiese ondervinding geskryf is, het dit ook geldigheid vir *alle* leksikograwe wat dikwels dieselfde of soortgelyke probleme ondervind.

Gerhard van Wyk
Buro van die Woordeboek van die Afrikaanse Taal
Stellenbosch
Republiek van Suid-Afrika
(gjvw@maties.sun.ac.za)

A. Moerdijk en R. Tempelaars (redactie). *Van A tot Z en verder ...: Lezingen bij de voltooiing van het WNT*, 1999, 112 pp. ISBN 90 75566 79 4. URL: <http://www.sdu.nl>. Den Haag: SDU Uitgevers/ Antwerpen: Standaard Uitgeverij. Prys f24.90.

In die voorwoord vermeld die redaksie dat die bundel saamgestel is uit lesings wat tydens 'n simposium ter viering van die voltooiing van die *Woordenboek der Nederlandsche Taal* (WNT) gelewer is. Dié feesgeleentheid het op 3 Desember 1998 in Antwerpen plaasgevind, en is deur die Instituut voor Nederlandse Lexicologie (INL) georganiseer. Tydens hierdie simposium het drie hoofredakteurs, drie vakkundiges en twee skrywers as sprekers opgetree.

Die skrywers, Kees Fens en Monika van Paemel, bied uiteenlopende weergawes van hul onderskeie kennismakings met die WNT aan.

Fens skilder 'n enigsins somber prentjie. Hy begin deur die leed en droefheid wat die WNT vroeg in die 20ste eeu moes verduur, weer te gee, en konsentreer verder veral op die verskil tussen die gebruik van woorde in 'n gedig en die aanbod van woorde in 'n woordeboek. Volgens hom toon 'n woordeboek, anders as 'n stuk poësie, nie 'n samehang nie: "Een verzameling woorden is niets." Hy stel dit onomwonde dat hy om hierdie rede nie 'n begeerte het om 'n woordeboek aan te skaf nie. Hy sou graag alle woorde in 'n woordeboek "tot een eindeloze ketting die tenslotte alles met alles verbindt tot een prachtig samenhangend geheel" met mekaar in verband wou bring.

Anders as in die geval van Fens, is Van Paemel se ervaring van die WNT baie gunstig. Sy het ná haar eerste kennismaking daarmee nie net 'n gebruiker daarvan geword nie, maar 'n "verslaafde", iemand wat maklik weggevoer kan raak wanneer sy 'n woord in die woordeboek naslaan.

Desondanks die "leesplezier" wat so 'n tasbare, gedrukte woordeboek verskaf, stel Van Paemel dit tereg dat 'n logiese ontwikkeling in die leksikografie sal wees dat woorde in 'n al hoe kleiner mate slegs in gedrukte vorm en in 'n al hoe groter mate in elektroniese vorm, bv. op CD-ROM of in 'n databasis, beskikbaar gestel sal word.

Hierdie verskuiwing van 'n gedrukte weergawe na 'n elektroniese wyse van publikasie is dan ook die vernaamste besprekingspunt van die ander lesings wat in hierdie bundel verskyn.

In sy lesing sê A. Moerdijk, hoofredakteur van die WNT, dat die toekoms, noudat die WNT voltooi is, nie met wanhoop tegemoet gegaan moet word nie. Hy sien die huidige leksikografiese situasie as 'n boeiende, ontwikkelende fase, en noem onder andere die gebruik van die rekenaar, elektroniese woordeboeke op CD-ROM en elektroniese databasisse as redes vir sy optimistiese toekoms-perspektief.

Volgens Moerdijk kan die leemtes wat in die gedrukte weergawe van die WNT voorkom, aangevul word met die elektroniese weergawe wat in 1995 op CD-ROM verskyn het. Deur gebruik te maak van die CD-ROM is aanvullende gewens oor byvoorbeeld die woord *boeken* verkry deurdat soektogte buite

die artikel **boeken** self en deur die hele woordeboek gedoen kon word. Op dié wyse is meer sitate, ouer én nuwer sitate, 'n addisionele betekenis, asook meer verbindingsmoontlikhede met *boeken* verkry.

Verder beskou Moerdijk elektroniese databasisse as 'n onmisbare bron van taalmateriaal vir die samestelling van woordeboeke in die toekoms. So 'n databasis lewer massas woorde in verskeie kontekste, gee 'n volledige beeld van 'n woord, en stel dus die leksikograaf in staat om 'n groot verskeidenheid betekenisse van 'n woord te abstraher uit die diverse verbindingsmoontlikhede van daardie woord. Die gebruik van 'n elektroniese databasis sal volgens Moerdijk die inkonsekwente of teenstrydige werkswyses wat in verskillende handwoordeboeke voorkom, oplos, en die daarmee gepaardgaande verwarring uitskakel.

Die hoofredakteur van die *Deutsches Wörterbuch (DWB)* in Göttingen, Michael Schläefer, skets as agtergrond vir sy lesing die geskiedenis van die Grimmsches Wörterbuch vanaf 1838. Hy fokus veral op die omvattendheid van die woordeboek, dat woorde té breedvoerig en in té veel besonderhede bewerk is en dat te veel sitate aangebied is. Dit strook met die leuse van hierdie woordeboek, naamlik dat alles daarin opgeteken moet word. Hierdie breedvoerige hantering het onder andere daartoe gelei dat dit baie lank geneem het om 'n letter af te handel. Die redaksie kon gevolglik nie tred hou met die veranderende wetenskap nie, wat tot gevolg gehad het dat die woordeboek se status ondermyn is. Daar is in 1986 met hervorming begin deurdat die omvang van die woordeboek gereduseer is en oortollige inligting weggelaat is. Slegs 'n beperkte mate van hervorming was egter moontlik, aangesien daar steeds aan die eis van omvattendheid voldoen moes word.

Volgens Schläefer sal daar in die toekoms 'n klemverskuiwing plaasvind. Woordeboeke sal byvoorbeeld gerig moet wees op spesifieke gebruikers, en elektroniese middele soos die rekenaar en elektroniese databasisse sal in die leksikografiese praktyk vir die skep van elektroniese woordeboeke benut moet word.

Dit is wat Schläefer en R. Bergmann met 'n historiese woordeboek van die moderne Hoogduitse standaardtaal, die *Neuhochdeutsches Wörterbuch (Göttingen)*, beoog. Hierdie woordeboek sal van die begin af as 'n elektroniese woordeboek beplan word. Soektogte sal elektronies binne 'n databasis gedoen kan word, waarna bewerking van die materiaal sal plaasvind. Die bewerkte materiaal sal ook elektronies beskikbaar wees, byvoorbeeld vir vakkundiges om afvoere te maak.

Schläefer voorspel dat hierdie elektroniese werkswyse gaan oorneem by die tradisionele manier van woordeboek maak, en dat woordeboeke in boekvorm dan hul belangrikheid gaan verloor.

J. Simpson, hoofredakteur van die *Oxford English Dictionary (OED)*, gee in sy lesing 'n paar riglyne oor wat 'n historiese woordeboek te doen staan nadat dit voltooi is. Die eerste riglyn word in die vorm van 'n waarskuwing aangebied, naamlik dat die kontinuïteit van 'n woordeboekprojek nie versteur moet word nadat 'n woordeboek voltooi is nie. Taal bly voortbestaan en ontwikkel,

en volgens hom moet 'n woordeboek ook ten einde die ontwikkelende taal te kan opteken. Indien 'n woordeboek gestaak word, is dit baie moeilik en duur om dit weer op die been te bring.

Die tweede riglyn wat Simpson aanbied, is in die vorm van 'n vraag, naamlik of 'n woordeboek, ná voltooiing van die eerste druk, hersien moet word, en of daar heeltemal van vooraf begin moet word. Simpson gee voorkeur aan eersgenoemde, en baseer dus sy benadering vir die toekoms op hersiening.

Hy gee verder 'n oorsig van hoe tegnologiese veranderinge ook veranderinge op leksikografiese terrein teweeg gebring het. Waar woordeboeke vroeër slegs in boekvorm verskyn het, is dit nou ook elektronies beskikbaar, bv. op CD-ROM, magneetband of aanlyn. Hy stel dit as vooruitsig dat laasgenoemde toenemend die wyse van publikasie vir woordeboeke sal word, alhoewel baie gebruikers voorkeur sal bly gee aan woordeboeke in boekvorm.

Hierdie aanlyn wyse van publikasie en bogenoemde hersieningsbenadering vul mekaar goed aan. Woordeboeke wat aanlyn gepubliseer word, is deurlopend beskikbaar vir kommentaar. Indien hierdie kommentaar op 'n konstante basis bygewerk word, word die woordeboek as 't ware deurlopend hersien.

Soos Moerdijk en Schlaefer heg Simpson ook waarde aan die gebruik van 'n elektroniese databasis vir die samestelling van toekomstige woordeboeke. Simpson sou graag wou sien dat daar skakels moet wees tussen so 'n databasis en 'n elektroniese woordeboek, sodat bykomende inligting vir die gebruiker beskikbaar gestel kan word. 'n Gebruiker sou byvoorbeeld vanaf 'n woord in 'n literêre teks met behulp van 'n skakel na die elektroniese woordeboek gelei word, waar die betrokke woord dan verklaar word.

Een van die vakkundiges wat as spreker opgetree het, die dialektoloog Magda Devos, tref in haar lesing 'n vergelyking tussen die *WNT* in boekvorm, die *WNT* op CD-ROM en 'n elektroniese databasis, die sogenaamde "Taalbank" wat deur die INL beplan word. Volgens haar is die *WNT* in boekvorm en die *WNT* op CD-ROM albei "afgeslote" en "geslote", onderskeidelik omdat die inhoud van elkeen die taal op 'n spesifieke oomblik weerspieël en omdat geen direkte toegang tot ander inligtingsbronne bevat nie. Die enigste opsig waarin die CD-ROM-weergawe van die weergawe in boekvorm verskil, is dat eersgenoemde nie 'n liniêre karakter soos laasgenoemde het nie. Die liniêre volgorde wat op papier bestaan, word met behulp van soektogte in die elektroniese weergawe opgehef. Deur van sulke soektogte binne die elektroniese teks gebruik te maak, kan, soos wat Moerdijk ook opgemerk het, aanvullende gegevens verkry word.

Volgens Devos is 'n elektroniese databasis ook nie afgesluit of gesluit nie. Deurdat inligting wat aanlyn aangebied word op 'n deurlopende basis bygewerk kan word, kan so 'n databasis nooit afgesluit wees nie. Verder het 'n elektroniese databasis 'n oop struktuur deurdat daar skakels na ander inligtingsbronne, soos opnames en ensiklopedieë, bestaan wat deur die leser opgevolg kan word. Hier toon Devos se toekomspektief weer ooreenstemming met dié van Simpson.

In hierdie resensie is daar hoofsaaklik gekonsentreer op die "en verder ..."-gedeelte van die titel, naamlik op die verskillende sprekers se toekomspektiewe en op wat 'n woordeboekprojek te doen staan ná die voltooiing van 'n woordeboek. Die klem is gelê op die woordeboek in elektroniese vorm, óf op CD-ROM óf aanlyn, en op die gebruik van elektroniese databasisse vir die samestelling van woordeboeke in die toekoms.

Deur die elektroniese aspekte te beklemtoon, wil ek geensins die "tradisionele" woordeboeke, naamlik dié in boekvorm, daaraan ondergeskik stel nie. Inteendeel, ek hoop Schlaefer se siening dat woordeboeke in boekvorm hulle belangrikheid in die toekoms gaan verloor, word verkeerd bewys. Ek wil eerder met Van Paemel saamstem wanneer sy sê dat die twee "werkinstrumente" aanvullend gebruik behoort te word.

Wat ook al die formaat waarin woordeboeke oorwegend in die toekoms gepubliseer gaan word, kan 'n mens net die hoop uitspreek dat, soos wat die historikus A. Th. van Deursen se ervaring met die *WNT* was, woordeboeke ons steeds die "gehoopte zekerheid" sal bied wanneer ons die betekenis van 'n woord nagaan.

Gerhard van Wyk
Buro van die Woordeboek van die Afrikaanse Taal
Stellenbosch
Republiek van Suid-Afrika
(gjvw@maties.sun.ac.za)

Publikasieaankondigings / Publication Announcements

- Juri Apresjan. *Systematic Lexicography*. Translated by Kevin Windle. 2000, xviii + 304 pp. ISBN 0-19-823780-4. Oxford/New York: Oxford University Press.
- B.T. Sue Atkins. *Using Dictionaries: Studies of Dictionary Use by Language Learners and Translators*, 1998, 214 pp. ISBN 3-484-30988-1. Tübingen: Max Niemeyer.
- Henri Béjoint. *Modern Lexicography: An Introduction*, 2000, xii + 276 pp. ISBN 0-19-829951-6. Oxford: Oxford University Press.
- Julie Coleman and Christian J. Kay (Editors). *Lexicology, Semantics and Lexicography: Selected Papers from the Fourth G.L. Brook Symposium, Manchester, August 1998*, 2000, xiv + 249 pp. ISBN 90 272 3701 8 (Eur.), 1 55619 972 4 (US). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- A.P. Cowie. *English Dictionaries for Foreign Learners: A History*, 1999, xiii + 232 pp. ISBN 0-19-823506-2. Oxford: Clarendon Press.
- Hanny Demeersseman (leiding en eindredactie), Piet Vermeer (eindredactie), Roy Pheiffer, Fred Pheiffer, Anneli Terre Blanche en Sief Veltkamp-Visser (redactie). *Kramers woordenboek Zuid-Afrikaans-Nederlands Nederlands-Zuid-Afrikaans*, 2000. 414 pp. ISBN 90-6882-378-7. Amsterdam: Elsevier. Prijs HFL16,90. (Resensie in hierdie uitgawe.)
- Thomas Herbst and Kerstin Popp. *The Perfect Learners' Dictionary (?)*, 1999, XII + 320 pp. ISBN 3-484-30995-4. Tübingen: Max Niemeyer Verlag.
- F. Heyvaert, A. Moerdijk, M. Mooijaart, M. Smits en R. Tempelaars (Redakteurs). *Het grootste woordenboek ter wereld: Een kijkje achter de kolommen van het Woordenboek der Nederlandsche Taal (WNT)*, s.j., 434 pp. ISBN 90 75566 96 4. URL: <http://www.sdu.nl>. Den Haag: SDU Uitgevers/ Antwerpen: Standaard Uitgeverij. Prijs: f49.90. (Resensie in hierdie uitgawe)
- P.A. Joubert. *Tweetalige Frasewoordeboek. Afrikaans – Engels. 'n Praktiese gebruiksgids. / Bilingual Phrase Dictionary. English – Afrikaans. A Practical Usage Guide*. 1997, 503 pp. ISBN 1 86890 004 5 (hardeband), 1 86890 006 1 (sagte-

band). Kaapstad: Pharos Woordeboeke. (Resensieartikel in hierdie uitgawe)

Dorothea Mantzel and Bernd Schulz. *Francolin Illustrated School Dictionary for Southern Africa*. Developed by The Dictionary Unit for South African English in consultation with The Molteno Project and READ Education Trust. 1997, xvi + 336 pp. ISBN 1-86859-015-1. Cape Town: Francolin Publishers. (Review in this issue.)

A. Moerdijk en R. Tempelaars (Redakteurs). *Van A tot Z en verder ...: Lezingen bij het voltooiing van het WNT*, 1999, 112 pp. ISBN 90 75566 79 4. URL: <http://www.sdu.nl>. Den Haag: SDU Uitgevers / Antwerpen: Standaard Uitgeverij. Prijs: f24.90 (Resensie in hierdie uitgawe)

Lynda Mugglestone. *Lexicography and the OED: Pioneers in the Untrodden Forest*, 2000, x + 288 pp. ISBN 0-19-823784-7. Oxford: Oxford University Press.

Hilary Nesi. *The Use and Abuse of EFL Dictionaries: How Learners of English as a Foreign Language Read and Interpret Dictionary Entries*, 2000, VI + 156 pp. ISBN 3-484-30998-9. Tübingen: Max Niemeyer.

F.F. Odendal en R.H. Gouws. *HAT: Verklarende Handwoordeboek van die Afrikaanse Taal*, 4de uitgawe, 2000, xx + 1387 pp. ISBN 0 628 03769 4. Midrand: Perskor Uitgewers.

Pharos. *Pharos Speller: Speltoets en Woordafbreker vir Afrikaans*. 2000. Pharos and Polderland Language and Speech Technology. Prys R299. (Resensie in hierdie uitgawe.)

Pharos. *Pharos Woordeboeke / Dictionaries 5 in 1: 5 ten volle geïntegreerde woordeboeke / Fully Integrated Dictionaries*. Afrikaans-Engels / English-Afrikaans 1. *Groot Woordeboek / Major Dictionary* 2. *Tweetalige Frasewoordeboek / Bilingual Phrase Dictionary* 3. *Nuwe Woorde / New Words Afrikaans* 4. *Verklarende Afrikaanse Woordeboek* 5. *Groot Tesaurus van Afrikaans*. 2000. Pharos and Logos Information Systems. Prys R299. (Resensie in hierdie uitgawe.)

Juan C. Sager (Selector and Editor). *Essays on Definition*. Introduction by Alain Rey. 2000, viii + 256 pp. ISBN 90 272 2370 0 (Eur.), 1 55619773 X (US). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Rita Temmerman. *Towards New Ways of Terminology Description: The Sociocognitive Approach*, 2000, xv + 258 pp. ISBN 90 272 2326 2 (Eur.), 1 55619772 1 (US). Amsterdam / Philadelphia: John Benjamins Publishing Company.

Yukio Tono. *Research on Dictionary Use in the Context of Foreign Language Learning: Focus on Reading Comprehension*, 2001, xii + 257 pp. ISBN 3-484-39106-5. Tübingen: Max Niemeyer Verlag.

K.H. van Dalen-Oskam, K.A.C. Depuydt, W.J.J. Pijnenburg and T.H. Schoonheim (Editors). *Dictionaries of Medieval Germanic Languages: A Survey of Current Lexicographical Projects. Selected Proceedings of the International Medieval Congress, University of Leeds, 4-7 July, 1994, 1997*, viii + 197 pp. ISBN 2-503-50601-1. Turnhout: Brepols. E-mail: publishers@brepols.com. Price: Belgian f1613. (Review in this issue)

Karina van Dalen-Oskam en Marijke Mooijaart: *Bijbels lexicon: Woorden en uitdrukkingen uit de bijbel in het Nederlands van nu*, 2001, 443 pp. ISBN 90 5333 923 X. Amsterdam: Prometheus.

Herbert Ernst Wiegand. *Wörterbuchforschung: Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie. Teilband 1*. 1998, xix + 1162 pp. ISBN 3-11-013584-1. Berlin/New York: De Gruyter. (Review article in this issue)

Herbert Ernst Wiegand. *Kleine Schriften: Eine Auswahl aus den Jahren 1970 bis 1999 in zwei Bänden*. 2000. *Band 1 1970-1988*, xxx + 1-876 pp. *Band 2 1988-1999*, vii + 877-1716 pp. ISBN 3-11-015665-2. Berlin/New York: De Gruyter.

VOORSKRIFTE AAN SKRYWERS

(Tree asseblief met die Buro van die WAT in verbinding vir 'n uitvoeriger weergawe van hierdie instruksies, of besoek ons webblad: <http://www.sun.ac.za/wat/index.html>)

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) **Navorsingsartikels:** Grondige oorspronklike wetenskaplike navorsing wat gedoen en die resultate wat verkry is.
- (2) **Beskouende artikels:** Bestaande navorsingsresultate en ander feite wat op 'n oorspronklike wyse oorsigtelik, interpreterend, vergelykend of krities evaluerend aangebied word.
- (3) **Resensieartikels:** Navorsingsartikels wat in die vorm van 'n kritiese resensie van een of meer gepubliseerde wetenskaplike bronne aangebied word. Bydraes in kategorieë (1)-(3) word aan streng anonieme keuring deur onafhanklike akademiese vakgenote onderwerp ten einde die internasionale navorsingsgehalte daarvan te versterk.
- (4) **Resensies:** 'n Ontleding en kritiese evaluering van gepubliseerde wetenskaplike bronne en produkte, soos boeke en rekenaarprogramme.
- (5) **Projekte:** Besprekings van leksikografiese projekte.
- (6) **Leksikonotas:** Enige artikel wat praktykgerigte inligting, voorstelle, probleme, vrae, kommentaar en oplossings betreffende die leksikografie bevat.
- (7) **Leksikovaria:** Enigeen van 'n groot verskeidenheid artikels, aankondigings en nuusvystellings van leksikografiese verenigings wat veral vir die praktiserende leksikograaf van waarde sal wees.
- (8) **Verslae:** Verslae van konferensies en werksessies.

Bydraes in kategorieë (4)-(8) 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* gepubliseer 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 navorsings-, beskouende of resensieartikel van hulle wat gepubliseer is asook een gratis eksemplaar van die uitgawe waarin sodanige artikel(s) verskyn het. Skrywers van suiwer evaluerende resensies en van bydraes tot die rubrieke Leksikovaria, Projekte en Verslae ontvang vyf gratis oordrukke van hulle bydraes. In laasgenoemde vier kategorieë kan die redaksie egter, afhangend van die aard en omvang van die bydraes, besluit om ook 'n eksemplaar van die betrokke uitgawe aan 'n skrywer toe te ken.

7. Uitnodiging en redaksionele adres

Alle belangstellende skrywers is welkom om bydraes vir opname in *Lexikos* te lewer en aan die volgende adres te stuur:

Die Redakteur
LEXIKOS
Buro van die WAT
Posbus 245
7599 STELLENBOSCH
Republiek van Suid-Afrika

B. VOORBEREIDING VAN MANUSKRIP

Die manuskrip van artikels moet aan die volgende redaksionele vereistes voldoen:

1. Lengte en formaat van artikels

Bydraes moet verkieslik nie 20 getikte A4-bladsye met teks in dubbelspasiëring en ruim kantlyne (ongeveer 2,5 cm) oorskry nie. Manuskrip moet verkieslik in elektroniese formaat as ASCII-teks, as volledig geformateerde Microsoft Word (DOS of Windows) lêers of as WordPerfect (DOS of Windows) lêers op rekenaarskyf (360 KB tot 1.44 MB) voorgelê word. 'n Rekenaardrukstuk van die artikel moet die skyf vergesel. Elke artikel moet voorsien wees van 'n Engelse opsomming van tussen 150 en 250 woorde, sowel as tussen 10 en 30 Engelse sleutelwoorde.

2. Grafika

Een stel duidelike oorspronklike illustrasies, tabelle, grafieke, diagramme, of kwaliteitsafdrukke daarvan, moet voorgelê word. Die plasing van grafika binne die teks moet duidelik aangedui word.

3. Bibliografiese gegewens en verwysings binne die teks

Kyk na onlangse nommers van *Lexikos* vir meer inligting.

INSTRUCTIONS TO AUTHORS

(For a more detailed version of these instructions, please contact the Bureau of the WAT or refer to our web page: <http://www.sun.ac.za/wat/index.html>)

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) **Research articles:** Fundamentally original scientific research that has been done and the results that have been obtained.
- (2) **Contemplative articles:** Reflecting existing research results and other facts in an original, synoptic, interpretative, comparative or critically evaluative manner.
- (3) **Review articles:** Research articles presented in the form of a critical review of one or more published scientific sources.

Contributions in categories (1)-(3) are subjected to strict anonymous evaluation by independent academic peers in order to ensure the international research quality thereof.

- (4) **Reviews:** An analysis and critical evaluation of published scientific sources and products, such as books and computer software.
- (5) **Projects:** Discussions of lexicographical projects.
- (6) **Lexiconotes:** Any article containing practice-oriented information, suggestions, problems, questions, commentary and solutions regarding lexicography.
- (7) **Lexicovaria:** Any of a large variety of articles, announcements and press releases by lexicographic societies which are of particular value to the practising lexicographer.
- (8) **Reports:** Reports on conferences and workshops.

Contributions in categories (4)-(8) 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 Woordboek 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 research, contemplative 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 Lexicovaria, Projects, and Reports receive five free offprints of their contributions. In the case of the latter four categories, the editors may, however, depending on the nature and scope of the contributions, decide to grant the author a copy of the issue concerned.

7. Invitation and editorial address

All interested authors are invited to submit contributions for publication in *Lexikos* to:

The Editor
LEXIKOS
Bureau of the WAT
P.O. Box 245
7599 STELLENBOSCH
Republic of South Africa

B. PREPARATION OF MANUSCRIPTS

Manuscripts of articles must meet the following editorial requirements:

1. Length and format

Contributions should not exceed more than 20 typewritten A4 pages with double spacing and ample margins (about 2,5 cms). Manuscript should preferably be in electronic form on a (360 KB to 1.44 MB) floppy disk as either ASCII text, fully-formatted Microsoft Word (DOS or Windows) or WordPerfect (DOS or Windows) files. A computer printout of the article should accompany the disk. Each article must be accompanied by an English abstract of 150 to 250 words, and between 10 and 30 English keywords.

2. Graphics

One set of clear original drawings, tables, graphs, diagrams or quality prints thereof must be submitted. The locations of graphics must be clearly indicated in the text.

3. Bibliographical details and references in the text

Examine recent issues of *Lexikos* for details.

HINWEISE UND RICHTLINIEN FÜR AUTOREN

(Nehmen Sie bitte mit dem Büro des WAT Kontakt auf für eine ausführlichere Wiedergabe dieser Hinweise, oder besuchen Sie unsere Webseite: <http://www.sun.ac.za/wat/index.html>)

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) **Forschungsartikel**, die grundlegend über neue Forschungsansätze und deren Ergebnisse berichten.
- (2) **Kontemplative Artikel**, die bestehende Forschungsergebnisse und andere Informationen selbständig, interpretativ, vergleichend oder kritisch bewertend wiedergeben.
- (3) **Rezensionsartikel**, die in der Form eines Forschungsartikels eine oder mehrere veröffentlichten wissenschaftlichen Quellen kritisch rezensieren.

Beiträge in Kategorien (1)-(3) werden streng anonym von unabhängigen wissenschaftlichen Experten begutachtet, um ein internationales fachliches Niveau in *Lexikos* zu gewährleisten.

- (4) **Rezensionen**, die veröffentlichte wissenschaftliche Quellen und Produkte, wie z.B. Bücher und Software, analysieren und kritisch bewerten.
- (5) **Lexikographische Projekte**, die vorgestellt werden.
- (6) **Notizen zum Lexikon**, die praxisbezogene Informationen, Vorschläge, Probleme, Fragen, Kommentare und Lösungen hinsichtlich der Lexikographie enthalten.
- (7) **Lexikovaria**, die unterschiedliche Beiträge, Ankündigungen und Pressemitteilungen lexikographischer Vereinigungen, die dem praktischen Lexikographen wichtig sein können, einschließen.
- (8) **Berichte** über Konferenzen und Workshops.

Beiträge in Kategorien (4)-(8) 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 Forschungsartikel, kontemplativen Artikel oder Rezensionsartikel gratis sowie ein Freixemplar der betreffenden Ausgabe. Rezensenten und Autoren von Beiträgen zu den Kategorien 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.