

Lexikos 12

Lexikos 12

Redakteur

Editor

J.C.M.D. du Plessis

Resensieredakteur

Review Editor

T. Harteveld



African Association for Lexicography

AFRILEX-REEKS 12:2002

AFRILEX SERIES 12:2002



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 © 2002 deur die uitgewer
Alle regte streng voorbehou
Eerste uitgawe 2002

Tipografie en uitleg
deur Tanja Harteveld en Riette Ruthven
Bandontwerp deur Piet Grobler

Geset in 10 op 12 pt Palatino

Gedruk en gebind deur Onyx Press
Stewartstraat 61 Goodwood

ISBN 0-9584401-2-3
ISSN 1684-4904

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

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

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

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

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. 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. R. Finlayson (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. Mfana (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	ix
Foreword	xi
<i>J.C.M.D. du Plessis</i>	
'n Woord van AFRILEX	xiii
A Few Words from AFRILEX	xv
<i>Danie Prinsloo</i>	
Redaksionele doelstellings	xvii
Editorial Objectives	xviii
Redaktionelle Ziele	xix
Politique éditoriale	xx
Navorsingsartikels / Research Articles	
Loan Words versus Indigenous Words in Northern Sotho — A Lexicographic Perspective	1
<i>Salmina Nong, Gilles-Maurice de Schryver and D.J. Prinsloo</i>	
The Lemmatization of Copulatives in Northern Sotho	21
<i>D.J. Prinsloo</i>	
Semi-automatic Term Extraction for the African Languages, with Special Reference to Northern Sotho	44
<i>Elsabé Taljard and Gilles-Maurice de Schryver</i>	
Beskouende artikels / Contemplative Articles	
Les tons dans les dictionnaires de langues gabonaises: situation et perspectives	75
<i>Thierry Afane Otsaga</i>	
E-terminology	90
<i>Mariëtta Alberts</i>	
An Advanced Dictionary? Similarities and Differences between <i>Duramazwi ReChiShona</i> and <i>Duramazwi Guru ReChiShona</i>	105
<i>Emmanuel Chabata</i>	

'n Praktiese benadering tot die samestelling van 'n tweerigting- vakwoordeboek met Afrikaans en Italiaans as taalpaar <i>F. de Foglio en H.J. Lubbe</i>	117
Niching as a Macrostructural Procedure <i>Rufus H. Gouws</i>	133
The Ndebele Language Corpus: A Review of Some Factors Influencing the Content of the Corpus <i>Samukele Hadebe</i>	159
Business Metaphors in a Bilingual Business Lexicon <i>Li Lan and Grahame T. Bilbow</i>	171
Synopsis Articles in the Planning of a Trilingual Dictionary: Yilumbu–French–English <i>P.A. Mavoungou</i>	181
The Cause of Urban Slang and its Effect on the Development of the Northern Sotho Lexicon <i>V.M. Mojela</i>	201
Le corpus lexicographique dans les langues à tradition orale: le cas du dialecte fang-mekè <i>Yolande Nzang-Bié</i>	211
Logistiek: Die oorsprong, konseptuele ontwikkeling en betekenis daarvan in hedendaagse gebruik <i>Wessel J. Pienaar</i>	227
Equivalence in Bilingual Lexicography: Criticism and Suggestions <i>Herbert Ernst Wiegand</i>	239
Projekte / Projects	
The African Proverbs Project and After <i>John S. Mbiti</i>	256
Leksikonotas / Lexiconotes	
Defining Formats and Corpus-based Examples in the General Ndebele Dictionary, <i>Isichazamazwi SesiNdebele</i> <i>Langa Khumalo</i>	264

Resensieartikels / Review Articles

- Les proverbes dans *Kongo Proverbs and the Origins of Bantu Wisdom* 275
par Mukumbuta Lisimba
Ludwine Mabika Mbokou
- The Systematic Development of Wiegand's Metalexigraphy as 290
Demonstrated in *Kleine Schriften*
Maria Smit

Resensies / Reviews

- Igor Burkhanov. *Linguistic Foundations of Ideography: Semantic 311*
Analysis and Ideographic Dictionaries
Werner Hillen
- Julie Coleman and Christian J. Kay (Eds.). *Lexicology, Semantics and 319*
Lexicography: Selected Papers from the Fourth G.L. Brook Symposium,
Manchester, August 1998
Adelia Carstens and Gerhard B. van Huyssteen
- G.J. van Wyk (Tegn. Red.), A.E. Cloete, A. Jordaan, H.C. Liebenberg 329
en H.J. Lubbe. *Etimologiewoordeboek van Afrikaans (eerste fase)*
Fritz Ponelis

Publikasieaankondigings / Publication Announcements

- Voorskrifte aan skrywers 336
Instructions to Authors 337
Richtlinien für Autoren 338
Instructions aux auteurs 339

Voorwoord

Lexikos is deur die Buro van die WAT beplan as 'n internasionale leksikografiese tydskrif met sy hoofdoel die bevordering van die praktiese leksikografie van die tale wat in Afrika gebruik word. Dit wou egter ook internasionale rigtings in die metaleksikografie aantoon.

Sedert *Lexikos* die lyfblad van AFRILEX geword het, is die oorspronklike beleid van *Lexikos* bevestig en verstewig. Dit is belangrik dat *Lexikos* nie net die kompleksiteit van die taalsituasie in Afrika weerspieël nie, maar ook die huidige belangstelling in die ontwikkeling en studie van die Afrikatale. In hierdie strewe speel die leksikografie 'n belangrike rol.

Om in staat te wees om met die res van die wêreld oor die leksikografie 'n gesprek te voer moet artikels in algemeen verstaanbare tale gepubliseer word. Sedert sy stigting het *Lexikos* artikels in vyf tale verwelkom: Engels, Frans, Duits, Nederlands en Afrikaans. Die artikels moet van twee opsommings versamel wees: een in Engels, Frans, Duits of Nederlands, en die ander in enige ander taal wat in Afrika gebruik word. Op hierdie manier kan die hoofargument van die artikels sowel plaaslik as internasionaal verstaan word.

AFRILEX wil die hele leksikografiese gemeenskap in Afrika verenig. En in hierdie doelstelling vervul *Lexikos* 'n belangrike funksie. As gevolg van groter kontak met Franstalige lande in Sentraal- en Wes-Afrika, begin Frans 'n belangriker rol vervul as taal van kommunikasie in die leksikografie. Ons is daarom bly om in hierdie uitgawe drie artikels in Frans te kan publiseer.

Redaksiekomitee

Prof. A. Carstens het as lid van die Redaksiekomitee bedank. Ons is jammer om so 'n skerpsinnige keurder te verloor, en wens haar die beste toe in die navolging van haar nuwe belangstellings. Ons is egter bly om prof. R. Finlayson te verwelkom wat ons uitnodiging aanvaar het om lid van die Redaksiekomitee te word. Ons hoop sy sal 'n lonende verbintenis met *Lexikos* hê.

Dankbetuigings

Soos in vorige uitgawes van *Lexikos*, wil ek weer eens al die keurders bedank wat artikels vir vanjaar se uitgawe beoordeel het. Hul bydrae is belangrik om die gehalte van *Lexikos* te handhaaf. *Lexikos* kan egter nie gepubliseer word sonder die bydraers van artikels nie. Ek wil hulle almal bedank, ook vir hul bereidwilligheid om dikwels hul bydraes aan te pas ooreenkomstig voorstelle van die keurders. Hierdie vrugbare wisselwerking tussen bydraers en keurders verbeter die standaard van die gepubliseerde artikels dikwels aan-

sienlik.

Soos in die verlede, was mes. Tanja Harteveld en Riette Ruthven in beheer van die administrasie en setwerk. Hulle moet altwee bedank word vir hul insette. Vanjaar het me. Riette Ruthven 'n aansienlike deel gehad in die voltooiing en finalisering van *Lexikos*. Haar noukeurigheid en kreatiwiteit het gelei tot die lewering van gesette artikels wat sowel tipografies as visueel bevredigend is.

J.C.M.D. du Plessis

Buro van die Woordeboek van die Afrikaanse Taal

Foreword

Lexikos was conceived by the Bureau of the WAT as an international lexicographic journal with its main aim the advancement of the practical lexicography of the languages used in Africa. However, it also wanted to show international trends in metalexicography.

Since *Lexikos* has become the mouthpiece of AFRILEX, the original policy of *Lexikos* has been confirmed and consolidated. It is important that *Lexikos* reflects not only the complexity of the language situation in Africa, but also the present interest in the development and study of African languages. In this endeavour, lexicography plays an important role.

To be able to discourse with the rest of the world on lexicography, articles should be published in widely understood languages. Since its inception, *Lexikos* has welcomed articles in five languages: English, French, German, Dutch and Afrikaans. The articles are to be accompanied by two summaries: one in English, French, German or Dutch, and the other in any other language used in Africa. In this way, the main argument of the articles can be understood both locally and internationally.

AFRILEX wants to unite the whole African lexicographic community. And in this aim *Lexikos* fulfils an important function. As a result of greater contact with Francophone countries in Central and West Africa, French starts playing a more important role as language of communication in lexicography. We are therefore happy to be able to publish three articles in French in this issue.

Editorial Committee

Prof. A. Carstens has resigned from the Editorial Committee. We are sorry to lose such a discerning referee, and wish her well in the pursuit of her new interests. We are however happy to welcome Prof. R. Finlayson who has accepted our invitation to become a member of the Editorial Committee. We hope she will have a rewarding association with *Lexikos*.

Acknowledgments

As in previous issues of *Lexikos*, I again wish to thank all the referees who evaluated articles for this year's issue. Their contribution is important for maintaining the quality of *Lexikos*. However, *Lexikos* cannot be published without the contributors of articles. I would like to thank all of them, also for their willingness often to adjust their contributions according to suggestions from the referees. This fruitful interaction between contributors and referees often substantially improves the standard of the articles published.

As in the past, Mss Tanja Harteveld and Riette Ruthven were in charge of the administration and typesetting. They must both be thanked for their input. This year Ms Riette Ruthven took a considerable share in the completion and finalization of *Lexikos*. Her accuracy and creativity led to the production of typeset articles which are both typographically and visually satisfying.

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

'n Woord van AFRILEX

Met die 12de uitgawe van *Lexikos* en die Sewende Internasionale Konferensie van die African Association for Lexicography (AFRILEX) agter die rug kry die suksesverhaal van AFRILEX en *Lexikos* steeds meer gestalte. Bykans 100 lede en belangstellendes het die Konferensie in Grahamstad bygewoon en 40 voordrage is gelewer. *Lexikos* verwerf toenemend nasionaal en internasionaal bekendheid as 'n leksikografietydskrif van besonder hoë gehalte. AFRILEX tel eweneens onder die toonaangewende leksikografieverenigings in die wêreld AFRILEX bied 'n tuiste aan almal wat belangstel in een of ander aspek van die leksikografie en speel, met die hulp van die Pan Suid-Afrikaanse Taalraad, 'n belangrike ondersteunende rol in die ontwikkeling van woordeboekeenhede vir die Afrikatale. Benewens die jaarlikse internasionale konferensie, fasiliteer AFRILEX 'n aantal seminare, slypskole en tutoriale en kommunikeer met lede deur sy nuusbriewe en verskeie omsendbriewe.

AFRILEX is besondere dank verskuldig aan Kathy Kavanagh en haar span van die Dictionary Unit for South African English vir die puik organisasie van die Sewende Internasionale Konferensie. 'n Hartlike woord van dank ook aan die organiseerder, Gilles-Maurice de Schryver, vir sy reusebydrae tot die vereniging deur die instandhouding van die tuisblad, die samestelling van die brosjure met opsommings van die kongresreferate en die dag-tot-dag-bestuur van die vereniging.

AFRILEX is terdeë bewus van die leidende rol wat die organisasie in die bevordering van die leksikografie in Suid-Afrika kan en moet speel en met voortgesette entoesiastiese ondersteuning van ons lede asook belangstellendes in die leksikografie kan dit wel vermag word. Uiteraard is daar bedreigings en swak punte soos finansiële beperkings, die behoefte aan opleiding, die dag-tot-dag-begeleiding van nuutgestigte woordeboekeenhede, ens., maar die entoesiasme hoofsaaklik by jonger leksikograwe is inspirerend en aansteeklik.

AFRILEX is ook verheug oor die toenemende belangstelling uit die Afrikakontinent vir AFRILEX-aktiwiteite wat die organisasie inderdaad in staat stel om sy rol as 'n ware organisasie vir Afrika te vervul. So is die Agtste Internasionale Konferensie geskeduleer vir Namibië en die Negende vir Gaboen.

Deur die Kernerman Woordeboeknavorsingstoekenning is 'n unieke geleentheid geskep vir ons lede om finansiële ondersteuning ten bedrae van \$1 000 vir navorsingsprojekte te verkry. AFRILEX spreek dan ook sy dank uit aan Kernerman vir hierdie ruim jaarlikse toekenning.

Die Raad van AFRILEX glo dat AFRILEX-lidmaatskap, wat vanaf 2003 'n R100 per jaar beloop, en boonop 'n gratis kopie van *Lexikos* insluit, uitstekende

waarde vir geld is. Ons spreek ook 'n woord van dank aan 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

Departement Afrikatale, Universiteit van Pretoria, Pretoria 0002

Tel.: +27(0)12-420-2320; Faks: +27(0)12-420-3163

E-pos: prinsloo@postino.up.ac.za

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

A Few Words from AFRILEX

With the 12th edition of *Lexikos* and the Seventh International Conference of the African Association for Lexicography over and done with, it is clear that the success story of AFRILEX and *Lexikos* is taking shape. Approximately 100 members and interested parties attended the Conference in Grahamstown where 40 lectures were delivered. *Lexikos* gains more and more prominence, nationally as well as internationally, as a lexicographic journal of high quality. AFRILEX also proved itself to be one of the leading lexicographic associations in the world. AFRILEX provides a home for everyone interested in some aspect or other of lexicography and, with the help of the Pan South African Language Board, it plays an important supportive role in the development of dictionary units for the African languages. In addition to the annual international conference, AFRILEX facilitates a number of seminars, workshops and tutorials and communicates with its members through its newsletters and various circulars.

AFRILEX owes a debt of gratitude to Kathy Kavanagh and her team from the Dictionary Unit of South African English for their superb organization of the Seventh International Conference. A heartfelt word of thanks also to the organizer, Gilles-Maurice de Schryver, for his huge contribution to the association through the maintenance of the web page, the composition of the booklet containing the abstracts of conference papers and the day-to-day management of the association.

AFRILEX is well aware of the leading role the organization can and should play in the promotion of lexicography in South Africa, and with the continued enthusiastic support of our members as well as other parties interested in lexicography, it may well be attainable. Naturally, threats and weaknesses do exist: financial restrictions, the need for training, the day-to-day mentoring of newly founded dictionary units, etc., but the enthusiasm, especially from younger lexicographers, proves infectious and serves as an inspiration.

AFRILEX also rejoices in the growing interest emerging from the African continent for AFRILEX activities, which enables the organization to fulfil its role as a truly African association. In this regard the Eighth International Conference is scheduled for Namibia and the Ninth for Gabon.

A unique opportunity has been created through the Kernerman Dictionary Research Grants for our members to obtain financial support totalling \$1 000 for research projects. AFRILEX herewith expresses thanks to Kernerman for this generous annual grant.

The Board of AFRILEX believes that AFRILEX membership is true value

for money. It will amount to R100 for 2003 and includes a free copy of *Lexikos*. We also take this opportunity to thank the WAT for the publication of *Lexikos* and special thanks to Dr J.C.M.D. du Plessis for his excellent work as editor.

Danie Prinsloo

Chairperson: AFRILEX

Department of African Languages, University of Pretoria, Pretoria 0002

Tel.: +27(0)12-420-2320 Fax: +27(0)12-420-3163

E-mail: prinsloo@postino.up.ac.za

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

Redaksionele doelstellings

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

Die strewe van die AFRILEX-reeks is:

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

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

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

Politique éditoriale

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

Les objectifs de la collection AFRILEX sont de :

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

Pour atteindre ces objectifs, la collection AFRILEX

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

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

Loan Words versus Indigenous Words in Northern Sotho — A Lexicographic Perspective*

Salmina Nong, *Department of African Languages, University of Pretoria, Pretoria, Republic of South Africa (snong@postino.up.ac.za)*,
Gilles-Maurice de Schryver, *Research Assistant of the Fund for Scientific Research — Flanders (Belgium) and Department of African Languages, University of Pretoria, Pretoria, Republic of South Africa (gillesmaurice.deschryver@rug.ac.be)*, and
D.J. Prinsloo, *Department of African Languages, University of Pretoria, Pretoria, Republic of South Africa (prinsloo@postino.up.ac.za)*

Abstract: The aim of this article is to investigate, from a lexicographic perspective, the preferences of Northern Sotho mother-tongue speakers for loan words versus so-called 'traditional' or 'original' counterparts in the language. Results obtained from a survey conducted among 100 randomly selected mother-tongue speakers from different age and gender groups, backgrounds, places of residence, etc. will be analysed. It is shown that although the overwhelming preference of the respondents lies with the use of (more) indigenous words in comparison to loan words, lexicographers should be alerted to possible, even rapid, changes in this preference pattern. The results from the survey are compared throughout with frequency counts derived from a corpus as well as with current dictionary treatment.

Keywords: LEXICOGRAPHY, DICTIONARY, LEMMATISATION, NORTHERN SOTHO (SEPEDE), LOAN WORD, SOTHOISED WORD, INDIGENOUS WORD, QUESTIONNAIRE, CORPUS, DESCRIPTIVENESS, PROSCRIPTIVENESS, PRESCRIPTIVENESS, PREFERENCE PATTERN

Senaganwa: Maadingwa ge a bapetšwa le Mantšu a Setlogo go Sesotho sa Leboa — Kgopolo ya Bangwalapukuntšu. Maikemišetšo a taodišwana ye ke go nyakišiša, go ya ka kgopolo ya bangwalapukuntšu, ka fao baboleli ba Sesotho sa Leboa ba dirago kgetho ya mantšu magareng ga maadingwa le mantšu a setlogo polelong ye. Dipolelo tše di hweditšwego go tšwa go bakgathatema ba e lego baboleli ba Sesotho sa Leboa, banna le basadi, ba lekgolo (100) ba mengwaga ya go fapana, maemo a a fapanego a thuto, ba ba dulago mafelong ao a fapafapanego, bj.bj. di tla fetlekwa. Go ipontšha gore le ge dipolelo tša nyakišišo ye di laetša gore bontši bja bakgathatema bo kgetha go šomiša mantšu a setlogo go ena le maadingwa, bangwadi ba

* An earlier version of this article was presented as a paper at the Seventh International Conference of the African Association for Lexicography, organised by the Dictionary Unit for South African English, Rhodes University, Grahamstown, 8–10 July 2002.

dipukuntšu ba swanetše go phafošwa mabapi le diphetogo tše di ka bago gona pateroneng ya kgetho ya tšhomišo ya mantšu. Dipoele tša nyakišišo ye di bapetšwa le ka moo mantšu a tšwelelago kgafetšakgafetša go tšwa khophaseng gammogo le ka fao dipukuntšu tše di šetšego di le gona di šomišitšego mantšu ao ka gona.

Mantšu a bohlokwa: GO HLAMA PUKUNTŠU, PUKUNTŠU, TSELA YA GO NGWALA MANTŠU KA PUKUNTŠUNG, SESOTHO SA LEBOA (SEPEDİ), LEADINGWA, LENTŠU LE LE SOTHO FADITŠWEGO, LENTŠU LA SETLOGO, LENANEOPOTŠIŠO, KHOPHASE, GO HLALOŠA POLELO, GO HLALOŠA POLELO KA GO FA KGETHO, GO LAELA KA GA TŠHOMIŠO YA POLELO, PATERONE YA KGETHO YA MANTŠU

Loan words versus their (more) indigenous counterparts

The use of loan words versus their (more) indigenous counterparts is studied in various disciplines such as science and technology, socio-linguistics, syntax and semantics, morphology and phonology, and not the least, lexicography. From a lexicographic perspective the issue to be investigated links well with one of the fundamental dichotomies in dictionary compilation, namely prescriptiveness versus descriptiveness. The coinage of loan words by language boards, for instance, is a prescriptive activity in itself. Within a descriptive approach towards dictionary compilation, it is imperative to know to what extent loan words in contrast to their so-called 'traditional' or 'original' counterparts are actually and actively used, and to study preferences of the target-user group of an envisaged dictionary in this regard. Not only should the lexicographer strive to lemmatise and lexicographically treat those words most likely to be looked for by the target users, the lexicographer should also be sensitive towards potential changes in preferences regarding the use of loan words versus their (more) indigenous counterparts.

In order to study this preference for Northern Sotho, and this with the compilation of a Northern Sotho dictionary in mind, a survey was conducted among 100 randomly selected mother-tongue speakers of varying ages, professions, regions, etc.¹ From the above it should be clear that only loan words with Northern Sotho counterparts are studied in this article, with the main focus on their macrostructural treatment.²

Prescriptiveness versus descriptiveness, and proscriptiveness

Traditionally dictionaries were regarded as a source of linguistic information. A good example is Samuel Johnson's *A dictionary of the English language* published in 1755. Wells (1973), Al-Kasimi (1977), Gouws (1989), Landau (2001) and many others agree that this dictionary is generally regarded as the reference work which normalised the English language. For centuries the lexicographer was seen as judge or ruler of the language who was to prescribe good and

to point out unacceptable use. According to this view the dictionary should rectify and cleanse the language, preserve its purity, lengthen its duration, correct or ban improprieties and absurdities, sensor faulty usage and repress anomaly. This authoritarian tradition, however, collapsed mainly due to one simple principle, viz. *language change*. In the words of Philip Gove (1961³: 4a), the Editor-in-Chief of *Webster's Third*:

English like other living languages is in a metabolic process of constant change. The changes affect not only word stock but meaning, syntax, morphology and pronunciation.

Indeed, in contrast to the prescriptive approach stands the descriptive approach with the focus on actual language usage. Gove (1961a: 13, quoted in Al-Kasimi (1977: 84)) writes in a letter to *Life Magazine*:

The responsibility of a dictionary is to record the language, not set its style. For us to attempt to prescribe the language would be like *Life* reporting the news as its editors would prefer it to happen.

Prinsloo (1992: 10) rightfully emphasizes that the then Northern Sotho Language Board made an invaluable contribution towards the clarification, systematisation, standardization and coining of new terms for Northern Sotho in, for example, religion, news broadcasting, mathematics, general science, etc. The last Terminology and Orthography for Northern Sotho, for short T&O, produced by this language board was published in 1988 (Departmental Northern Sotho Language Board 1988⁴). The Language Board adopted a sensible approach in being prescriptive in the coinage and approval/disapproval of terminology on the one hand, while still placing a high premium on actual usage as criterion for acceptability on the other hand. The Language Board (T&O: 3) allows for more than one option rather than attempting to enforce just one term while suppressing others:

It is anticipated that practical usage of the terms offered will prove that some of them can be replaced by other more commendable ones.

The same holds true for the Language Board's attitude towards certain entries in existing dictionaries (T&O: 1):

[C]ertain new terms and concepts are included which appear in some dictionaries but which are not generally accepted in the language yet. In such instances the Northern Sotho Language Board devised a term of their own which in their opinion is more appropriate.

Let there be no doubt that *actual usage* and not the sentiments of a language board will eventually determine whether a word should be included into or

omitted from a dictionary. Zgusta (1971: 187) strongly argues:

Lexicographers can coin new expressions, they can normalize their form and meaning, they can systematize and clarify the old ones, they can help in an endless number of such exceedingly useful and necessary tasks. The real life of a language, however, is in its use; and the definitive, full-fledged stabilization of the standard national language is brought about by its being really and extensively used in literature and in oral communications of all types.

It is interesting to note, for example, that as far as the months of the year are concerned, the then Language Board prescribes the use of "Sothoised" terms like **Matšhe** 'March' instead of **Hlakola**, **Aprele** 'April' instead of **Moranang**, etc. If the actual usage of the traditional terms would gain in importance or even prevail, a normalising board would have to back down and take a more descriptive approach.

As will be indicated below, the names for the months of the year seem to be an exception in the general preference pattern, namely that in this case loan words are preferred to the traditional words. Note, however, that there might be other factors influencing the choice in respect of months of the year, such as the lack of a one-to-one correlation between the traditional name and the actual month to which it refers. As a matter of fact, Ziervogel and Mokgokong (1975: 828, 1022, 1039) state that **Moranang** can refer to both April and June, **Phato** to both August and October, **Pherekgong** to both January and March, etc. which is of course unacceptable in real life situations where activities are punctually scheduled in terms of date and time. In simple terms it means that two people can agree on meeting each other on, say, the 1st of **Pherekgong** but then miss each other by two months.

More recently, metalexigraphers such as Bergenholtz (2002: 12) have introduced the term proscriptiveness:

I wish to suggest a specification and the introduction of a new term, proscription, which in actual fact is only new as a term, since the phenomenon itself is known in many dictionaries around the world. What is meant is the suggested use of a certain variant based on an exact analysis of an empirical basis without prohibiting other existing variants.

Within a proscriptive framework, the paradox that even a descriptive dictionary has a prescriptive effect on the target users is taken into account (compare Bergenholtz 2001). With reference to the current study, the task of the proscriptive compiler of dictionaries for Northern Sotho in terms of loan words versus their (more) indigenous counterparts, is thus to reflect user *preferences* in the selection of lemma signs on the macrostructural level as well as in the extent of treatment on the microstructural level, while still allowing for other existing variants. Consequently, within a proscriptive approach towards the lemmatisation of loan words in contrast to their 'traditional' or 'original' coun-

terparts, it is imperative for the lexicographer to know what the preferences of the target-user group are in this regard. This is exactly what will be pursued in the following paragraphs.

The survey

The scope of the research was limited to a random sample of 100 respondents, all mother-tongue speakers of Northern Sotho. The survey was conducted in May 2002, and the breakdown in terms of gender, age, birthplace and education/job is presented in condensed format in (1).

(1) Loan word survey — Basic respondents' data (N = 100)

Gender	Male: 47 Female: 53
Age	16: 1 17: 2 18: 8 19: 6 20: 5 21: 6 22: 6 23: 7 24: 5 25: 7 26: 4 27: 7 28: 5 29: 8 30: 4 31: 2 32: 5 34: 1 35: 2 36: 1 37: 1 38: 1 48: 1 53: 1 54: 2 65: 2
Birthplace	Atteridgeville, Bochum, Botlokwa, Burgersfort, Buschbuckridge, Ga-Dikgale, Ga-Kgoroshi, Ga-Ledwaba, Ga-Maja, Ga-Mamabolo, Ga-Masemola, Ga-Mashabela, Ga-Mashashane, Ga-Matlala, Ga-Molepo, Ga-Moraba, Ga-Mphahlele, Ga-Nkwana, Ga-Rankuwa, Ga-Sekhukhune, Giyani, Glencowie, Kgapanne, Lebowa-Kgomo, Lenenye, Magatle, Mahwelereng, Mohodi, Moletjie, Mphakane, Pietersburg, Potgietersrus, Pretoria, Seshogo, Siyabuswa, Steilloop, Turfloop
Education/Job	Accountant, Assistant Lecturer, Bank teller, Business person, Cashier, Consultant, Co-ordinator, Dental Assistant, Dietician, Driver, Engineer, Fashion designer, Housewife, Lawyer, Learner, Lexicographer, Librarian, Marketing official, Nurse, Professor, Student, Teacher, Unemployed

A total number of 64 single words were presented in pairs to the respondents, thus 32 pairs each containing a loan word and a (more) indigenous counterpart, e.g. **radio** versus **seyalemoya** 'radio', or **dimonamonane** versus **malekere** 'sweets', etc. Respondents were asked to mark the alternative(s) which they would like to see included in a Northern Sotho dictionary. A third column was added for comments and suggestions of other words considered to be still better than the two choices offered. Respondents were also invited to report spelling errors or to suggest improvement of spelling, and even to motivate why a word should be included or excluded from the dictionary. Finally, an informal conversation was conducted with each respondent in order to obtain additional information and an overall impression. A typical example of a completed questionnaire is reproduced in the Appendix. Compare also the translated version of the questionnaire in (2).

- (2) Loan word survey — Translation of the questionnaire (with in Columns **A** and **B**: **L** = loan word, and **I** = (more) indigenous word; and in Column **C**: the English translation)

Hello!		
Choose from A and B those words which, according to you, should be included in a Northern Sotho dictionary. You may choose either A or B , or both, and in C you may write any relevant comment, such as: any other word that you think is better than the two words already given in A and B , if you see a spelling mistake for the words given you may provide the correct spelling, or if you wish you may give reasons why you say the word you have chosen or given is the one to be included in the dictionary, or, conversely, you may give the reason why you say the other words should not be included in the dictionary, etc.		
A	B	C
L radio	I seyalemoya	<i>a radio</i>
I sefatanaga	L mmotoro	<i>a car</i>
I mmila	L seterata	<i>a street</i>
L watšhe	I sešupanako	<i>a watch</i>
L terene	L setimela	<i>a train (distracter)</i>
I Pherekgong	L Janeware	<i>January</i>
L dikšinare	I pukuntšu ³	<i>a dictionary</i>
I šoma	L bereka	<i>to work</i>
I fofo	L teye	<i>tea</i>
L sekerete	I motsoko	<i>a cigarette; tobacco (distracter)</i>
L khomphutha	L computer	<i>a computer (distracter)</i>
L amalantshe	L ambulantshe	<i>an ambulance (distracter)</i>
L tonki	I molahlwaleboya	<i>a donkey</i>
L sebara	I molamo	<i>a brother-in-law</i>
I nkadingala	L sankoma	<i>a witchdoctor</i>
L nalete	I lemao	<i>a needle</i>
L sekoloto	I mokitlana	<i>debt</i>
I bookelo	L sepetlela	<i>a hospital</i>
I letsikangope	L lefasetere	<i>a window</i>
L tšhoko	I motaga	<i>chalk</i>
I phapoši	L kamora	<i>a room</i>
I dimonamonane	L malekere	<i>sweets</i>
L reila	I otlela	<i>to drive</i>
I befelwa	L kwata	<i>become angry</i>
forminal	sephekgo	<i>(distracter)</i>
I ntlanya	L paesekela	<i>a bicycle</i>
L pata	I tsela	<i>a road</i>
L sepete	L garafo	<i>a spade (distracter)</i>
I sephuthelwana	L phasela	<i>a parcel</i>
L polane	I leano	<i>a plan</i>
L ranta	I lesome	<i>Rand</i>
L newspaper	L kuranta	<i>a newspaper (distracter)</i>
We assure that all the information you have provided will be processed anonymously. You are kindly requested to provide us with the following for research purposes:		
Age (e.g.: 29)		
Home language (e.g. Selobedu)		
Birthplace (e.g. Ga-Matlala)		
THANK YOU FOR YOUR COOPERATION!		

Note that the pairs were not presented in a fixed order. In some cases the loan word **L** is given first followed by the (more) indigenous word **I**, sometimes it is the reverse. The loan words themselves are all adopted from Afrikaans, English or Zulu — occasionally Sothoised, at other times direct borrowings. The corresponding (more) indigenous forms are either new coinages in Northern Sotho, shifts in meaning from existing Northern Sotho words, or simply traditional/original words.

Analysing the survey

Basically, there are three levels on which the questionnaires were analysed. Firstly, each respondent's input was analysed in isolation. This resulted in 100 user profiles (Level 1). Secondly, these 100 profiles were summed, upon which general tendencies became evident (Level 2). Thirdly, each **L-I** pair was studied separately, which provided interesting and highly specific data for each specific **L-I** pair. For each of these pairs, frequency counts derived from a corpus as well as the treatment in all currently available dictionaries were also taken into account (Level 3). Space considerations unfortunately do not allow us to present the outcome on these three levels exhaustively. Rather, some representative findings will be singled out.

The 100 filled-in questionnaires were processed in spread-sheet format, reflecting each pair and the full statistical response for every pair — viz. a score for **L**, a score for **I**, and a score for both — and this for Respondent 1, 2, ... up to Respondent 100. See (3).

(3) Analysing the loan word survey — Levels 1 and 2

	L-I	1	2	3	...	36	...	98	99	100	only L only I both	L or both I or both
radio	L		1			1	1	17%	42
seyalemoya	I	1	1	1	...	1	...	1	1	1	57%	82
		b	ab	b	...	b	...	b	ab	ab	25%	
...
malekere	L	1	1	1	...	1	...			1	24%	50
dimonamonane	I	1	1		1	1	1	45%	71
		ab	ab	a	...	a	...	b	b	ab	26%	
...
Age		22	32	27	...	18	...	29	37	24		
Home language		S	S	S	...	S	...	S	S	S		
Birthplace		Siy	Ml	Mt	...	Mt	...	Mj	Sk	Pta		
Gender		M	F	M	...	F	...	F	M	F		
Education/Job		S	L	S	...	S	...	U	T	S		
Address		M	S	U	...	Pta	...	Ptb	Ptb	Pta		

From (3) one can for instance see that 57% of the respondents opted for **seyalemoya**, 17% for **radio**, and as many as 25% for both. (Only one respondent did

R slightly favour **L** (42% vs. 38%). C seemingly (cf. below) only has **I**, and D agree on **I**. However, **L** appears as **Janaware** (19/0) in New E, T&O, Pop. and Grb., which is thus the form to be included in the D (and not **Janeware** (0/0)). The plural form for **Janaware** is **diJanaware** and for **Pherekgong** it is **diPherekgong**, yet C does not contain any of the two plurals. 16% of R agree on both **L** and **I**, while 4% suggest nothing. The preferred form by both R and C is thus **L**, but with spelling **Janaware**. D are therefore very wrong, as they mainly focus on **I**. Klein has nothing. T&O (p. 23): "The names of the months of the year are Sothoised instead of using the Sotho names, e.g.: *Janaware* (instead of *Pherekgong*) ...". One of the respondents, aged 21, who is a student, female and a Northern Sotho speaker wrote: "Go ya ka nna re swanetše re hlalose mantšu ao a šomišwago tšatši ka tšatši re se ke ra šomiša Sepedi sa kgale seo e le go [sic] gore bana ba matšatši a ba ka se se kwešiše, go swana le bo 'Pherekgong' [sic]." ("According to me we should only include those words that are used every day and avoid using old Sepedi words like **Pherekgong**, which children of today do not understand.") 15% of the respondents say the correct spelling is **Janaware** (19/0) not **Janeware** (0/0) which means that the correct word to be included in D is indeed **Janaware**. **Pherekgong** may also be included, but with a note pointing out the potential confusion (January vs. March), and a cross-reference to **Janaware**.

The extract shown in (5) is taken verbatim from the survey analysis. Even from this single (cryptic) **L–I** sketch the proscriptive approach should be evident: the target users and the corpus favour the loan word **L** (descriptiveness), and this happens to correspond with the Language Board's suggestion (prescriptiveness). However, as the target users and the corpus also indicate that the traditional word **I** has a right to be included in the dictionary, this 'variant' should also be treated on both the macro- and microstructural levels, albeit with the necessary cross-references and cursory notes (proscriptiveness).

Analysing the survey — Distracters

The **Janeware–Pherekgong** pair is one example of an **L–I** pair where the one-to-one correlation does not truly hold, yet where there is still a large overlap. Another example is shown in (6).

(6) Analysing the loan word survey — **mmila–seterata** pair

Items	%	Σ	Fsg Fst	Fpl -	Z&M	New E	Klein	T&O	V Wyk	P&S	P et al.	Pop.	Grb.
mmila	53	79	226	57	✓	✓	✓	-	✓	✓	✓	✓	✓
seterata	16	42	9	8	✓	✓	✓	✓	✓	✓	✓	✓	✓
both	26	-	-	-	-	-	-	-	-	-	-	-	-
∅	5	5	-	-	-	-	-	-	-	-	-	-	-

Taken at face value, the two options seem semantically different with **mmila** 'footpath' versus **seterata** 'street'. However, when the respondents' considerations are studied, it becomes clear that **mmila** and **seterata** can indeed be semantically linked via 'road' since some respondents and dictionaries seem to

imply a semantic continuum for the end points **mmila** versus **seterata** in foot-path ↔ road ↔ street. This overlap is obvious in the treatment of **mmila** and **seterata** in, e.g., Ziervogel and Mokgokong (1975):

(7)(a) **MMILÁ** ... footpath, trail (of game), road, street, side-walk

(7)(b) **-TÉRÁTA, se-** ... street

Pairs such as **Janeware–Pherekgong** and **mmila–seterata** were purposely inserted into the survey in an attempt to trigger comments that would enable the deduction of the overriding sentiment with regard to loan words versus (more) indigenous words. For the latter pair, several respondents aged between 21 and 36, from different areas in both the Limpopo and Gauteng provinces, state that **seterata** is equivalent to **mokgotha** (13/47). This is (partly) true and clearly points in the direction of a preference for indigenous words.

Apart from pairs which only partly overlap, carefully selected *distracters* were also built into the questionnaire. These have been marked as such in (2) above. Their main aim was to verify the quality of the respondents' feedback. Some distracter-pairs consist of two loan words **L**, for others the two options have very different meanings, and a third category was added just to find out if the respondents themselves were consistent. Consider (8) and (9) as examples of the first category.

(8) Analysing the loan word survey — **newspaper–kuranta** pair

Items	%	Σ	Fsg Fst	Fpl -	Z&M	New	E	Klein	T&O	V	Wyk	P&S	P et al.	Pop.	Grb.
newspaper	1	7	0	0	-	-	-	-	-	-	-	-	-	-	-
kuranta	82	88	168	84	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
both	6	-	-	-	-	-	-	-	-	-	-	-	-	-	-
∅	11	11	-	-	-	-	-	-	-	-	-	-	-	-	-

(9) Analysing the loan word survey — **khomphutha–computer** pair

Items	%	Σ	Fsg Fst	Fpl -	Z&M	New	E	Klein	T&O	V	Wyk	P&S	P et al.	Pop.	Grb.
khomphutha	61	68	18	4	-	-	-	✓	-	-	-	-	-	✓	✓
computer	10	17	0	0	-	-	-	-	-	-	-	-	-	-	-
both	7	-	-	-	-	-	-	-	-	-	-	-	-	-	-
∅	22	22	-	-	-	-	-	-	-	-	-	-	-	-	-

In these examples the respondents were presented with a direct borrowing (not the purpose of the investigation) and a (Sothoised) loan word. In (8) just one respondent preferred **newspaper** as the only option and 6 allowed for both, while in (9) 10 opted for **computer** and another 7 for both. These distracter-pairs thus managed to discriminate well. Compare also the following remark by one of the respondents:

Dealing with two non-words, still half of the respondents opted for what they perceived to be an original word.⁵ At first glance this might question the reliability of the feedback of a huge number of the respondents. However, it is a clear confirmation of the respondents' general preference for indigenous words over loan words to be treated in a dictionary. Compare the following telling remark from one of the respondents in this regard:

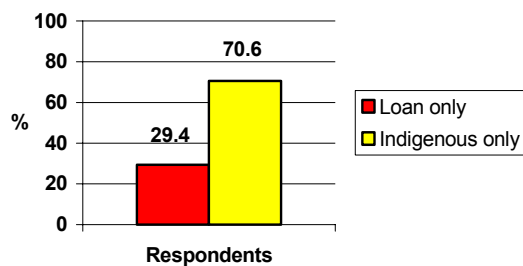
The word **sephkego** is unknown to me. I chose it because I want to know what it means and I think that if it is included in the dictionary, most of the people who are like me and don't know anything about the word will know and understand it once it is explained in the dictionary. I therefore request that you include it.

General findings of the survey

The general findings of the survey will now be summarised. For the statistics presented below, the data for the randomly-interspersed distracters have been excluded.

Firstly, one can calculate the percentage of respondents in favour of loan words as opposed to the respondents in favour of their (more) indigenous counterparts. The results are shown in (13).

(13) Respondents' preference for loan words vs. their (more) indigenous counterparts



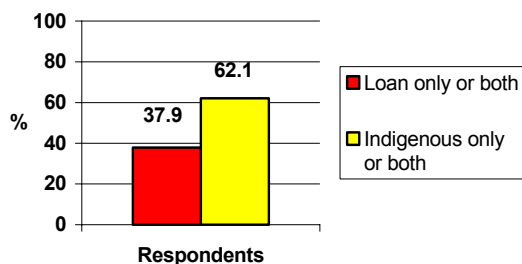
This graph clearly indicates that an overwhelming more than two-thirds majority of those who only accept one option are in favour of the indigenous word. The following remark by one of the respondents is quite revealing in this respect:

I personally believe that our language will lose value if more and more words from other languages are accepted in Northern Sotho. If we have original Northern Sotho words, why do we have to loan words from other languages? I don't see the necessity for us to loan words from other languages if we have our own original words. Only those words which do not exist in our language could be

loaned from other languages, for example AIDS. The Zulus have managed to formulate their own word for AIDS and called it **nxolazi** [sic], why can't we the Northern Sotho people do the same and stop loaning from other languages?

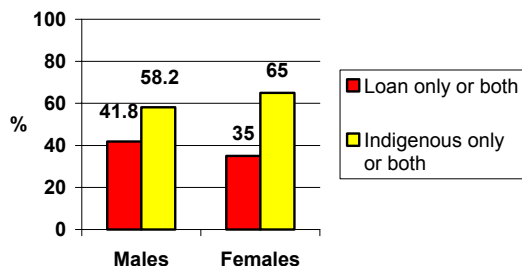
If one studies the overall pattern for 'loan + both' versus 'indigenous + both', then one arrives at the pattern shown in (14).

(14) Respondents' preference for 'loan words or both' vs. their '(more) indigenous counterparts or both'



Secondly, differentiating between preferences of males versus females results — with the data shown in (14) as a point of departure — in the percentages shown in (15).

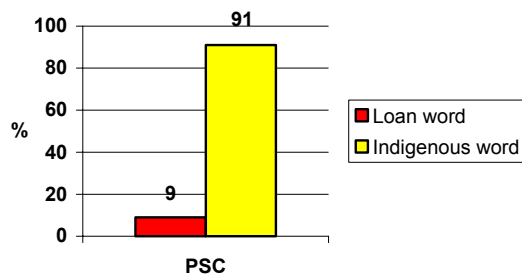
(15) Males' and females' preference for 'loan words or both' vs. their '(more) indigenous counterparts or both'



From (15) one can clearly see that males tend to prefer the inclusion of loan words in dictionaries more than females do, as they suggest 6.8% more of them on average.

Thirdly, the above findings should be compared with the occurrence of loan words versus (more) indigenous counterparts in the corpus. The latter is done in (16).

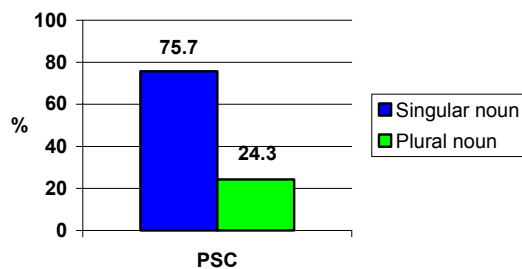
(16) Loan words vs. their (more) indigenous counterparts in PSC



The Pretoria Sepedi Corpus (PSC) is based for 100% on written sources. Ten times more indigenous words appear in PSC than their loan counterparts. This is a significant and clear indication that indigenous words take overwhelming preference over loan words in written texts. If a dictionary were solely based on corpus data, chances would be good that loan words would turn out to be undertreated compared to the respondents' preferences. The difference between the respondents' preferences for loan words and the corpus attestations for those same loan words is as high as 20.4% (= 29.4 – 9.0). In order for a Northern Sotho dictionary to reflect the true needs of the community, it is thus clear that this research reveals an important hiatus in an approach *solely* based on corpus data.

The graph shown in (17) indicates that the nouns that were the topic of this study appear more often in their singular form than in their plural form in PSC.

(17) Singular vs. plural nouns in PSC

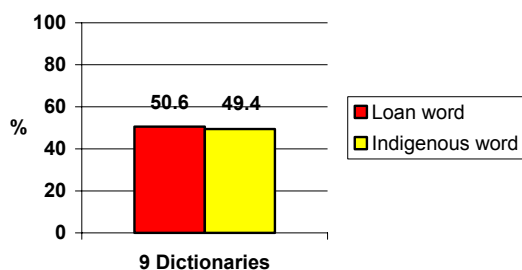


This simply means that, as far as the microstructure of a Northern Sotho dictionary is concerned, the studied nouns should rather be treated and exemplified in the singular. A clear exception is of course **malekere** / **dimonamonane**

'sweets' discussed above.

Fourthly, the respondents' input and the corpus data should also be compared with the overall treatment in the 9 currently-available dictionaries. The latter is summarised in (18).

(18) Loan words vs. their (more) indigenous counterparts in 9 Northern Sotho dictionaries



From (18) one can see that, on average, the 9 available dictionaries for Northern Sotho treat both loan words and their (more) indigenous counterparts on a par. In order to see whether or not this situation would suit future dictionary users, the next series of analyses needs to be presented first.

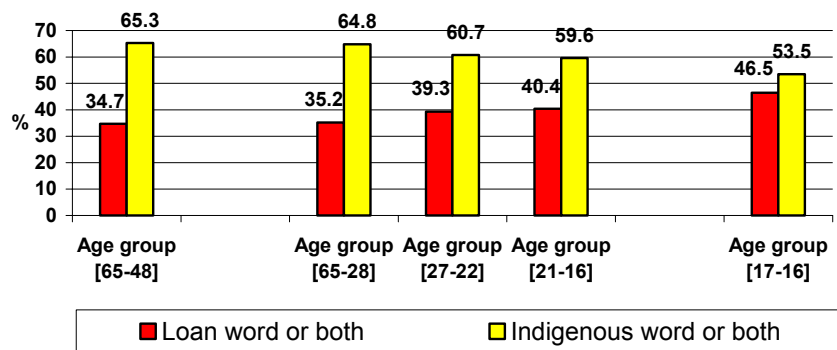
Fifthly, and lastly, the sum of the user profiles should also be brought into account. As far as the profiles for rural versus urban birthplaces are concerned, the results are unfortunately inconclusive, probably as a result of the fact that most respondents commute frequently between, and live in, both places. This is an area which should be researched further. When the age profiles are summed, however, a clear pattern emerges. Indeed, the most significant finding of the survey is obtained when the respondents' answers are broken down according to age groups. In the shaded columns of (19) the respondents' feedback is divided into three major age groups, namely 16-21, 22-27 and 28-65 (each containing roughly the same number of respondents). On both sides of this shaded block the preferences for L versus I of the extreme age groups 16-17 and 48-65 are also indicated.

(19) Distribution across age groups of the respondents' preference for 'loan words or both' vs. the respondents' preference for their '(more) indigenous counterparts or both' — 1. data

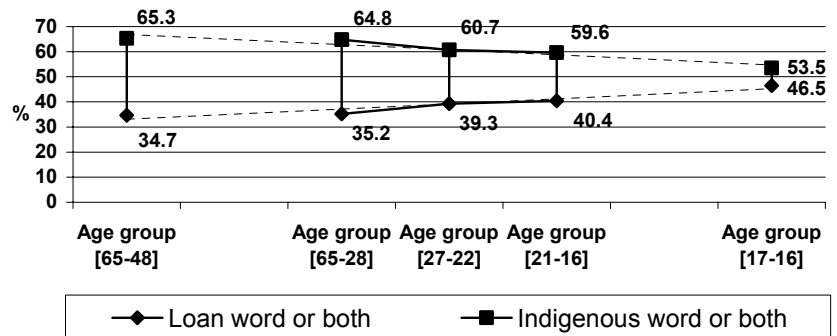
	<i>upper extreme</i> Age group [65-48]	Age group [65-28]	Age group [27-22]	Age group [21-16]	<i>lower extreme</i> Age group [17-16]
Loan word or both	34.7	35.2	39.3	40.4	46.5
Indigenous word or both	65.3	64.8	60.7	59.6	53.5

Thus, read from left to right, the preference for loan words gradually increases from 34.7% for the age group 48-65 to 46.5% for the age group 16-17. Conversely, the preference for their (more) indigenous counterparts decreases from 65.3% to 53.5% across the same spectrum. This is shown graphically in (20) and (21) by means of respectively histograms and trendlines.

(20) Distribution across age groups of the respondents' preference for 'loan words or both' vs. the respondents' preference for their '(more) indigenous counterparts or both' — 2. histograms



(21) Distribution across age groups of the respondents' preference for 'loan words or both' vs. the respondents' preference for their '(more) indigenous counterparts or both' — 3. trendlines



Graph (21) clearly indicates that the gap between preference for indigenous words and preference for loan words narrows as the age of the respondents decreases. This discovery is of crucial importance to dictionary compilers, especially for the compilation of dictionaries for younger mother-tongue speak-

ers of Northern Sotho as target users. At the same time this result renders justification for the current dictionary situation as indicated in (18) above, where an almost equal treatment of loan words versus indigenous counterparts was found.

A typical manifestation of this phenomenon could be the preference pattern for **teye** versus **fofo**, both meaning 'tea', where it is clear that most of the respondents who chose the indigenous **fofo** are aged between 40 and 53, whereas those who opted for the loan word **teye** belong to the younger generation (compare in this regard the findings of Slabbert and Finlayson 1999).

Conclusion

In this article, loan words versus their (more) indigenous counterparts in Northern Sotho were studied from a lexicographic perspective. Within a prescriptive approach to dictionary compilation, the selection of the main variant is based on the analysis of a sound empirical basis, whilst that empirical basis also provides the lesser-important items to be treated. It was shown that the by-now standard empirical basis consisting of data derived from an electronic corpus is not good enough for the treatment of loan words, and that fieldwork is imperative. The latter was achieved by means of a survey conducted among 100 mother-tongue speakers of Northern Sotho.

By studying the respondents' comments as a whole, it is clear that they prefer the (more) indigenous words to be treated in dictionaries, and that loan words should only be used if there is no good alternative in Northern Sotho. Quite a number of respondents even suggest that words should be coined in order to have a Northern Sotho word instead of an adoptive from other languages. Where offered a direct borrowing and a (Sothoised) loan as only options, the (Sothoised) loan is preferred to the direct borrowing. This thus suggests the following preference hierarchy: *indigenous word* → *Sothoised loan word* → *direct borrowing*.

The most important finding of this study is that younger respondents seem to accept loan words much more easily than the older generation. This might be a result of the intensified influence from other languages in both rural and urban areas, as well as a direct consequence of the fact that most teenagers are no longer enrolling for Northern Sotho as a school subject. Also, older people tend to favour so-called 'old words' that are no longer known by the youth of today.

All in all, however, first preference is still given to (more) indigenous words over loan words whenever there is a choice. This pattern is rapidly changing, and today's dictionaries should definitely pay more attention to loan words than the dictionaries compiled half a century ago. A watchful eye will have to be kept on this evolving preference pattern.

Endnotes

1. Nowadays, one often finds so-called politically correct phraseologies such as "primary language" and "home language" for "mother tongue".
2. We are aware of the fact that a great deal of linguistic research has been devoted to what linguists variously call 'loan words', 'borrowings', 'adoptives', 'transliterations', etc. In this article, however, the focus is on lexicography, and we only use those terms *in contrast* to their traditional/original counterparts. We refer to the latter as '(more) indigenous counterparts'.
3. Strictly speaking **pukuntšu** is a combination of **puku** 'book' (a loan word) and **lentšu** 'word' (an indigenous word).
4. The noun **setimela**, e.g., has a class prefix of class 7 **se-**, while **terene** has no class prefix.
5. The non-words **forminal** and **sephkego** were 'created' in such a way that they resemble genuine English and Northern Sotho words respectively. Actually, **forminal** was derived from a permutation of sections of the English word 'informal'. Unfortunately, an Internet search with Google (<http://www.google.com/>) returns 56 hits, revealing that **forminal** is a technical neurology term used in for instance 'Forminal Stenosis of Cervical Spine' or 'Extra Forminal Non-discogenic Lumbar Nerve Entrapment as a Cause of Sciatica'. We can however safely assume that no respondent knew this technical sense. The non-word **sephkego** seems to be a Northern Sotho noun belonging to class 7 **se-**. Fortunately, no Internet pages were found containing the non-word **sephkego**.

List of Dictionary and Corpus Abbreviations

Grb	Gerber 2000
Klein	Ziervogel and Mokgokong 1988 ⁴
New E	Kriel 1976 ⁴
P et al.	Prinsloo, Sathekge and Kapp 1997
P&S	Prinsloo and Sathekge 1996
Pop.	Kriel, Prinsloo and Sathekge 1997 ⁴
PSC	Pretoria Sepedi Corpus (5.8 million running words)
T&O	Departmental Northern Sotho Language Board 1988 ⁴
V Wyk	Kriel, Van Wyk and Makopo 1989 ⁴
Z&M	Ziervogel and Mokgokong 1975

References

- Al-Kasimi, Ali M. 1977. *Linguistics and Bilingual Dictionaries*. Leiden: E.J. Brill.
- Bergenholtz, Henning. 2001. Proskription, oder: So kann man dem Wörterbuchbenutzer bei Textproduktionsschwierigkeiten am ehesten helfen. Lehr, Andrea, Matthias Kammerer, Klaus-Peter Konerding, Angelika Storrer, Caja Thimm and Werner Wolski (Eds.). 2001. *Sprache im Alltag. Beiträge zu neuen Perspektiven in der Linguistik. Herbert Ernst Wiegand zum 65. Geburtstag gewidmet*: 499-520. Berlin: de Gruyter.

- Bergenholtz, Henning.** 2002. Towards a User-oriented Understanding of Descriptive, Proscriptive and Prescriptive Lexicography. De Schryver, Gilles-Maurice (Ed.). 2002. *AFRILEX 2002, Culture and Dictionaries, Programme and Abstracts*: 12-13. Pretoria: (SF)² Press.
- Departmental Northern Sotho Language Board.** 1988^a. *Northern Sotho Terminology and Orthography No. 4 / Noord-Sotho terminologie en spelreëls No. 4 / Sesotho sa Leboa mareo le mongwalo No. 4*. Pretoria: Government Printer.
- Gerber, Hendri H.** 2000. *Woordeboek Afrikaans-Noord-Sotho / Pukuntšu Seburu-Sesotho sa Leboa*. Eldoraigane: Arbeidsprestasie BK.
- Gouws, Rufus H.** 1989. *Leksikografie*. Pretoria: Academica.
- Gove, Philip B.** (Ed.). 1961³. *Webster's Third New International Dictionary of the English Language*. Springfield: Merriam-Webster.
- Gove, Philip B.** 1961a. Letter to the Editor. *Life Magazine*, 17 November 1961: 13.
- Johnson, Samuel.** 1755. *A dictionary of the English language: in which the words are deduced from their originals, and illustrated in their different significations by examples from the best writers; to which are prefixed, A history of the language, and an English grammar*. London: Printed by W. Strahan, for J. and P. Knapton.
- Kriel, Theunis J.** 1976^a. *The New English-Northern Sotho Dictionary, English-Northern Sotho, Northern Sotho-English*. Johannesburg: Educum Publishers.
- Kriel, Theunis J., D.J. Prinsloo and Bethuel P. Sathekge.** 1997^a. *Popular Northern Sotho Dictionary, Northern Sotho-English, English-Northern Sotho*. Cape Town: Pharos.
- Kriel, Theunis J., Egidius B. van Wyk and S.A. Makopo.** 1989^a. *Pukuntšu woordeboek, Noord-Sotho-Afrikaans, Afrikaans-Noord-Sotho*. Pretoria: J.L. van Schaik.
- Landau, Sidney I.** 2001. *Dictionaries: The Art and Craft of Lexicography (2nd edition)*. Cambridge: Cambridge University Press.
- Prinsloo, D.J.** 1992. Prescriptiveness versus Descriptiveness — an Approach for Northern Sotho Lexicography. *South African Journal of African Languages* 12(1): 7-11.
- 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.
- Slabbert, Sarah and Rosalie Finlayson.** 1999. The Future of the Standard African Languages in the Multilingual South African Classroom. Rosalie Finlayson (Ed.). 1999. *African Mosaic: A Festschrift for J.A. Louw on his 75th Birthday*: 235-249. Pretoria: Unisa Press, University of South Africa.
- Wells, Ronald A.** 1973. *Dictionaries and the Authoritarian Tradition. Study in English Usage and Lexicography*. Janua Linguarum. Series Practica, 196. Berlin: Mouton de Gruyter.
- Zgusta, Ladislav.** 1971. *Manual of Lexicography*. The Hague: Mouton.
- 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.** 1988^a. *Klein Noord-Sotho woordeboek, N.-Sotho-Afrikaans-English, Afrikaans-N.-Sotho, English-N.-Sotho*. Pretoria: J.L. van Schaik.

Appendix: Reproduction of a randomly-selected filled-in questionnaire (# 36)

36

Thobela!

Kgetha go tšwa go A le B mantšu ao o naganago gore a loketšwe go ba gona ka mo gare ga pukuntšu ya Sesotho sa Leboa. O ka no kgetha ka mo go A le ka go B gomme o ngwale se sengwe ka mo go C, bjalo ka; lentšu leo o naganago gore le kaone go phala a mabedi ao a filwego ka go A le B, ge o bona phošo ya mopeleto mo manišung ao a filwego ka go A le B o ngwale mopeleto wa nnete wa mantšu ao, ge o rata o ka no fa mabaka a gore ke ka lebaka la eng o re lentšu leo le loketšwe go ba ka gare ga pukuntšu goba ga la lokelwa, bj. bj.

A	B	C
radio	seyalemoya ✓	Radio ke lentšu leo le atlingwilewego.
sefatanaga ✓	mmotoro	mmoto le tšwa go English.
mmila ✓	seterata	Tsela e ka ba le lengwe leo le ka lokelwa.
watšhe	sešupanako ✓	
terene	setimela ✓	Tšhutšumakgale le tona e ka šomišwa
Pherekgong ✓	Janeware	
dikšinare	pukuntšu ✓	
šoma ✓	bereka	
fofo	teye ✓	Teye fofo ke sona sesotho ešwa.
sekerete	motsoko ✓	
khomphutha ✓	computer	Computer ke English ga sa Sesotho.
amalantshe	ambulantshe	
tonki	molahlwaleboya ✓	
sebara	molamo ✓	
nkadingala ✓	sankoma	
nalete	lemao ✓	
sekoloto	mokitlana	Ka moka a ka šomišwa.
bookelo ✓	sepetlela	
letsikangope ✓	lefasetere	
tšhoko	motaga	Nke e tšeke (a e gona ko gae).
phapoši ✓	kamora	
dimonamonane	malekere ✓	
reila	otlela ✓	
befelwa ✓	kwata	Kwata ga se sesotho se nnete.
forminal	sepheko ✓	
ntlanya	paesekela ✓	Le ge e e gore paesekela ke lead ingwana.
pata	tsela ✓	
sepete	garafo	Garafo e swanetše go ba kgarafo gape.
sephuthelwana ✓	phasela	
polane	leano ✓	
ranta ✓	lesome	lesome ga le kwagala ka go re R ke currency ya SA.
newspaper	kuranta ✓	

Re go netefaleša gore tšohle tše o di ngwadilego fa e tla ba sephiri. O kgopelwa gore o tlatše ka mo fase: Mengwaga (Mohlala: 29) 1.8
 Leleme la ka gae (Mohlala: Selobedu) Sepedi
 O belegetšwe kae (Mohlala: Ga-Matlala) Ga-Matlala (C.phofu)

RE LEBOGA GO KGATHA TEMA GA GAGO!

The Lemmatization of Copulatives in Northern Sotho*

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

Abstract: For learners of Northern Sotho as a second or even foreign language, the copulative system is probably the most complicated grammatical system to master. The encoding needs of such learners, i.e. to find enough information in dictionaries in order to actively use copulatives in speech and writing, are poorly served in currently available dictionaries. The aim of this article is to offer solutions to the lemmatization problems regarding copulatives in Northern Sotho and to propose guiding entries for paper and electronic dictionaries which could serve as models for future dictionaries. It will be illustrated that the maximum utilisation of macrostructural and microstructural strategies as well as the mediostructure is called for in order to reach this objective. Prerequisites will be to reconstruct the entire copulative system in a user-friendly way, to abstract the rules governing the use of copulatives and to isolate the appropriate lemmas. The treatment of copulatives in Northern Sotho dictionaries will also be critically evaluated, especially in terms of frequency of use and target users' needs.

Keywords: LEXICOGRAPHY, LEMMATIZATION, COPULATIVES, INFORMATION RETRIEVAL, ACCESS STRUCTURE, ELECTRONIC DICTIONARY, MACROSTRUCTURE, MICROSTRUCTURE, CROSS-REFERENCING, MEDIOSTRUCTURE, DICTIONARY, AFRICAN LANGUAGES

Opsomming: Die lemmatisering van kopulatiewe in Noord-Sotho. Vir aanleerders van Noord-Sotho as tweede of vreemde taal is die kopulatief waarskynlik die mees komplekse grammatiese sisteem om te bemeester. Die enkoderende behoeftes van sulke aanleerders, dit is om genoegsame inligting in woordeboeke te verkry ten einde kopulatiewe in spraak en skrif aktief te kan gebruik, word nie bevredig in beskikbare woordeboeke nie. Die doel van hierdie artikel is om oplossings aan die hand te doen vir die lemmatiseringsprobleme ten opsigte van kopulatiewe in Noord-Sotho en om gidsinskrywings voor te hou wat as modelle kan dien vir toekomstige papier en elektroniese woordeboeke. Daar sal aangetoon word dat maksimale benutting van makrostrukturele en mikrostrukturele strategieë asook die mediostruktuur 'n vereiste is vir bereiking van hierdie doelwit. 'n Voorvereiste is dat die struktuur van die totale kopulatiewe sisteem op gebruikersvriendelike wyse voorgestel moet word. Die reëls wat die gebruik van die kopulatief bepaal kan dan geabstraheer word. Daarna moet die toepaslike lemmas geïdentifiseer word. Die bewerking van kopulatiewe in Noord-Sothowoordeboeke sal ook krities geëvalueer

* An earlier version of this article was presented as a paper at the Seventh International Conference of the African Association for Lexicography, organised by the Dictionary Unit for South African English, Rhodes University, Grahamstown, 8–10 July 2002.

word, veral in terme van gebruiksfrekwensie en gebruikersbehoefes.

Sleutelwoorde: LEKSIKOGRAFIE, LEMMATISERING, KOPULATIEWE, INLIGTINGS-ONTSluitING, TOEGANGSTRUKTUUR, ELEKTRONIESE WOORDEBOEK, MAKROSTRUKTUUR, MIKROSTRUKTUUR, KRUISVERWYSING, MEDIOSTRUKTUUR, WOORDEBOEK, AFRIKATALE

Introduction

From a learner's point of view, the copulative in Northern Sotho is probably the most complicated grammatical system to master. Northern Sotho grammars and dictionaries often fail in many respects to treat copulatives in a satisfactory way. Grammar books generally do not give learners a *comprehensive overview* of the copulative *in a logical way*.¹ Copulatives are either described in bits and pieces or treated in such an elaborate and *ad hoc* manner that the learner fails to get an overall picture that is essential for comprehension of such a complicated system. Currently available dictionaries for Northern Sotho also do not treat copulatives in a systematic way. Copulatives in these dictionaries are lemmatized and treated at random and in an *ad hoc* manner as they come under the compiler's attention.

For the lexicographer it is imperative to depart from giving an overall account or the full scope of copulatives when planning the macrostructure and microstructure in respect of their treatment. An attempt will firstly be made to construct such a comprehensive account in a user-friendly way by means of an *example-driven* approach which as such is suitable to form the back matter entry for copulatives. From this detailed (re)construction it will be attempted to abstract the rules underlying or governing the use of copulatives in Northern Sotho. Once these rules have been deducted/defined, candidates for lemmatization will be isolated and the compilation of model entries for both paper and electronic dictionaries will be attempted. It will be clearly illustrated that maximum utilization of access structures in terms of the macrostructure, microstructure and mediostructure is essential for the successful treatment of copulatives. The construction of back matter entries in particular, and the required mediostructural strategies required to successfully link central text entries with the back matter is crucial.

A brief theoretical perspective

Theoretical aspects pertaining to the lemmatization of copulatives in Northern Sotho in respect of macrostructure, microstructure, mediostructure and access structure will firstly be outlined briefly. Definitions of the two key concepts that constitute the title of this article, namely *copulatives* and *lemmatization* will be given first.

Louwrens (1994: 40) defines the term *copulative* as follows:

copulative (leba, kopulatief)

A term used to refer to structures which incorporate a variety of *prefixes* (called particles in some grammatical descriptions of Northern Sotho) of which some resemble the *subject concord* (e.g. *o bohlae*, *le bohlae*, *e bohlae*, etc. 'he/it is clever'), whereas others have a unique form like *ke* (e.g. *ke monna* 'it is a man') and *ga se* (e.g. *ga se monna* 'it is not a man'), as well as a series of verb forms with the *copulative verb stems* *-bê* (e.g. ... *gore a bê bohlae* 'so that he may become clever'), *-ba* (e.g. ... *a ba bohlae* 'and then he became clever'), *-le* (e.g. ... *gê a le bohlae* 'if he is clever'), *-se* (e.g. ... *gê a se bohlae* 'if he is not clever') and *-na* (e.g. *o na le lehufa* 'she is jealous'). The copulative expresses the English verb *to be*, and is therefore translated with 'is', 'was', 'will be', 'has been', etc.

Taljarid (1999: 3) justifiably remarks that linguists disagree on terminology in respect of copulatives:²

Wat die terminologie rakende die kopulatief en sy samestellende dele betref, is dit by nadere ondersoek duidelik dat die linguïste hier in 'n taalkundige mynvelde bevind. Aan die een kant word verskillende terme deur grammatici gebruik om na dieselfde begrippe te verwys. Aan die ander kant word dieselfde terme gebruik, maar verskillend geïnterpreteer deur verskillende grammatici.

Louwrens (1991: 63) states that clear terminological distinctions should be drawn between the terms 'copulative', 'subject', 'copula' and 'complement'. He cites the example: *Madika ke morutiši* 'Madika is a teacher' where 'copulative' is the umbrella term for the entire construction, *Madika* = subject, *ke* = copula and *morutiši* = complement. In Poulos and Louwrens (1994: 290) they, however, decide not to draw a distinction between the terms 'copula' and 'copulative'. From a lexicographic point of view, the value of such a distinction is questionable anyhow and therefore 'copulative' will be used throughout in this article.

Hartmann and James (1998) defines *lemmatization* as 'the reduction of a paradigm of variant word forms to a canonical form e.g. the inflected forms (-s, -ed, -ing etc.) of English verbs to the infinitive'. They define *canonical form* as 'the base form under which several variants of a word or phrase can be cited as a headword for a dictionary entry'. Thus, in terms of this definition *walk*, *walks*, *walked*, *walking*, etc. can be lemmatized as *walk*. For more advanced users of a Northern Sotho dictionary the words *moreki* 'buyer', *morekiši* 'seller', *rekela* 'buy for', *theko* 'price', etc. can all be lemmatized as *reka*.

Their use of the term *headword* focuses the attention on a 'word' being the key element as lemma in the macrostructure of a dictionary. Gouws (1989: 84), however, rightly emphasizes that such a traditional focus on the word as representative of the lexicon should be shifted to lemmas representing the lexical items of the particular language.

Die leksikon [bevat] veel meer as net woorde ... Leksikale items ... sluit ook elemente wat groter as woorde en elemente wat kleiner as woorde is, in. Die

struktuur van die leksikon moet dus gesien word as gebaseer op leksikale items. Hierdie leksikaalgebaseerde struktuur van die leksikon staan in opposisie tot die tradisionele woordgebaseerde struktuur van woordeboeke. ... Die leksikograaf [moet] in sy keuse van makrostruktuurelemente die fokus verskuif vanaf trefwoorde wat die woorde van die taal verteenwoordig, na lemmas wat die leksikale items van daardie taal verteenwoordig. ... Die lemmas van 'n woordeboek [moet] die hele reeks leksikale items verteenwoordig ... woorde asook elemente kleiner en groter as woorde [moet] lemmastatus kry.

For the lemmatization of copulatives in Northern Sotho the final remark, namely that lexical items often consist of elements bigger or smaller than orthographic words, is of particular importance. It will be argued that the copulatives *ga se, e le, e se, e ba*, etc. are all excellent examples of multiword lemmas which have not been entered as such in Northern Sotho dictionaries.³

The lemmatization of copulatives in modern dictionaries cannot be described without focusing on the so-called *user-perspective* which has emerged as an all important criterion in the selection and lexicographical treatment of lexical items. Lexicographers such as Gouws and Prinsloo (1998), Hartmann and James (1998), Prinsloo and De Schryver (1999), Gouws (2000), etc. emphasize the importance of the user-perspective in modern dictionary compilation. A dictionary should not primarily reflect the attitude of the lexicographer but it should rather be aimed at specific needs of a well-defined target user. Dictionary makers are constantly striving to improve the quality of their reference works, especially in respect of enhanced information retrieval strategies. The aim should be to present the dictionary text as user-friendly as possible. Gouws and Prinsloo (1998: 18) write:

The user-perspective, so prevalent in modern-day metalexigraphy, compels lexicographers to compile their dictionaries according to the needs and research skills of well-defined target user groups. The dominant role of the user has had a definite effect on the compilation of dictionaries as well as on the evaluation of their quality. Good dictionaries do not only display a linguistically sound treatment of a specific selection of lexical items. They are also products that can be used as linguistic instruments by their respective target user groups. The better they can be used, the better dictionaries they are.

Lemmatizing the complex copulative system in a user-friendly way poses a major challenge to the lexicographic skills of a dictionary compiler and can be regarded as a text-book example of the lexicographer's duty as a *mediator* between a complicated linguistic issue on the one hand and the dictionary user on the other. His/her task as mediator is to find ways to lexicographically treat copulatives in such a way that it is user-friendly. According to Prinsloo and Gouws (1996: 103), lexicographers compiling Northern Sotho dictionaries have to rely on insights from the field of metalexigraphy:

As the lexicographer is the mediator between theoretical linguistics and the everyday language user, it is important that a dictionary should convey enough information to satisfy the needs of the user and it should present the information in such a way that the relevant linguistic patterns and peculiarities of the target language can be emphasized. Within the boundaries of metalexicographical criteria a lexicographer should also be innovative enough to apply and adapt the theoretical insights, which are not language specific, to a specific language.

In addition to satisfying the *decoding* needs of the target user, namely to correctly interpret copulatives, particular attention should also be given to the *encoding* needs of the user, i.e. providing sufficient assistance to the user to produce correct copulative structures in speech or writing. Fox (1987: 138-139) states in this regard:

It is important that we realize that learners' needs for encoding are at least as urgent as their needs for decoding. More so, perhaps, in these days when such strong emphasis is placed on communication. ... We must ensure that the information we give students will genuinely help them to produce language of their own which is as near as possible to that produced by native speakers. In the past, dictionaries have been seen more as aids to decoding than aids of encoding. There is no reason why they should not succeed in doing both. ... Dictionaries have traditionally concentrated more on the grammatical, partly perhaps because they have been seen as tools for interpreting language rather than for creating it.

In order to achieve these objectives the lexicographer should optimise the three major structural components of dictionaries, namely the *macrostructure*, *microstructure* and *access structure* (cf. Gouws 1996: 19). On the *macrostructural* level the lexicographer must ensure that all copulative forms, be they particles, conCORDS or verb forms, are lemmatized. On the *microstructural* level the treatment of each copulative lemma should be done with the decoding and encoding needs of the target user in mind. In terms of Prinsloo and Gouws (1996: 104), the linguistic and communicative needs of the target user should determine both the macrostructural and microstructural selection.

The dictionary should be a practical linguistic instrument. Hartmann and James (1998: 2) define *access* as 'the relative ease with which information can be located in a reference work' and state that 'in a general dictionary this may be the alphabetical order in which headwords appear'. Compare also Gouws and Prinsloo (1998: 18):

The quality of dictionary use, that is the degree of success a user experiences when consulting a dictionary and employing the retrieved information, is determined by a variety of features but one of the most important characteristics of a good dictionary is its accessibility. Accessibility leads to an unambiguous retrieval of the information presented on both the macro- and microstructural levels.

The *outer access structure* (cf. Gouws 1996: 19 and Louw 1999: 110) should successfully guide the user to the lemmatized canonical form or multiword lemma of each and every copulative likely to be looked for by the target user. The *inner access structure* then guides the user to the specific information categories within the article of each lemmatized copulative. As will be illustrated below, a dedicated, preferably corpus-based, approach towards the compilation of the latter microstructural categories is required for copulatives. However, it will be emphasized that, even rich and well-constructed articles for copulatives are at best insufficient for the encoding needs of the target user and should be *strongly* supported by the *mediostructure* of the dictionary. The *mediostructure* or *system of cross-referencing* can be described in terms of Wiegand (1996) as referring the dictionary user from a reference position to a reference address where additional information can be obtained. This is done by a variety of reference markers such as *see, compare, →*, etc. (See Gouws and Prinsloo 1998: 19 for a detailed discussion.) Of particular importance for copulatives are the so-called *external reference addresses* where the cross-reference exceeds the boundaries of the article, referring the user from the central text to the *back matter* or even to an external source such as a grammar book. For copulatives, it will be argued, the entire route of information retrieval, from the outer access structure right up to one (or more) external reference address(es), should entail a comprehensive and logical *build-up of information* for the user. Here cross-reference to well-compiled tables and additional information reflecting the *total scope* of the copulative is essential. It is here that the user must be presented with the 'overall picture' of the copulative system, or more formally, that the *mediostructure* should be employed to counter the decontextualisation which is brought about by the alphabetical ordering of copulative lemmas in the central text.⁴

Complexity of the copulative

It has been claimed above that the copulative system in Northern Sotho is complicated. Poulos and Louwrens (1994) devote 37 pages to the discussion of copulatives and summarises it as follows (Poulos and Louwrens 1994: 320-326):

CONSOLIDATION OF COPULATIVE PATTERNS⁵

The various patterns for the different basic tense forms of the copulative are listed below, for ease of reference. It must be remembered that *CB*, i.e. the *complement* of the copula represents different entities or word categories, depending on the type of copulative that is used. The following abbreviations are used: *SC* *subject concord*, *CP* *copulative prefix*, *CB* *copulative base* which is also termed the *complement*.

The identifying copulative

The indicative series The present tense Principal Identifying pos. 1st and 2nd persons: *SC - CB* Classes: *CP - CB* neg. 1st and 2nd persons: *ga - SC - CB* Classes: *ga - se - CB* Participial pos. 1st and 2nd person: *SC - le - CB* Classes: *CP - le - CB* neg. 1st and 2nd person: *SC - se - CB* Classes: *CP - se - CB*

The future tense Principal pos. 1st and 2nd person: SC - tlô/tla - ba + CB Classes: CP - tlô/tla - ba + CB neg. 1st and 2nd person: SC - ka - se - bê + CB SC Classes: CP - ka - se - bê + CB *Participial* pos. 1st and 2nd person: SC - tlô/tla - ba + CB Classes: CP - tlô/tla - ba + CB neg. 1st and 2nd person: SC - ka - se - bê + CB Classes: CP - ka - se - bê + CB *The past tense Principal* pos. 1st and 2nd person: SC - bilê + CB Classes: CP - bilê + CB neg. 1st and 2nd person: ga - se - SC - be + CB ga - se - SC2 - a - ba + CB ga - SC2 - a - ba + CB Classes: ga - se - CP - bê + CB ga - se - SC2 - a - ba + CB1 ga - SC2 - a - ba - CB *Participial* pos. 1st and 2nd person: SC - bilê + CB Classes: CP - bilê + CB neg. 1st and 2nd person: SC - sa - ba + CB Classes: CP - sa - ba + CB

The potential Principal and participial 1st and 2nd person: pos. SC - ka - ba + C neg. SC - ka - se - bê + CB Classes: pos. CP - ka - ba + CB neg. CP - ka - sê - bê + CB

The subjunctive 1st and 2nd person: pos. SC - bê + CB neg. SC - se - bê + CB Classes: pos. CP - bê + CB neg. CP - se - bê + CB Note also the compound negative SC/CP - se - kê + SC2 - a - ba + CB

The consecutive 1st and 2nd person: pos. SC2 - a - ba + CB neg. SC2 - a - se - bê + CB Classes: pos. SC2 - a - ba + CB neg. SC2 - a - se - bê + CB Note also the compound negative SC2 - a - se - ke + SC2 - a - ba + CB

The habitual 1st and 2nd person: pos. SC - be + CB neg. SC - se - be + CB - be + CB Classes pos. CP - be + CB neg. CP - se - be + CB

The infinitive pos. go - ba + CB neg. go - se - bê + CB

The imperative pos. e - ba - ng + CB or ba - a - ng + CB neg. se - bê - ng + CB

The descriptive and locational copulative

The indicative series Principal Persons and classes: pos. SC + CB neg. ga - SC + CB *Participial* Persons and classes: pos. SC - le + CB neg. SC - se + CB *The future tense Principal and participial* Persons and classes: pos. SC - tlô/tla - ba + CB neg. SC - ka - se - bê + CB *The past tense Principal* Persons and classes: pos. SC - bilê + C neg. ga - se - SC - ba + CB ga - se - SC - be + CB ga - SC2 - a - ba + CB *Participial* Persons and classes: pos. SC - bilê + CB neg. SC - sa - ba + CB

The potential Principal and participial Persons and classes: pos. SC - ka - ba + CB neg. SC - ka - se - bê + CB

The subjunctive Persons and classes: pos. SC - bê + CB neg. SC - se - bê + CB Note the compound negative SC - se - kê + SC2 - a - ba + CB

The consecutive Persons and classes: pos. SC2 - a - ba + CB neg. SC2 - a - se - bê + CB Note the compound negative SC2 - a - se - kê + SC2 - a - ba + CB

The habitual Persons and classes: pos. SC - be + CB neg. SC - se - be + CB

The infinitive Persons and classes: pos. go - ba + CB neg. go - se - bê + CB

The imperative Persons and classes: pos. e - ba - ng + CB or ba - a - ng + CB neg. se - bê - ng + CB

The associative copulative

The indicative series Principal Persons and classes: pos. SC - na - le + CB neg. ga - SC + na - (le) + CB *Participial* Persons and classes: pos. SC - na - le + CB neg. SC - se - na - (le) + CB The stem remains -na in both the positive and the negative. *The future tense Principal and participial* Persons and classes: pos. SC - tlô/tla - ba - le + CB neg. SC - ka - se - bê - le + CB *The past tense Principal* Persons and classes: pos. SC - bilê - le + CB neg. ga - se - SC2 - a - ba - le + CB ga - se - SC - be - le + CB ga - SC2 - a - ba - le + CB *Participial* Persons and classes: pos. SC - bilê - le + CB neg. SC - sa - ba - le + CB

The potential Principal and participial Persons and classes: pos. SC - ka - ba - le + CB neg. SC - ka - se - bê - le + CB

The subjunctive Persons and classes: pos. SC - bê - le + CB neg. SC - se - bê - le + CB Note the compound negative SC - se - kê + SC2 - a - ba - le + CB

The consecutive Persons and classes: pos. SC2 - a - ba - le + CB neg. SC2 - a - se - bê - le + CB Note the compound negative SC2 - a - se - kê + SC2 - a - ba - le + CB

The habitual Persons and classes: pos. SC - be - le + CB neg. SC - se - be - le + CB

The infinitive Persons and classes: pos. go - ba - le + CB neg. go - se - bê - le + CB

The Imperative Persons and classes: pos. e - ba - ng - le + CB or ba - a - ng - le + CB neg. se - bê - ng - le + CB

This effort to summarise or consolidate copulative patterns firstly proves that copulatives in Northern Sotho is a *complicated issue*. Secondly, it is clear that such a summary is, unfortunately, user-unfriendly and not useful in such a format for, e.g., inclusion in the back matter of a dictionary. The lexicographer in his/her role as mediator between a complex grammatical pattern on the one hand and the inexperienced target user on the other, must find ways to present these copulatives in a logical and user-friendly way in dictionaries. Before lexicographers can consider appropriate lemmatization strategies for copulatives in Northern Sotho, they themselves must have absolute clarity on the nature and full scope of the occurrences of copulatives. An example-driven, oversimplified approach will now be taken in order to compile a comprehensive overview of copulatives in Northern Sotho. This will form the basis of, or input to, suggested lemmatization strategies for these copulatives in subsequent paragraphs. Prospective compilers of dictionaries for African languages should take note of the fact that in many cases it will be necessary to do work and even research on a linguistic level before lemmatization strategies can be considered for a specific phenomenon. Compare in this regard the lemmatization of reflexives (Prinsloo 1992) and the lemmatization of adjectives (Gouws and Prinsloo 1997).

An example-driven reconstruction of copulatives in Northern Sotho

In order to master the copulative it is crucial for learners to understand that:

- two types of copulatives can be distinguished, namely *static* (in a state of rest) and *dynamic* (in motion or changing);
- copulatives express three different semantic *relations* between a subject and a complement, namely *identification/equality*, *descriptive* or *associative*;
- copulatives can consist of a particle e.g. *ké*, or a copulative subject concord, *ke, re, o, le, ba*, etc. or a copulative verb, *e le, e se, o ba, o na*, etc.⁶

Consider the following examples of a copulative particle, a copulative subject concord and copulative verbs of *static* and *dynamic* copulatives in *identifying*, *descriptive* and *associative* relations. These relations are expressed between *mon-na* 'man' and *morutiši* 'teacher' or *bohlale* 'wisdom' or *mpša* 'dog' in Table 1, A–C respectively.

Table 1

A Identifying copulative: The relation is one of <i>identification/equality</i> , i.e. subject = complement			
Static:	monna	ke	morutiši
	the man	is	a teacher
(ke = copulative particle)			
Dynamic:	monna	e ba	morutiši
	the man	becomes	a teacher
(e ba = copulative verb)			
B Descriptive copulative: The relation is one of <i>description</i> , i.e. complement describes subject			
Static:	monna	o	bohlale
	the man	is	clever
(o = copulative subject concord)			
Dynamic:	monna	o ba	bohlale
	the man	becomes	clever
(o ba = copulative verb)			
C Associative copulative: The relation is one of <i>association</i> , i.e. subject is associated with complement			
Static:	monna	o na	le mpša
	the man	is	with a dog (has a dog)
(o na = copulative verb)			
Dynamic:	monna	o tlo ba	le mpša
	the man	will be	with a dog (will have a dog)
(o tlo ba = copulative verb containing the future prefix)			

Dynamic copulatives occur in all moods.⁷ It is advisable for the lexicographer (and the compiler of a basic Northern Sotho grammar) to use the user's presumed basic knowledge of the noun class system and the moods, tenses and aspects of common verbs as point of departure. Learners normally master the nominal and verbal systems first when studying an African language.

Consider firstly copulatives such as *o* in Table 1, B, which Poulos and Louwrens (1994: 307) describe as 'a form resembling the subject concord'. The lexicographer should immediately link such copulatives to the target user's presumed knowledge of subject concords as summarized in an example-driven layout in Table 2.

Table 2: Subject Concords

Person or noun class	Example	Sc.
1st Person singular	nna 'I'	ke
1st Person plural	rena 'we'	re
2nd Person singular	wena 'you' (singular)	o
2nd Person plural	lena 'you' (plural)	le
Class 1	monna 'man'	o/a
Class 2	banna 'men'	ba

Class 3	monwana 'finger'	o
Class 4	menwana 'fingers'	e
Class 5	lesogana 'young man'	le
Class 6	masogana 'young men'	a
Class 7	selepe 'axe'	se
Class 8	dilepe 'axes'	di
Class 9	nku 'sheep (singular)'	e
Class 10	dinku 'sheep (plural)'	di
Class 14	bogobe 'porridge'	bo
Class 15	go sepela 'to walk'	go
Class 16	fase 'below'	go
Class 17	godimo 'above'	go
Class 18	morago 'behind'	go
invariable ⁹	ge e le ... 'if he/she/it ... is ...'	e
indefinite ¹⁰	go tseba mang? 'who knows?'	go

Secondly, the lexicographer should link the dynamic copulative verb *ba* (*be*, *bile*) to the user's knowledge of the moods, tenses, etc. of ordinary verbs. This is exactly what is suggested for columns 5, 6 and 7 in comparison with columns 1-4 in Table 3. Note the degree of similarity between common verbs and the copulative verb *ba*. In an attempt to save space, the complete initial part, *mosadi* or *ge mosadi*, or *mosadi yo*, etc. is not repeated for the copulative in columns 5, 6 and 7.

In Table 3 no fewer than 34 copulative forms for 3 different copulative relations were given, covering only class 1. Multiplied by the roughly 20 different sets of concords for persons and classes in Table 1, this means about $34 \times 3 \times 20 = 2,040$ possible candidates for lemmatization of the dynamic copulative. It stands to reason that a *drastic reduction strategy* is required for dynamic copulatives in order to stay within the physical limitations of a paper dictionary. The same is applicable to static copulatives for which Table 4 is a condensed, example-based layout.

Thus, for instance, the final example in the last column of Table 4, the copulative *se ... nago* can be described as a static, associative, copulative verb, in the relative mood in the present tense, negative.

Static copulatives occur in the indicative, situative and relative moods only. Once again, it has advantages for the user to link it to his/her presumed knowledge of common verbs,¹¹ and now also building on his/her knowledge of the copulative subject concords and dynamic copulative verbs.

Table 4 reflects, in a simplistic way, the use of first, second and third persons (the latter is divided into noun classes in Northern Sotho), for the positive and negative forms of the three different relations for the indicative, situative and relative moods, including the progressive (+prog.).

Table 3: Dynamic Copulatives

Column 1: MD. = MOOD, IND. = INDICATIVE, SIT. = SITUATIVE, REL. = RELATIVE, SUB. = SUBJUNCTIVE, CON. = CONSECUTIVE, INF. = INFINITIVE, IMP. = IMPERATIVE, HAB. = HABITUAL	Column 2: PRES. = PRESENT, FUT. = FUTURE, PAS. = PAST +Pot. = containing the Potential	Column 3: ACT. = ACTUALITY (p. = positive, n. = negative)	Identifying	Descriptive	Associative
MD. IND.		Common verb			
+PRES.	p.	mosadi o reka dipuku	e ba morutiši	o ba bohlahe	o ba le mpša
+Pot.	n.	mosadi ga a reke dipuku	ga e be morutiši	ga a be bohlahe	ga a be le mpša
FUT.	p.	mosadi a ka reka dipuku	e ka ba morutiši	a ka ba bohlahe	a ka ba le mpša
	n.	mosadi a ka se reke dipuku	e ka se be morutiši	a ka se be bohlahe	a ka se be le mpša
PAS.	p.	mosadi o tlo/ta reka dipuku	e tlo/ta ba morutiši	o tlo/ta ba bohlahe	o tlo/ta ba le mpša
	n.	mosadi a ka se reke dipuku	e ka se be morutiši	a ka se be bohlahe	a ka se be le mpša
SIT.	p.	mosadi o rekile dipuku	e bile morutiši	o bile bohlahe	o bile le mpša
	n.	mosadi ga se a reka dipuku	ga se a ba morutiši	ga se a ba bohlahe	ga se a ba le mpša
	p.	ge mosadi a reka dipuku	e eba morutiši	e eba bohlahe	e eba le mpša
+Pot.	n.	ge mosadi a sa reke dipuku	e sa be morutiši	a sa be bohlahe	a sa be le mpša
FUT.	p.	ge mosadi a ka reka dipuku	e ka ba morutiši	a ka ba bohlahe	a ka ba le mpša
	n.	ge mosadi a ka se reke dipuku	e ka se be morutiši	a ka se be bohlahe	a ka se be le mpša
PAS.	p.	ge mosadi a tlo/ta reka dipuku	e tlo/ta ba morutiši	a tlo/ta ba bohlahe	a tlo/ta ba le mpša
	n.	ge mosadi a ka se reke dipuku	e ka se be morutiši	a ka se be bohlahe	a ka se be le mpša
	p.	ge mosadi a rekile dipuku	e bile morutiši	a bile bohlahe	a bile le mpša
REL.	p.	ge mosadi a sa reka dipuku	e sa ba morutiši	a sa ba bohlahe	a sa ba le mpša
	n.	mosadi yo a reka dipuku	e bago morutiši	a bago bohlahe	a bago le mpša
+Pot.	p.	mosadi yo a sa rekego dipuku	e sa bego morutiši	a sa bego bohlahe	a sa bego le mpša
	n.	mosadi yo a ka rekego dipuku	e ka bago morutiši	a ka bago bohlahe	a ka bago le mpša
FUT.	p.	mosadi yo a ka se rekego dipuku	e ka se bego morutiši	a ka se bego bohlahe	a ka se bego le mpša
	n.	mosadi yo a tlo/ta rekego dipuku	e tlo/ta bago morutiši	a tlo/ta bago bohlahe	a tlo/ta bago le mpša
PAS.	p.	mosadi yo a ka se rekego dipuku	e ka se bego morutiši	a ka se bego bohlahe	a ka se bego le mpša
	n.	mosadi yo a rekilego dipuku	e bilego morutiši	a bilego bohlahe	a bilego le mpša
SUB.	p.	mosadi yo a sa rekego dipuku	e sa bago morutiši	a sa bago bohlahe	a sa bago le mpša
	n.	(gore) mosadi a reke dipuku	e be morutiši	a be bohlahe	a be le mpša
CON.	p.	(gore) mosadi a se reke dipuku	e se be morutiši	a se be bohlahe	a se be le mpša
	n.	mosadi a reke dipuku	ya se be morutiši	a se be bohlahe	a se be le mpša
INF.	p.	go reka dipuku	go ba morutiši	go ba bohlahe	go ba le mpša
	n.	go se reke dipuku	go se be morutiši	go se be bohlahe	go se be le mpša
IMP.	p.	reka dipuku!	eba morutiši!	eba bohlahe!	eba le mpša!
	n.	se reke dipuku!	se be morutiši!	se be bohlahe!	se be le mpša!
HAB.	p.	mosadi a reke dipuku	e be morutiši	a be bohlahe	a be le mpša
	n.	mosadi a se reke dipuku	e se be morutiši	a se be bohlahe	a se be le mpša

Reduction of copulative subject concords, static and dynamic copulatives

The next step in isolating lemmas for copulatives should be directed at reducing the number of possible candidates for lemmatization in a logical, user-friendly way. The reduction of *subject concords* can be done purely logically by eliminating duplicate forms i.e. lemmatising the subject concord for classes 8 and 10 as a single lemma *di*, then by treating it for all of its copulative functions as well as its normal use as subject concord of classes 8 and 10 in the micro-structure. Table 5 suggests such a reduction from 22 (grammatical) candidates to 11 logical lemmas.

Table 5

1st Person singular	ke	Class 6, 1	a
1st Person plural	re	Class 7	se
2nd Person singular, Cl. 1, 3	o	Class 8, 10	di
2nd Person plural	le	Class 14	bo
Class 2	ba	Class 15-18, Indefinite Sc.	go
Class 4, 9, Invariable Sc.	e		

Thus the subject concords *ke, re, o, le, ba, e, a, se, di, bo* and *go* are to be entered as macrostructural elements, firstly for their ability to express a copulative meaning (cf. Table 4 above) and for their occurrence in dynamic and static copulative verbs with *ba, le*, etc.

Based on the assumption that users have mastered the basic verbal system, the entire Table 3 for *dynamic copulatives* can be collapsed into the following structures where *Sc.* represents the entire series of subject concords in Table 5.

- (1) Structure of Dynamic Copulatives
 - (a) (**Sc.**) + (tlo/tla/ka) + (e)ba(go) + (le)
 - (b) ga se + **Sc.** + ba + (le)
 - (c) **Sc.** + be
 - (d) ga + **Sc.** + be
 - (e) **Sc.** + ka se + be(go)
 - (f) **Sc.** + bile(go)

Finally the whole of the complex Table 4 for *static copulatives* can be reduced to the rules formulated in (2).

- (2) Structure of Static Copulatives
 - (a) ké
 - (b) ga se
 - (c) **Sc.**
 - (d) ga + **Sc.**

- (e) Sc. + (sa) + le(go)
- (f) Sc. + se(go)
- (g) Sc. + (sa) + na(go)
- (h) ga + Sc. + (sa) + na
- (i) Sc. + se + (sa) + na(go)

Selecting the lemmas

The list of subject concords (Table 5) as well as the rules given in (1) and (2) can be very useful as input strings to morphological analysers and machine translation programmes with the aim of generating — or breaking down — copulative constructions in Northern Sotho. However, for the lexicographer a further step is required in the lemmatization process, namely to isolate the final macrostructural elements, i.e. copulative lemmas, from these rules.

For the dynamic copulative *ba(go)*, *be(go)* and *bile(go)* should be lemmatized. Here it is assumed that *subject concords*, the future tense prefix *tlo/tla*, the potential *ka* and the conjunctive *le* have all been lemmatized in their own right with the necessary cross-references to the copulative tables in the back matter.

For the static copulative, *ké*, *ga se*, the reduced list of *subject concords*, this time for their *copulative use*, as well as *le(go)*, *se(go)*, and *na(go)* should be lemmatized. The progressive *sa* is assumed to be lemmatized with a cross-reference to the copulative tables in the back matter.

Consider the number of occurrences of each of these copulatives in a 6-million-word corpus of Northern Sotho:

Table 6

ké	±8,000	ga se	±4,000		
ba(go)	±1,200	be(go)	±6,000	bile(go)	30
na(go)	±12,000	le(go)	±30,000	se(go)	±2,000

These enormous overall counts clearly indicate not only that they should be included as lemmas but also that *exhaustive treatment* is required/justified especially for the encoding needs of inexperienced target users. The success or failure of a selected number of Northern Sotho dictionaries, *Pukunišu* (Kriel 1983), *Pukunišu* (Kriel and Van Wyk 1989), *New English–Northern Sotho Dictionary* (Kriel 1967) and *Groot Noord-Sotho woordeboek* (Ziervogel and Mokgokong 1975), to satisfactorily treat these copulatives will now be evaluated. These dictionaries are the most comprehensive commercially published dictionaries for Northern Sotho.

Critical evaluation of the treatment of copulatives in Northern Sotho dictionaries

Let us start with what these dictionaries have to offer the learner in terms of

English or Afrikaans dictionaries with Northern Sotho as the target language in selecting the word *is*. For users departing from English as source language, no guidance is available since *is* is not lemmatized in *New English–Northern Sotho Dictionary* (Kriel 1967) and cannot be looked up in *Groot Noord-Sotho woordeboek* (Ziervogel and Mokgokong 1975) since it is a mono-directional Northern Sotho → English/Afrikaans dictionary. For Afrikaans, it is lemmatized in both versions of *Pukuntšu*.

- (3) *Pukuntšu* (Kriel 1983)
is, ke; go; daar –, go na le; **dit –**, ke; **dit – so**, go bjalo; **daar –**, go na le; **dit – nie so nie**, ga go bjalo; **daar – nie**, ga go, ga dio; ga di gona.
- (4) *Pukuntšu* (Kriel and Van Wyk 1989)
is, (dit) –, ke; **(dit) – nie**, ga se; **daar –**, go na le; **dit – so**, go bjalo.

At least *is* is included as a lemma in these dictionaries. Treatment does not, however, go beyond giving basic decoding information. It is furthermore limited to the third person only and, in terms of the basic relations in Table 1 above, covers only the first part of A. Note also that in (3), the information '*daar –*, go na le' is given twice.

A suggested improvement could be (5):

- (5) **is** ke [id. cop.], **ke lengwalo, ga se sephuthana** it is a letter, it is not a parcel; **bohodu ke sebe**. theft is a sin; o/e/le ... [des. cop.], **aowa, mosadi yo o bohlae, ga a bogale!** No, this woman is clever, she's not cruel!; o/e/le ... na le [ass. cop.], **Satsope o na le Sara** Satsope is with Sara. ► BM 1.1-1.3

In this relatively short entry, basic distinctive grammatical information in A–C of Table 1 is given. Suitable examples and translations are included which are designed in such a way that the positive and negative forms are contrasted. Contrasting *ke* with *ga se* in the first example is valuable information for learners. Appropriate cross-reference to the back matter is also given where more information can be found. Note also the employment of a new convention *o/e/le ...* indicating to the user that a whole range of concords can be used. Similar entries should be compiled for English lemmas *am, are, be, become* or Afrikaans *word, wees*, etc.

When we turn to the treatment of copulatives in Northern Sotho with English or Afrikaans as the target languages, *ke* will be taken as a first example:

- (6) *Pukuntšu* (Kriel 1983)
ke, ke, dit is; **se ke**, moenie; – **go re**, dis om te sê, dit beteken.
ke, – mo, dit is dan wanneer. Ke nna yo ke lego Yena, Ek is wat Ek is.
- (7) *Pukuntšu* (Kriel and Van Wyk 1989)
ke¹, identifiserende kopula, ind. imp. H: (dit) is, deur
ke², osk. 1 p.e. L: ek.

- (8) *New English–Northern Sotho Dictionary* (Kriel 1967)
ke, 1st person singular, I
ké, pron. copulative, it is
- (9) *Groot Noord-Sotho woordeboek* (Ziervogel and Mokgokong 1975)
KE (ke-) ... [1.p.s. cop. formative] I am; *ke mogwêra wa gagwê* ... I am your friend
KE (ké-) ... [cop. formative] it is, he is, she is, they are; *ké lehôdu* ... he/she/it is a thief; *ké mahôdu* ... they are thieves; *ké moka* ... that is all

These entries are technically correct but only offer limited decoding information and very little encoding information. The amount of dictionary space could have been utilized much better. For example, the value of the entry **Ke nna yo ke lego Yena** is highly questionable. Firstly, it does not occur even once in the corpus. Then, the user should be informed that the first *ke* is the copulative particle 'it is' and that the second *ke* means 'I' occurring as part of a copulative verb *ke lego* 'I who am'. The inexperienced user is also misguided by the occurrence of *yena* 'he/him' which is translated in this context as 'I'. All information given in Kriel's first entry deals with the copulative except *se ke* 'must' (in which *ke* is used as an auxiliary verb stem). It could at least have been separated from the copulatives. It is also unclear why three separate entries were made. Kriel (1983), Kriel and Van Wyk (1989) and Kriel (1967) do not treat the copulative subject concord of the first person singular. Ziervogel and Mokgokong (1975) do much better in this regard in at least including both the copulative particle *ke* and the copulative subject concord of the first person singular as lemmas with typical examples of usage. They also indicate by means of an example that, in contrast to English, a single strategy exists in Northern Sotho expressing the notions 'he is', 'she is', and 'it is'. This is a sign of good lexicography since learners of Northern Sotho tend to have difficulty with this concept. Note, however, that they incorrectly translated *gagwe* 'his/her friend' as 'your friend'. The most serious shortcoming in the treatment of the copulative particle *ke* is that *none* of these dictionaries lemmatize the negative copulative particle *ga se*, of which the total occurrence in the corpus is estimated at more than 4,000. Consider (10) for *ga se* and (11) as a suggested improved entry for *ke*.

- (10) **ga se** [cop. part. neg.] it is not, **ga se phošo ya gago** it is not your fault; he/she/it is not, **Satsope ga se morutiši**, **ke mongwaledi** Satsope is not a teacher, she is a secretary; they are not, **dingaka ga se mahodu** doctors are not thieves ► **ke** ► BM 1.1; ■■■■■
se, ga ~ ► ga se
- (11) **ke¹ (ké)** [id. cop. part.], it is, **ke lengwalo**, **ga se sephuthana** it is a letter, it is not a parcel, **bohodu ke sebe** theft is a sin; he/she is, **Madika ke morutiši** Madika is a teacher; they are **dikgoši ke baetapele ba ditšhaba** kings are leaders of the nations ► BM 1.1; ► **ke² ■■■■■**

Consider as a second example the treatment of *le* in these dictionaries.

- (12) *Pukuntšu* (Kriel 1983)
le, en, plus, ook, mede –; *ge go* –, as daar is; *mogolegwa* –, medegevangene; ...
- (13) *Pukuntšu* (Kriel and Van Wyk 1989)
le², osk. kl 5, H: dit.
le⁴, osk. 2 p.m. L: julle, u.
le⁷, sit./rel. kop. ww. H: synde.
- (14) *New English–Northern Sotho Dictionary* (Kriel 1967)
le, copulative, le-class.
- (15) *Groot Noord-Sotho woordeboek* (Ziervogel and Mokgokong 1975)
LE (le-) ... [cc. 2 p.p.] ... you; ... [cc. cl. le-] ... he, she, it
LE ... [cop. d.v. of part.] ... being; *e sa le ka bošego* ... while it is still night/early; *e bê e le la mang?* ... whose was it?; *e bê e le mang?* ... who was it?; *e bê e le ma/Matome* ... it was I/Matome; *o bê a le gôna* ... he was present

The treatment in *Pukuntšu* (1989) and *New English–Northern Sotho Dictionary* is once again technically correct, but not exhaustive enough for a copulative occurring more than 30,000 times in the corpus. Ziervogel and Mokgokong do much better in at least including typical examples of usage but could have varied the choice of subject concord more instead of using *e* five times. The copulative verb *e le* with an estimated overall count of 20,000 or an average of 2 occurrences per page in a Northern Sotho literary work should seriously be considered for inclusion as a multiword lemma with treatment analogous to (11) above.

The same line of argumentation could be followed for the treatment of *se*, *na*, etc. In the case of *se* only one dictionary treats it.

- (16) *Pukuntšu* (Kriel and Van Wyk 1989)
se⁵, kop. ww. H; *gê a se gôna*, as hy nie aanwesig is nie.
- (17) *Pukuntšu* (Kriel 1983)
na, met, saam met, het, ...; *se/ga* - nie hê nie, het nie; ... *ke na le tšhêlêtê*, ek het geld (lett., ek is saam met geld); ...
- (18) *Pukuntšu* (Kriel and Van Wyk 1989)
na³, kop. ww. H: wees (met); *ke na le tšhêlêtê* ek is met/het geld,
na⁴, part. H: met; *nabô*, met hulle; *nago*, met jou.
- (19) *New English–Northern Sotho Dictionary* (Kriel 1967)
na, aux., verb, to be (with), to have; *o na le*, (lit., he is with), he has.
- (20) *Groot Noord-Sotho woordeboek* (Ziervogel and Mokgokong 1975)
NA (-na) (part. **ena**) ... [associative copulative verb] ... have, be; *a na naso* ... he having it; *o na le bana* ... she has children; *yô o nago bana* ... she who has children; *gê a na nabô* ... if she is with them; *ga re na selô* ... we have nothing; *go bê go se na motho* ... nobody was there; *gê a ena le tsebô* ... if he has knowledge; ...

From these examples it is clear that the treatment offered in the majority of cases is barely sufficient to fulfil the decoding needs of advanced users. The learner gets no guidance for encoding purposes from any Northern Sotho dictionary.

Lacking in all these dictionaries is treatment of copulatives in the back matter of the dictionary and appropriate use of the mediostucture. Ziervogel and Mokgokong, for example, include a detailed mini-grammar in the front matter of the dictionary, but hardly refer to copulatives, let alone giving an exhaustive discussion.

As far as the dynamic copulatives are concerned, fairly detailed entries are offered for the copulative verb stem *ba*. The need for more encoding information, however, remains. Note also that Kriel (1983) confuses the auxiliary verb stem *ba* with the copulative one but that this has been rectified in Kriel and Van Wyk (1989).

- (21) *Pukuntšu* (Kriel 1983)
ba, ... hulpww., sal wees; selfs, ook, eindelijk, totdat; *go – le*, het, besit; lett., om te wees met; – *gona*, daar wees, *bile*, was gewees; *eba*, sal wees, *go –*, om te wees.
- (22) *Pukuntšu* (Kriel and Van Wyk 1989)
ba⁶, kop. ww. L: wees, word; *go ba le*, om te wees/word met, om te hê/kry; *go ba gôna*, om aanwesig/teenwoordig/wel te wees; *go ba le dinala tšê ditêlêlê*, om lui te wees, om graag te slaan.
- (23) *New English–Northern Sotho Dictionary* (Kriel 1967)
ba v.n., to be, to become.
- (24) *Groot Noord-Sotho woordeboek* (Ziervogel and Mokgokong 1975)
BA (-ba, -bile) ... be, become; *go ba monna ...* to be/become a man; *lesogana, gê le nyala le ba monna ...* when a young man marries he becomes a man; *go ba setekeng ...* be pursued doggedly.

Be and *bile* occur frequently in the corpus, but are not substantially treated in any of the four dictionaries. The following examples show the complete treatment in all four dictionaries combined, clearly reflecting insufficient treatment.

- (25) *New English–Northern Sotho Dictionary* (Kriel 1967)
bê, subj., of **go ba**, to be; **e se bê mohlomong**, lest.
- (26) *Pukuntšu* (Kriel 1983)
bile, *bi.le*, (*ba*), was (gewees), reeds, selfs; geword het; *e – kaone*, dit was beter.
- (27) *Pukuntšu* (Kriel and Van Wyk 1989)
bilê¹, kop. ww. HH (perf. van **ba**): was gewees, het geword.
- (28) *New English–Northern Sotho Dictionary* (Kriel 1967)
bile, aux. v., even, pft., has been.

Constructing the back matter

It has been argued so far that the data types on microstructural level should be designed in such a way that they (a) fulfil the needs of the decoding user, and (b) go some way in assisting the encoding user in his/her information retrieval process. Formulated differently, the article should be constructed in such a way that it contains sufficient decoding information for mother-tongue speakers and advanced users. Thus, in most cases, it will not be necessary for them to follow up on the article-external cross references to the back matter or to an outside source like a grammar book. The learner, on the other hand, will in most cases benefit from following up on all references to the back matter and dictionary external sources.

The back matter should be constructed in such a way as to act as a bridge between the article and the outside source. This means firstly that the relevant section should not be as comprehensive as the discussion in a grammar book. Compare in this regard Ziervogel and Mokgokong (1975) who dedicated 39 pages in small print in the front matter to an all-in-one mini-grammar and user's guide where no exhaustive discussion of copulatives is given anyhow. Secondly, the discussion of copulatives in the back matter should fulfil the basic purpose of cross-reference, namely to be the reference address where the user would indeed find more information on copulatives, structured in such a way that it extends the information the user has obtained in consulting the article of the copulative in the central text. For the encoding user it should thus be 'the next logical step' in explaining the correct use of the copulative. Likewise, the back matter should also be the logical step/link to the outside source — thus a *comprehensive process from dictionary article to back matter to outside source*. In paper dictionaries this does not narrow the gap between dictionary and grammar but at least offers logical steps to the user in the information retrieval process, namely dictionary article to back matter to grammar book. In an electronic dictionary, this information retrieval process is simplified as will be briefly outlined below.

Excellent candidates for inclusion in the back matter are Tables 1–5 above. For the learner who has mastered basic skills regarding nouns and verbs in Northern Sotho, these tables will immediately put the copulative into perspective. Table 1 gives the user a full perspective on the relations, types and variant forms of copulatives. Table 2 brings together all the possible subject concords that can occur in copulative constructions. Table 3 gives the overall picture for the dynamic copulative and at the same time links it directly and visually to the use of ordinary verbs. Table 4, which should be expanded to include examples from all classes, will probably be the most useful table, reflecting the much-needed overall picture for the static copulatives. Translations should be added where applicable. The use of the invariable subject concord *e-* and the so-called indefinite subject concord *go* could also be explained in the back matter or, alternatively, catered for with a dictionary-external cross-reference to one or more grammar books. These tables could also form the basis for what is offered to the encoding user in electronic dictionaries.

Electronic dictionary entries

In contrast to the paper dictionary, an electronic dictionary can offer the user an exciting new range of data-access routes to the contents. (See Prinsloo 2001 for a detailed discussion.) The encoding needs of users who look up copulatives in electronic dictionaries for Northern Sotho can, for example, be satisfied by means of pop-up screens. Compare the following example for (10) above, now presented for an electronic dictionary as (29):

Indicative:	Identifying	
1ps	(Nna) <i>ke morutiši</i>	<i>ga ke morutiši</i>
+prog.	(Nna) <i>ke sa le morutiši</i>	<i>ga ke sa le morutiši</i>
1pp-2pp -----		-----
1	Monna <i>ke morutiši</i>	<i>ga se morutiši</i>
+prog.	Monna <i>e sa le morutiši</i>	<i>ga e sa le morutiši</i>
2-18 -----		-----

Click here for *Complete Table*

ga se phošo ya gago *it is not your fault; he/she/it is not*
Satsope ga se morutiši, ke mongwaledi *Satsope is not a teacher, she is a secretary; they are not*, **dingaka ga se mahodu** *doctors are not thieves*

(29) **ga se** ... [cop. part. Neg] *it is not, n structure; examples; pronunciation; combination; frequency; concords; idioms; expressions; picture; copulative relations; ■■■■□*

A Identifying copulative: The relation is one of <i>identification/equality</i> , i.e. subject = complement Click here for <i>Complete Table</i>
B Descriptive copulative: The relation is one of <i>description</i> , i.e. complement describes subject Click here for <i>Complete Table</i>
C Associative copulative: The relation is one of <i>association</i> , i.e. subject is associated with complement Click here for <i>Complete Table</i>

The only effort required of the user is to rest the cursor momentarily on the lemma. A click of the mouse takes him/her down to a next level of information. Such dictionary articles can be short but multifunctional in that they simultaneously serve the decoding needs of the more experienced user, and the encoding needs of the less experienced one. For decoding purposes the skeleton entry might suffice, therefore information boxes only appear if the user wants to see them. For the inexperienced user, the text boxes in this example offer encoding information which could at best be offered in a paper dictionary as a reference address in the back matter. One of the major advantages of such entries in electronic dictionaries is that the gap between dictionary and grammar, which is generally believed to be 'unbridgeable', is starting to close, as suggested by Geeraerts (2000: 77):

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.

Conclusion

Compiling user-friendly dictionaries of a high lexicographic standard for African languages poses a great challenge to prospective lexicographers. They are the mediators between complicated grammatical structures and the decoding and encoding needs of their target users. Complicated structures such as nouns, verbs, copulatives, etc. should not be tackled haphazardly as they come under the attention of the compiler. They should be carefully studied and even researched to obtain a comprehensive overview of the relevant structures. Only then can the lexicographer proceed to plan the macrostructure and microstructure for the lemmatization of a specific construction. On the macrostructural level, candidates for inclusion (or omission) should carefully be considered, preferably based on corpus data. On the microstructural level, data should be presented in such a way that it satisfies both the needs of encoding and decoding users. The mediostructure should be employed in a sensible way to refer the user to reference addresses where more information can be found. Special attention should be given to references to a well-compiled back matter where cohesion of decontextualised items is restored, thus rendering the 'full picture' to the user. In terms of Laufer (1992: 71), this means to really 'know' the copulative:

When a person "knows" a word, he/she knows the following: the word's pronunciation, its spelling, its morphological components, if any, the words that are morphologically related to it, the word's syntactic behaviour in a sentence, the full range of the word's meaning, the appropriate situations for using the word, its collocational restrictions, its distribution and the relation between the word and other words within a lexical set.

Endnotes

1. An exception is the clear, systematic way in which copulatives are discussed in Louwrens (1991).
2. See Taljard (1999) for a detailed overview of the different standpoints.
3. The term *multiword* in this context refers to more than one *orthographic* word.
4. See Gouws and Prinsloo (1998: 24) for a detailed discussion of this problem and the overuse of mediostructure in this regard to maintain structural links in the African languages.
5. The original layout has been changed to save space.
6. Since copulative verb stems are orthographically written as separate words, they will be indicated as such in this article, namely *le*, *ba*, *be*, *bile* and not as verb stems *-le*, *-ba*, *-be*, *-bile*. Note also that Kriel (1983), and Kriel and Van Wyk (1989) opted for lemmatizing these stems

without hyphens.

7. The Van Wyk approach (Van Wyk et al. 1992) will be favoured in this article. The debate in respect of moods and alternative classifications fall outside the scope of this article. For the latter, the reader is referred to Poulos and Louwrens (1994).
8. In this article the term 'copulative subject concords' will be used.
9. "It is said to be invariable since it's form is not influenced by the particular noun class to which the subject noun belongs." (Louwrens 1994: 91)
10. "A term used to refer to the concord *go-* which occurs in verbs when the subject of the sentence is *indefinite*." (Louwrens 1994: 84)
11. In some instances 'linguistic impurity' should be tolerated in favour of user-friendliness. Strictly speaking, the copulative *particle ke* cannot occur in a mood but in this case fruitfully fills an empty slot and should be tolerated as natural scientists tolerate the scientific incorrect statement: 'the sun rises in the east and sets in the west'.

References

- Fox, Gwyneth.** 1987. The Case for Examples. Sinclair, John M. (Ed.). *Looking Up, An Account of the COBUILD Project in Lexical Computing and the Development of the Collins COBUILD English Language Dictionary*: 137-49. London: Collins ELT.
- 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 on Lexicography, Stuttgart, Germany, August 8th-12th, 2000*: 75-84. Stuttgart: Stuttgart University.
- Gouws, R.H.** 1989. *Leksikografie*. Cape Town: Academica.
- Gouws, R.H.** 1996. Bilingual Dictionaries and Communicative Equivalence for a Multilingual Society. *Lexikos* 6: 14-31.
- Gouws, R.H.** 2000. Toward the Formulation of a Metalexigraphic 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*: 109-33. Tübingen: Max Niemeyer.
- Gouws, R.H. and D.J. Prinsloo.** 1997. Lemmatisation of Adjectives in Sepedi. *Lexikos* 7: 45-57.
- Gouws, R.H. and D.J. Prinsloo.** 1998. Cross-referencing as a Lexicographic Device. *Lexikos* 8: 17-36.
- Hartmann, R.R.K. and G. James.** 1998. *Dictionary of Lexicography*. London: Routledge.
- Kriel, T.J.** 1967. *The New English-Northern Sotho Dictionary, English-Northern Sotho, Northern Sotho-English*. Johannesburg: Educum.
- Kriel, T.J.** 1983. *Pukuntšū*. Pretoria: J.L. van Schaik.
- Kriel, T.J. and E.B. Van Wyk.** 1989^a. *Pukuntšū woordeboek, Noord-Sotho-Afrikaans, Afrikaans-Noord-Sotho*. Pretoria: J.L. van Schaik.
- Laufer, Batia.** 1992. Corpus-based versus Lexicographer Examples in Comprehension and Production of New Words. Tommola, Harnu et al. (Eds.). 1992. *Euralex '92 Proceedings*: 71-76. Tampere: University of Tampere.
- Louw, P.A.** 1999. Access Structures in a Standard Translation Dictionary. *Lexikos* 9: 108-118.
- Louwrens, L.J.** 1991. *Aspects of Northern Sotho Grammar*. Pretoria: Via Afrika
- Louwrens, L.J.** 1994. *Dictionary of Northern Sotho Grammatical Terms*. Pretoria: Via Afrika
- Poulos, G. and L.J. Louwrens.** 1994. *A Linguistic Analysis of Northern Sotho*. Pretoria: Via Afrika.

- Prinsloo, D.J.** 1992. Lemmatization of Reflexives in Northern Sotho. *Lexikos* 2: 178-191.
- Prinsloo, D.J.** 2001. The Compilation of Electronic Dictionaries for the African Languages. *Lexikos* 11: 139-159.
- Prinsloo, D.J. and G.-M. de Schryver.** 1999. The Lemmatization of Nouns in African Languages with Special Reference to Sepedi and Cilubà. *South African Journal of African Languages* 19(4): 258-275.
- Prinsloo, D.J. and R.H. Gouws.** 1996. Formulating a New Dictionary Convention for the Lemmatization of Verbs in Northern Sotho. *South African Journal of African Languages* 16(3): 100-107.
- Taljad, Elizabeth.** 1999. *Die Kopulatie van Noord-Sotho: 'n Nuwe Perspektief*. Unpublished D.Litt. Thesis. Pretoria: University of Pretoria.
- Van Wyk, E.B. et al.** 1992. *Northern Sotho for First-Years*. Pretoria: J.L. van Schaik.
- Wiegand, H.E.** 1996. Über die Mediostrukturen bei gedruckten Wörterbüchern. Zettersten, A. and V.H. Pedersen (Eds.). 1996. *Symposium on Lexicography VII*: 11-43. Tübingen: Max Niemeyer.
- Ziervogel, D. and Mokgokong, P.C.** 1975. *Groot Noord-Sotho-woordeboek*. Pretoria: J.L. van Schaik.

Les tons dans les dictionnaires de langues gabonaises: situation et perspectives

Thierry Afane Otsaga, *Département d'afrikaans et de néerlandais, Université
de Stellenbosch, Stellenbosch, République d'Afrique du Sud*
(*thiafane@yahoo.fr*)

Résumé: La plupart des dictionnaires des langues gabonaises, parus jusqu'à ce jour, n'ont pas vraiment tenu compte de la tonalité dans la rédaction des articles. Or, à présent, il est unanimement reconnu que le ton joue un rôle important dans le fonctionnement des langues bantoues en général et des langues gabonaises en particulier. Les dictionnaires, outils d'apprentissage, de standardisation, de préservation et de l'emploi correct d'une langue, ne pourront plus négliger cet aspect dans leur traitement lexicographique. Malgré le grand nombre de tons qui existent dans les langues gabonaises et la difficulté de les noter tous dans la présentation des articles, il est néanmoins nécessaire de trouver des moyens de mettre en valeur chaque ton, sans néanmoins rendre trop compliqué l'emploi du dictionnaire. L'objectif de notre exposé est de formuler, pour ces dictionnaires, certaines directives qui permettront de prendre en considération la tonalité des langues gabonaises, tout en gardant accessible l'emploi du dictionnaire. Avant de formuler ces directives, notre exposé tente tout d'abord de clarifier le problème par l'analyse et l'énumération des différents tons identifiés jusqu'à ce jour dans les langues gabonaises. En outre, nous prenons en considération certains points de vue qui, dans les dictionnaires, ont voulu identifier les tons, en donnant des exemples du traitement de la tonalité dans certains dictionnaires des langues tonales autres que les langues gabonaises.

Mots-clés: DICTIONNAIRES, LANGUES GABONAISES, TONS, TONALITÉ, RÉDACTION DES ARTICLES, PERSPECTIVES, LANGUES BANTOUES, TRAITEMENT LEXICOGRAPHIQUE, OUVRAGES DE RÉFÉRENCE LEXICOGRAPHIQUE, COURBE MÉLODIQUE, SIGNES DIACRITIQUES, ORTHOGRAPHE, PRONONCIATION

Abstract: Tones in Dictionaries of the Gabonese Languages: State and Perspectives. Until now most dictionaries that have appeared in the Gabonese languages have not really taken tonality into account in the editing of articles. But at present it is unanimously recognized that tone plays an important role in the functioning of Bantu languages in general and Gabonese languages in particular. Dictionaries which are tools in the learning, standardization, preservation and correct usage of a language can no longer neglect this aspect in their lexicographic treatment. In spite of the large number of tones present in the Gabonese languages and the difficulty of attending to all of them in the presentation of the articles, it is nevertheless necessary to find means to account for each tone, without however rendering the use of the dictionary too complicated. The objective of this article is to formulate certain directions for these dictionaries, which

will make allowance for the tonality of the Gabonese languages, keeping the dictionary completely accessible to users. After having formulated these directions, our article firstly attempts to clarify the problem by analysing and enumerating the different tones identified in the Gabonese languages up till now. Moreover, we consider certain angles from which, in dictionaries, tones can be identified, by giving examples of the treatment of tonality in certain dictionaries of tonal languages other than the Gabonese languages.

Keywords: DICTIONARIES, GABONESE LANGUAGES, TONES, TONALITY, EDITING OF ARTICLES, PERSPECTIVES, BANTU LANGUAGES, LEXICOGRAPHIC TREATMENT, LEXICOGRAPHIC REFERENCE WORKS, MELODIC CURVE, DIACRITIC SIGNS, ORTHOGRAPHY, PRONUNCIATION

Introduction

Les langues gabonaises sont dans leur ensemble des langues bantoues et comme la plupart des langues bantoues elles comportent des éléments supra-segmentaux telles que l'accent, l'intonation, la longueur, les tons, etc. qui jouent un rôle important dans le fonctionnement de ces langues. De tous ces éléments supra-segmentaux, les tons sont, sans conteste, ceux dont l'implication est le plus déterminant dans le comportement des dites langues (Afane Otsaga 2001: 155). Les tons sont des oppositions de hauteurs qui dans certaines langues sont utilisées comme unités distinctives (Baylon et Fabre 1990: 101). Tous les travaux de description réalisés sur les langues gabonaises reconnaissent unanimement le rôle majeur que jouent les tons dans celles-ci. De ces différents travaux, il ressort que les langues gabonaises sont non seulement des langues à tons puisqu'elles utilisent la hauteur mélodique à des fins distinctives, mais sont aussi des langues à type tonal parce que la structure tonale de certains termes varient selon qu'ils sont en isolation ou en contexte. Pour ce qui concerne les travaux lexicographiques sur les langues gabonaises, particulièrement pour la réalisation des dictionnaires, une question importante demeure sans réponse: doit-on oui ou non noter les tons dans les dictionnaires de langues gabonaises? L'objectif de cet article est non seulement de faire le point en ce qui concerne le traitement de la tonalité dans les dictionnaires de langues gabonaises jusqu'à présent, mais surtout de proposer quelques méthodes qui pourraient aider à résoudre ce problème dans le traitement lexicographique des langues gabonaises.

Un mot sur les tons dans les langues gabonaises

Les tons sont généralement répartis en deux catégories. D'une part les tons ponctuels où seul un aspect de la courbe mélodique est tenu en compte (le plus haut ou le plus bas) ainsi que les tons modulés qui se distinguent par des directions successives de la courbe mélodique (Baylon et Fabre 1990: 101). Dans

les langues à tons, la variation mélodique est généralement associée à une syllabe (Matthews 1997: 379).

Les différentes descriptions¹ réalisées jusqu'à ce jour sur les langues gabonaises ont identifié dans celles-ci l'existence de trois catégories de tons.

Les tons principaux

Par tons principaux, j'entends les tons qui sont les plus récurrents et les mieux répartis dans les langues gabonaises. Il s'agit

— du *ton ponctuel haut* noté avec un accent aigu (´). C'est le point le plus haut de la courbe mélodique.

Langues	Exemples ²	Traduction française
Fan	mvám	générosité
Meka	nké	plantation
Yipunu	mvúlà	pluie

— du *ton ponctuel bas* noté avec un accent grave (`). C'est le point le plus bas de la courbe mélodique.

Langues	Exemples	Traduction française
Gisira	bilimè	les années
Ikota	èlùngá	panier
Latege	jùrù	tirer

— du *ton modulé montant* noté avec un accent circonflexe inversé (ˇ). Il s'agit d'une courbe mélodique qui prend deux directions successives. Elle descend d'abord vers le point le plus bas, avant de remonter vers le point le plus haut.

Langues	Exemples	Traduction française
Seki	dikõn	lance
Sake	yă	marécage
Yilumbu	dĩmbú	village

— du *ton modulé descendant* noté avec un accent circonflexe (^). Il s'agit d'une courbe mélodique qui prend deux directions successives. Elle monte d'abord vers le point le plus haut, avant de redescendre vers le point le plus bas.

Langues	Exemples	Traduction française
Lekanini	yâ	autrefois
Liwanzi	tôlô	sommeil
Omyene	íbôbò	peur

Les tons intermédiaires

Par tons intermédiaires, j'entends les tons qui sont moins récurrents et moins répartis dans les langues gabonaises que les tons principaux, mais dont la présence dans ces langues n'est pas négligeable puisque facilement remarquable. Il s'agit

— du *ton ponctuel moyen* noté avec une barre horizontale (¨). C'est le point situé au milieu de la courbe mélodique, c'est-à-dire entre le point le plus bas et le point le plus haut de la courbe mélodique.

Langues	Exemples	Traduction française
Fa <u>ṅ</u>	ā <u>sōṅ</u>	dent
Liwanzi	ḵi <u>wānī</u>	frapper
Omye <u>ṅ</u>	ts <u>ólō</u>	oiseau

— du *ton haut-abaisse* noté avec une barre verticale ('). Notons que ce ton ne se place pas au-dessous d'un segment syllabique comme c'est généralement le cas des tons et est aussi appelé *faille tonale*. La faille tonale a pour résultat de produire des tons que l'on appelle haut-abaisse. Ces tons sont réalisés phonétiquement à un niveau qui peut être assimilable au ton moyen, mais contrairement à un véritable ton moyen, ils ne peuvent être suivis par un ton plus haut (Hombert 1990: 102). Ce ton est suivi d'un terrassement tonal, c'est-à-dire que la série de tons hauts qui suit la faille tonale est réalisée au même niveau que le ton haut-abaisse.

Langues	Exemples	Traduction française
Fa <u>ṅ</u>	á <u>wi</u> 'ny	tuer
Lembaama	mè <u>ma</u> 'áta'	j'ai tiré
<u>Ṅ</u> gubi	í <u>ṅ</u> gu'yí	mère

— du *ton supra-haut* noté avec deux accents aigus ("). Ce ton a pour point de départ une valeur plus élevée qu'un ton haut et qui ensuite chute très rapidement (Hombert 1990: 103).

Langues	Exemples	Traduction française
Yí <u>p</u> unu	é <u>l</u> á <u>b</u> ì	il voit
Lembaama	v <u>ú</u> gh <u>i</u>	miel
Leka <u>n</u> í <u>n</u> i	s <u>á</u> s <u>i</u>	appât

Les tons secondaires

Par tons secondaires, j'entends les tons assez rares dans les langues gabonaises et dont la répartition entre ces dernières est véritablement infime. Il s'agit

— du *ton infra-bas* noté par deux accents graves ("). Sa définition est identique à celle du ton supra-haut, avec la différence que sa courbe est descendante. Il s'agit donc d'un ton bas réalisé plus bas que le ton bas normal. Il existe parfois un contraste entre le ton bas ponctuel et le ton infra-bas. Il est fréquent que cette distinction ne se produise que dans certains contextes, par exemple avant la pause (Hombert 1990: 102).

Langues	Exemples	Traduction française
Lenduma	wǎ́lú	ped
Meka	wǎ́má	combattre

— du *ton très haut descendant* noté par un accent aigu suivi d'un accent circonflexe (´ ^). Je n'ai pas de définition précise de ce ton, mais on pourrait penser qu'il s'agit d'un ton supra-haut qui chute par la suite. Le seul exemple que j'ai de ce ton vient du Yipunu.

Langue	Exemple	Traduction française
Yipunu	ngáânzì	racines

Le recensement des tons révélés ainsi dans les langues gabonaises n'est certainement pas exhaustif, car toutes les langues gabonaises n'ont pas encore été décrites. Mais il permet de constater qu'il en existe un grand nombre et que par conséquent leur influence dans le fonctionnement de ces langues est sans conteste à prendre en compte.

Le traitement de la tonalité dans les dictionnaires existants

Les premiers travaux lexicographiques dans les langues gabonaises ont été exécutés par des anciens administrateurs coloniaux et par les missionnaires. L'objectif premier de ces œuvres à caractère lexicographique (glossaires, lexiques, dictionnaires ...) était d'établir une communication efficace entre ces administrateurs et missionnaires, et les populations gabonaises (Mavoungou 2001). De même, les réalisateurs de ces travaux n'étaient ni des linguistes de formation, encore moins des lexicographes attitrés. Malgré ce fait, ces travaux, dans leur globalité, répondent effectivement aux normes métalexigraphiques requises (Mihindou 2001: 161), bien que certaines améliorations y soient nécessaires.

Bien que n'ayant pas eu à ma disposition tous les dictionnaires³ réalisés sur les langues gabonaises, je peux estimer que ceux que j'ai pu consulter constituent entre 70 et 80% environ de ce qui existe. De cette consultation, j'ai pu remarquer, en ce qui concerne le traitement de la tonalité, les situations suivantes:

— Deux tiers des dictionnaires consultés ne font pas état de la tonalité. Les articles des dictionnaires sont présentés sans structure tonale et cette

absence n'est, ni mentionnée, ni expliquée dans les introductions (prétextes) des dictionnaires. La majorité de ces dictionnaires sont plutôt des glossaires ou des lexiques, car il s'agit dans l'ensemble de listes alphabétiques de mots dans une langue-source suivie de leurs traductions équivalentes dans une langue-cible. Le *Dictionnaire Français-Yipounou / Yipounou-Français* (1966) de l'Église Évangélique du Gabon est une très bonne illustration de cette situation. Dans ce dernier aucun ton n'est noté, alors que le Yipunu est une langue à tons avérée.

— Le deuxième scénario est celui des dictionnaires qui utilisent des diacritiques dans la transcription. Mais ces diacritiques sont utilisés non pas comme marques de tonalités, mais plutôt comme éléments des symboles alphabétiques utilisés. Le *Dictionnaire Français-Mpongwé*, suivi d'éléments de grammaire (1995) de Raponda-Walker nous donne une excellente illustration de ce type de démarches. Dans ce dernier par exemple le « è » transcrit le son phonétique [ɛ], alors que le « ó » y figure pour transcrire le son phonétique [ɔ]. En d'autres termes, les tons n'y sont pas pris en considération. Mais contrairement au premier cas, celui-ci a la particularité de prendre déjà en compte certaines situations d'intonation ou d'accentuation en ce qui concerne la prononciation du « è » et du « ó ».

— Le dernier scénario est à l'image du second, car il ne tient compte que des phénomènes d'intonation et d'accentuation qui sont clairement mis en évidence au niveau de la transcription orthographique. Or, nous pouvons constater que dans ces dictionnaires, plusieurs traits accentuels (ou d'intonation) attribués aux langues gabonaises sont en réalité des manifestations tonales. Ce qui peut s'observer dans *l'Encyclopédie Pahouine* (1901) de V. Largeau où ce dernier présente par exemple trois réalisations différentes du « a » dans la *Fañ* (a, ā, â). Le « a » sans accent pour lui est équivalent à celui qui existe en français comme dans « ami ». Celui avec la barre horizontale « ā » est décrit comme un son ayant la même qualité que le premier, mais avec une quantité plus importante, alors que celui avec l'accent circonflexe « â » augmente aussi bien sur le plan qualitatif que quantitatif et se prononce comme le « â » français dans « pâte ». Cette même application est faite pour d'autres sons comme « e » ou « o ». Il semble que Largeau, qui était un administrateur colonial et non pas un linguiste ou lexicographe de formation, confond l'intonation, l'accentuation et la tonalité en *Fañ*. Les descriptions intonatives ou accentuelles qu'il donne de ces sons correspondent en réalité aux situations où ils portent le ton bas, moyen, haut, montant ou descendant comme mis en évidence dans différentes descriptions du *Fañ* par Andeme Allogo (1980), Voltz (1990), Cinnamon (1990), Ondo Mebiame (1992), etc.

La confusion de Largeau entre tons et d'autres phénomènes suprasegmentaux a pu se glisser dans d'autres dictionnaires qui, j'insiste sur ce point, ont été réalisés (pour la plupart) par des non-spécialistes en matière de description linguistique. Cela peut donc expliquer l'absence des tons dans le plus grand nombre de dictionnaires réalisés dans les langues gabonaises.

Signalons également que dans un petit nombre de dictionnaires et d'ouvrages⁴ (tel que le dictionnaire Fang–Français en Français–Fang de Galley 1964) réalisés dans les langues gabonaises, les tons sont notés non pas sur les segments, mais entre parenthèse à côté de ces derniers. Cette méthode donne l'impression que pour les réalisateurs de ces ouvrages, les tons ne font pas parties intégrantes de la langue et sont des phénomènes situationnels qui apparaissent çà et là dans le fonctionnement de celles-ci.

Quelques avis informels au sujet de la notation des tons dans les dictionnaires

Le point de vue le plus général rencontré chez beaucoup d'auteurs en ce qui concerne l'absence de tons dans les dictionnaires de langues gabonaises est que ceux-ci ne sont pas indispensables. La raison souvent avancée est que les premiers ouvrages à caractère lexicographique ne tenaient pas compte des tons de façon générale, mais ils ont été les principaux supports qui ont aidé beaucoup de missionnaires et d'administrateurs coloniaux à apprendre les langues gabonaises. Dans mes discussions sur ce problème avec certains locuteurs gabonais profanes de la linguistique, la plupart m'ont ressorti l'exemple des prêtres européens qui disaient les messes dans les langues gabonaises, sans pourtant avoir une prononciation tonale très correcte de ces langues. Pour eux une langue peut d'abord s'acquérir sur le plan segmental, la structure suprasegmentale de celle-ci s'acquérant par la pratique régulière au contact des locuteurs.

Je voudrais rappeler à ceux qui parlent de l'expérience des prêtres européens, qu'ils ne doivent pas non plus oublier les incidents survenus dans certaines églises du Gabon au cours de messes dites par ces derniers. L'incident le plus courant était que les fidèles sortaient de l'église parce qu'ils se sentaient choqués ou insultés par les propos d'un prêtre qui en prononçant mal certains termes (sans la bonne tonalité) aurait signifié des choses qu'il ne souhaitait pas dire en réalité. Cela montre que la tonalité est primordiale dans les langues gabonaises et que par conséquent ne pas en tenir compte dans les dictionnaires pourrait occasionner, pour les apprenants de ces langues, des incidents identiques à ceux dont je viens de faire état.

Pour autant, je ne suis pas totalement en désaccord avec ceux qui pensent que l'acquisition d'une langue se réalise d'abord au niveau de sa structure segmentale, car elle est la plus palpable, autant sur le plan acoustique que sur le plan visuel. Le dictionnaire doit d'abord mettre l'accent sur la transcription (orthographique) simple et accessible des formes segmentales de la langue. Par conséquent, je ne partage pas l'avis de ceux qui estiment qu'un livre d'appren-

tissage linguistique qui ne fait pas état des marques de tonalité est à rejeter complètement⁵. J'estime que la présentation de la structure segmentale est une étape ou une partie de l'entière structure (segmentale et suprasegmentale) d'une langue. Il n'y a donc pas de raison à ce qu'elle soit rejetée. Beaucoup de descriptions purement linguistiques sur les langues gabonaises se restreignent bien souvent aux aspects phonétiques, phonologiques et morphologiques. Rares sont celles qui font des descriptions sémantiques ou syntaxiques de ces langues. Mais tout le monde est d'accord pour reconnaître que ces descriptions, qui ne mettent en évidence que certains aspects de la langue, sont des travaux scientifiquement valables. Présenter une langue sans sa structure tonale devrait être vu dans le même esprit. C'est un aspect de la langue qui est présenté (la structure segmentale), il faut donc l'accepter comme tel.

Pour autant, dans le cas des langues à types tonals comme le sont certaines langues gabonaises, il ne serait pas inutile pour l'utilisateur du dictionnaire d'ajouter les marques tonales afin de faciliter la tâche aux apprenants de ces langues. Au lieu que l'apprenant d'une langue attende d'être au contact des usagers de celle-ci pour la pratiquer et ainsi acquérir sa structure suprasegmentale, il pourrait l'apprendre directement à travers le dictionnaire ou du moins l'avoir à sa disposition dans ce dernier. Mon avis est que les tons devraient être présentés dans les dictionnaires de langues gabonaises, mais cette présentation devrait se faire de façon méthodique et en tenant compte d'un certain nombre de paramètres.

Les tons et l'orthographe

Le problème le plus important des langues gabonaises en ce moment est leur transcription orthographique. Tous les travaux réalisés sur ces langues sont transcrits, soit phonétiquement, soit avec des symboles orthographiques différents les uns des autres. Aucune langue au monde ne s'est développée sans une forme orthographique unique utilisée par tous (Afane Otsaga 2001: 156). L'orthographe rappelons-le est l'art et la technique de la formation des mots par les lettres selon l'usage accepté (Gove 1961). Le choix d'utiliser une forme orthographique spécifique pour l'écriture d'une langue donnée est une convention qui naît d'un accord (généralement entre experts) et qui est accepté (ou imposé dans certains cas) pour être utilisé dans les écoles, les administrations (gouvernementales et non gouvernementales), la presse et tous les autres champs d'utilisation de la langue (Mdee 1999: 123). L'acceptation d'une forme scripturale d'une langue se fait généralement par la création d'un modèle d'imitation et la promotion de ce modèle.

Dans la situation des langues gabonaises, l'acceptation du modèle orthographique proposé au cours d'un colloque d'experts en avril 1999 à Libreville sous l'égide du Ministère de l'Éducation Nationale (Afane Otsaga 2001: 156) permettra de faire un pas important vers la lecture et l'écriture de ces langues par le plus grand nombre. En effet, tout débat sur la mise en valeur ou non des

tons dans la transcription des langues gabonaises resterait vain tant que celles-ci n'auraient pas une forme orthographique acceptée et utilisée par tous. Toute notation de tons serait inutile si elle ne servait pas à aider l'utilisateur dans l'amélioration de la prononciation. Mais si ce dernier ne se retrouve même pas dans la structure segmentale (par la forme orthographique) des termes qui lui sont proposées, il lui serait difficile d'avoir accès à la structure suprasegmentale de ceux-ci. La structure segmentale d'une langue est le support de sa structure suprasegmentale. Si le support est instable, ce qu'il supporte l'est aussi forcément. Le problème de tonalité dans l'écriture des langues gabonaises ne peut donc trouver de solution tant que celui de l'orthographe n'est pas résolu.

Je pourrais ajouter que la vulgarisation de l'orthographe de langues gabonaises devrait aussi inclure la vulgarisation des marques de tonalité. Un certain nombre de diacritiques (pas trop nombreux) pourrait être retenu pour transcrire les tons les plus réguliers et les plus répandus dans les langues gabonaises. La promotion de ces diacritiques devrait se faire en même temps que les symboles retenus pour l'orthographe. Cette promotion consistera essentiellement à habituer les usagers à l'utilisation de ces diacritiques, mais surtout à leur permettre d'associer clairement chaque diacritique à un ton spécifique, en précisant la manière dont ce ton doit être réalisé.

Quelques exemples de notation des tons dans les dictionnaires de langues bantoues

Les langues gabonaises font parties de l'ensemble bantou. Toutes les langues appartenant à ce groupe rencontrent le même problème en ce qui concerne la prise en compte de la tonalité dans les dictionnaires. Ce problème est abordé de différentes façons selon le choix des lexicographes.

La situation la plus courante est celle des dictionnaires qui présentent les marques de tons dans la macrostructure du dictionnaire, mais sans donner aucune explication ni par rapport à la présence de ces tons, ni par rapport à leur rôle dans le fonctionnement de la langue. C'est la situation générale que nous avons observée pour les dictionnaires des langues gabonaises. Deux hypothèses peuvent être émises par rapport à cette attitude:

— La première hypothèse est que certains lexicographes prétendent que puisque les usagers-cible des dictionnaires sont des locuteurs natifs de ces langues, par la maîtrise intuitive qu'ils ont de la langue, ils n'ont pas besoin d'explications particulières sur la structure suprasegmentale d'une langue qu'ils maîtrisent déjà. C'est l'occasion ici de rappeler à ceux qui pourraient se reconnaître dans cette hypothèse que le dictionnaire même s'il est réalisé avant tout selon les besoins du groupe-cible, peut dans certaines situations tomber entre les mains d'un individu n'appartenant pas au groupe-cible du dictionnaire, comme c'est souvent le cas aujourd'hui dans un monde de plus en plus polyvalent. Les informations

sur la structure suprasegmentale pourraient, dans ce cas, être utiles à un non-locuteur de la langue.

— La deuxième hypothèse est que certains lexicographes travaillant sur les langues bantoues estiment que dans les dictionnaires des langues indo-européennes, qui ont pourtant une longue tradition lexicographique, aucune explication particulière n'est faite au sujet des phénomènes suprasegmentaux comme l'accent ou la longueur par exemple. Par rapport à cet argument, je voudrais souligner que les langues indo-européennes ont une très longue tradition écrite. Les problèmes liés à l'orthographe et à la prononciation ont depuis longtemps trouvé des solutions dans ces langues, contrairement aux langues bantoues qui se cherchent encore sur tous les plans. Les phénomènes accentuels en français par exemple sont clairement identifiés par un certain nombre de diacritiques qui ôtent toute ambiguïté dans la prononciation de cette langue. La comparaison entre les dictionnaires des langues indo-européennes et les dictionnaires des langues bantoues est donc une méprise qu'il est souhaitable d'éviter.

Dans d'autres dictionnaires, les lexicographes ont choisi de présenter la structure tonale de façon complète et sans restriction telle qu'elle apparaît dans les termes de la langue. Tous les tons y sont marqués. C'est le cas du dictionnaire Tsivenda⁶-English (Van Warmelo 1937) dans lequel les tons hauts, bas, moyens et haut abaissés sont notés. Selon les rédacteurs de ce dictionnaire, ces quatre tons ont été retenus parce qu'ils sont les seuls qui apparaissent dans les termes en isolation, les autres tons étant de simples variations de ces derniers en situation de discours. Ils représentent donc la tonalité de base de la langue, c'est pourquoi ils sont les seuls à être notés dans le dictionnaire. Les tons de base sont ainsi notés au niveau des lemmes à tous les niveaux où ils apparaissent.

Cette présentation, bien que visant l'exhaustivité, a l'inconvénient de trop alourdir les articles du dictionnaire par la présence d'un grand nombre de signes diacritiques. La vue d'un trop grand nombre de signes, qu'ils ne maîtrisent pas forcément, est décourageant pour beaucoup d'utilisateurs de dictionnaires.

Un autre cas est celui du dictionnaire English-Zulu⁷ et Zulu-English (Doke, Malcolm, Sikakana et Vilakazi 1990) dans lequel il est indiqué que fondamentalement le Zulu est une langue qui comporte neuf variétés tonales. Mais les rédacteurs estiment (comme ceux du dictionnaire Venda) que seuls les tons ponctuels haut et bas sont les véritables tons de la langue; les autres sont des variations et/ou des associations de ces deux tons. Mais à la différence du dictionnaire Venda où les marques de tonalité sont bien visibles dans la présentation de la macrostructure, le dictionnaire Zulu se contente de décrire, dans les pré-textes, les différents contextes d'apparition ou de prononciation de chaque ton. Le ton descendant par exemple est décrit comme étant le résultat d'un ton haut qui est articulé avec une dépression occasionnée généralement en Zulu par la présence des sons tels que *bh, d, g, gc, gx, gq, v, z, dl, hh* et *j*. Dans ce dic-

tionnaire aucune tonalité n'est présentée dans la macrostructure. L'utilisateur doit donc régulièrement consulter les pré-textes pour comprendre le comportement de la langue au niveau suprasegmental.

Cette méthode, bien qu'allégeant la macrostructure du dictionnaire en faisant une économie en terme d'espace (dans le dictionnaire) et de temps (pour le travail du dictionnaire), a l'inconvénient de ne présenter aucun ton dans la macrostructure. Toutes les explications relatives aux tons sont concentrées dans les pré-textes. Cela risque de ne pas être très utile aux usagers, surtout que ces derniers n'ont pas toujours la culture de la bonne utilisation du dictionnaire et par conséquent ils consultent assez peu ou rarement les pré-textes.

Perspectives

Le problème de la tonalité dans les dictionnaires est donc traité différemment selon le choix des lexicographes. Pour ce qui est des dictionnaires de langues gabonaises, la problématique reste entière, puisque n'ayant pas encore été véritablement abordée par les spécialistes (linguistes et lexicographes). Le Gabon est entré dans un processus irréversible de développement de ses langues. Ce processus a comme priorités les plus immédiates, la réalisation de dictionnaires dont le contenu devrait non seulement satisfaire les besoins des usagers, mais prévoir les problèmes de ces derniers. Il ne fait aucun doute que l'une de ces difficultés serait la réalisation tonale. Il n'est donc pas inutile dès aujourd'hui de proposer des voies qui pourraient aider à mieux aborder le problème de tons dans les futurs dictionnaires de langues gabonaises.

Quelques propositions

L'un des problèmes majeurs rencontré dans les langues gabonaises est celui de la pluralité des tons. La trop grande variété de tons dans les langues gabonaises est l'une des principales raisons de la réticence d'un grand nombre de chercheurs à les noter dans les dictionnaires ou des ouvrages à référence lexicographique. Les noter non seulement alourdirait les textes, mais on n'est jamais certain de pouvoir tous les noter convenablement comme ils apparaissent réellement dans les langues. L'une des solutions que l'on pourrait proposer est que pour la notation des tons dans les dictionnaires et autres ouvrages à référence lexicographique, on se limite aux tons les moins récurrents et à deux tons au maximum. Cette option ferait appel à une description linguistique⁸ détaillée de la langue choisie. En vue de la description tonologique de cette langue, le lexicographe pourra clairement y distinguer la fréquence des différents tons. Ma proposition est donc que seuls les tons les moins fréquents soient notés. Mais la prise en compte des tons à noter doit se faire en fonction de chaque terme. Si un terme n'a que deux types de tons, seul un ton doit être noté. Mais si le terme contient plus de trois tons, seuls deux devraient être notés. La notation des tons

ne devrait pas aller au-delà de deux. Cela permettra de ne pas trop surcharger les articles dans le dictionnaire, à condition que le lexicographe informe les usagers de cette disposition dans les pré-textes.

La solution de noter les tons les moins fréquents ne pourrait être efficace que si la variété des tons retenus est minime. S'il existe plus de trois variétés tonales dans une langue, il serait difficile pour les usagers de deviner quels sont les tons qui ne sont pas notés et comment les différencier de ceux qui le sont pas. La solution à cette situation serait que seuls les deux tons les plus fréquents de la langue soient retenus. Les descriptions linguistiques permettraient d'identifier les deux tons les plus réguliers dans chaque langue. De ces deux tons, le moins régulier devrait être noté dans le dictionnaire.

Une autre approche serait de ne tenir compte que des tons ponctuels principaux (haut, bas et moyen). Il semble que dans la majorité des langues à tons, le ton haut et le ton bas soient les plus faciles à identifier sur le plan auditif, leur courbe mélodique étant constante et courte. Ainsi, le ton moyen qui n'est pas aussi facile à identifier que les deux autres, serait obligatoirement noté accompagné par le moins fréquent des deux autres (le haut ou le bas). Dans cette configuration, seuls deux tons seraient notés, tout en ayant la certitude que les articles du dictionnaire ne seraient pas trop surchargés par les marques tonales. Cette démarche s'inspire en partie de celle adoptée dans les dictionnaires Venda et Zulu dans lesquels les autres tons (montant, descendant, faillotonale, etc.) sont considérés comme de simples variations des tons ponctuels haut et bas.

Marquer les tons dans un dictionnaire peut aussi se faire d'une autre manière que par les signes diacritiques habituels. Dans le cas où le lexicographe estime ne pas vouloir (ou pouvoir) se passer de certains tons, il pourrait, comme je l'ai proposé plus tôt, identifier les deux tons les plus réguliers et ne noter que le moins fréquent des deux. À cela il pourrait ajouter la transcription en italique par exemple, des syllabes portant un autre ton qu'il estime important. Le caractère italique serait ainsi une marque tonale, mais à la différence qu'il ne s'agirait pas d'une diacritique spécifique. Une fois de plus, l'objectif est de marquer le mieux possible la structure suprasegmentale des langues gabonaises, sans pour autant rendre le dictionnaire illisible ou décourageant pour les usagers.

Conclusion

L'intérêt pour les dictionnaires en langues gabonaises va de pair avec le développement de ces langues. Les principaux problèmes auxquels sont confrontés les lexicographes travaillant sur ces langues, est celui de l'orthographe et de la notation de la tonalité. L'objectif de cet exposé était de proposer quelques voies en vue d'aider à résoudre le problème de la notation des tons dans les dictionnaires de langues gabonaises. Ces propositions ne sont pas exhaustives et ne prétendent pas être les seules possibles pour la résolution du problème, mais

elles pourraient fournir une issue qui permettrait de mieux tenir compte, dans l'avenir, de la tonalité dans les dictionnaires de langues gabonaises, contrairement à ce qui s'est fait jusqu'à ce jour.

Il est important de souligner que, quelles que soient les démarches choisies par les lexicographes en ce qui concerne la notation des tons dans les dictionnaires, celles-ci doivent être expliquées en détails dans les pré-textes afin d'informer les usagers et leur permettre d'utiliser au mieux le dictionnaire. Il est évident que sans ces explications, toute démarche serait vouée à l'échec, du moins en ce qui concerne l'objectif de permettre aux usagers de mieux saisir la bonne prononciation tonale des langues gabonaises.

Notes

1. Les références de ces descriptions sont disponibles dans la bibliographie.
2. Les exemples sont transcrits en orthographe des langues gabonaises proposée à Libreville par un colloque d'expert en avril 1999.
3. Les titres de tous les dictionnaires de langues gabonaises consultés sont disponibles dans la bibliographie.
4. Ces dictionnaires et ouvrages parcourus il y a longtemps n'étaient pas à ma disposition au moment où j'ai rédigé cet article.
5. Une polémique existe au Gabon entre certains linguistes et la Fondation Raponda Walker auteur des livres « Rapido-langue ». Les linguistes estiment que ces ouvrages ne sont pas scientifiques parce que la tonalité n'y est pas marquée, alors que la Fondation Raponda pense que mettre les tons pourrait rendre les ouvrages difficilement accessibles pour le commun des usagers non-initié à la linguistique.
6. Le Tsivenda est une langue bantoue d'Afrique du Sud et s'écrit généralement avec un accent circonflexe renversé en dessous du « d ». Cette diacritique n'a malheureusement pas pu être reproduite, mais cette absence ne change pas grand chose pour sa prononciation.
7. Le Zulu est une langue bantoue d'Afrique du Sud.
8. Une collaboration plus étroite et plus efficace entre linguistes et lexicographes gabonais reste, dans ce contexte, plus que nécessaire. Les lexicographes pourront se servir des résultats des descriptions linguistiques pour améliorer le traitement lexicographique des langues gabonaises.

Bibliographie

- Adam, J.J.** 1969. *Dictionnaire Ndumu-Mbebe-Français et Français-Ndumu-Mbebe. Petite flore de la région de Franceville (Gabon). Grammaire Ndumu-Mbebe.* Archevêché de Libreville (Bar-le-Duc): Imprimerie St. Paul.
- Afane Otsaga, T.** 2001. Les dictionnaires bilingues dans les langues gabonaises: Approche méthodologique. *Éléments de lexicographie gabonaise Tome I: 137-159.* New York: Jimacs-Hillman.
- Andeme Allogo, M.F.** 1980. *Esquisse Phonologique du Nzaman de Makokou.* Mémoire de Licence. Libreville: Université Omar Bongo.

- Baylon, C. et P. Fabre.** 1990. *Initiation à la linguistique*. Paris: Nathan
- Bessieux, J.R. (Mgr).** 1847. *Dictionnaire Français-Pongoué / Pongoué-Français*. 2 volumes. Amiens: Lenoël-Hereaurt.
- Biton, A.** 1907. *Dictionnaire Français-Ndumu et Ndumu-Français, précédé d'éléments de grammaire*. Nantes: Imprimerie Bourgeois.
- Cinnamon, J.** 1990. *Faj (Nzaman)*. *Revue Gabonaise des Sciences de l'Homme* 2: 175-188.
- Doke, C.M., D.M. Malcolm, J.M.A. Sikakana et B.W. Vilakazi.** 1990. *English-Zulu / Zulu-English Dictionary*. Johannesburg: Witwatersrand University Press.
- Église Évangélique du Gabon.** 1960. *Dictionnaire Français-Yipounou / Yipounou-Français*. Mouila.
- Fishman, Joshua A. (Éd.).** 1968. *Readings in the Sociology of Language*. La Haye/Paris: Mouton.
- Galley, S.** 1964. *Dictionnaire fang-français et français-fang, suivi d'une grammaire fang*. Neuchâtel: Henri Messeiller.
- Gove, P.B. (Éd. en chef).** 1961. *Webster's Third New International Dictionary of the English Language*. London: G. Bell/ Springfield, Mass.: G. & C. Merriam.
- Hombert, J.M.** 1990. Problèmes phonétiques et phonologiques rencontrés dans les langues du Gabon. *Revue Gabonaise des Sciences de l'Homme* 2: 97-103.
- Kwenzi-Mikala, J.T.** 1980. La phonologie du Yipunu. *Éléments de description du Yipunu*. Lyon: CRLS. Univ. Lyon II.
- 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/H. Teillard.
- Marichelle, C.** 1900. *Dictionnaires Français-Vili*. Loango.
- Marling (R.P.).** 1872. *Dictionnaire fang-français*. New York.
- Martrou, L.** 1924. *Lexique fān-français*. Paris: Procure Générale (des Pères du St Esprit) / Abbeville: Imprimerie Paillard.
- Matthews, P.H.** 1997. *Concise Dictionary of Linguistics*. New York: Oxford University Press.
- Mavougou, P.A.** 2001. La mondialisation et la lexicographie trilingue ou plurilingue au Gabon. *Éléments de lexicographie gabonaise, Tome I*: 160-183. New York: Jimacs-Hillman.
- Mdee, J.S.** 1999. Dictionaries and the Standardization of Spelling in Swahili. *Lexikos* 9: 119-134.
- Mihindou, G.R.** 2001. Apports des missionnaires à la lexicographie gabonaise: Dictionnaires bilingues Fang-Français/Français-Fang; Français-Yipounou/Yipounou-Français; Français-Mpongwé/Mpongwé-Français. *Éléments de lexicographie gabonaise, Tome I*: 7-37. New York: Jimacs-Hillman.
- Mission Catholique du Gabon.** 1925. *Syllabaire Fang*. Mellville (Nord): Imprimerie Ceugnard Le sage.
- Mission Évangélique de Paris.** 1927. *Catéchisme Élémentaire en Fang*. Réédité en 1931 et 1958 par la Société des Missions Évangéliques de Paris.
- Mission Protestante du Congo Français.** 1912. *Ntèn Fan Ôsua: Premier Livre de Lecture en Fan*. Saint-Blaise (Suisse): Samuel Robert.
- Mission Protestante de Talagouga.** 1889. *Nteni osu Nteni Fanwe*. Talagouga/Ndjolé.
- Mission Protestante de Talagouga.** 1898. *Nten Misili Ye biya bi fan Ôsua*. Talagouga.
- Mission Protestante Française.** 1920. *Ezangô Elônga Zi Galwa: Premier Livre de Lecture en Galoa*. London: A.E.F.

- Muroni, J.-M.** 1989. *Petit dictionnaire du bantu du Gabon*. Paris: l'Harmattan.
- Nsuka Nkutsi, F. (Éd.)**. 1980. *Éléments de Description du Punu*. Lyon: Université de Lyon II.
- Ondo Mebiame, P.** 1992. *De la phonologie à la morphologie du Fang parlé à Aboumezok (langue Bantu A.78)*. 2 volumes. Thèse de doctorat. Tervuren: Université Libre de Bruxelles.
- Puech, G.** 1990. Bekwil. *Revue Gabonaise des Sciences de l'Homme* 2: 127-128.
- Raponda Walker, A. et R. Sillans.** 1961. *Plantes utiles: essai d'inventaire et de concordance des noms vernaculaires et scientifiques des plantes spontanées et introduites du Gabon*. Paris: Lechevalier.
- Raponda Walker, A.** 1995². *Dictionnaire Français-Mpongwé*. Class. Africaines.
- Raponda Walker, A.** 1995². *Dictionnaire Mpongwé-Français*. Class. Africaines.
- Raponda Walker, A.** 1995. *Éléments de grammaire Fang*. Libreville: Éd. Raponda.
- Raponda Walker, A.** 1995. *Éléments de grammaire Gisira*. Libreville: Éd. Raponda.
- Raponda Walker, A.** 1995. *Essai de grammaire Tsogo*. Libreville: Éd. Raponda.
- Voltz, M.** 1990. Faŋ (Ntumu). *Revue Gabonaise des Sciences de l'Homme* 2.
- Warmelo, N.J. van.** 1937. *Tshivenda-English Dictionary*. Pretoria: Government Printer.

Semi-automatic Term Extraction for the African Languages, with Special Reference to Northern Sotho *

Elsabé Taljard, *Department of African Languages, University of Pretoria, Pretoria, Republic of South Africa (e.taljard@freemail.absa.co.za)*, and Gilles-Maurice de Schryver, *Research Assistant of the Fund for Scientific Research — Flanders (Belgium) and Department of African Languages, University of Pretoria, Pretoria, Republic of South Africa (gillesmaurice.deschryver@rug.ac.be)*

Abstract: Worldwide, semi-automatically extracting terms from corpora is becoming the norm for the compilation of terminology lists, term banks or dictionaries for special purposes. If African-language terminologists are willing to take their rightful place in the new millennium, they must not only take cognisance of this trend but also be ready to implement the new technology. In this article it is advocated that the best way to do the latter two at this stage, is to opt for computationally straightforward alternatives (i.e. use 'raw corpora') and to make use of widely available software tools (e.g. WordSmith Tools). The main aim is therefore to discover whether or not the semi-automatic extraction of terminology from untagged and unmarked running text by means of basic corpus query software is feasible for the African languages. In order to answer this question a full-blown case study revolving around Northern Sotho linguistic texts is discussed in great detail. The computational results are compared throughout with the outcome of a manual excerption, and *vice versa*. Attention is given to the concepts 'recall' and 'precision'; different approaches are suggested for the treatment of single-word terms *versus* multi-word terms; and the various findings are summarised in a Linguistics Terminology lexicon presented as an Appendix.

Keywords: TERMINOLOGY, TERMINOGRAPHY, MANUAL EXCERPTION, READING AND MARKING, SEMI-AUTOMATIC TERM EXTRACTION, RETRIEVAL, AFRICAN LANGUAGES, NORTHERN SOTHO (SEPEDE), RAW CORPORA, PRETORIA SEPEDEI CORPUS (PSC), WORDSMITH TOOLS, WEIRDNESS RATIO, KEY WORD, LOG-LIKELIHOOD, RECALL, PRECISION, MOTHER TERM, SINGLE-WORD TERM, MULTI-WORD TERM, STEM, ROOT, KEY-WORD-IN-CONTEXT (KWIC), COLLOCATION, COLLOCATE, LEXICAL GAP, CLUSTER, LINGUISTICS TERMINOLOGY LEXICON

Senaganwa: Go ntšhwa ga mareo ka tirišo ya seripa sa semotšhene malebana le maleme a Afrika, šedi ye kgolo e lego Sesotho sa Leboa (Sepedi). Go ntšhwa ga mareo ka tirišo ya seripa sa semotšhene go tšwa ka gare ga dikhophase go thomile go ba

* An earlier version of this article was presented as a paper at the Seventh International Conference of the African Association for Lexicography, organised by the Dictionary Unit for South African English, Rhodes University, Grahamstown, 8–10 July 2002.

setlwaedi go hlangweng ga mananeo a mareo, dipanka tša mareo goba dipukuntšu mererong yeo e itšego lefaseng ka bophara. Ge e le gore boramareo ba maleme a Afrika ba ikemišeditše go tšea madulo a bona mo mileneamong wo mofsa, ga ba swanela go hlokomela fela tsela ye, eupša ba swanetše gape ke go ikemišetša go diriša theknolotši ye mphsa. Mo taodišwaneng ye go hlalošwa gore mo nakong ye, tsela ye kaone ya go dira dilo tše pedi tše go boletšwego ka tšona ke go kgetha ditlhamolo tša thwii tšeo di dirišago khomphutha (se se ra gore tšhomišo ya khophase) le go šomiša ditlabakelo tša *software* (bj.k. *WordSmith Tools*) tšeo di lego gona gohle. Ka fao maikemišetšo a magolo ke go humana ge e ka ba go ntšhwa ga mareo ka seripa sa semotšhene go tšwa ka gare ga khophase yeo e se nago ditlaleletšo tšeo di tšeneletšego ka mašakaneng, tša go hlhla, go ka dirišwa malemeng a Afrika goba aowa. Gore re kgone go araba potšišo ye, go hlalošitšwe ka tsinkelo mohlala wa taba ya go nyakišišwa yeo e amanego le diteng tša thutapolelo tša Sesotho sa Leboa. Dipoelo tšeo di humanwego ka go diriša khomphutha di bapetšwa ka gohle le dipoelo tšeo di humanwego ge go dirišwa kgetho ya mantšu ka matsogo. Šedi e fiwa dikgopolo tša kgakologelo (*recall*) le nepagalo (*precision*); mekgwa yeo e fapafapanego e a akanywa gore e kgone go hlatholla mareo a lentšu le tee ge a bapetšwa le mareo a mantšu a mantši; gomme dikhumano tšeo di fapanego di akaretšwa ka gare ga pukuntšu ya Mareo a Thutapolelo yeo e tšweletšwago bjalo ka Mamatletšo.

Mantšu a bohlokwa: MAREO, MONGWALO WA MAREO, KGETHO YA MANTŠU KA MATSOGO, GO BALA LE GO SWAYA, GO NTŠHWA GA MAREO KA SERIPA SA SEMOTŠHENE, GO HWETŠA GAPE, MALEME A AFRIKA, SESOTHO SA LEOBOA (SEPEDI), DIŠEGONTŠU (DIKHOPHASE), KHOPHASE YA SESOTHO SA LEOBOA YA TSHWANE (KST), WORDSMITH TOOLS, WEIRDNESS RATIO, LENTŠU LA BOHLOKWA, LOG-LIKELIHOOD, KGAKOLOGELO, NEPAGALO, LEREO LA MOTHEO, LEREO LA LENTŠU LE TEE, LEREO LA MANTŠU A MANTŠI, KUTU, MODU, LENTŠU LA BOHLOKWA KA GARE GA KAMANO (LBGK), PEAKANYO, BEAKANYA, TLHOKEGO YA LEREO, SEHLOPHA, PUKUNTŠU YA MAREO A THUTAPOLELO

1. Semi-automatic Term Extraction — A Brief Theoretical Conspectus

On the international front, the use of electronic corpora for general lexicographical purposes has for the past two decades become a firmly entrenched procedure. According to Ahmad and Rogers (2001: 729) "it is common practice these days in many different types of dictionary to use the systematic evidence of corpora rather than the more *ad hoc* selection of citations by readers more traditionally used in lexicography". De Schryver and Prinsloo (2000: 292) state that "[t]he intensified systematic exploitation of electronic corpora for lexicographic purposes has unmistakably revolutionised dictionary making" and point for example out that during the compilation of all the recent British English learners' dictionaries "electronic corpora were used very actively in order to produce reference works of a standard hitherto simply unimaginable".

The use of electronic corpora for terminological purposes has however been accepted much more slowly on both the theoretical and practical levels. The onomasiological approach in terminology as opposed to the semasiological

bias of general lexicography is offered by Ahmad and Rogers (2001: 729) as one possible reason for the lack of corpora utilisation in terminology management. Terminology as a scientific discipline is concept-driven, the basic objectives of terminological work being: (a) delimiting or identifying the concepts of a subject field or domain, (b) naming these concepts by means of terms, and (c) fixing the referential scope of each term by means of a definition. Thus, on a strictly theoretical level, the concept-oriented nature of terminology logically excludes the possible use of text corpora, since a concept is an abstract entity, not to be found in textual material. If, however, it is accepted that terminology also includes a terminographical dimension, which has the compilation and/or publication of a terminology list, a term bank or a dictionary for special purposes as its final objective, the use of corpora and the subsequent computational management of terminology becomes a relevant issue. In this regard, Ahmad and Rogers (2001: 730) indicate that "there are clear signs that corpus-based terminology management, including the identification of terms and translation-oriented terminology, as well as the whole concept of terminology management, is now being discussed". Sager (1990: 130) is even more emphatic in stating that systematic term compilation is firmly corpus-based, which implies that terms are no longer manually excerpted from previous lists or by individual searches, but from a corpus of material.

The process whereby computer software is used to automatically detect and extract potential terms from electronic corpora, is known as *(semi-)automatic term extraction*. In the great majority of the current approaches, characteristics of a special-language corpus are compared to those of a general-language corpus. In *all* approaches, humans remain the final arbiters, and must decide whether or not the terms suggested by the software do indeed have term status. Broadly speaking, the approaches themselves are either purely statistical, purely linguistic, or they are hybrid, i.e. they combine features of the two extremes. Moreover, different methods are often used to extract single-word terms as compared to the extraction of multi-word terms.

2. Electronic Special-field Corpora for the African Languages and Query Software

When it comes to the use of electronic corpora for the computational management of terminology for the African languages spoken in South Africa,¹ no headway has hitherto been made, and this for obvious reasons. A corpus is based on available written texts in a specific language. As such, this does not pose a problem for these languages, since they all have a written tradition. Granted, some of these languages have only been reduced to writing during the last decade (e.g. Ndebele), but other languages such as Zulu or Northern Sotho have a relatively long literary tradition. The compilation of an electronic corpus for general lexicographical purposes is therefore not problematic, pro-

vided of course that the necessary technological support is available.²

However, if an electronic database is to be compiled for terminological purposes, it presupposes the availability of text material revolving around specific fields. Due to the historically disadvantaged situation of the African languages, even today virtually no subject-specific texts which could be used to build an electronic database are available. As a result of the pre-1994 political and educational system, the vast majority of subject-specific material is written in either English or Afrikaans, with textbooks on literature and grammar of the African languages a possible exception. The African-language terminologist therefore has very little, if any, access to special-field texts which can be used to compile an electronic special-field corpus. This does not only have implications for the compilation of corpora, but also determines the methodology which has hitherto been used by African-language terminologists. Due to the lack of special-field texts, terminologists compiling terminology lists would make use of texts written in English and/or Afrikaans to select the relevant terms. After having isolated the terms which characterise a specific subject field, these terms would then be translated into the African languages, resulting in a multilingual terminology list. Alberts (2000: 236-237) refers in this regard to the compilation of technical dictionaries for the African languages by terminologists from the National Language Service (NLS) of the Department of Arts, Culture, Science and Technology (DACST) on a variety of special-field subjects.

Nonetheless, the dependency of African-language speakers and terminologists on textbooks and other subject-specific sources produced in English and/or Afrikaans will hopefully decrease in the times to come, seeing that the African languages are starting to take their rightful place in the South African educational landscape. As a matter of fact, it is the authors' belief that special-language texts will soon be produced on a *large* scale in the African languages. The evolution on the Internet is a good case in point. Indeed, more and more texts of a technical nature already make their appearance in African languages online. It is this realisation that prompted the current research. Since the trend worldwide is increasingly towards the semi-automatic extraction of terminology from corpora, African-language terminologists must not only be aware of this development, but they must also be fully prepared. The main aim of this article is thus to research the *feasibility* of semi-automatic term extraction for the African languages.

We purposely opted for computationally straightforward alternatives and insisted on using widely available software tools. The rationale behind this is simply that we wish to reach out to as many colleagues as possible. On the corpus level, this implies that we chose to work with *raw corpora*, i.e. just plain running text without any tags or mark-up whatsoever. As far as the software is concerned, we selected WordSmith Tools (WST), "an integrated suite of programs for looking at how words behave in texts" (Scott 1999: WST help).³ WST is inexpensive, easy to acquire, user-friendly, and already in use at several National Lexicography Units (see e.g. De Schryver and Lepota 2001: 3).

3. Case Study: Northern Sotho Linguistics Terminology

In order to evaluate the success rate and usefulness of the suggested approach (i.e. the analysis of raw corpora with WST, with the aim of semi-automatically extracting terminology), the computational result has to be compared to the outcome of the current method of manual term excerption (i.e. the physical reading of texts and marking of relevant terms). A list of manually excerpted terms will therefore serve as a terminological benchmark against which the success rate of the computational extraction of terms will be measured. Conversely, since manual term excerption is of necessity subject to human error, the results of the computational processing will also be compared to the results of the manual excerption in order to ascertain whether the semi-automatic processing might have succeeded in identifying terms which were overlooked during the process of manual excerption.

For the purpose of this investigation, a number of texts on Northern Sotho linguistics were taken as the textual material from which terms are to be retrieved. These texts were kindly provided in electronic format by Prof. L.J. Louwrens and are an integral part of the study material used at the University of South Africa (UNISA). After conversion to a text-only format, a simple count with WST's WordList tool revealed that this special-field corpus contains 74,251 tokens (running words) and 4,744 types (unique words).

4. Reading and Marking — A Case Study

Term excerption as a conscious activity forming part of terminology management is influenced by a number of aspects, e.g. the target users and their specific needs, the exact purpose for which the special-field corpus is to be created, the literacy levels of the discourse participants, etc. These aspects do, however, not form the focus of the current investigation. The linguistic texts from which terms are to be retrieved, form part of the study material written for pre-graduate students who are mother-tongue speakers of Northern Sotho. For the sake of the argument, it can be assumed that these students would also be the target users of a basic terminology list containing all the relevant linguistic terms appearing in the texts, and their English equivalents.

The initial phase of the investigation consisted of a manual excerption of terms from the linguistic texts. Manual excerption implies close scrutiny of a text in order to identify terms which are relevant to a specific subject field; in this case, linguistics. This manual reading and marking was performed by a professional terminologist, and the terms were entered into a preliminary term list. Locativised nouns (i.e. nouns displaying the locative suffix **-ng**), as well as relative verbs (i.e. verbs containing the relative suffixes **-go/-ng**), were excluded from this list. This decision was based on the fact that the meaning of these derivations can regularly be inferred from the base forms entered in the list. Furthermore, terms were listed as they appeared in the text, so if a term for

example appeared in its plural form, it was listed as such.

The process of manual term excerption resulted in a term list containing 350 'raw' (i.e. 'unlemmatised') terms. Of these terms, 309 are single-word terms, the rest being made up of 41 multi-word terms. Lemmatisation of the initial list produced a term list containing 285 terms, lemmatised under their singular form in the case of nouns, followed by an indication of gender affiliation. Verbs with verbal extensions were lemmatised as such, and not according to the stem of the verb. Each term was then provided with a translation equivalent in English. The result of this endeavour can be found in the Appendix, where all the articles preceded by ☐ and ☑ constitute the 285 manually excerpted terms after lemmatisation.

5. Semi-automatic Term Extraction — A Case Study

As was pointed out in par. 2, our aim is to investigate how well a simple yet powerful and versatile program such as WST fares in semi-automatic term extraction when this software is fed with *untagged* and *unmarked* corpora. This procedure is therefore by definition language-independent and purely statistical. Consequently, the *unlemmatised* list of 350 manually excerpted terms must be used as a benchmark, as also WST processes raw data. As is generally done in the field of semi-automatic term extraction, we will first look into various ways to computationally extract single-word terms, and only then into the extraction of multi-word terms, for which the benchmarks are 309 and 41 terms respectively.

5.1. Semi-automatic Extraction of Single-word Terms

Ironically, the only publicised attempt to automatically extract African-language terminology from corpora, is the report by Sewangi (2000, 2001) for Swahili. He (Sewangi 2000: 67-68) states:

[T]he identification of single-word terms in a text corpus is difficult because there are no structural criteria that can be used to separate term-words from non-term-words in the text. [...] the identification of single-word terms should involve subject specialists and language experts. [...] This should be done manually on the basis of the knowledge of the subject-domain and of the language.

Although Sewangi could make free use of the computational tools that have been developed for over a decade by Hurskainen (1992, 1995, 1996, 1999, Hurskainen and Halme 2001), and although he thus had access to corpora with full descriptions of Swahili morphological patterns and constraints, he nonetheless effectively marks single-word terms *manually*, which, in view of the highly-technical corpus annotations at his disposal, is disappointing.

Up to this day Swahili remains the only African language for which an

efficient morphological analyser has been built. Several human-language technology projects are under way however — in Pretoria for Zulu and Northern Sotho, in Harare for Shona and Zimbabwean Ndebele — and it is expected that in less than a decade there will be a handful of African-language morphological analysers. As this is not yet the case, investigating the possibility of analysing *raw* corpora at this stage is defensible.

5.1.1. Top Ranks in a Frequency List

An undemanding operation in the computational processing of a corpus is a simple frequency count, in which each type in the corpus can be listed according to its rank, or in order of its frequency. Obviously, since a frequency count lists all types, i.e. unique (orthographic) words, which appear in the text according to their frequency of occurrence, such a procedure will produce a lot of 'noise', in that most of the items appearing amongst, say, the 100 most frequent words, are not terms at all, but other lexical items such as concords, conjunctions, etc. and general vocabulary not related to the subject field.

The linguistic texts were selected in WST and on comparing the frequency list created by WordList with the list of manually excerpted terms, it was found that only 20 items appearing in the top 100, appeared on the list of manually excerpted terms and could thus be regarded as terms. Scanning the frequency list down to rank 500 revealed only an average of 18 terms per 100 items (viz. for the first 100, 20; the second 100, 20; the third 100, 18; the fourth 100, 14; and the fifth 100, 17).

5.1.2. Stop-list Constrained Top Ranks in a Frequency List

Much of the 'noise' produced by a simple frequency count can be reduced substantially by filtering out all items which are generally known as function words and closed-class words. In WST, this filtering can be done automatically by making use of stop lists (also known as 'exclude lists'). For Northern Sotho, the following can be regarded as function words, i.e. lexical items with little or no lexical content: all agreement morphemes, demonstratives, particles, conjunctions, copulative verb stems and auxiliary verb stems. The second group of lexical items that can also be included in a stop list are those items that belong to so-called closed classes, i.e. classes which contain a very limited number of items. The following closed classes can be identified for Northern Sotho: adverbs, interrogative words, adjectives (class prefix + adjective stem, which number only a handful in Northern Sotho), locative nouns, pronouns and ordinal numbers.

Even if, in addition, locativised nouns, relative verbs and nouns with diminutive suffixes are also disregarded, the top 100 items on the 'cleansed' frequency list still contain only 39 items (as compared to 20 in par. 5.1.1) that also appear on the manually excerpted term list. Even with a stop list, it is clear that a simple frequency list tends to over-generate, i.e. to identify items which

are not terms relevant to the specific subject field. Also, even though stop lists have the advantage that they can be made once and then be used for numerous extractions, they, conversely, hold the risk that interesting data is nonetheless cut away.

5.1.3. Weirdness Ratios, Key Words and Recall and Precision

Weirdness ratios

Reading through *top-frequency words*, no matter whether these were automatically extracted with or without stop lists from running text, is obviously an unrefined procedure. Ahmad and Rogers (2001: 744-745) suggest that the comparison of the frequency distribution of items in special-language texts to that of items in general-language texts might be a next option worth investigating. When the relative frequency of each item in special-language texts is divided by the relative frequency of the same item in general-language texts, the 'weirdness ratio' of those items, and thus their potential as term candidates, can be measured.

Key words — How they can be extracted semi-automatically

A much more sophisticated and statistically sounder approach to weirdness ratios is offered by the KeyWord tool of WST. Since 1997 Mike Scott, the creator of WST, has extensively studied — and widely published on — the computational treatment of key words and related concepts (e.g. Scott 1997, 1997a, 2000, 2000a, 2001). He rightly notes that the term 'key word' has remained undefined in linguistics, despite being in common use by non-linguists, and despite the fact that the notion itself features strongly in fields such as Content Analysis, Information Retrieval, and Corpus Linguistics at large (Scott 2000: 51).

In his work on text schemata and stereotypes, Scott (1997: 236) defines the term 'key word' as 'a word which occurs with unusual frequency in a given text'. Unusual frequency can be related to outstandingness and implies that a word has an unusually high (or unusually low) frequency in a text (or sub-corpus), in comparison to its occurrence in a *reference corpus* of some kind. In this specific study, Scott's aim was to make certain culturally significant inferences about schemata, i.e. the socially determined networks of links between ideas, by identifying key words which appear across feature stories taken from the *Guardian* newspaper.

The basic procedure for identifying key words in one or more texts is to compare the frequency of every distinct word-type in those texts with the frequency of the same word-type in a reference corpus, the reference corpus being the bigger of the two corpora. As a first step, a wordlist is compiled of the reference corpus by making use of the WordList tool. The wordlist contains all the different types in the reference corpus and lists them according to their frequency of occurrence. Secondly, a similar, but much smaller, wordlist is drawn

up for the specific text(s) for which the key words are to be identified. The third step, which is done by means of the KeyWord tool, consists of comparing the frequency of each item in the smaller of the two wordlists with the frequency of the same item in the reference wordlist. Items which display a great disparity in frequency are identified as key words, since the disparity would imply that that specific item occurs with unusual frequency in the smaller corpus. Note that KeyWord throws up items with outstanding frequencies and not 'top frequencies'. Thus, even if a given item appears with, for example, an extremely high frequency of 5% in the smaller corpus, and if that item would have a similar percentage in the reference corpus, such an item will not turn out to be 'key', even though it might perhaps be the 'most frequent' item. Every word which appears in the smaller corpus is taken into account when 'keyness' is calculated, except if it has been excluded by entering it into a stop list.⁴

For the actual calculation of keyness, WST offers two statistical tests, viz. the classic χ^2 (chi-square) test of significance with Yates' correction for a 2 x 2 table, and Dunning's (1993) log-likelihood test "which gives a better estimate of keyness, especially when contrasting long texts or a whole genre against [a] reference corpus" (Scott 1999: WST help). We opted for the log-likelihood test, a choice generally made by corpus linguists today (compare Scott 1997a: 238). Two more parameters are important when calculating key words, and must be set: (a) the minimal frequency, and (b) the level of outstandingness. The first parameter specifies the minimum frequency with which a potential key word must occur in the text(s) from which the key words are to be extracted. The value was set at 3, yet values of 2 and 1 resulted in near-identical findings. The second parameter, also known as p value, establishes a minimum probability. The standard $p \leq 0.000001$ was used, meaning that each key word's appearance has a danger of only 1 in a million of not being statistically significant. Reformulated, in our study an item is said to be a 'key word' if: (a) it occurs in the text(s) at least 3 times, and (b) its frequency in the text(s) when compared with its frequency in a reference corpus is such that the statistical probability as computed by the log-likelihood procedure is smaller than or equal to one in a million.

Key words — An illustration of how they can be extracted semi-automatically

It is clear that the identification of key words is a purely mechanical process, based on a comparison of patterns of frequency. To illustrate the procedure, a Northern Sotho text on the new South African coat of arms, taken from the Internet,⁵ was randomly selected to serve as an example of a specific text for which key words are to be identified. This text, henceforth CoA, contains 1,038 tokens and 356 types. As a reference corpus, a selection of the Pretoria Sepedi Corpus (PSC), consisting of 5,175,686 tokens and 136,567 types, was used. As far as the reference corpus is concerned, Scott (1997: 244, endnote 9) observes that "as long as the reference corpus is fairly sizeable" — and he suggests at least a million running words — "results are quite similar even if the reference

corpus is altered" (compare also Scott 2000a: 115). This observation was found to be true.⁶ Scott (2001: 126, endnote 2) further suggests that the text(s) from which key words are to be extracted can even form a tiny sub-set of the reference corpus. This is also what was done, i.e. CoA is part of PSC. Wordlists were drawn up for CoA and PSC with the WordList tool, and these lists were fed into the KeyWord tool. All the key words suggested by KeyWord are shown in (1).

- (1) Semi-automatic key-word extraction from a text on the new South African coat of arms

N	Key word	Translation	CoA Count	CoA %	PSC Count	PSC %	Keyness
1	sefoka	coat of arms	10	0.96	231		87.3
2	ditirelo	services	8	0.77	114		77.3
3	bontšha	show	11	1.06	640	0.01	76.1
4	seswa ⁷	(something) new cl. 7	6	0.58	18		75.2
5	Afrika	Africa(n)	9	0.87	941	0.02	51.9
6	tlhame	secretary-bird	4	0.39	19		46.9
7	barulaganyi	designers	3	0.29	3		42.8
8	Borwa	South	8	0.77	1,137	0.02	41.4
9	badiriši	users	3	0.29	9		37.6
10	setšhaba	nation	10	0.96	3,204	0.06	36.2
11	lebišitšwe	is / are aimed at	3	0.29	14		35.3
12	batho	people	16	1.54	12,232	0.24	33.1
13	se	subj. conc. cl. 7; dem. cl. 7; ...	39	3.76	73,986	1.43	27.6
14	mmušo	government	6	0.58	1,214	0.02	26.9
15	manaka	tusks	3	0.29	72		25.9
16	leswa	(something) new cl. 5	3	0.29	73		25.9
17	emela	represent(s)	4	0.39	308		25.5
18	tshedimošo	information	3	0.29	85		25.0
19	mabapi	with regard to, regarding	5	0.48	830	0.02	24.4
20	<i>a</i>	<i>subj. conc. cl. 6; poss. conc. cl. 6; ...</i>	26	2.50	301,005	5.82	26.1

Columns 2 and 3 in (1) list the key words the KeyWord tool extracted entirely automatically. As stressed at the outset, human beings remain the final arbiter, which is why Column 3 was added here. Columns 4 and 5 show the occurrence (as a count and percentage respectively) of the suggested key words in CoA. The count and percentage of those same items in PSC, the reference corpus, is shown in Columns 6 and 7. The last column, Column 8, lists the keyness values.

From (1), it is clear that **sefoka** 'coat of arms' occurs 10 times in CoA, compared to an occurrence of 231 times in the bigger reference corpus, yet *proportionally* its frequency is many times higher in the smaller corpus than in the 5.2-million-word reference corpus. **Sefoka** is, as a result of this large disparity in frequency, the item with the highest keyness value. In (1) all suggested key words are 'positively key', except for the last item (in italics) **a** 'subject concord of class 6; possessive concord of class 6; ...' which is 'negatively key'. The latter simply means that it occurs less often than would be expected by chance in comparison with the reference corpus. By simply scrolling over the key words in (1) one can deduce that the text in question must provide information

regarding the new South African coat of arms. It also seems as if the designers aimed at showing symbols such as a secretary bird and tusks to represent the government's attempt to provide new services to the nation's people / users. Note that all the relevant⁸ key words were used in the previous description, clearly suggesting that key words — which, lest we forget, are proffered *fully automatically* by the KeyWord tool — do indeed pinpoint the 'aboutness' (Scott 2000a: 107-109) of a text.

Key words — Semi-automatically extracting single-word terminology

In his discussion of possible applications of the KeyWord tool, Scott (1997: 243) makes no mention of its potential value for terminological purposes. However, Ahmad and Rogers (2001: 744) claim that "[c]omputing the 'ratio' of word forms in special-language and general-language texts also allows a provisional distinction to be made between general-language open-class words on the one hand, and special-language open-class words on the other, i.e., term candidates". This is exactly what KeyWord does, albeit in a more sophisticated way. In terms of Scott's KeyWord procedure, this will imply that the frequency of items appearing in the linguistic texts (being special-language 'texts') be compared to their frequency in the reference corpus (being general-language 'texts'), in order to identify term candidates. Again, the result of such an investigation will be compared to the outcome of the manual term excerption in order to evaluate the efficacy of the computational procedure.

In carrying out the KeyWord procedure as described above, 654 key words are suggested by KeyWord. From item 586 onwards, though, one moves into negative keyness, where there are obviously no linguistic terms to be found. Reading through the 585 terms that are positively key, one quickly finds out that an amazing 189 of them also appear on the manual list. 61% of the manually excerpted single-word terms (189 out of 309) are thus thrown up *entirely automatically* by the software. The only required human intervention is to read through the suggested list and to decide on term status.

Moreover, in doing so, an *extra* 18 terms that had been missed during the manual excerption, are revealed. Besides the intrinsic advantage of a computational approach with which a large percentage of the terminology can be extracted automatically, the fact that a computational approach also reveals items that are missed during a manual pass, might even be of greater value. Actually, following the various semi-automatic approaches for the extraction of single-word terms (see also par. 5.1.4 below), a total of 33 new single-word terms were revealed. The benchmark that will therefore henceforth be used for comparisons will be $309 + 33 = 342$ single-word terms.

Recall and Precision

Two concepts now need to be introduced that are central to Information Retrieval, viz. 'recall' and 'precision'. We list two sets of definitions — first Streh-

low's (2001: 428-429) from a document-retrieval perspective, and then Ahmad and Rogers' (2001: 748) from a terminology-retrieval perspective — to show their wide application:

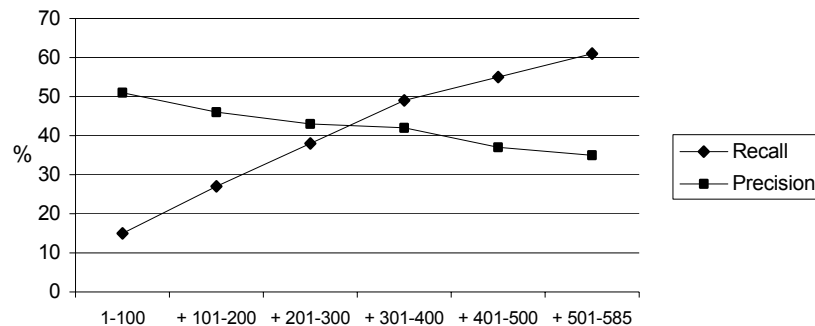
Retrieval effectiveness is usually described by the *precision* and *recall* that are associated with a retrieval operation. *Recall* is the fraction of relevant documents retrieved from a collection, and *precision* is the fraction of documents that are relevant in a retrieved set of documents. [...] In general, recall and precision are found to be inversely related, i.e., an increase of one results in a decrease of the other [...]

"Recall" is the proportion of relevant materials retrieved from a text collection given a set of terms. "Precision" is the proportion of retrieved materials that are relevant.

In other words, given a running text or sub-corpus in which there are a number of terms, *recall* is the percentage of terms actually retrieved as compared to the total number of terms in that text or sub-corpus. When all the retrieved items as a whole, i.e. both the retrieved terms and the 'noise' (i.e. retrieved non-terms) are considered, the actual percentage of terms in that body of retrieved items is called *precision*.

Both recall and precision will now be calculated for the semi-automatic single-word extraction achieved with the KeyWord tool above. In the top 100 KeyWord list, 51 so-called key words are effectively terms, of which 1 is new compared to the manual excerption, i.e. 50 + 1. Analogously, in the second 100, there are 36 + 4 terms; in the third 100, 36 + 3; in the fourth 100, 31 + 5; in the fifth 100, 19 + 2; and in the last stretch up to 585, 17 + 3. As one proceeds through the KeyWord list, recall and precision, expressed in %, are therefore as shown in (2).

(2) Recall and Precision for the semi-automatic extraction of single-word linguistic terms with KeyWord



With a benchmark of 342, the *recall* for the first 100 is 51 divided by 342 or 15%; when the next 100 are added, it is 51 + 40 divided by 342 or 27%; etc. As one proceeds through the KeyWord list, the recall thus *increases*. The *precision* for the first 100 is 51 out of 100 items, or thus 51%; when the next 100 are added it is 51 + 40 out of 200 items, or thus 46%; etc. As one proceeds through the Key-Word list, the precision thus *decreases*. From the graph in (2) one can see that the claimed inverse relationship between recall and precision holds rather well for our case study.

It should be clear from the discussion so far that the KeyWord procedure can be used successfully for the semi-automatic extraction of a large proportion (here over 60%) of the single-word terms in running text. When compared to the use of (stop-list constrained) top ranks in a frequency list, the KeyWord analysis represents a substantial increase as regards the success rate of computational processing. The question now remains, however, as to how the rest of the terms can be extracted in a semi-automatic way.

5.1.4. Mother Terms and the Generation of Single-word (Compound) Terms — Stems / Roots

When one studies the first 25 unique *terms* on the KeyWord list, one sees that the core terminology of the special-field at hand has been identified. To use Ahmad and Rogers' (2001: 742) terminology, one could say that these terms are "the 'mother' terms of a given specialism or the signature terms of a specialist subject". Like mothers, they can generate other terms through *compounding*, or whatever term formation process is valid for the language at hand.

By making use of the Concord tool of WST, single-word (compound) terms built up around the stem or root of any mother term can quickly be identified. For example, if the stem of the key word **tlhalošo** 'meaning', i.e. **tlhaloš-**, is used as the search node in concordance lines, the terms shown in (3) are but a few of the terms generated by the mother term.

- (3) A sample of the single-word compound terms generated by the stem of the mother term **tlhalošo**

swana le theksi, bese, setimela, paesekele, mmotoro, bj. bj. **Tlhaloš-** **-okamanyi** e dira gore mantšu a a amanego ka tlhalošo a lebopi la mathomo mo lentšung, bj. bj. Re hlalošitše gore **tlhaloš-** **-okatološo** ga e kwešišege ge lentšu le eme le le nnoši ka ge lentšu, di bitšwa tlhalošokelello ya lentšu leo. Lemoga ge **tlhaloš-** **-okelello** e fapana go ya ka batho, ka ge e theilwe godimo ga

The first two terms in (3), namely **tlhalošokamanyi** 'associative meaning' and **tlhalošokatološo** 'extended meaning', were also marked manually, but the third term **tlhalošokelello** 'cognitive meaning' was only found computationally. By doing a search with the stem and/or root of each of the top 25 mother terms as search item, all (compound) terms containing those stems and/or roots can be identified. The Search function furthermore allows for automatic exclusion of certain items, thus further refining the search procedure. Although this procedure is less automatic than the KeyWord tool, it is still a rather swift

way to detect terms that were not included in the KeyWord list. Column 5 in (4) lists the number of those (new and extra) single-word terms each mother term generated.

- (4) Generation of *new* and *extra* (compound) terms from the first 25 unique mother terms on the KeyWord list (new = not listed by KeyWord; extra = missed during manual excerption and not listed by KeyWord)

N	Term	Translation	Stem / Root	New + Extra single-word terms	Extra multi-word terms
1	lediri	verb	-dir-	6 + 3 = 9	0 + 1 = 1
2	lefoko	sentence	-fok-	2 + 2 = 4	1 + 2 = 3
3	mantšu	(linguistic) words	-ntšu	3 + 1 = 4	3 + 3 = 6
4	tlhalošo	meaning	tlhaloš-	3 + 1 = 4	0 + 1 = 1
5	serewa	topic	-rew-	0	0
6	popego	morphology, form, structure	-popeg-	1 + 0 = 1	0
7	mmoledi	first person	-boled- / -moled-	0 + 1 = 1	0
8	kgokagano	discourse	kgokagan-	0	1 + 2 = 3
9	legoro	(noun) class	-goro	2 + 2 = 4	0
10	togaganyo	cohesion	togagan- / logagan-	0	0
11	legokedi	agreement morpheme	-kgoked-	0	0
12	tswalano	relationship, association	tswalan-	0	1 + 0 = 1
13	tiro	predicate, action, process	tir-	1 + 1 = 2	0 + 1 = 1
14	mmoledišwa	second person, addressee	bolediš- / -molediš-	0	0 + 3 = 3
15	leina	noun	-ina	3 + 1 = 4	0 + 1 = 1
16	lethuši	auxiliary verb	-thuš-	0	0
17	lešala	pronoun	-šal-	3 + 0 = 3	0
18	lebopi	morpheme	-bop-	1 + 0 = 1	1 + 1 = 2
19	lehlathi	adverb	-hlath-	1 + 0 = 1	0
20	kganetšo	negation, negative	-ganetš-	0	0 + 1 = 1
21	lereo	term	-reo	0	0
22	makopanyi	conjunctions	-kopany-	4 + 1 = 5	0
23	kamano	(inter)relationship	-aman-	0	0
24	matlema	prepositions	-tlem-	1 + 2 = 3	0
25	kgatelelo	emphasis	-kgatelel-	0	0
Σ				31 + 15 = 46	7 + 16 = 23

In all, the top 25 mother terms in (4) generated 31 *new* single-word terms (i.e. terms that were not listed by KeyWord) and 15 *extra* single-word terms (i.e. terms that were missed during the manual excerption, and were not picked up by KeyWord either). For example, the root **-tlem-** of the mother term **matlema** 'prepositions', generated 1 new term (**tlemagano** 'cohesion') and 2 extra terms (**tlemagantšha** 'link, connect' and **tlemaganya** 'link, connect'), as shown in (5).

- (5) New and extra single-word terms generated by the root of the mother term **matlema**

mo temaneng ka ge bobedi bja ona bo hlola kgokagano goba **-tlem-** **-agano** gare ga mantšu, dikafoko le mafoko. Ke go re
 mo nomorong ye: (vii). Le tlemagantšha mantšu a fe? Le **-tlem-** **-agantšha** melamo le marumo. Go bonala ka eng ge e le
 ke lefe? Ngwala nomoro ya maleba ... Le **-tlem-** **-aganya** mantšu a fe? ... le ...

As far as the single-word terms are concerned, the KeyWord tool followed by the Concord tool for just the top 25 unique KeyWord terms, throws up as many

as 189 + 18 and 31 + 15 single-word terms respectively, or thus 253 terms in all. With a benchmark of 342 items, this thus means that 74% of the single-word terms were extracted semi-automatically.

5.2. Semi-automatic Extraction of Multi-word Terms

Up to this point, the investigation has centred around single-word terms only. Terms often consist of multi-word units — in the list of manually excerpted terms, 41 were multi-word units — and any computational term-extraction process should also be able to isolate multi-word terms. For Northern Sotho, as for any other language, the computational identification of multi-word terms is many times more complex than the identification of single-word terms. Given the fact that we are working with raw corpora, the extraction will still be purely statistical, yet linguistics will have to come into play to make more informed decisions as to the term status of the computational suggestions. The need for this linguistic, and thus language-dependent, support will be apparent in the discussion below.

5.2.1. Top Ranks in a Frequency List

With the WordList tool it is possible to make multi-word wordlists, i.e. 2-word wordlists, 3-word wordlists, 4-word wordlists, etc. up to 8-word wordlists. From the manual excerption, one knows that at least three 2-word terms, thirty-five 3-word terms, two 4-word terms, and one 5-word term — totalling 41 multi-word terms in all — are to be extracted semi-automatically. Reading through the first few hundred items in each of those multi-word wordlists, quickly indicates that this process does not produce very significant results. It was thus decided to immediately move to the KeyWord process.

5.2.2. Key Words

Using the KeyWord tool on multi-word level is analogous to using it on single-word level. Multi-word wordlists for the linguistic texts were already compiled in the previous phase (par. 5.2.1), so only multi-word reference wordlists (based on the 5.2-million-word reference corpus) had to be compiled in addition. Once done, KeyWord was requested to calculate the 2-word key words, the 3-word key words, etc.

The two manually identified 2-word terms were listed in the 2-word KeyWord list, while no extra terms were found. The recall is thus 100% for the 2-word terms. 18 of the 35 manually excerpted 3-word terms were listed in the 3-word KeyWord list (recall = 51%), together with 5 extra 3-word terms. The two 4-word terms were listed in the 4-word KeyWord list (recall = 100%), together with 1 extra 4-word term. Finally, the sole manually identified 5-word term was not listed in the 5-word KeyWord list (recall = 0%), yet 2 extra 5-word terms were identified instead. Precision values are extremely low in all cases.

Taken together, 22 multi-word terms (out of 41, recall = 54%) were thrown up semi-automatically, and a surprisingly high number of 8 extra multi-word terms were discovered. Together with the extras that were additionally extracted with the methods described below (in the last section of par. 5.2.3 and in par. 5.2.4), the benchmark for the multi-word terms rises to 72. This simply means that over 40% (31 out of 72) of the multi-word terms were missed during the manual pass — confirming not only the value of a computational approach, but also pointing at the difficulty of a manual excerption of especially multi-word terms from running text.

5.2.3. Mother Terms and the Generation of Multi-word (Compound) Terms — KWIC Lines

Key-word-in-context (KWIC) searches

One simple method which could be used to isolate the multi-word terms that were missed with KeyWord, is to use single-word mother terms in key-word-in-context (KWIC) searches. Such searches will reveal the collocations in which these terms are involved, which often turn out to be multi-word terms. When used for terminological purposes, this procedure is however not without its problems. The purpose of a concordance is primarily to identify collocations, i.e. — to use Scott's (1999: WST help) metaphor — to provide 'information on the company words keep'. It cannot simply be assumed that all collocations showing up in concordance lines are multi-word terms. From a multi-word perspective, KWIC lines tend to over-generate, producing false positives as well as true positives. In this regard Heid (2001: 791) states that "[t]he relationship and the borderline between collocations [...] and 'multiword terms' is not easy to describe". The combination of a term and its collocate(s) seems to be a linguistic issue rather than a terminological one. From the existing literature, it would seem that there are two basic principles which provide guidance as to the distinction between collocations and multi-word terms.

Term : Collocate(s) combinations and The denomination of new concepts

In the first instance, the terminological status of the *term : collocate(s)* combination depends on whether the combination of a term and its collocate(s) can be seen as the denomination of a new concept in its own right. If this is the case, such a collocational combination will qualify as a multi-word term; if not, it would be described as a false positive. In some cases, false positives in a concordance are quite obvious and easily identifiable, whereas others seem to be on the borderline between multi-word terms on the one hand, and simple collocations on the other. The concordance for the word **lediri** 'verb' serves as an example. There are 391 instances of **lediri** in the linguistic texts. With WST's concordance function, Concord, a concordance line can be drawn up for each, and one finds 15 incidences of **lethuši le lediri** 'auxiliary verb and verb'. This

collocation is obviously a false positive and cannot be regarded as a multi-word term, since it does not comply with the basic requirement for multi-word term status, i.e. the combination of the term **lediri** 'verb' and its collocate **lethuši (le)** 'auxiliary verb (and)' does not refer to a single and/or new concept. Other cases are, however, more problematic: 26 incidences of the combination **modirišo wa lediri** 'mood of the verb' were found — the collocate **modirišo (wa)** 'mood (of)' having the highest co-occurrence frequency with the term **lediri**. Even with the guidance of the principle of 'denomination of a new concept', it is difficult to decide whether **modirišo wa lediri** is to be regarded as a multi-word term, or whether it simply indicates the company which **lediri** typically keeps.

Term : Collocate(s) combinations and Lexical gaps across languages

A second principle which could be useful for the terminologist trying to decide on the terminological status of collocations, is to ascertain whether the combination of term and collocate(s) in one language fills a lexical gap when compared to a dominant language such as English; in other words, does the term and its collocate(s) have a *term* as translation equivalent in another language? This might lead the terminologist to decide that **modirišo wa lediri** 'mood of the verb' should not be regarded as a multi-word term, since its equivalent in English does not represent a term. On the other hand, a collocation such as **tatelano ya mantšu** 'word order' (where **tatelano (ya)** 'order (of)' was found as collocate for the term **mantšu** 'words') may then be regarded as a multi-word term, since it does: (a) refer to a concept in its own right, and (b) have a term, i.e. 'word order', as an equivalent in English. The downside of this principle is of course that in some cases it simply shifts the decision as regards the terminological status of a multi-word unit from one language to another.

Language-specific guidelines

It is clear that the two principles as formulated above are not sufficient as regards the distinction between collocations and multi-word terms. It is therefore understandable that Heid (2001) recognises the need for the establishment of some internal guideline as regards the distinction between collocations and multi-word terms in any given language. He does this by first identifying frequent collocational patterns appearing in the Indo-European languages, and classifies these patterns according to the part-of-speech category to which the two lexical items making up the collocation belong. Using the structure of these recurrent collocational patterns as basis, he attempts to formulate an internal and language-specific guideline in order to distinguish between mere collocational patterns and true multi-word terms. The syntactic categories which he distinguishes for Indo-European languages are as follows: (a) noun + verb, (b) noun + adjective, (c) noun + noun, (d) verb + adverb, and (e) adjective + adverb. For terminological purposes, only the first three categories (a)–(c) seem to

be of importance, since they are much more frequent in special-field languages than (d) and (e).

With specific reference to German, Heid (2001: 791) mentions the possibility of classifying sub-type denoting noun-adjective collocations as multi-word terms, and noun-verb collocations simply as combinatory properties of the nominal term. He uses the following noun-adjective type collocations taken from maintenance literature for automobiles as illustrative examples: (a) **pneumatische Leuchtweitenregulierung**, (b) **elektrische Leuchtweitenregulierung**, and (c) **automatische Leuchtweitenregulierung**. These combinations are regarded as being sub-type denoting noun-adjective collocations and each should therefore be regarded as a multi-word term. The adjectives **pneumatische** 'pneumatic', **elektrische** 'electric' and **automatische** 'automatic' denote sub-types of **Leuchtweitenregulierung** 'light-distance regulation' in that they define or describe different types of light-distance regulation. A noun-verb collocation such as **Parameter festlegen** 'fix a parameter' (Heid 2001: 790) would then not be regarded as a multi-word term — **festlegen** 'fix' would simply be regarded as a combinatory property of the term **Parameter** 'parameter'.

Internal guidelines for Northern Sotho

The point that needs to be emphasised at this stage is that no attempt has hitherto been made to formulate internal guidelines for Northern Sotho that would enable terminologists to distinguish mere collocations from multi-word terms. Since such internal guidelines are dependent on the structure of frequently occurring collocations, a preliminary investigation was consequently done by making use of Concord in order to identify the typical collocational *patterns* that are found in Northern Sotho concordance lines. (Note that it is only on this level that Sewangi (2000, 2001), see par. 5.1, starts to use the power of the tagged Swahili corpora he had at his disposal. In his approach a 'pattern matching program' semi-automatically extracts potential multi-word terms, using *pre-defined* term-formation patterns.)

Within WST, a concordance can be drawn up from either a KeyWord list or a WordList list. The output of such a query can then be sorted in a multitude of ways so as to see different collocational patterns emerge. In the case of English, sorting the output to the left shows up modifier-head patterns where the mother term is the *head*, e.g. **infinitive verb**, **singular verb**, **transitive verb**, etc., whereas sorting to the right results in modifier-head patterns in which the mother term is the *modifier*, e.g. **verb phrase**, **verb root**, etc. These patterns are language-specific and will have to be formulated for any particular language in which any such investigation is to be done.

For the purpose of the current discussion, the term **mantšu** 'words', which appears high up in both the keyness-ordered KeyWord list and the frequency-ordered WordList list, can be used as a search term. Concord displays 392 concordance lines for **mantšu**. By sorting to the left (primary sort on *L1*, i.e. first position to the left of the search item, and secondary sort on *L2*, i.e. second

position to the left) two typical patterns can be identified. The pattern which occurs most frequently is the noun-determiner pattern, in which the search item **mantšu** is the nucleus of the determiner. Compare the examples shown in (6), taken from Concord.

- (6) Noun-determiner pattern, with the search item the nucleus of the determiner

Thutišo ye e leka go go lemoša bohlokwa bja *tatelano* ya **mantšu** mo lefokong. Go ya ka melao ya tatelanontšu ya Sesotho sa ke gore lefoko le lengwe le le lengwe le na le *tatelanotseo* ya **mantšu**. Gape re tlišitše taba ya go re lentšu la mathomo mo lefokong, mmadi goba mmoledišwa mo kgokaganong, ka *ge tlhalošo* ya **mantšu** a mohuta wo e ka ba e na le maphakga a mantši. Eupša,

The second most frequent pattern is the verb-noun pattern in which the search item typically appears as the object of a verb. Examples are shown in (7).

- (7) Verb-noun pattern, with the search item typically as the object of a verb

mokgwa woo. Ge batho ba boledišana, ba swanetše *go kgetha* **mantšu** le go bopa mafoko semeetseng. Gantši mmoledi o timelelwa (iii) 3. Bopa mafoko a mahlano ka *go šomiša* **mantšu** a mangwe bakeng sa mantšu a a dirišitšwego mo lefokong le: di a fapanā? Na e ka ba ke eng se se dirago gore *ba tswalanye* **mantšu** a a filwego le ditlhalošokelello tše di fapanego? MODIRO 3

When sorting takes place to the right of the search word, the typical pattern which emerges is again a noun-determiner pattern, but in this case the search item forms the nucleus of the nominal part of the combination. This can be seen from the examples in (8).

- (8) Noun-determiner pattern, with the search item the nucleus of the nominal part

e hlolwa ke tlhalošo ya deiktiki. Gore re kwešiše taba ye ya **mantšu** a *deiktiki* gabotse, a re fetelele poledišano gare ga mmoledi go šetša molao wo, gomme a ka se kgone go tswalanya **mantšu** a *lefoko* ka boikgethelo. Ge mmoledi goba mongwadi a ka

A pattern which also frequently occurs, is the noun-verb pattern, but in this instance, the search term appears as the subject of the verb. Compare (9) in this regard.

- (9) Noun-verb pattern, with the search item typically as the subject of a verb

a e šomišago e šašarakane. 2.3.4 Tswalano ka tlhalošo Ka ge **mantšu** a *fapana* ka tlhalošo, ga a šome go swana mo lefokong. Ge re mantšu ka gona ge re bopa dikafoko goba mafoko. Ge **mantšu** a *hlatlamantšhwa* mo lefokong, tthatlamano (tatelano) yeo e , o tšweletšwa ka mafoko ao a bopilwego ka mantšu, gomme **mantšu** a *latelana* ka tsela ye e itšego mo mafokong. Ge re šetša

With regard to Northern Sotho, the principles regarding the term status of certain collocational patterns as formulated by Heid, would for example imply that the collocations **lediri la modirišopego** 'indicative verb', **lediri la modirišopegotlhaodi** 'participial / situative verb' and **lediri la modirišogore** 'subjunctive verb' should all be regarded as multi-word terms, since the function of the determiners **la modirišopego**, **la modirišopegotlhaodi**, etc. is similar to that of the adjectives cited in Heid's example, i.e. to denote sub-types of **lediri**. A general principle for Northern Sotho could thus be to regard sub-type denoting noun-determiner collocations as multi-word terms, provided of course that these

collocations also meet the requirement of referring to an independent concept. Such an approach seems to make sense from a theoretical point of view, but the practical implications it would have for the full-scale compilation of a terminology list and/or an LSP dictionary remain to be seen.

If the principle to *regard noun-verb collocations as combinatory properties of the noun* is applied to one of the collocations found for **lediri**, viz. **lediri le tšwelela ...** 'a/the verb generates ...', this would imply that this combination should thus not be regarded as a multi-word term, which seems sound. One could however assume that in a dictionary on linguistic terms, the frequent combination of **lediri** 'verb' and **tšwelela** 'generate' should in some way or other be reflected in the article of **lediri**. This argument is further supported by the statistical analysis of collocates, which can be done computationally from the concordance given for any search item. The statistical analysis of collocates gives an indication of the significance of certain items appearing within a specified span or horizon of the node term. A calculation of the collocates of **lediri** 'verb' indicates that the verb **tšwelela** 'generate' is the verb which appears most often (26 times) as a collocate of the noun **lediri**. Therefore, even though the combination **lediri le tšwelela ...** 'a/the verb generates ...' cannot be regarded as a multi-word term, it is clear that it should be treated in a more detailed terminology list and/or LSP dictionary, either as a combinatory property of the noun, or, alternatively, used in an explanatory example in the microstructural treatment of the specific noun.

It should however be kept in mind that the two guidelines formulated above, i.e. to regard sub-type denoting noun-determiner collocations as potential multi-word terms and noun-verb collocations as combinatory properties of the noun in Northern Sotho, should at this stage be seen as suggestions, since the formulation of such guidelines presupposes an in-depth study of the linguistic structure of collocations in Northern Sotho — an endeavour which falls outside the scope of the current article.

Semi-automatic generation of multi-word (compound) terms

The results of the above discussion can now be used to semi-automatically generate multi-word terms by means of KWIC lines where the node is a single-word term. Recall that for the generation of single-word (compound) terms (par. 5.1.4), the stems / roots of the top 25 unique KeyWord mother terms were used. Those same 25 mother terms can now function as node terms for the KWIC lines. For comparison purposes, the results are also shown in (4) above.

From (4) one sees that, compared to the various multi-word terms thrown up by the multi-word KeyWord lists, the 25 mother terms generated 7 multi-word terms that had also been marked manually, but in addition also an astonishing 16 multi-word terms that were missed manually (and were not picked up by KeyWord either). Together, KeyWord and Concord for the top 25 unique single-word mother terms, thus throw up 22 + 8 and 7 + 16 multi-word terms respectively, or thus 53 terms in all. With a benchmark of 41 + 8 + 16 =

65, this means that 82% of the multi-word terms (known so far) were extracted semi-automatically.

5.2.4. Clusters

If kept in mind that the multi-term words are the hardest to retrieve, a recall of 82% seems exceptional. Nonetheless, besides the unsuccessful study of the multi-word top ranks on the one hand, and the very successful study of KeyWord and Concord on the other, a fourth approach to extract even more multi-words was experimented with, viz. the Cluster function. In the words of Scott (1999: WST help):

Clusters are words which are found repeatedly in each others' company. They represent a tighter relationship than collocates, more like groups or phrases (but I call them clusters because these terms already have uses in grammar).

Clusters can be identified from either a wordlist or a concordance, the difference being that Concord only processes concordance lines, whereas WordList processes whole texts. As we are primarily interested in clusters containing the search item itself, we opted for Clusters in Concord. The 2-, 3-, 4- and 5-word clusters for each of the top 25 unique single-word KeyWord mothers were calculated. The outcome of this Cluster procedure was truly surprising, as yet another extra 7 multi-word terms, shown in (10), were extracted.

- (10) Generation of *extra* multi-word linguistic terms with the Cluster function (extra = not marked manually, and not found with KeyWord nor KWIC lines)

mmoledišwa ka botee second person singular
peakanyo ya mantšu arrangement of words, word arrangement
popogo ya lediri morphology of the verb, verbal morphology
tlhalošo ya lefoko sentence meaning
tlhalošo ya lentšu word meaning
tlhaloštheo ya lefoko basic sentence meaning
tumanoši ya mafelelo ya lediri verbal ending (lit. final vowel of the verb)

None of the multi-word terms listed in (10) had been marked in the manual pass, nor were they picked up with KeyWord or the study of KWIC lines. With these 7 extras, the recall for the multi-word terms becomes 60 out of 72, which thus means that as many as 83% of the multi-word terms were extracted semi-automatically!

6. Outlook: Manual Excerption *versus* Semi-automatic Extraction of Terminology

A very significant result of the research presented above is the fact that over

half the multi-word terms identified by computational means (8 + 16 + 7 = 31, out of 60) were overlooked during the process of manual excerption. The manual excerption of multi-word terms therefore merits some further discussion. Identification of multi-word terms is a problematic issue for terminologists, since the terminological status of especially multi-word units is not always clear-cut.⁹ The process of manual term excerption is subject to the influence of a number of factors, and rightly so. According to the principles of good terminological practice, term excerption cannot be done without taking into account practical issues such as *inter alia* the potential target user and that user's encyclopaedic knowledge regarding the specific subject field. Taking the target user into account, the terminologist might argue that the conceptual meaning of a multi-word unit can be derived from the meaning of its constituent parts and that it is therefore unnecessary to list the particular unit as a term. For a different target user, the terminologist might deem it necessary to indeed list the multi-word unit as a term. An unexpected, and certainly underestimated, value of computational term extraction therefore seems to lie in its ability to *identify* term candidates. Taking all relevant information as regards target user, etc. into account, the terminologist can then make an informed decision as to whether a particular multi-word unit should be included as a multi-word term or not.

Apart from the research reported on above, experiments were also done with other concepts that were introduced by Scott (1997), viz. *key key words* (i.e. items that are key across numerous texts), *associates* (i.e. items that are key in the same texts as a given key key word), and *clumps* (i.e. co-occurring associates). With these concepts (and their computational implementations) we hoped to be able to automatically group terms into sub-fields. So far, however, the results of these attempts have not been satisfying.

7. In Conclusion

The main aim of this article was to discover whether or not the semi-automatic extraction of terminology from *untagged* and *unmarked* running text by means of *basic* corpus query software would be feasible for the African languages. In order to answer this question a full-blown case study revolving around Northern Sotho linguistic texts was undertaken. Upon comparison of the manual outcome with the computational results, it was found that 74% of the single-word linguistic terms, and an astonishing 83% of the multi-word linguistic terms could indeed be extracted semi-automatically. These high figures were obtained with basically just three software tools: WordList, KeyWord and Concord, all part of WordSmith Tools (Scott 1999). Based on this case study one is thus bound to conclude that the semi-automatic extraction of terms for the African languages is indeed a *viable* endeavour.

It was also pointed out that human beings will always remain the final judges in any terminological activity, whether that endeavour be manual or computational. The terms proffered by the software will always need to be

scrutinised by the terminologist. Conversely, however, the research revealed rather surprisingly that the software can *isolate* potential terms and *force* the terminologist to *consider* term status in ways that are less obvious when wading manually through running text. This turned out to be especially valid for multi-word terms, as more than 40% of the multi-word linguistic terms were seemingly missed during manual excerption. Viewed from this angle, the semi-automatic extraction of terms for the African languages is not only viable, but even *crucial* in order to counteract inevitable human errors.

Finally, at the start of this article we pointed out that a terminographical approach to terminology would enable the computational management of terminology. Terminography, whether pursued manually or computationally, always has the creation of a terminology list, a term bank or a dictionary for special purposes as its final objective. The various outcomes of the research presented in this article are therefore summarised in a tiny special-field *lexicon*. This lexicon is the Linguistics Terminology presented in the Appendix, in which all the terms that were retrieved in the current study are listed in alphabetical order and in their *lemmatised* form. In addition, the linguistic terms that were only excerpted manually are preceded by ☞, those that were only extracted computationally are preceded by ☛, and the ones that were retrieved both manually and computationally are marked with ☞☛. There are respectively 98, 50 and 187 of them. This means that, out of the 335 lemmatised terms, 285 or 85% were excerpted manually, and 237 or 71% were extracted computationally. The difference between the two approaches (14%) is smaller than the number of items not retrieved in either approach. There can thus be no doubt that, when looking at the end product, semi-automatically extracting terminology for and in the African languages is indeed a worthwhile venture.

Endnotes

1. Since this article is being submitted for publication in a South African journal, necessary sensitivity with regard to the term 'Bantu' languages is exercised in our choice rather to use the term *African* languages. Keep in mind, however, that the latter includes more than just the 'Bantu Language Family'.
2. General-language corpora for all South African languages have indeed been built at the Department of African Languages of the University of Pretoria. The sizes of these corpora are in constant evolution. For the latest developments, we would therefore like to refer the reader to the home page of ELC for ALL (Electronic Corpora for African-Language Linguistics): <http://www.up.ac.za/academic/libarts/afrilang/elcforall.htm>
3. For more information on WordSmith Tools, we would like to refer the reader to the home page of Mike Scott, the creator of the software: <http://www.lexically.net> (or else: <http://www.liv.ac.uk/~ms2928>).
4. The term 'keyness' too remains undefined. Scott and Thompson (2001a: 109) state rather vaguely that "[k]eyness relates to the frequency of particular lexical items within a text as compared with their frequency in a reference corpus". Unfortunately, even under the head-

ing 'definition of keyness' in WST's help section (Scott 1999), 'key words' are discussed and not 'keyness'.




5. The text on the new South African coat of arms can be found at: http://www.gov.za/symbols/coa_sepedi.htm
6. Compare in this respect Prinsloo and De Schryver (2001) which deals extensively with corpus-stability issues, with special reference to Northern Sotho and Tsonga.
7. The spelling of the adjective **seswa** is incorrect in the CoA text. The correct orthography is **sefsa**, but we have chosen to keep the spelling as it appears on the cited web page. The same goes for **leswa** which should be **lefsa**.
8. Due to the morphology of Northern Sotho the subject concord and/or demonstrative **se**, which refers to **sefoka** 'coat of arms', is also thrown up by KeyWord but does of course not appear as such in the English-language paraphrase.
9. This is of course also true of single-word terms, but in this particular instance the relatively high correspondence between the manually excerpted single-word terms and those extracted by semi-automatic means, indicates that manual identification of single-word terms is less of a problem for terminologists.

References



- Ahmad, Khurshid and Margaret Rogers.** 2001. Corpus Linguistics and Terminology Extraction. Wright, Sue Ellen and Gerhard Budin (Eds.). 2001: 725-760.
- Alberts, Mariëtta.** 2000. Terminology Management at the National Language Service. *Lexikos* 10: 234-251.
- De Schryver, Gilles-Maurice and B. Lepota.** 2001. The Lexicographic Treatment of Days in Sepedi, or When Mother-Tongue Intuition Fails. *Lexikos* 11: 1-37.
- De Schryver, Gilles-Maurice and D.J. Prinsloo.** 2000. 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.
- Dunning, Ted.** 1993. Accurate Methods for the Statistics of Surprise and Coincidence. *Computational Linguistics* 19(1): 61-74.
- Heid, Ulrich.** 2001. Collocations in Sublanguage Texts: Extraction from Corpora. Wright, Sue Ellen and Gerhard Budin (Eds.). 2001: 788-808.
- Hurskainen, Arvi.** 1992. A Two-Level Computer Formalism for the Analysis of Bantu Morphology. An Application to Swahili. *Nordic Journal of African Studies* 1(1): 87-122.
- Hurskainen, Arvi.** 1995. Information Retrieval and Two-directional Word Formation. *Nordic Journal of African Studies* 4(2): 81-92.
- Hurskainen, Arvi.** 1996. Disambiguation of Morphological Analysis in Bantu Languages. *COLING-96. The 16th International Conference on Computational Linguistics, Center for Sprogteknologi, Copenhagen, Denmark, August 5-9, 1996. Proceedings, Volume 1: 568-573.*
- Hurskainen, Arvi.** 1999. SALAMA. Swahili Language Manager. *Nordic Journal of African Studies* 8(2): 139-157.
- Hurskainen, Arvi and Riikka Halme.** 2001. Mapping between Disjoining and Conjoining Writing Systems in Bantu Languages: Implementation on Kwanyama. *Nordic Journal of African Studies* 10(3): 399-414.

- Prinsloo, D.J. and Gilles-Maurice de Schryver.** 2001. 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: 85-129.
- Sager, Juan C.** 1990. *A Practical Course in Terminology Processing*. Amsterdam: John Benjamins.
- Scott, Mike.** 1997. PC Analysis of Key Words — and Key Key Words. *System* 25(1): 233-245.
- Scott, Mike.** 1997a. The Right Word in the Right Place: Key Word Associates in Two Languages. *AAA — Arbeiten aus Anglistik und Amerikanistik* 22(2): 235-248.
- Scott, Mike.** 1999. *WordSmith Tools version 3*. Oxford: Oxford University Press. See for this software also: <http://www.lexically.net/wordsmith/index.html>.
- Scott, Mike.** 2000. Reverberations of an Echo. Lewandowska-Tomaszczyk, Barbara and Patrick J. Melia (Eds.). 2000. *PALC'99: Practical Applications in Language Corpora. Papers from the International Conference at the University of Lodz, 15-18 April 1999*: 49-65. *Lodz Studies in Language* 1. Frankfurt am Main: Peter Lang.
- Scott, Mike.** 2000a. Focusing on the Text and its Key Words. Burnard, Lou and Tony McEnery (Eds.). 2000. *Rethinking Language Pedagogy from a Corpus Perspective. Papers from the Third International Conference on Teaching and Language Corpora*: 103-121. *Lodz Studies in Language* 2. Frankfurt am Main: Peter Lang.
- Scott, Mike.** 2001. Mapping Key Words to *problem* and *solution*. Scott, Mike and Geoff Thompson (Eds.). 2001: 109-127.
- Scott, Mike and Geoff Thompson (Eds.).** 2001. *Patterns of Text: In Honour of Michael Hoey*. Amsterdam: John Benjamins.
- Scott, Mike and Geoff Thompson.** 2001a. Editors' Introduction. Scott, Mike and Geoff Thompson (Eds.). 2001: 109-110.
- Sewangi, Selemán S.** 2000. Tapping the Neglected Resource in Kiswahili Terminology: Automatic Compilation of the Domain-Specific Terms from Corpus. *Nordic Journal of African Studies* 9(2): 60-84.
- Sewangi, Selemán S.** 2001. *Computer-assisted Extraction of Terms in Specific Domains: The Case of Swahili*. Academic Dissertation, December 2001. Helsinki: University of Helsinki, Faculty of Arts, Institute of Asian and African Studies. Also available at: <http://ethesis.helsinki.fi/julkaisut/hum/aasia/vk/sewangi/>.
- Strehlow, Richard A.** 2001. The Role of Terminology in Retrieving Information. Wright, Sue Ellen and Gerhard Budin (Eds.). 2001: 426-441.
- Wright, Sue Ellen and Gerhard Budin (Eds.).** 2001. *Handbook of Terminology Management. Volume 2. Application-Oriented Terminology Management*. Amsterdam: John Benjamins.








Appendix: Manually-excerpted and/or Semi-automatically Extracted Linguistics Terminology

- Legend:  = only excerpted manually
 = only extracted computationally
 = retrieved both manually and computationally
x/y = gender (singular and plural class) of noun
v. = verb
⇒ = cross-reference to synonym(s) and/or variant(s)


A

-  **-a- ya lebjaletelelele** 9/- imperfect tense -a-
 **amana** v. relate



B

-  **boemo** 14/6 position
 **boemo bja motheo** 14/6 basic position ⇒ **boemotheo**
 **boemotheo** 14/6 basic position ⇒ **boemo bja motheo**
 **boikakanyetšo** 14/6 idea, thought, concept
 **boinaganelo** 14/6 consciousness, imagination, pre-knowledge
 **bontši** 14/- plural
 **botee** 14/- singular


D

-  **deiktiki** 9/- deictic





F

-  **fonetiki** 9/- phonetics
 **fonolotši** 9/- phonology




G

-  **gatelela** v. emphasise

H

-  **hlogo** 9/10 prefix
 **hlogo ya leina** 9/10 nominal prefix, noun prefix ⇒ **hlogoina**
 **hlogoina** 9/10 nominal prefix, noun prefix ⇒ **hlogo ya leina**
 **hlopha** v. categorise, arrange

K

-  **kago** 9/10 structure
 **kamano** 9/10 (inter)relationship
 **kamanotlhalošo** 9/10 semantic relationship

-  **karolo ya lefoko** 9/10 sentence part ⇒ **karolofoko**
 **karolofoko** 9/10 sentence part ⇒ **karolo ya lefoko**
 **karolopolelo** 9/10 word class, part of speech ⇒ **legorontšu**
 **kganetšo** 9/10 negation, negative
 **kgatelelo** 9/10 emphasis
 **kgetho ya mantšu** 9/10 word choice
 **kgoeletšo** 9/10 exclamation
 **kgokagano** 9/10 discourse
 **kgokagano ka go ngwala** 9/10 written discourse
 **kgokagano ka molomo** 9/10 spoken / oral discourse
 **kgokagano ya linkwistiki** 9/10 linguistic discourse / communication
 **kgokaganya** v. connect, link
 **kgopolotheo** 9/10 main idea
 **khuduego** 9/10 excitement, enthusiasm
 **khutlo** 9/10 full stop
 **khutsofatša** v. abbreviate
 **khutsofatšo** 9/10 abbreviation
 **kopanyo ya mantšu** 9/10 combination of words, word combination
 **kutu** 9/10 stem
 **kutu ya lediri** 9/10 verb stem ⇒ **kutudiri**
 **kutudiri** 9/10 verb stem ⇒ **kutu ya lediri**
 **kwagatšo** 9/10 sound production
 **kwana** v. agree
 **kwano** 9/10 agreement
 **kwano ka popego** 9/10 morphological agreement

L

- 📖 **leadingwa** 5/6 adoptive, loan word
 🗣️ **leamanyi** 5/6 relative
 🗣️ **leamanyidiri** 5/6 verbal relative, relative verb
 🗣️ **leamanyi-ina** 5/6 nominal relative
 📖 **learogi** 5/6 exception
 📖 **leba** 5/6 copula(tive)
 📖 **lebad** 5/6 numeral
 📖 **lebadiri** 5/6 copulative verb
 📖 **lebahlogwana** 5/6 copulative prefix
 🗣️ **lebaka la lefetile** 5/- past tense ⇒ **lefetile**
 📖 **lebaka la letlago** 5/- future tense ⇒ **letlago**
 🗣️ **lebjale** 5/- present tense
 📖 **lebjale le lekopana** 5/- short present tense ⇒ **lebjalekopana**
 📖 **lebjale le letelele** 5/- long present tense ⇒ **lebjaletelele**
 📖 **lebjalekopana** 5/- short present tense ⇒ **lebjale le lekopana**
 🗣️ **lebjaletelele** 5/- long present tense ⇒ **lebjale le letelele**
 🗣️ **lebopi** 5/6 morpheme
 📖 **lebopi la kganetšo** 5/6 negative morpheme ⇒ **lebopikganetši**
 🗣️ **lebopi la lebaka la letlago** 5/6 future tense morpheme
 📖 **lebopi la lebjaletelele** 5/6 present tense morpheme
 🗣️ **lebopikganetši** 5/6 negative morpheme ⇒ **lebopi la kganetšo**
 📖 **lebotšiši** 5/6 interrogative
 🗣️ **lediri** 5/6 verb
 📖 **lediri la modirišogore** 5/6 subjunctive verb
 📖 **lediri la modirišopego** 5/6 indicative verb
 📖 **lediri la modirišopegotlhaodi** 5/6 participial / situative verb
 📖 **lediri la modirišotaelo** 5/6 imperative verb
 🗣️ **ledirifelopedi** 5/6 verb which can combine with a subject and one object, two-place verb
 🗣️ **ledirifelotharo** 5/6 verb which can combine with a subject and two objects, three-place verb
 🗣️ **leekiši** 5/6 ideophone
 🗣️ **lefeledi** 5/6 intransitive verb
 🗣️ **lefetedi** 5/6 transitive verb
 🗣️ **lefetile** 5/- past tense ⇒ **lebaka la lefetile**
 🗣️ **lefoko** 5/6 sentence
 📖 **lefoko le le feleletšego** 5/6 full sentence
 🗣️ **lefokofoko** 5/6 complete sentence
 🗣️ **lefokofokwana** 5/6 complex sentence ⇒ **lefokontši**
 📖 **lefokofokwanapego** 5/6 complex declarative sentence
 🗣️ **lefokonolo** 5/6 basic / simple sentence ⇒ **lefokotho**
 🗣️ **lefokontši** 5/6 complex sentence ⇒ **lefokofokwana**
 🗣️ **lefokotho** 5/6 basic / simple sentence ⇒ **lefokonolo**
 📖 **lefokothwi** 5/6 direct speech
 🗣️ **legoro** 5/6 (noun) class
 🗣️ **legorofelo** 5/6 locative noun class
 🗣️ **legoroina** 5/6 noun class
 🗣️ **legorontšu** 5/6 word class, part of speech ⇒ **karolopolelo**
 🗣️ **lehlalošagotee** 5/6 synonym ⇒ **lehlalošetšagotee**
 🗣️ **lehlalošantši** 5/6 polysemous word ⇒ **lehlalošetšagontši**
 🗣️ **lehlalošetšagontši** 5/6 polysemous word ⇒ **lehlalošantši**
 🗣️ **lehlalošetšagotee** 5/6 synonym ⇒ **lehlalošagotee**
 🗣️ **lehlao** 5/6 adjective
 🗣️ **lehlathafelo** 5/6 adverb of place
 🗣️ **lehlathamokgwa** 5/6 adverb of manner
 🗣️ **lehlathanako** 5/6 temporal adverb, adverb of time
 🗣️ **lehlathasedirišwa** 5/6 instrumental adverb
 🗣️ **lehlathi** 5/6 adverb
 🗣️ **leina** 5/6 noun
 🗣️ **leinaina** 5/6 proper name
 📖 **leinakgopolo** 5/6 abstract noun
 🗣️ **leinataodi** 5/6 head noun
 🗣️ **lekgoka-amanyi** 5/6 relative con-

- cord
- 📖 **lekgokahlaodi** 5/6 adjectival concord, qualificative concord ⇒ **lekgokatlhaodi**
 - 📖 **lekgokamong** 5/6 possessive concord ⇒ **lekgokarui**
 - 🗣️ **lekgokarui** 5/6 possessive concord ⇒ **lekgokamong**
 - 🗣️ **lekgokasediri** 5/6 subject concord
 - 📖 **lekgokasediri la mmoledišwa ka bontši** 5/- subject concord second person plural
 - 📖 **lekgokasediri la mmoledišwa ka botee** 5/- subject concord second person singular
 - 🗣️ **lekgokasedirwa** 5/6 object concord
 - 📖 **lekgokatlhaodi** 5/6 adjectival concord, qualificative concord ⇒ **lekgokahlaodi**
 - 🗣️ **lekgokedi** 5/6 agreement morpheme
 - 🗣️ **lekopanyi** 5/6 conjunction
 - 🗣️ **lekopanyibaka** 5/6 causal connector / conjunction
 - 🗣️ **lekopanyikoketšo** 5/6 additive connector / conjunction
 - 🗣️ **lekopanyinako** 5/6 temporal connector / conjunction
 - 🗣️ **lekopanyipeelano** 5/6 conditional connector / conjunction
 - 🗣️ **lekopanyipharologanyo** 5/6 adversative connector / conjunction
 - 🗣️ **lelahlelwa** 5/6 interjection
 - 🗣️ **lelatodi** 5/6 antonym, opposite
 - 🗣️ **lentšu** 5/6 (linguistic) word
 - 🗣️ **lentšu la deiktiki** 5/6 deictic word
 - 🗣️ **lentšugokwa** 5/6 compound word
 - 📖 **lentšutheo** 5/6 head word
 - 🗣️ **lereo** 5/6 term
 - 🗣️ **lereokakaretšo** 5/6 general term, umbrella term
 - 🗣️ **lerui** 5/6 possessive
 - 📖 **leruo** 5/6 possession
 - 🗣️ **lešala** 5/6 pronoun
 - 🗣️ **lešalagohle** 5/6 quantitative pronoun
 - 🗣️ **lešalapadi** 5/6 quantitative pronoun
 - 🗣️ **lešalašala** 5/6 absolute pronoun
 - 🗣️ **lešalašupi** 5/6 demonstrative pronoun
 - 📖 **lešupakarolo** 5/6 word referring to a part of a whole ⇒ **lešupetšakarolo**, **sešupetšakarolo**
 - 🗣️ **lešupetšagotee** 5/6 coreferent
 - 📖 **lešupetšakarolo** 5/6 word referring to a part of a whole ⇒ **lešupakarolo**, **sešupetšakarolo**
 - 📖 **leswaodikga** 5/6 punctuation mark
 - 🗣️ **lethekgi** 5/6 stabiliser
 - 🗣️ **lethuši** 5/6 auxiliary verb
 - 🗣️ **letlago** 5/- future tense ⇒ **lebaka la letlago**
 - 🗣️ **letlema** 5/6 preposition
 - 🗣️ **letšwalediring** 5/6 deverbative
 - 🗣️ **logaganya** *v.* integrate
- M**
- 📖 **maikemišetšo** -/6 purpose, aim, intention ⇒ **malebiša**
 - 📖 **malebiša** -/6 purpose, aim, intention ⇒ **maikemišetšo**
 - 📖 **malebišatheo** -/6 basic purpose / aim / intention
 - 🗣️ **mmoledi** 1/2 (*pl. is baboledi*) first person
 - 🗣️ **mmoledišani** 1/2 (*pl. is baboledišani*) interlocutor
 - 🗣️ **mmoledišwa** 1/2 (*pl. is baboledišwa*) second person, addressee
 - 📖 **mmoledišwa ka bontši** 1/- second person plural
 - 📖 **mmoledišwa ka botee** 1/- second person singular
 - 📖 **mmolelwa** 1/2 (*pl. is babolelwa*) third person
 - 📖 **moanegwa** 1/2 person who is being described, character
 - 🗣️ **modirišo** 3/4 mood
 - 🗣️ **modirišogo** 3/- infinitive mood
 - 🗣️ **modirišogore** 3/- subjunctive mood
 - 🗣️ **modirišokanegelo** 3/- consecutive mood
 - 🗣️ **modirišopego** 3/- indicative mood
 - 🗣️ **modirišopegotlhaodi** 3/- participial form of indicative mood, situative mood

- 🗨️ **modirišotaelo** 3/- imperative mood
- 🗨️ **modirišotlwaelo** 3/- habitual mood
- 🗨️ **modirišotona** 3/- main mood
- 🗨️ **modiro** 3/4 function
- 🗨️ **modu** 3/4 root
- 🗨️ **modu wa lediri** 3/4 verb root
- 🗨️ **modumo** 3/4 speech sound
- 🗨️ **mofolotši** 3/- morphology ⇒ **popegopolelo**
- 🗨️ **mokgwapolelo** 3/4 language use
- 🗨️ **molaetšatebanyo** 3/4 intended message
- 🗨️ **molaetšatheo** 3/4 main idea
- 🗨️ **molao wa kgokagano** 3/4 discourse rule
- 🗨️ **molao wa polelo** 3/4 linguistic / language rule ⇒ **molaopolelo**
- 🗨️ **molao wa popafoko** 3/4 syntactic rule
- 🗨️ **molao wa popegopolelo** 3/4 morphological rule
- 🗨️ **molao wa tlhalošo** 3/4 semantic rule
- 🗨️ **molao wa tlhatlamano wa tatanontšu** 3/4 word order rule
- 🗨️ **molao wa tšhomišano** 3/4 cooperative principle
- 🗨️ **molaokgokagano** 3/4 discourse / communication rule
- 🗨️ **molaopolelo** 3/4 linguistic / language rule ⇒ **molao wa polelo**
- 🗨️ **molaotheo** 3/4 basic rule
- 🗨️ **momagana** *v.* coalesce
- 🗨️ **momagano** 3/4 coalescence
- 🗨️ **mong** 1/2 (*pl. is beng*) possessor
- 🗨️ **mongwalo wa fonetiki** 3/4 phonetic orthography
- 🗨️ **mosela** 3/4 suffix
- 🗨️ **mosela wa bontši** 3/4 plural suffix
- 🗨️ **motheeletši** 1/2 addressee
- 🗨️ **motheo** 3/4 basic element, basis

N

- 🗨️ **ngangego** 9/10 disagreement
- 🗨️ **noko** 9/10 syllable
- 🗨️ **nokotee** 9/10 monosyllable
- 🗨️ **NP** noun phrase
- 🗨️ **nyenyefatšo** 9/10 diminution

P

- 🗨️ **palo** 9/10 number
- 🗨️ **peakanyo ya mantšu** 9/10 arrangement of words, word arrangement
- 🗨️ **peakanyofoko** 9/10 syntactic arrangement (of constituents)
- 🗨️ **peelano** 9/10 condition
- 🗨️ **pego** 9/10 statement
- 🗨️ **peobakeng** 9/10 replaceability; substitution
- 🗨️ **pharologantšho** 9/10 (distinctive) feature
- 🗨️ **pharologantšho ya linkwistiki** 9/10 (distinctive) linguistic feature
- 🗨️ **pharologantšho ya popego** 9/10 (distinctive) morphological feature ⇒ **pharologantšhopopego**
- 🗨️ **pharologantšho ya seemotikologo** 9/10 (distinctive) discourse feature, discourse characteristic
- 🗨️ **pharologantšho ya tlhalošo** 9/10 (distinctive) semantic feature ⇒ **pharologantšhotlhalošo**
- 🗨️ **pharologantšhopopego** 9/10 (distinctive) morphological feature ⇒ **pharologantšho ya popego**
- 🗨️ **pharologantšhotlhalošo** 9/10 (distinctive) semantic feature ⇒ **pharologantšho ya tlhalošo**
- 🗨️ **pharologanyo** 9/10 distinction
- 🗨️ **phetleko** 9/10 analysis
- 🗨️ **phetleko ya kgokagano** 9/10 discourse analysis ⇒ **phetlekokgokagano**
- 🗨️ **phetlekokgokagano** 9/10 discourse analysis ⇒ **phetleko ya kgokagano**
- 🗨️ **phetlekotaodišo** 9/10 discourse analysis
- 🗨️ **poledišano** 9/10 dialogue
- 🗨️ **popafoko** 9/10 syntax
- 🗨️ **popego** 9/10 morphology, form, structure
- 🗨️ **popego ya lediri** 9/10 morphology of the verb, verbal morphology
- 🗨️ **popego ya lefoko** 9/10 sentence structure

- 🗨️ **popegopolelo** 9/10 morphology ⇨
mofolotši
 📖 **popegopolelo ya Sesotho sa Leboa**
 9/- Northern Sotho morphology
S
- 🗨️ **sebopego** 7/8 (relating to) form
 🗨️ **sediri** 7/8 subject
 🗨️ **sedirwa** 7/8 object
 📖 **sedirwa sa bobedi** 7/8 direct object
 (lit. second object)
 📖 **sedirwa sa pele** 7/8 indirect object
 (lit. first object)
 📖 **seema** 7/8 proverb
 🗨️ **seemotikologo** 7/8 context
 📖 **seemotikologo sa kgokagano** 7/8
 discourse context
 📖 **segagabo** 7/- (their) mother-tongue
 📖 **segageno** 7/- (your) mother-tongue
 📖 **segalo** 7/8 tone
 📖 **seka** 7/8 idiomatic expression
 🗨️ **sekafoko** 7/8 phrase
 🗨️ **sekafokodiri** 7/8 verb phrase
 🗨️ **sekafokoina** 7/8 noun phrase
 📖 **senaganwa** 7/8 idea, thought, con-
 cept
 🗨️ **serewa** 7/8 topic
 📖 **serewa sa poledišano** 7/8 discourse
 topic
 🗨️ **serewakgolo** 7/8 main topic
 🗨️ **serewana** 7/8 sub-topic
 🗨️ **serewapoeletšo** 7/8 re-introducing
 (discourse) topic
 🗨️ **serewatiego** 7/8 delayed discourse
 topic
 🗨️ **serewatirišano** 7/8 collaborating
 (discourse) topic
 🗨️ **serewatšhielano** 7/8 introducing
 (discourse) topic
 🗨️ **serewatswalano** 7/8 incorporating
 (discourse) topic
 📖 **sešupetšakarolo** 7/8 word referring
 to a part of a whole ⇨ **lešu-**
pakarolo, lešupetšakarolo
 🗨️ **sešupša** 7/8 referent
 📖 **swantšhiša** *v.* compare
T
- 🗨️ **taelo** 9/10 command
 📖 **taodišo** 9/10 essay
 📖 **tatelano** 9/10 succession, sequence
 🗨️ **tatelano ya mantšu** 9/10 word order
 ⇨ **tatelanontšu, tthatlamano ya**
mantšu, tthatlamanontšu
 🗨️ **tatelanontšu** 9/10 word order ⇨
tatelano ya mantšu, tthatlamano
ya mantšu, tthatlamanontšu
 📖 **tatelanotseo** 9/10 dominant / basic
 order
 🗨️ **thabe** 9/10 clause
 🗨️ **thabekutu** 9/10 main clause
 🗨️ **thabenyana** 9/10 subordinate clause
 📖 **thulano ya tlhalošo** 9/10 semantic
 incompatibility
 🗨️ **thutamedumo** 9/- study of (speech)
 sounds, study of phonetics
 🗨️ **thutapolelo** 9/- study of linguistics
 🗨️ **thutapopofoko** 9/- study of syntax
 🗨️ **thutapopontšu** 9/- study of mor-
 phology
 🗨️ **thutatlhalošo** 9/- study of semantics
 📖 **tirišano** 9/10 cooperation
 🗨️ **tiro** 9/10 predicate, action, process
 🗨️ **tirotona** 9/10 main predicate /
 action / process
 📖 **tirwa** 9/10 passive
 🗨️ **tirwana** 9/10 subordinate predicate
 🗨️ **tlaleletša** *v.* determine, qualify
 🗨️ **tlaleletšadiri** 9/10 verbal determiner
 🗨️ **tlaleletšaina** 9/10 nominal deter-
 miner
 📖 **tlaleletšatiro** 9/10 verbal adjunct ⇨
tlaleletšo ya tiro
 🗨️ **tlaleletšo** 9/10 complement, adjunct
 🗨️ **tlaleletšo ya tiro** 9/10 verbal adjunct
 ⇨ **tlaleletšatiro**
 📖 **tlami** 9/10 hyphen
 🗨️ **tlamo** 9/10 connection
 🗨️ **tlemagano** 9/10 cohesion ⇨ **toga-**
ganyo
 📖 **tlemagantšha** *v.* link, connect
 📖 **tlemaganya** *v.* link, connect
 📖 **tlhakakgolo** 9/10 capital letter
 📖 **tlhalošišo** 9/10 definition
 🗨️ **tlhalošo** 9/10 meaning
 📖 **tlhalošo ya deiktiki** 9/10 deictic
 meaning
 📖 **tlhalošo ya lefoko** 9/10 sentence

- meaning
- 📖 **tlhalošo ya lentšu** 9/10 word meaning
- 🗨️ **tlhalošo ya medirišo** 9/10 modal meaning
- 🗨️ **tlhalošofoko** 9/10 sentence meaning
- 🗨️ **tlhalošokamanyi** 9/10 associative meaning
- 🗨️ **tlhalošokatološo** 9/10 extended meaning
- 📖 **tlhalošokelello** 9/10 cognitive meaning
- 📖 **tlhalošokhuduego** 9/10 emotive meaning
- 🗨️ **tlhalošontši** 9/- polysemy
- 📖 **tlhalošotebanyo** 9/10 intended meaning
- 🗨️ **tlhalošotheo** 9/10 basic meaning
- 📖 **tlhalošotheo ya lefoko** 9/10 basic sentence meaning
- 📖 **tlhaodi** 9/10 qualificative, modifier
- 📖 **tlhatlagano** 9/10 hierarchy
- 📖 **tlhatlamano** 9/10 succession
- 📖 **tlhatlamano ya ditiro** 9/10 consecutive actions
- 🗨️ **tlhatlamano ya mantšu** 9/10 word order ⇨ **tatelano ya mantšu, tatanontšu, tlhatlamanontšu**
- 🗨️ **tlhatlamanontšu** 9/10 word order ⇨ **tatelano ya mantšu, tatanontšu, tlhatlamano ya mantšu**
- 🗨️ **tlhatlamanotheo** 9/10 basic (word) order
- 📖 **tlhatlamanotheo ya mantšu** 9/10 basic / dominant word order
- 📖 **tlhopho** 9/10 categorization
- 🗨️ **tlhopollo** 9/10 analysis
- 🗨️ **tlogelo** 9/10 deletion
- 🗨️ **togaganyo** 9/10 cohesion ⇨ **tlemagano**
- 📖 **tsebo** 9/10 knowledge
- 🗨️ **tsebo ye e feleletšego** 9/10 full knowledge
- 🗨️ **tšhalafatšo** 9/10 pronominalisation
- 🗨️ **tshedimošo** 9/10 information
- 🗨️ **tšhomišo** 9/10 function
- 🗨️ **tšhomišo ya polelo** 9/10 function of language, language function
- 🗨️ **tšhupetšogotee** 9/- coreference
- 🗨️ **tšhupetšokarolo** 9/- part-whole relationship, interreference
- 📖 **tšhutišo** 9/10 shifting
- 📖 **tshwantšhišo** 9/10 simile
- 🗨️ **tswalana** *intransitive v.* relate, associate, link
- 🗨️ **tswalano** 9/10 relationship, association
- 🗨️ **tswalanya** *transitive v.* relate, associate, link
- 🗨️ **tswalanyo** 9/10 association, link, connection
- 🗨️ **tswalanyo ya mantšu** 9/10 association / linking / connection of words
- 📖 **tumagwaša** 9/10 fricative
- 📖 **tumammogo** 9/10 consonant
- 📖 **tumammogokodu** 9/10 voiced consonant
- 🗨️ **tumanoši** 9/10 vowel
- 📖 **tumanoši ya mafelelo** 9/10 final vowel
- 📖 **tumanoši ya mafelelo ya lediri** 9/10 verbal ending (lit. final vowel of the verb)
- 📖 **tumathu** 9/10 plosive consonant
- 📖 **tumatshwano** 9/10 homonym
- 🗨️ **tumelo** 9/10 affirmative
- 📖 **tummogotu** 9/10 voiceless consonant
- 🗨️ **tumotshwano** 9/10 homonymy
- V**
- 📖 **VP** verb phrase

Les tons dans les dictionnaires de langues gabonaises: situation et perspectives

Thierry Afane Otsaga, *Département d'afrikaans et de néerlandais, Université
de Stellenbosch, Stellenbosch, République d'Afrique du Sud*
(*thiafane@yahoo.fr*)

Résumé: La plupart des dictionnaires des langues gabonaises, parus jusqu'à ce jour, n'ont pas vraiment tenu compte de la tonalité dans la rédaction des articles. Or, à présent, il est unanimement reconnu que le ton joue un rôle important dans le fonctionnement des langues bantoues en général et des langues gabonaises en particulier. Les dictionnaires, outils d'apprentissage, de standardisation, de préservation et de l'emploi correct d'une langue, ne pourront plus négliger cet aspect dans leur traitement lexicographique. Malgré le grand nombre de tons qui existent dans les langues gabonaises et la difficulté de les noter tous dans la présentation des articles, il est néanmoins nécessaire de trouver des moyens de mettre en valeur chaque ton, sans néanmoins rendre trop compliqué l'emploi du dictionnaire. L'objectif de notre exposé est de formuler, pour ces dictionnaires, certaines directives qui permettront de prendre en considération la tonalité des langues gabonaises, tout en gardant accessible l'emploi du dictionnaire. Avant de formuler ces directives, notre exposé tente tout d'abord de clarifier le problème par l'analyse et l'énumération des différents tons identifiés jusqu'à ce jour dans les langues gabonaises. En outre, nous prenons en considération certains points de vue qui, dans les dictionnaires, ont voulu identifier les tons, en donnant des exemples du traitement de la tonalité dans certains dictionnaires des langues tonales autres que les langues gabonaises.

Mots-clés: DICTIONNAIRES, LANGUES GABONAISES, TONS, TONALITÉ, RÉDACTION DES ARTICLES, PERSPECTIVES, LANGUES BANTOUES, TRAITEMENT LEXICOGRAPHIQUE, OUVRAGES DE RÉFÉRENCE LEXICOGRAPHIQUE, COURBE MÉLODIQUE, SIGNES DIACRITIQUES, ORTHOGRAPHE, PRONONCIATION

Abstract: Tones in Dictionaries of the Gabonese Languages: State and Perspectives. Until now most dictionaries that have appeared in the Gabonese languages have not really taken tonality into account in the editing of articles. But at present it is unanimously recognized that tone plays an important role in the functioning of Bantu languages in general and Gabonese languages in particular. Dictionaries which are tools in the learning, standardization, preservation and correct usage of a language can no longer neglect this aspect in their lexicographic treatment. In spite of the large number of tones present in the Gabonese languages and the difficulty of attending to all of them in the presentation of the articles, it is nevertheless necessary to find means to account for each tone, without however rendering the use of the dictionary too complicated. The objective of this article is to formulate certain directions for these dictionaries, which

will make allowance for the tonality of the Gabonese languages, keeping the dictionary completely accessible to users. After having formulated these directions, our article firstly attempts to clarify the problem by analysing and enumerating the different tones identified in the Gabonese languages up till now. Moreover, we consider certain angles from which, in dictionaries, tones can be identified, by giving examples of the treatment of tonality in certain dictionaries of tonal languages other than the Gabonese languages.

Keywords: DICTIONARIES, GABONESE LANGUAGES, TONES, TONALITY, EDITING OF ARTICLES, PERSPECTIVES, BANTU LANGUAGES, LEXICOGRAPHIC TREATMENT, LEXICOGRAPHIC REFERENCE WORKS, MELODIC CURVE, DIACRITIC SIGNS, ORTHOGRAPHY, PRONUNCIATION

Introduction

Les langues gabonaises sont dans leur ensemble des langues bantoues et comme la plupart des langues bantoues elles comportent des éléments supra-segmentaux telles que l'accent, l'intonation, la longueur, les tons, etc. qui jouent un rôle important dans le fonctionnement de ces langues. De tous ces éléments supra-segmentaux, les tons sont, sans conteste, ceux dont l'implication est le plus déterminant dans le comportement des dites langues (Afane Otsaga 2001: 155). Les tons sont des oppositions de hauteurs qui dans certaines langues sont utilisées comme unités distinctives (Baylon et Fabre 1990: 101). Tous les travaux de description réalisés sur les langues gabonaises reconnaissent unanimement le rôle majeur que jouent les tons dans celles-ci. De ces différents travaux, il ressort que les langues gabonaises sont non seulement des langues à tons puisqu'elles utilisent la hauteur mélodique à des fins distinctives, mais sont aussi des langues à type tonal parce que la structure tonale de certains termes varient selon qu'ils sont en isolation ou en contexte. Pour ce qui concerne les travaux lexicographiques sur les langues gabonaises, particulièrement pour la réalisation des dictionnaires, une question importante demeure sans réponse: doit-on oui ou non noter les tons dans les dictionnaires de langues gabonaises? L'objectif de cet article est non seulement de faire le point en ce qui concerne le traitement de la tonalité dans les dictionnaires de langues gabonaises jusqu'à présent, mais surtout de proposer quelques méthodes qui pourraient aider à résoudre ce problème dans le traitement lexicographique des langues gabonaises.

Un mot sur les tons dans les langues gabonaises

Les tons sont généralement répartis en deux catégories. D'une part les tons ponctuels où seul un aspect de la courbe mélodique est tenu en compte (le plus haut ou le plus bas) ainsi que les tons modulés qui se distinguent par des directions successives de la courbe mélodique (Baylon et Fabre 1990: 101). Dans

les langues à tons, la variation mélodique est généralement associée à une syllabe (Matthews 1997: 379).

Les différentes descriptions¹ réalisées jusqu'à ce jour sur les langues gabonaises ont identifié dans celles-ci l'existence de trois catégories de tons.

Les tons principaux

Par tons principaux, j'entends les tons qui sont les plus récurrents et les mieux répartis dans les langues gabonaises. Il s'agit

— du *ton ponctuel haut* noté avec un accent aigu (´). C'est le point le plus haut de la courbe mélodique.

Langues	Exemples ²	Traduction française
Fan	mvám	générosité
Meka	nké	plantation
Yipunu	mvúlà	pluie

— du *ton ponctuel bas* noté avec un accent grave (`). C'est le point le plus bas de la courbe mélodique.

Langues	Exemples	Traduction française
Gisira	bilimè	les années
Ikota	èlùngá	panier
Latege	jùrù	tirer

— du *ton modulé montant* noté avec un accent circonflexe inversé (ˇ). Il s'agit d'une courbe mélodique qui prend deux directions successives. Elle descend d'abord vers le point le plus bas, avant de remonter vers le point le plus haut.

Langues	Exemples	Traduction française
Seki	dikõn	lance
Sake	yă	marécage
Yilumbu	dĩmbú	village

— du *ton modulé descendant* noté avec un accent circonflexe (^). Il s'agit d'une courbe mélodique qui prend deux directions successives. Elle monte d'abord vers le point le plus haut, avant de redescendre vers le point le plus bas.

Langues	Exemples	Traduction française
Lekanini	yâ	autrefois
Liwanzi	tôlô	sommeil
Omyene	íbôbò	peur

Les tons intermédiaires

Par tons intermédiaires, j'entends les tons qui sont moins récurrents et moins répartis dans les langues gabonaises que les tons principaux, mais dont la présence dans ces langues n'est pas négligeable puisque facilement remarquable. Il s'agit

— du *ton ponctuel moyen* noté avec une barre horizontale (¨). C'est le point situé au milieu de la courbe mélodique, c'est-à-dire entre le point le plus bas et le point le plus haut de la courbe mélodique.

Langues	Exemples	Traduction française
Fañ	āsō̄n̄	dent
Liwanzi	jìwānī	frapper
Omyengē	tsólō̄	oiseau

— du *ton haut-abaisse* noté avec une barre verticale ('). Notons que ce ton ne se place pas au-dessous d'un segment syllabique comme c'est généralement le cas des tons et est aussi appelé *faille tonale*. La faille tonale a pour résultat de produire des tons que l'on appelle haut-abaisse. Ces tons sont réalisés phonétiquement à un niveau qui peut être assimilable au ton moyen, mais contrairement à un véritable ton moyen, ils ne peuvent être suivis par un ton plus haut (Hombert 1990: 102). Ce ton est suivi d'un terrassement tonal, c'est-à-dire que la série de tons hauts qui suit la faille tonale est réalisée au même niveau que le ton haut-abaisse.

Langues	Exemples	Traduction française
Fañ	áwí'ny	tuer
Lembaama	mèma'áta'	j'ai tiré
Ngubi	ín̄gu'yí	mère

— du *ton supra-haut* noté avec deux accents aigus ("). Ce ton a pour point de départ une valeur plus élevée qu'un ton haut et qui ensuite chute très rapidement (Hombert 1990: 103).

Langues	Exemples	Traduction française
Yipunu	élábì	il voit
Lembaama	vúghì	miel
Lekaninì	sǎsì	appât

Les tons secondaires

Par tons secondaires, j'entends les tons assez rares dans les langues gabonaises et dont la répartition entre ces dernières est véritablement infime. Il s'agit

— du *ton infra-bas* noté par deux accents graves ("). Sa définition est identique à celle du ton supra-haut, avec la différence que sa courbe est descendante. Il s'agit donc d'un ton bas réalisé plus bas que le ton bas normal. Il existe parfois un contraste entre le ton bas ponctuel et le ton infra-bas. Il est fréquent que cette distinction ne se produise que dans certains contextes, par exemple avant la pause (Hombert 1990: 102).

Langues	Exemples	Traduction française
Lenduma	wǎ́lú	ped
Meka	wǎ́má	combattre

— du *ton très haut descendant* noté par un accent aigu suivi d'un accent circonflexe (´ ^). Je n'ai pas de définition précise de ce ton, mais on pourrait penser qu'il s'agit d'un ton supra-haut qui chute par la suite. Le seul exemple que j'ai de ce ton vient du Yipunu.

Langue	Exemple	Traduction française
Yipunu	ngáânzì	racines

Le recensement des tons révélés ainsi dans les langues gabonaises n'est certainement pas exhaustif, car toutes les langues gabonaises n'ont pas encore été décrites. Mais il permet de constater qu'il en existe un grand nombre et que par conséquent leur influence dans le fonctionnement de ces langues est sans conteste à prendre en compte.

Le traitement de la tonalité dans les dictionnaires existants

Les premiers travaux lexicographiques dans les langues gabonaises ont été exécutés par des anciens administrateurs coloniaux et par les missionnaires. L'objectif premier de ces œuvres à caractère lexicographique (glossaires, lexiques, dictionnaires ...) était d'établir une communication efficace entre ces administrateurs et missionnaires, et les populations gabonaises (Mavoungou 2001). De même, les réalisateurs de ces travaux n'étaient ni des linguistes de formation, encore moins des lexicographes attitrés. Malgré ce fait, ces travaux, dans leur globalité, répondent effectivement aux normes métalexigraphiques requises (Mihindou 2001: 161), bien que certaines améliorations y soient nécessaires.

Bien que n'ayant pas eu à ma disposition tous les dictionnaires³ réalisés sur les langues gabonaises, je peux estimer que ceux que j'ai pu consulter constituent entre 70 et 80% environ de ce qui existe. De cette consultation, j'ai pu remarquer, en ce qui concerne le traitement de la tonalité, les situations suivantes:

— Deux tiers des dictionnaires consultés ne font pas état de la tonalité. Les articles des dictionnaires sont présentés sans structure tonale et cette

absence n'est, ni mentionnée, ni expliquée dans les introductions (prétextes) des dictionnaires. La majorité de ces dictionnaires sont plutôt des glossaires ou des lexiques, car il s'agit dans l'ensemble de listes alphabétiques de mots dans une langue-source suivie de leurs traductions équivalentes dans une langue-cible. Le *Dictionnaire Français–Yipounou / Yipounou–Français* (1966) de l'Église Évangélique du Gabon est une très bonne illustration de cette situation. Dans ce dernier aucun ton n'est noté, alors que le Yipunu est une langue à tons avérée.

— Le deuxième scénario est celui des dictionnaires qui utilisent des diacritiques dans la transcription. Mais ces diacritiques sont utilisés non pas comme marques de tonalités, mais plutôt comme éléments des symboles alphabétiques utilisés. Le *Dictionnaire Français–Mpongwé*, suivi d'éléments de grammaire (1995) de Raponda-Walker nous donne une excellente illustration de ce type de démarches. Dans ce dernier par exemple le « è » transcrit le son phonétique [ɛ], alors que le « ó » y figure pour transcrire le son phonétique [ɔ]. En d'autres termes, les tons n'y sont pas pris en considération. Mais contrairement au premier cas, celui-ci a la particularité de prendre déjà en compte certaines situations d'intonation ou d'accentuation en ce qui concerne la prononciation du « è » et du « ó ».

— Le dernier scénario est à l'image du second, car il ne tient compte que des phénomènes d'intonation et d'accentuation qui sont clairement mis en évidence au niveau de la transcription orthographique. Or, nous pouvons constater que dans ces dictionnaires, plusieurs traits accentuels (ou d'intonation) attribués aux langues gabonaises sont en réalité des manifestations tonales. Ce qui peut s'observer dans *l'Encyclopédie Pahouine* (1901) de V. Largeau où ce dernier présente par exemple trois réalisations différentes du « a » dans la *Fañ* (a, ā, â). Le « a » sans accent pour lui est équivalent à celui qui existe en français comme dans « ami ». Celui avec la barre horizontale « ā » est décrit comme un son ayant la même qualité que le premier, mais avec une quantité plus importante, alors que celui avec l'accent circonflexe « â » augmente aussi bien sur le plan qualitatif que quantitatif et se prononce comme le « â » français dans « pâte ». Cette même application est faite pour d'autres sons comme « e » ou « o ». Il semble que Largeau, qui était un administrateur colonial et non pas un linguiste ou lexicographe de formation, confond l'intonation, l'accentuation et la tonalité en *Fañ*. Les descriptions intonatives ou accentuelles qu'il donne de ces sons correspondent en réalité aux situations où ils portent le ton bas, moyen, haut, montant ou descendant comme mis en évidence dans différentes descriptions du *Fañ* par Andeme Allogo (1980), Voltz (1990), Cinnamon (1990), Ondo Mebiame (1992), etc.

La confusion de Largeau entre tons et d'autres phénomènes suprasegmentaux a pu se glisser dans d'autres dictionnaires qui, j'insiste sur ce point, ont été réalisés (pour la plupart) par des non-spécialistes en matière de description linguistique. Cela peut donc expliquer l'absence des tons dans le plus grand nombre de dictionnaires réalisés dans les langues gabonaises.

Signalons également que dans un petit nombre de dictionnaires et d'ouvrages⁴ (tel que le dictionnaire Fang–Français en Français–Fang de Galley 1964) réalisés dans les langues gabonaises, les tons sont notés non pas sur les segments, mais entre parenthèse à côté de ces derniers. Cette méthode donne l'impression que pour les réalisateurs de ces ouvrages, les tons ne font pas parties intégrantes de la langue et sont des phénomènes situationnels qui apparaissent çà et là dans le fonctionnement de celles-ci.

Quelques avis informels au sujet de la notation des tons dans les dictionnaires

Le point de vue le plus général rencontré chez beaucoup d'auteurs en ce qui concerne l'absence de tons dans les dictionnaires de langues gabonaises est que ceux-ci ne sont pas indispensables. La raison souvent avancée est que les premiers ouvrages à caractère lexicographique ne tenaient pas compte des tons de façon générale, mais ils ont été les principaux supports qui ont aidé beaucoup de missionnaires et d'administrateurs coloniaux à apprendre les langues gabonaises. Dans mes discussions sur ce problème avec certains locuteurs gabonais profanes de la linguistique, la plupart m'ont ressorti l'exemple des prêtres européens qui disaient les messes dans les langues gabonaises, sans pourtant avoir une prononciation tonale très correcte de ces langues. Pour eux une langue peut d'abord s'acquérir sur le plan segmental, la structure suprasegmentale de celle-ci s'acquérant par la pratique régulière au contact des locuteurs.

Je voudrais rappeler à ceux qui parlent de l'expérience des prêtres européens, qu'ils ne doivent pas non plus oublier les incidents survenus dans certaines églises du Gabon au cours de messes dites par ces derniers. L'incident le plus courant était que les fidèles sortaient de l'église parce qu'ils se sentaient choqués ou insultés par les propos d'un prêtre qui en prononçant mal certains termes (sans la bonne tonalité) aurait signifié des choses qu'il ne souhaitait pas dire en réalité. Cela montre que la tonalité est primordiale dans les langues gabonaises et que par conséquent ne pas en tenir compte dans les dictionnaires pourrait occasionner, pour les apprenants de ces langues, des incidents identiques à ceux dont je viens de faire état.

Pour autant, je ne suis pas totalement en désaccord avec ceux qui pensent que l'acquisition d'une langue se réalise d'abord au niveau de sa structure segmentale, car elle est la plus palpable, autant sur le plan acoustique que sur le plan visuel. Le dictionnaire doit d'abord mettre l'accent sur la transcription (orthographique) simple et accessible des formes segmentales de la langue. Par conséquent, je ne partage pas l'avis de ceux qui estiment qu'un livre d'appren-

tissage linguistique qui ne fait pas état des marques de tonalité est à rejeter complètement⁵. J'estime que la présentation de la structure segmentale est une étape ou une partie de l'entière structure (segmentale et suprasegmentale) d'une langue. Il n'y a donc pas de raison à ce qu'elle soit rejetée. Beaucoup de descriptions purement linguistiques sur les langues gabonaises se restreignent bien souvent aux aspects phonétiques, phonologiques et morphologiques. Rares sont celles qui font des descriptions sémantiques ou syntaxiques de ces langues. Mais tout le monde est d'accord pour reconnaître que ces descriptions, qui ne mettent en évidence que certains aspects de la langue, sont des travaux scientifiquement valables. Présenter une langue sans sa structure tonale devrait être vu dans le même esprit. C'est un aspect de la langue qui est présenté (la structure segmentale), il faut donc l'accepter comme tel.

Pour autant, dans le cas des langues à types tonals comme le sont certaines langues gabonaises, il ne serait pas inutile pour l'usager du dictionnaire d'ajouter les marques tonales afin de faciliter la tâche aux apprenants de ces langues. Au lieu que l'apprenant d'une langue attende d'être au contact des usagers de celle-ci pour la pratiquer et ainsi acquérir sa structure suprasegmentale, il pourrait l'apprendre directement à travers le dictionnaire ou du moins l'avoir à sa disposition dans ce dernier. Mon avis est que les tons devraient être présentés dans les dictionnaires de langues gabonaises, mais cette présentation devrait se faire de façon méthodique et en tenant compte d'un certain nombre de paramètres.

Les tons et l'orthographe

Le problème le plus important des langues gabonaises en ce moment est leur transcription orthographique. Tous les travaux réalisés sur ces langues sont transcrits, soit phonétiquement, soit avec des symboles orthographiques différents les uns des autres. Aucune langue au monde ne s'est développée sans une forme orthographique unique utilisée par tous (Afane Otsaga 2001: 156). L'orthographe rappelons-le est l'art et la technique de la formation des mots par les lettres selon l'usage accepté (Gove 1961). Le choix d'utiliser une forme orthographique spécifique pour l'écriture d'une langue donnée est une convention qui naît d'un accord (généralement entre experts) et qui est accepté (ou imposé dans certains cas) pour être utilisé dans les écoles, les administrations (gouvernementales et non gouvernementales), la presse et tous les autres champs d'utilisation de la langue (Mdee 1999: 123). L'acceptation d'une forme scripturale d'une langue se fait généralement par la création d'un modèle d'imitation et la promotion de ce modèle.

Dans la situation des langues gabonaises, l'acceptation du modèle orthographique proposé au cours d'un colloque d'experts en avril 1999 à Libreville sous l'égide du Ministère de l'Éducation Nationale (Afane Otsaga 2001: 156) permettra de faire un pas important vers la lecture et l'écriture de ces langues par le plus grand nombre. En effet, tout débat sur la mise en valeur ou non des

tons dans la transcription des langues gabonaises resterait vain tant que celles-ci n'auraient pas une forme orthographique acceptée et utilisée par tous. Toute notation de tons serait inutile si elle ne servait pas à aider l'utilisateur dans l'amélioration de la prononciation. Mais si ce dernier ne se retrouve même pas dans la structure segmentale (par la forme orthographique) des termes qui lui sont proposées, il lui serait difficile d'avoir accès à la structure suprasegmentale de ceux-ci. La structure segmentale d'une langue est le support de sa structure suprasegmentale. Si le support est instable, ce qu'il supporte l'est aussi forcément. Le problème de tonalité dans l'écriture des langues gabonaises ne peut donc trouver de solution tant que celui de l'orthographe n'est pas résolu.

Je pourrais ajouter que la vulgarisation de l'orthographe de langues gabonaises devrait aussi inclure la vulgarisation des marques de tonalité. Un certain nombre de diacritiques (pas trop nombreux) pourrait être retenu pour transcrire les tons les plus réguliers et les plus répandus dans les langues gabonaises. La promotion de ces diacritiques devrait se faire en même temps que les symboles retenus pour l'orthographe. Cette promotion consistera essentiellement à habituer les usagers à l'utilisation de ces diacritiques, mais surtout à leur permettre d'associer clairement chaque diacritique à un ton spécifique, en précisant la manière dont ce ton doit être réalisé.

Quelques exemples de notation des tons dans les dictionnaires de langues bantoues

Les langues gabonaises font parties de l'ensemble bantou. Toutes les langues appartenant à ce groupe rencontrent le même problème en ce qui concerne la prise en compte de la tonalité dans les dictionnaires. Ce problème est abordé de différentes façons selon le choix des lexicographes.

La situation la plus courante est celle des dictionnaires qui présentent les marques de tons dans la macrostructure du dictionnaire, mais sans donner aucune explication ni par rapport à la présence de ces tons, ni par rapport à leur rôle dans le fonctionnement de la langue. C'est la situation générale que nous avons observée pour les dictionnaires des langues gabonaises. Deux hypothèses peuvent être émises par rapport à cette attitude:

— La première hypothèse est que certains lexicographes prétendent que puisque les usagers-cible des dictionnaires sont des locuteurs natifs de ces langues, par la maîtrise intuitive qu'ils ont de la langue, ils n'ont pas besoin d'explications particulières sur la structure suprasegmentale d'une langue qu'ils maîtrisent déjà. C'est l'occasion ici de rappeler à ceux qui pourraient se reconnaître dans cette hypothèse que le dictionnaire même s'il est réalisé avant tout selon les besoins du groupe-cible, peut dans certaines situations tomber entre les mains d'un individu n'appartenant pas au groupe-cible du dictionnaire, comme c'est souvent le cas aujourd'hui dans un monde de plus en plus polyvalent. Les informations

sur la structure suprasegmentale pourraient, dans ce cas, être utiles à un non-locuteur de la langue.

— La deuxième hypothèse est que certains lexicographes travaillant sur les langues bantoues estiment que dans les dictionnaires des langues indo-européennes, qui ont pourtant une longue tradition lexicographique, aucune explication particulière n'est faite au sujet des phénomènes suprasegmentaux comme l'accent ou la longueur par exemple. Par rapport à cet argument, je voudrais souligner que les langues indo-européennes ont une très longue tradition écrite. Les problèmes liés à l'orthographe et à la prononciation ont depuis longtemps trouvé des solutions dans ces langues, contrairement aux langues bantoues qui se cherchent encore sur tous les plans. Les phénomènes accentuels en français par exemple sont clairement identifiés par un certain nombre de diacritiques qui ôtent toute ambiguïté dans la prononciation de cette langue. La comparaison entre les dictionnaires des langues indo-européennes et les dictionnaires des langues bantoues est donc une méprise qu'il est souhaitable d'éviter.

Dans d'autres dictionnaires, les lexicographes ont choisi de présenter la structure tonale de façon complète et sans restriction telle qu'elle apparaît dans les termes de la langue. Tous les tons y sont marqués. C'est le cas du dictionnaire Tsivenda⁶-English (Van Warmelo 1937) dans lequel les tons hauts, bas, moyens et haut abaissés sont notés. Selon les rédacteurs de ce dictionnaire, ces quatre tons ont été retenus parce qu'ils sont les seuls qui apparaissent dans les termes en isolation, les autres tons étant de simples variations de ces derniers en situation de discours. Ils représentent donc la tonalité de base de la langue, c'est pourquoi ils sont les seuls à être notés dans le dictionnaire. Les tons de base sont ainsi notés au niveau des lemmes à tous les niveaux où ils apparaissent.

Cette présentation, bien que visant l'exhaustivité, a l'inconvénient de trop alourdir les articles du dictionnaire par la présence d'un grand nombre de signes diacritiques. La vue d'un trop grand nombre de signes, qu'ils ne maîtrisent pas forcément, est décourageant pour beaucoup d'utilisateurs de dictionnaires.

Un autre cas est celui du dictionnaire English-Zulu⁷ et Zulu-English (Doke, Malcolm, Sikakana et Vilakazi 1990) dans lequel il est indiqué que fondamentalement le Zulu est une langue qui comporte neuf variétés tonales. Mais les rédacteurs estiment (comme ceux du dictionnaire Venda) que seuls les tons ponctuels haut et bas sont les véritables tons de la langue; les autres sont des variations et/ou des associations de ces deux tons. Mais à la différence du dictionnaire Venda où les marques de tonalité sont bien visibles dans la présentation de la macrostructure, le dictionnaire Zulu se contente de décrire, dans les pré-textes, les différents contextes d'apparition ou de prononciation de chaque ton. Le ton descendant par exemple est décrit comme étant le résultat d'un ton haut qui est articulé avec une dépression occasionnée généralement en Zulu par la présence des sons tels que *bh*, *d*, *g*, *gc*, *gx*, *gq*, *v*, *z*, *dl*, *hh* et *j*. Dans ce dic-

tionnaire aucune tonalité n'est présentée dans la macrostructure. L'utilisateur doit donc régulièrement consulter les pré-textes pour comprendre le comportement de la langue au niveau suprasegmental.

Cette méthode, bien qu'allégeant la macrostructure du dictionnaire en faisant une économie en terme d'espace (dans le dictionnaire) et de temps (pour le travail du dictionnaire), a l'inconvénient de ne présenter aucun ton dans la macrostructure. Toutes les explications relatives aux tons sont concentrées dans les pré-textes. Cela risque de ne pas être très utile aux usagers, surtout que ces derniers n'ont pas toujours la culture de la bonne utilisation du dictionnaire et par conséquent ils consultent assez peu ou rarement les pré-textes.

Perspectives

Le problème de la tonalité dans les dictionnaires est donc traité différemment selon le choix des lexicographes. Pour ce qui est des dictionnaires de langues gabonaises, la problématique reste entière, puisque n'ayant pas encore été véritablement abordée par les spécialistes (linguistes et lexicographes). Le Gabon est entré dans un processus irréversible de développement de ses langues. Ce processus a comme priorités les plus immédiates, la réalisation de dictionnaires dont le contenu devrait non seulement satisfaire les besoins des usagers, mais prévoir les problèmes de ces derniers. Il ne fait aucun doute que l'une de ces difficultés serait la réalisation tonale. Il n'est donc pas inutile dès aujourd'hui de proposer des voies qui pourraient aider à mieux aborder le problème de tons dans les futurs dictionnaires de langues gabonaises.

Quelques propositions

L'un des problèmes majeurs rencontré dans les langues gabonaises est celui de la pluralité des tons. La trop grande variété de tons dans les langues gabonaises est l'une des principales raisons de la réticence d'un grand nombre de chercheurs à les noter dans les dictionnaires ou des ouvrages à référence lexicographique. Les noter non seulement alourdirait les textes, mais on n'est jamais certain de pouvoir tous les noter convenablement comme ils apparaissent réellement dans les langues. L'une des solutions que l'on pourrait proposer est que pour la notation des tons dans les dictionnaires et autres ouvrages à référence lexicographique, on se limite aux tons les moins récurrents et à deux tons au maximum. Cette option ferait appel à une description linguistique⁸ détaillée de la langue choisie. En vue de la description tonologique de cette langue, le lexicographe pourra clairement y distinguer la fréquence des différents tons. Ma proposition est donc que seuls les tons les moins fréquents soient notés. Mais la prise en compte des tons à noter doit se faire en fonction de chaque terme. Si un terme n'a que deux types de tons, seul un ton doit être noté. Mais si le terme contient plus de trois tons, seuls deux devraient être notés. La notation des tons

ne devrait pas aller au-delà de deux. Cela permettra de ne pas trop surcharger les articles dans le dictionnaire, à condition que le lexicographe informe les usagers de cette disposition dans les pré-textes.

La solution de noter les tons les moins fréquents ne pourrait être efficace que si la variété des tons retenus est minime. S'il existe plus de trois variétés tonales dans une langue, il serait difficile pour les usagers de deviner quels sont les tons qui ne sont pas notés et comment les différencier de ceux qui le sont pas. La solution à cette situation serait que seuls les deux tons les plus fréquents de la langue soient retenus. Les descriptions linguistiques permettraient d'identifier les deux tons les plus réguliers dans chaque langue. De ces deux tons, le moins régulier devrait être noté dans le dictionnaire.

Une autre approche serait de ne tenir compte que des tons ponctuels principaux (haut, bas et moyen). Il semble que dans la majorité des langues à tons, le ton haut et le ton bas soient les plus faciles à identifier sur le plan auditif, leur courbe mélodique étant constante et courte. Ainsi, le ton moyen qui n'est pas aussi facile à identifier que les deux autres, serait obligatoirement noté accompagné par le moins fréquent des deux autres (le haut ou le bas). Dans cette configuration, seuls deux tons seraient notés, tout en ayant la certitude que les articles du dictionnaire ne seraient pas trop surchargés par les marques tonales. Cette démarche s'inspire en partie de celle adoptée dans les dictionnaires Venda et Zulu dans lesquels les autres tons (montant, descendant, faillotonale, etc.) sont considérés comme de simples variations des tons ponctuels haut et bas.

Marquer les tons dans un dictionnaire peut aussi se faire d'une autre manière que par les signes diacritiques habituels. Dans le cas où le lexicographe estime ne pas vouloir (ou pouvoir) se passer de certains tons, il pourrait, comme je l'ai proposé plus tôt, identifier les deux tons les plus réguliers et ne noter que le moins fréquent des deux. À cela il pourrait ajouter la transcription en italique par exemple, des syllabes portant un autre ton qu'il estime important. Le caractère italique serait ainsi une marque tonale, mais à la différence qu'il ne s'agirait pas d'une diacritique spécifique. Une fois de plus, l'objectif est de marquer le mieux possible la structure suprasegmentale des langues gabonaises, sans pour autant rendre le dictionnaire illisible ou décourageant pour les usagers.

Conclusion

L'intérêt pour les dictionnaires en langues gabonaises va de pair avec le développement de ces langues. Les principaux problèmes auxquels sont confrontés les lexicographes travaillant sur ces langues, est celui de l'orthographe et de la notation de la tonalité. L'objectif de cet exposé était de proposer quelques voies en vue d'aider à résoudre le problème de la notation des tons dans les dictionnaires de langues gabonaises. Ces propositions ne sont pas exhaustives et ne prétendent pas être les seules possibles pour la résolution du problème, mais

elles pourraient fournir une issue qui permettrait de mieux tenir compte, dans l'avenir, de la tonalité dans les dictionnaires de langues gabonaises, contrairement à ce qui s'est fait jusqu'à ce jour.

Il est important de souligner que, quelles que soient les démarches choisies par les lexicographes en ce qui concerne la notation des tons dans les dictionnaires, celles-ci doivent être expliquées en détails dans les pré-textes afin d'informer les usagers et leur permettre d'utiliser au mieux le dictionnaire. Il est évident que sans ces explications, toute démarche serait vouée à l'échec, du moins en ce qui concerne l'objectif de permettre aux usagers de mieux saisir la bonne prononciation tonale des langues gabonaises.

Notes

1. Les références de ces descriptions sont disponibles dans la bibliographie.
2. Les exemples sont transcrits en orthographe des langues gabonaises proposée à Libreville par un colloque d'expert en avril 1999.
3. Les titres de tous les dictionnaires de langues gabonaises consultés sont disponibles dans la bibliographie.
4. Ces dictionnaires et ouvrages parcourus il y a longtemps n'étaient pas à ma disposition au moment où j'ai rédigé cet article.
5. Une polémique existe au Gabon entre certains linguistes et la Fondation Raponda Walker auteur des livres « Rapido-langue ». Les linguistes estiment que ces ouvrages ne sont pas scientifiques parce que la tonalité n'y est pas marquée, alors que la Fondation Raponda pense que mettre les tons pourrait rendre les ouvrages difficilement accessibles pour le commun des usagers non-initié à la linguistique.
6. Le Tsivenda est une langue bantoue d'Afrique du Sud et s'écrit généralement avec un accent circonflexe renversé en dessous du « d ». Cette diacritique n'a malheureusement pas pu être reproduite, mais cette absence ne change pas grand chose pour sa prononciation.
7. Le Zulu est une langue bantoue d'Afrique du Sud.
8. Une collaboration plus étroite et plus efficace entre linguistes et lexicographes gabonais reste, dans ce contexte, plus que nécessaire. Les lexicographes pourront se servir des résultats des descriptions linguistiques pour améliorer le traitement lexicographique des langues gabonaises.

Bibliographie

- Adam, J.J.** 1969. *Dictionnaire Ndumu–Mbebe–Français et Français–Ndumu–Mbebe. Petite flore de la région de Franceville (Gabon). Grammaire Ndumu–Mbebe.* Archevêché de Libreville (Bar-le-Duc): Imprimerie St. Paul.
- Afane Otsaga, T.** 2001. Les dictionnaires bilingues dans les langues gabonaises: Approche méthodologique. *Éléments de lexicographie gabonaise Tome I: 137-159.* New York: Jimacs-Hillman.
- Andeme Allogo, M.F.** 1980. *Esquisse Phonologique du Nzaman de Makokou.* Mémoire de Licence. Libreville: Université Omar Bongo.

- Baylon, C. et P. Fabre.** 1990. *Initiation à la linguistique*. Paris: Nathan
- Bessieux, J.R. (Mgr).** 1847. *Dictionnaire Français-Pongoué / Pongoué-Français*. 2 volumes. Amiens: Lenoël-Hereaurt.
- Biton, A.** 1907. *Dictionnaire Français-Ndumu et Ndumu-Français, précédé d'éléments de grammaire*. Nantes: Imprimerie Bourgeois.
- Cinnamon, J.** 1990. *Faj (Nzaman)*. *Revue Gabonaise des Sciences de l'Homme* 2: 175-188.
- Doke, C.M., D.M. Malcolm, J.M.A. Sikakana et B.W. Vilakazi.** 1990. *English-Zulu / Zulu-English Dictionary*. Johannesburg: Witwatersrand University Press.
- Église Évangélique du Gabon.** 1960. *Dictionnaire Français-Yipounou / Yipounou-Français*. Mouila.
- Fishman, Joshua A. (Éd.).** 1968. *Readings in the Sociology of Language*. La Haye/Paris: Mouton.
- Galley, S.** 1964. *Dictionnaire fang-français et français-fang, suivi d'une grammaire fang*. Neuchâtel: Henri Messeiller.
- Gove, P.B. (Éd. en chef).** 1961. *Webster's Third New International Dictionary of the English Language*. London: G. Bell/ Springfield, Mass.: G. & C. Merriam.
- Hombert, J.M.** 1990. Problèmes phonétiques et phonologiques rencontrés dans les langues du Gabon. *Revue Gabonaise des Sciences de l'Homme* 2: 97-103.
- Kwenzi-Mikala, J.T.** 1980. La phonologie du Yipunu. *Éléments de description du Yipunu*. Lyon: CRLS. Univ. Lyon II.
- 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/H. Teillard.
- Marichelle, C.** 1900. *Dictionnaires Français-Vili*. Loango.
- Marling (R.P.).** 1872. *Dictionnaire fang-français*. New York.
- Martrou, L.** 1924. *Lexique fān-français*. Paris: Procure Générale (des Pères du St Esprit) / Abbeville: Imprimerie Paillard.
- Matthews, P.H.** 1997. *Concise Dictionary of Linguistics*. New York: Oxford University Press.
- Mavougou, P.A.** 2001. La mondialisation et la lexicographie trilingue ou plurilingue au Gabon. *Éléments de lexicographie gabonaise, Tome I*: 160-183. New York: Jimacs-Hillman.
- Mdee, J.S.** 1999. Dictionaries and the Standardization of Spelling in Swahili. *Lexikos* 9: 119-134.
- Mihindou, G.R.** 2001. Apports des missionnaires à la lexicographie gabonaise: Dictionnaires bilingues Fang-Français/Français-Fang; Français-Yipounou/Yipounou-Français; Français-Mpongwé/Mpongwé-Français. *Éléments de lexicographie gabonaise, Tome I*: 7-37. New York: Jimacs-Hillman.
- Mission Catholique du Gabon.** 1925. *Syllabaire Fang*. Mellville (Nord): Imprimerie Ceugnard Le sage.
- Mission Évangélique de Paris.** 1927. *Catéchisme Élémentaire en Fang*. Réédité en 1931 et 1958 par la Société des Missions Évangéliques de Paris.
- Mission Protestante du Congo Français.** 1912. *Ntèn Fan Ôsua: Premier Livre de Lecture en Fan*. Saint-Blaise (Suisse): Samuel Robert.
- Mission Protestante de Talagouga.** 1889. *Nteni osu Nteni Fanwe*. Talagouga/Ndjolé.
- Mission Protestante de Talagouga.** 1898. *Nten Misili Ye biya bi fan Ôsua*. Talagouga.
- Mission Protestante Française.** 1920. *Ezangô Elônga Zi Galwa: Premier Livre de Lecture en Galoa*. London: A.E.F.

- Muroni, J.-M.** 1989. *Petit dictionnaire du bantu du Gabon*. Paris: l'Harmattan.
- Nsuka Nkutsi, F. (Éd.)**. 1980. *Éléments de Description du Punu*. Lyon: Université de Lyon II.
- Ondo Mebiame, P.** 1992. *De la phonologie à la morphologie du Fang parlé à Aboumezok (langue Bantu A.78)*. 2 volumes. Thèse de doctorat. Tervuren: Université Libre de Bruxelles.
- Puech, G.** 1990. Bekwil. *Revue Gabonaise des Sciences de l'Homme* 2: 127-128.
- Raponda Walker, A. et R. Sillans.** 1961. *Plantes utiles: essai d'inventaire et de concordance des noms vernaculaires et scientifiques des plantes spontanées et introduites du Gabon*. Paris: Lechevalier.
- Raponda Walker, A.** 1995². *Dictionnaire Français-Mpongwé*. Class. Africaines.
- Raponda Walker, A.** 1995². *Dictionnaire Mpongwé-Français*. Class. Africaines.
- Raponda Walker, A.** 1995. *Éléments de grammaire Fang*. Libreville: Éd. Raponda.
- Raponda Walker, A.** 1995. *Éléments de grammaire Gisira*. Libreville: Éd. Raponda.
- Raponda Walker, A.** 1995. *Essai de grammaire Tsogo*. Libreville: Éd. Raponda.
- Voltz, M.** 1990. Faŋ (Ntumu). *Revue Gabonaise des Sciences de l'Homme* 2.
- Warmelo, N.J. van.** 1937. *Tshivenda-English Dictionary*. Pretoria: Government Printer.

E-terminology*

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

Abstract: This article addresses computer applications in terminology and terminography. Terminologists aim at creating source language terms and target language term equivalents for new concepts and thereby enhancing communication in various subject areas and domains. The worldwide market for terminological products is gradually expanding. There is an ever increasing need for the elaboration of terminologies and the distribution of terminologically relevant information, either in computerized form or in printed form, to provide available terminological data to as many users as possible. This article provides various strategies for different users to access multilingual terminology resources. On-line availability of terminological data can empower South African language practitioners and subject specialists. Matters dealing with terminological management systems form a vital part of this article. Outcomes, advantages, constraints and requirements regarding e-terminology are given.

Keywords: COMPUTER APPLICATION, ELECTRONIC TERMINOLOGY, E-TERMINOLOGY, HUMAN-COMPUTER INTERFACE, INFORMATION, KNOWLEDGE, KNOWLEDGE-BASED SOCIETY, LANGUAGE BARRIER, MACHINE TRANSLATION, MULTILINGUAL TERM EQUIVALENTS, MULTILINGUAL TERMINOLOGY, TERMINOLOGY APPLICATION

Opsomming: E-terminologie. Hierdie artikel behandel rekenaartoeappings in terminologie en terminografie. Terminoloë stel hul ten doel om brontaalterme en doeltaalekwivalente vir nuwe begrippe te skep en daardeur kommunikasie in verskeie vakgebiede en terreine te verbeter. Die wêreldwye mark vir terminologiese produkte brei geleidelik uit. Daar is 'n steeds groter wordende behoefte aan die uitbouing van terminologieë en die verspreiding van terminologiese relevante inligting, in óf gerekenariseerde óf gedrukte vorm, ten einde die beskikbare terminologiese inligting aan soveel gebruikers moontlik te voorsien. Hierdie artikel bied verskeie strategieë aan verskillende gebruikers om toegang tot meertalige terminologiese bronne te verkry. Intydse beskikbaarheid van terminologiese inligting kan Suid-Afrikaanse taalpraktisyne en vakkundiges bemagtig. Aangeleenthede wat met terminologiese bestuurstelsels te make het, vorm 'n integrale deel van hierdie artikel. Uitkomstes, voordele, belemmerings en vereistes betreffende e-terminologie word aangebied.

Slutelwoorde: ELEKTRONIESE TERMINOLOGIE, E-TERMINOLOGIE, INLIGTING, KENNIS, KENNISGEBASEERDE SAMELEWING, MENS-REKENAARKOPPELVLAK, REKE-

* This article is a slightly adapted version of a paper 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.

NAARTOEPASSING, MASJIENVERTALING, TAALHINDERNIS, TERMINOLOGIETOEPASSING, VEELTALIGE TERMEKWIVALENTE, VEELTALIGE TERMINOLOGIE

1. Introduction

The terminology of each subject field or domain is increasing with each new invention. Information is distributed and knowledge is acquired by means of terminology. The TermNet slogan says it all: *No knowledge without terminology*. Terminology can therefore be regarded as one of the major language resources to assist with communication. The availability of multilingual and polythematic terminology can be an indicator for estimating a nation's scientific, technological and economical achievements and social, cultural and educational development. In the information- and knowledge-based society of the 21st century, knowledge and information are commodities that should be available to the community at large.

Terminology applications cover aspects such as knowledge and language engineering, machine translation, multilingual markets in education, business, science and technology. At present e-commerce (electronic commerce) is a matter of course. There is no reason why terminology practice cannot conduct most of its functions electronically. This article touches on various aspects regarding e-terminology (electronic terminology).

2. Terminological data collections

Terminology has its best tool in knowledge development through terminological and knowledge databases. Science and technology based on advanced knowledge continue to increase in both quantity and quality. In the lively scientific, technological, economic, political, social and academic environments, large numbers of documents are continuously produced. In order to ensure creative and efficient future research, stakeholders have to process a huge amount of information across a wide range of disciplines and domains. Most of these documents are written in languages foreign to the speakers of the indigenous languages in South Africa and many of these have to be translated into the various South African languages. Such an increase in foreign literature, coupled with the limited number of trained translators available, have led decision-makers to seek labour-saving, cost-effective solutions to the translation process. At the same time, developed and developing countries do business with South Africa and documents therefore also need to be translated into foreign languages.

In order to overcome the many difficulties involving language barriers, more attention should be given to the enormous task of the terminologist in supplying multilingual term equivalents for the various new concepts in different subject areas and domains. E-terminology can solve many communication

problems.

3. Computer applications in terminology

The constant generation of knowledge and the fast pace of technological development have created the need for understanding different knowledge organisations in each specific area through the use of technology. Specialised communication has a central hub in terminology.

The number of computing and communication systems is projected to grow continuously within the next few years. From a prospective point of view new applications of the National Termbank will increase. The combination of knowledge transfer, technology and specific knowledge with terminology is expected to create a national information structure situated in the National Termbank, that will enable new applications (e.g. translations, machine translations, etc.), and will open a new set of specialised communications. This will have an impact on the academic, business, and educational world, by advances making human-computer interfaces with environments available. Tools to support those with disabilities (e.g. the blind (Braille; speech recognition), the deaf (e.g. illustrations of the various signs of a sign language) or the handicapped (e.g. adapted keyboards)) will broaden the reach of the computer and communication revolution.

4. Standardised access to lexical and terminological data

In today's fast-paced global e-marketplace, a streamlined multilingual translation workflow will present government departments, institutions and companies with a solid competitive advantage. Standardised terminology contributes significantly to the quality of translations. As government departments, institutions and companies strive towards globalisation, consistent multilingual and polythematic terminology becomes a critical issue.

5. The National Termbank

The Terminology Coordination Section of the National Language Service, Department of Arts, Culture, Science and Technology, is the only language office in South Africa that devotes all its energy and time to terminology work. As the national office for terminology work, its aim is to coordinate all terminological activities in South Africa.

The Terminology Coordination Section renders a terminology service for the advancement of the official languages on a national level. Its vision is to develop, document, standardise and disseminate multilingual terminology to promote and facilitate communication in various subject fields and domains. For this purpose the Terminology Coordination Section maintains an electronic

National Termbank. Terminology work is done in close collaboration with stakeholders, subject specialists, linguists, academics, end-users and business partners (such as the national, provincial and local government, PANSALB and its structures, e.g. the National Lexicography Units (NLUs), the National Language Bodies (NLBs), Provincial Language Committees (PLCs), etc.).

In the 1950s, terminologists started documenting terminology on index cards. In the 1980s, the Coordinating Terminology Board (COTERM) initiated research on various aspects regarding the computerisation of terms and on the terminological management process. After thorough investigations, the national terminology office in 1996 decided to purchase licenses to use the MultiTerm database system. MultiTerm is a professional multilingual terminology management system developed by TRADOS. At present most of the functions and activities of the Terminology Co-ordination Section can be carried out via the electronic media.

In the Terminology Coordination Section, data is usually excerpted manually from documentation, but is immediately documented in the computerised MultiTerm database program. Terminological data that is obtained electronically from discs or compatible programs (such as the Translator's Workbench) can be entered into the database itself. Terminological data created by means of the TRADOS Translator's Workbench is automatically captured in the MultiTerm database.

The TRADOS Translator's Workbench for Windows represents state-of-the-art as far as powerful and flexible management of linguistic reference material is concerned. It offers access to previous translations not only at word level, but also at sentence level. It therefore helps translators who want to be relieved of time-consuming and repetitive tasks.

The Translation Memory (TM) of the Translator's Workbench is a database in which source language sentences are stored together with their target language equivalents. An additional system, WinAlign, matches source language terms or sentences with their target language equivalents. The memory learns unobtrusively while the translator works, and ensures that no repeated term or phrase in this database has to be kept twice or is used contradictorily. Fuzzy searches in the database give the user instant reference to text segments already translated. It can also manage additional information such as the client of the translation project or the domain in which translation units were created. The additional data can later be used to distinguish between different subsets of the data stored in the memory.

Active terminology recognition is one of the facilities built into the Translator's Workbench system. This system functions in collaboration with the MultiTerm database. Known terms already stored in the MultiTerm database are automatically highlighted on the word processor by the Translator's Workbench system. At the click of a mouse button or a keystroke, their translation equivalents can then be pasted into the document being translated. Compound words, morphologically modified forms, and even parts of sentences can be

found with a fuzzy-matching algorithm, even if they do not occur in the same form in the text to be translated. New source language terms and multilingual equivalents can be added to the terminological database through the Translator's Workbench.

The Translator's Workbench gives instant and flexible access to previously translated material because of its translation memory and the concordance features, and assists with active terminology recognition performed in cooperation with MultiTerm.

MultiTerm is integrated into the Translator's Workbench but is also an independent program. This means that terminologists can work on MultiTerm while a translator can also access the terminological database, even when he/she is not working in the Translator's Workbench.

By collecting multilingual and polythematic terminology, the Terminology Coordination Section builds very large corpora on specific subject areas and domains. From these corpora, multilingual terminology lists, technical dictionaries or CD-ROMs can be compiled for dissemination to end-users. Collaboration with related professional and academic institutions, subject specialists and linguists promotes quality control and the standardisation of terminologies. The Terminology Coordination Section aims to coordinate all terminological endeavours in South Africa and to be a clearinghouse for all terminological activities by sharing terminology and terminological knowledge with a multilingual content (see Diagram 1).

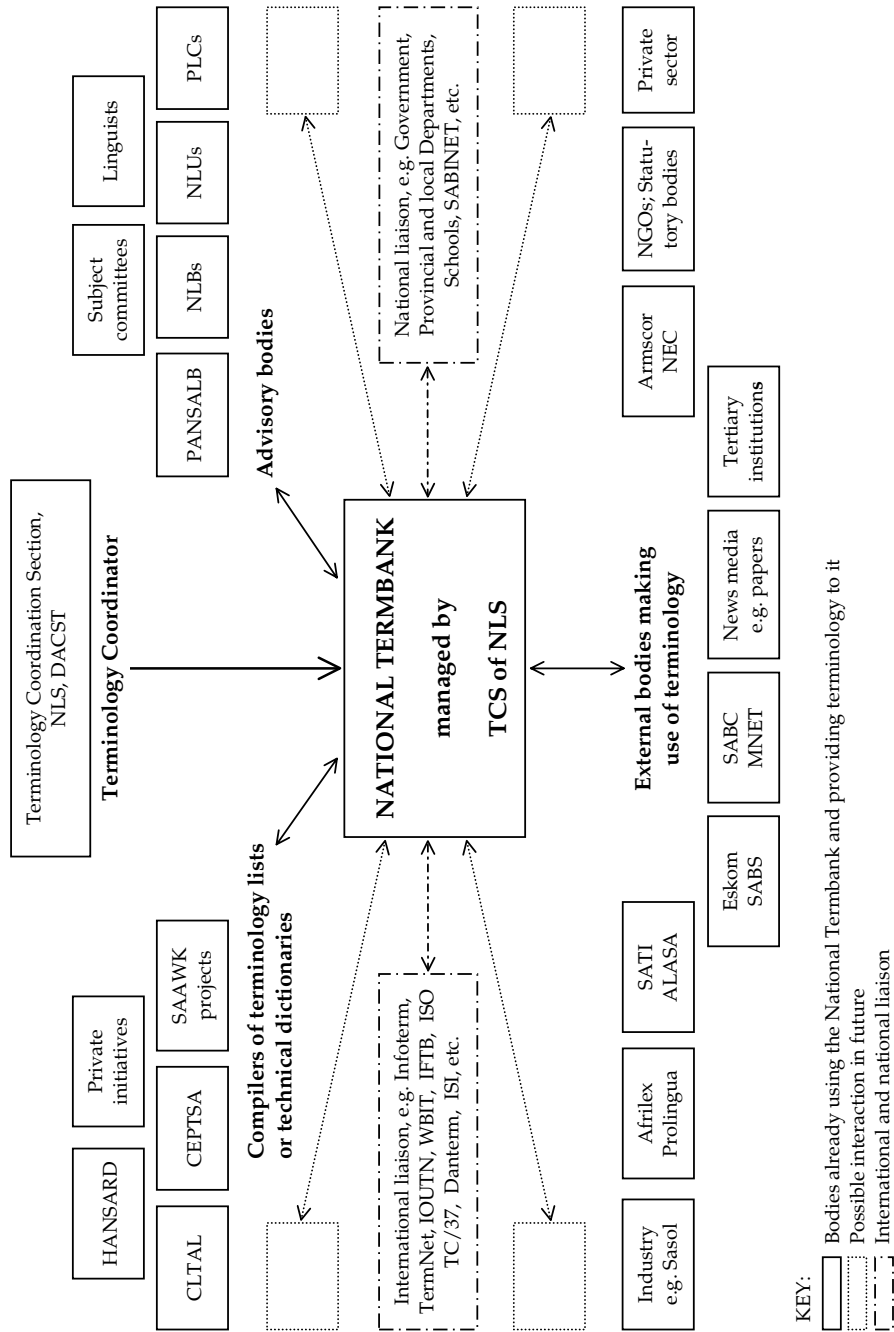
6. Terminology management on the Internet

Knowledge and information can only be transformed into a competitive differentiator when deployed and distributed wisely. Proper terminology management supplies such a tool. A terminology office that focuses on electronic devices as solutions for terminology management, capitalises on the added flexibility and efficiencies offered by proper client-server architectures and Web hosting in particular (Sanders 2001).

According to Crespel (2001), it is possible to provide a central and organised access point for the linguistic sector and to build and develop a linguistic portal to cover the complete range of multilingual activities, disciplines and needs of the sector. The set of objectives offered by a portal should be organised under the following sections:

- On-line consultation of linguistic resources:
 - all kinds of resources: terminology, lexicography, textual corpora, registration of terminology projects, etc.;
 - multilingual resources: no language limitation.
- On-line acquisition of linguistic resources, tools and services:
 - the possibility of viewing and/or testing the resources and tools before acquisition;

Diagram 1: Terminology Coordination



- needs assessment studies and results, e.g. priorities;
- on-line payment mechanisms.
- Technical, scientific and business information:
 - links to existing sites containing newspapers, publications, etc.;
 - "jobs on-line" directory;
 - market players' directory;
 - support for publications of research projects.
- Forum services:
 - inventory and links to existing newsgroups;
 - moderated newsgroups;
 - implementation of specific chats.
- Value-added services:
 - resource creation, validation and adaptation support;
 - consulting services;
 - project development analysis, publications of bids for tenders, etc.

7. Terminology dissemination

The focus of terminology dissemination is shifting from traditional dictionary publishing to efforts of content provision in termbanks. There are new challenges regarding terminology dissemination.

Implications for data management and the marketing and dissemination of terminology are the following:

- new challenges and related specialist dictionary publishing needs with regard to
 - terminological sources, and
 - the validation of dictionary entries;
- quality criteria when managing large heterogeneous terminographical contents;
- transition in the data capturing process;
- a single databank for multiple print and electronic devices;
- globalisation of the market;
- the necessity for co-operating with experts in the field such as information scientists, software developers, terminologists and publishers;
- business strategies to address the consumers' (dictionary users') present needs;
- new co-operation schemes for the dictionary publisher as content provider via intranet and Internet solutions with other partners;
- a considerable requirement for standardisation in the field of lexicography for the harmonisation of terminographical and lexicographical resources to enable data exchange (SQL configurations);
- a successful terminological data dissemination and an opening of new profitable markets without hidden threats concerning quality content

and protection of intellectual property.

8. Convergence of technologies and language engineering

Terminological practice is moving into a world where computers of all kinds, television, the Web, telephones, music media, graphic design and electronic books are evolving toward a seamless and yet heterogeneous environment in which information flows across platform boundaries to guarantee instantaneous terminological access anywhere anytime (Wright 2001).

Terminology is also in the process of becoming mobile: m-terminology is possible in an era of mobile computing (m-computing) and mobile commerce (m-commerce). The electronic means of creating, sharing, re-using and coordinating terminological knowledge for multilingual and transcultural content management paved the way for mobile terminology. It is already possible for terminology users abroad to access various termbanks by means of personal diaries and mobile phones (cell phones). Although this possibility has not been introduced in South Africa yet, there is no reason why it cannot be done. In South Africa one can already conduct business via cell phones (e-commerce) and one can listen to broadcasts from a radio station (e.g. Radiosondergrense) on the cell phone. It is therefore possible to connect the National Termbank with mobile devices to enable access to the extensive terminological variety.

9. Envisaged Web-enabled management model (TRADOS MuWA and TermCo)

The NLS aims to make the terminological information in the National Termbank available to all end-users via the Internet. The NLS started utilising the Trados MultiTerm system in 1996. With this version of the program terminological data can only be displayed in flat files on the Internet. An updated version of MultiTerm with an Internet interface has been available since the end of 2001 and the NLS system has been upgraded to enable users to consult the live data in the National Termbank.

9.1 TRADOS MultiTerm Web Access (MuWA)

The TRADOS MuWA form is used together with MultiTerm to present organisations with access to the National Termbank on the Internet or intranet for wider use. It fits perfectly with the current use of MultiTerm by the NLS.

MuWA offers a robust infrastructure for distributing, accessing and managing terminology over the Internet/intranet. Using a standard Web browser (such as Microsoft Internet Explorer or Netscape Navigator), users can be granted instantaneous, effortless access to the contents of the National Termbank. The termbase is hosted on an SQL database, which is accessed by the cli-

ent browser through the Web server.

MuWA enables centralised administration and management on the Web server. This administration model offers significant benefits to language workers by giving them access to the multilingual and polythematic terminology available in the National Termbank. It also presents mobile engineers or traveling executives with a valuable multilingual communication tool.

The read-only client access enables the terminology administrator to have complete, real-time control over content and revisions.

9.2 TermCo

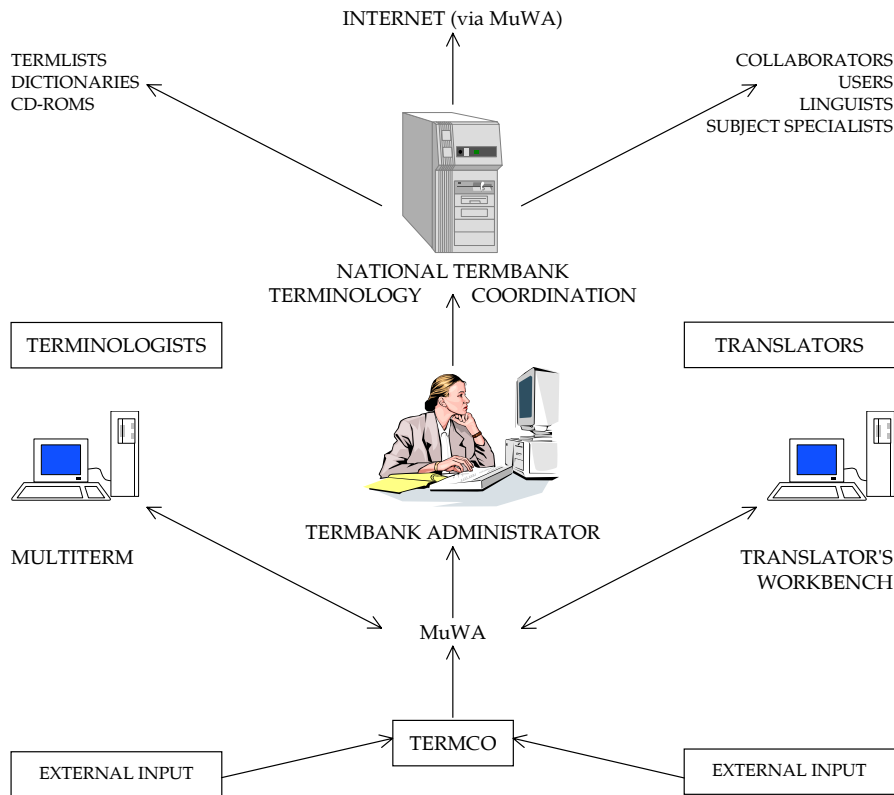
The Unit for Language Facilitation and Empowerment at the University of the Free State holds the TRADOS distribution license in South Africa. This unit recognized the need for terminology contributions to the National Termbank by terminology users. These contributions will have to proceed via an interface between MultiTerm and the Internet. The NLS commissioned the Unit to do a pilot study to establish the actual Internet-based software for a Web-based terminology management system. The idea with this pilot study was to showcase the system's functionality, and a detailed system design will follow to ensure the full and successful implementation of the system.

The envisaged model is based on the expansion of the existing electronic management of terminology by the NLS with TRADOS MultiTerm. The easy use and benefits of the Internet make this model the obvious choice to incorporate as part of a terminology management system. Although not all terminology users have access to the Internet, this is a valuable way of disseminating terminological information. Users may find access through normal information technology support channels, e.g. at libraries.

The envisaged Web-enabled Terminology management model consists of the following components which, when combined, form the model schematically represented in Diagram 2:

- The National Termbank is managed by the Terminology Coordination Section at the NLS.
- Queries are made via TRADOS MultiTerm Web Access (MuWA).
- The user accepts or rejects the terminological information.
- Contributions are made via TermCo or through other methods (e.g. e-mail, fax, telephone) in cases where the information is not available in the National Termbank or the user is not satisfied with the terminological information.
- Data manipulation is done by the terminologists.
- Data is available in the National Termbank.

Diagram 2: The Envisaged Web-enabled Terminology Management Model



9.3 Electronic terminology contributions to the National Termbank via the TermCo interface

The biggest disadvantage of MuWA as used in the context of the envisaged Web-enabled Terminology Management Model is the fact that MuWA only allows query facilities. With the TermCo Interface users will be able to comment on and supply new equivalents for consideration as additions to the National Termbank.

Persons (such as language workers, linguists and subject specialists) who will use and contribute to the envisaged Web-enabled Terminology Management Model will largely determine the success or failure of the model. The contributions will assist in the development of multilingual polythematic terminologies, which will enable the official languages to become functional languages

in all spheres of life. The Terminology Coordination Section of the NLS forms the hub around which the envisaged interface model will turn. The databank manager and the terminologists will actively manage the envisaged model and approve all contributions for inclusion in MultiTerm, MuWA and the National Termbank. There is a dire need for terminology co-ordination and the dissemination of terminological information in South Africa. This model enables proper co-ordination and sharing of existing terms and contributions. The model can also contribute towards the establishment of a multilingual society.

9.4 Manipulation of data

The Terminology Coordination Section of the NLS is the national coordinating body for terminology and will, through liaison with various stakeholders, ensure that the terminology entered in the National Termbank for dissemination is approved by the various governing bodies such as the different PANSALB structures, i.e. the 14 National Language Bodies (NLBs), the 11 National Lexicography Units (NLUs), the 9 Provincial Language Committees (PLCs), the national, provincial and local language offices, the Hansard offices, other linguists and subject specialists. The Terminology Coordination Section aims to be a clearinghouse for terminology documentation, systematisation and dissemination.

On reaching the Terminology Coordination Section via the Internet, the terminological data will have to be manipulated before it can be entered into the National Termbank. The terminology will have to go through the normal processing procedures and will have to be accessed by terminologists, subject specialists and linguists.

Various information fields are available in MultiTerm. When a contribution is made via the TermCo interface, the termbank manager will know who contributed and whether the contribution was made locally, from neighbouring countries or from abroad. Terminology for instance from Lesotho where different spelling and orthography rules for Sesotho apply, will have to be scrutinised by local Sesotho linguists and the National Language Body for Sesotho before it can be disseminated via the National Termbank. When Setswana terminology for instance is contributed, the terminologists responsible for Setswana will have to ensure that first language speakers of Setswana from the North West, the Free State and the Northern Cape and even from Botswana are consulted to ensure consensus on the term equivalents provided.

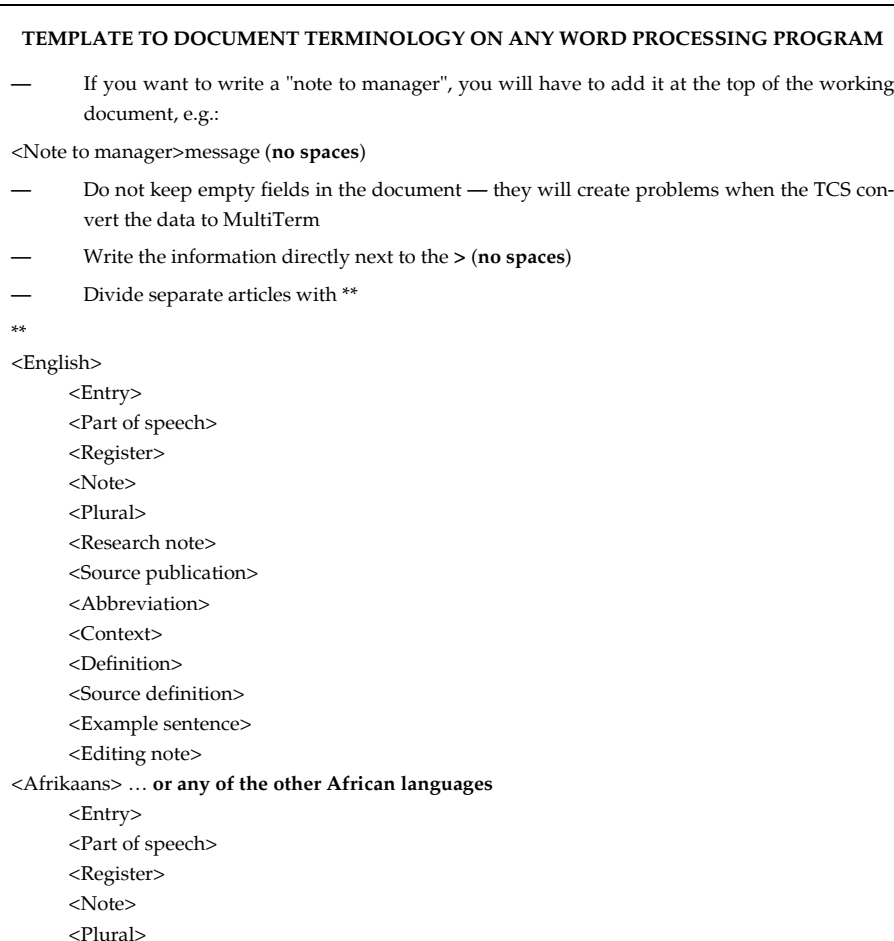
9.5 Contributions via word-processing programs

This interface cannot be implemented as yet. A new version of the Trados MultiTerm program has only been available since the end of 2001. The NLS will now be able to upgrade its terminological management system. In the mean-

time terminology contributions can be made electronically.

People who want to contribute terminology to the National Termbank and who do not have Internet access may also make their contributions electronically by using any standard word-processing software (e.g. Microsoft or Word-Perfect). They have only to use a template (see Diagram 3) supplied by the NLS with an indication of the various information fields to be presented. The terminological information can then be entered next to the relevant information field and a disc with the information can be presented to the Terminology Coordination Section for further manipulation of the data. The data can be utilised without being retyped.

Diagram 3: Template



<Research note>
<Source publication>
<Abbreviation>
<Context>
<Definition>
<Source definition>
<Example sentence>
<Editing note>

**

Contributions can further be made via purpose-specific lexicographical or terminographical databases such as Omnipage which is being used by the National Lexicography Units. Since MuWA and Omnipage are both SQL-based, there would be no problem to exchange data. Any other lexicographical or terminological database program with a SQL configuration will enable data exchange.

10. Outcomes regarding e-terminology

- Linguists, language workers, subject specialists, laypeople, students and learners have easy access to multilingual, polythematic terminology.
- Terminology in original and translated documents can be documented by utilising the TRADOS Translators' Workbench system. Terminology can be excerpted from source language documents and the translated versions in the various target languages. With the aid of the WinAlign program the source language term and target language term can be aligned for documentation in the MultiTerm database.
- Terminological corpora can be used for the parsing of language systems and the development and testing of machine translation programs.
- Multilingual terminology in the National Termbank can be used for machine translation (e.g. the Lexica program).
- Existing published dictionaries, which may already be out of print, can be recirculated by means of CD-ROM publications.
- Terminology can be used for artificial intelligence purposes (e.g. speech recognition).
- The various concepts represented in the National Termbank can be utilised by the deaf or hearing-impaired community since diagrams or illustrations of signs or pictures of the signs can be included in the MultiTerm database.
- Corpora can be built for utilisation by lexicographers (e.g. NLU's). The building of corpora is a very important basic data-creation instrument in an information society. It is considered as the first level infrastructure of language information. A parallel corpus that shows the co-relations between two or more languages can be used for machine translation,

-
- multilingual information retrieval, foreign and indigenous language training and comparative philology.
- Spelling and orthography rules can be tested against the available terminological data (e.g. by NLBs).
 - Spelling control systems can be developed by utilising the electronic data.
 - Studies can be conducted to determine the word-forming and term-creation principles of the various languages.
 - On-line terminological data will be available via mobile devices (e.g. cell phones and personalised diaries).
 - The advantages of on-line access via mobile hand-held devices:
 - they are designed for personalised information;
 - they almost always accompany the users;
 - they are almost always switched on;
 - they make terminological information accessible anytime;
 - users have access to them in real time (updated information is available).
 - Constraints:
 - their screen-size, input capabilities and bandwidth are limited.
 - Requirements:
 - interaction would have to be minimalised;
 - short, meaningful headlines would have to carry relevant information;
 - the screen lay-out would have to be simplistic.
 - Electronic language management through human language technology (HLT) can benefit from e-terminology (e.g. speech recognition, signal recognition, multilingual speech processing, automatic language identification, machine translation, automatic analysis, multilingual information retrieval and computer assistance in text creation and editing).

11. Conclusion

Given the rapid development of economy, culture, science and technology in the world and the value of terminology as communication resource, it is becoming increasingly important to document terminology in such a way that terminological information can be easily retrieved. The multilingual and polythematic terminology should be collected and documented systematically and managed in a coordinated way for proper distribution, publication and application in language and knowledge engineering. The input of collaborators can empower the Terminology Coordination Section to standardise available terminology and to give access to high-quality and reliable terminology. The advantages of e-terminology can promote the terminology management process, the infrastructure and the communication process.

Bibliography

- Baca, M.** 1997. *International Terminology Working Group Minutes, September 5, 1997*. Paris: Offices of the Inventaire Général.
- Crespel, Alexis.** 2001. *GEMA: Gates for Enhanced Multilingual Resource Access*. Paper read at the Fifth TermNet Symposium TAMA 2001, 1–2 February 2001, Antwerp, Belgium.
- Derouin, Marie-Jeanne.** 2001. *From Traditional Dictionary Publisher to Content Provider. Implications for Data-management and Marketing in a New Publishing Era*. Paper read at the Fifth TermNet Symposium TAMA 2001, 1–2 February 2001, Antwerp, Belgium.
- Nyangintsimbi, L.L.T. and J.H. Hattingh.** 2001. *Report on the Findings of the Terminology Survey at the Provincial Language Units (TermCo Project)*. Bloemfontein: University of the Free State.
- Plested Avarez, María Cecilia, Díaz Castrillón and Adriana Lucía.** 2001. *Terminology in the Information Age — Sharing Terminological Knowledge. Colterm Databank: From Concept Systems to Terminological Entry Management*. Paper read at the Fifth TermNet Symposium TAMA 2001, 1–2 February 2001, Antwerp, Belgium.
- Sanders, G.** 2001. *Sharing Terminology: The Benefits of a Web-based Workflow*. Paper read at the Fifth TermNet Symposium TAMA 2001, 1–2 February 2001, Antwerp, Belgium.
- Sonneveld, Helmi.** 1997. *Terminological Principles and Methods in the Translation of Thesauri. International Terminology Working Group Minutes, September 5, 1997*. Paris: Offices of the Inventaire Général.
- Wright, Sue-Ellen.** 2001. *Convergence Technologies and Language Engineering*. Paper read at the Fifth TermNet Symposium TAMA 2001, 1–2 February 2001, Antwerp, Belgium.

An Advanced Dictionary? Similarities and Differences between *Duramazwi ReChiShona* and *Duramazwi Guru ReChiShona**

Emmanuel Chabata, *African Languages Research Institute, University of Zimbabwe, Harare, Zimbabwe (echabata@arts.uz.ac.zw)*

Abstract: In this article a comparative analysis of *Duramazwi ReChiShona* (DRC) and *Duramazwi Guru ReChiShona* (DGC) is made. Both DRC and DGC are monolingual Shona dictionaries compiled by a team of researchers under the African Languages Lexical (ALEX) Project, now the African Languages Research Institute (ALRI). During the compilation process, DRC was known as the General Shona Dictionary and DGC as the Advanced Shona Dictionary. A simple analysis of these titles shows that the dictionaries are similar in some ways and also different in others. The writer tries to show the ways in which DGC is regarded as a more advanced dictionary when compared to DRC. Although the argument of the article is mainly built on those differences which make DGC the more advanced, attention is also paid to the similarities between the dictionaries.

Keywords: ALEX PROJECT, ALRI, DURAMAZWI RECHISHONA, DURAMAZWI GURU RECHISHONA, DICTIONARY, SHONA, HEADWORD, SENSE, MONOLINGUAL DICTIONARY, CORPUS, IDIOM, PROVERB, PITHY SAYING

Opsomming: 'n Gevorderde woordeboek? Ooreenkomste en verskille tussen *Duramazwi ReChiShona* en *Duramazwi Guru ReChiShona*. In hierdie artikel word 'n vergelykende ontleding van *Duramazwi ReChiShona* (DRC) en *Duramazwi Guru ReChiShona* (DGC) gemaak. Sowel die DRC en DGC is eentalige Sjonawoordeboeke, saamgestel deur 'n span navorsers by die African Languages Lexical (ALEX) Project, tans die African Languages Research Institute (ALRI). Gedurende die samestellingsproses was DRC bekend as die Algemene Sjonawoordeboek en DGC as die Gevorderde Sjonawoordeboek. 'n Eenvoudige ontleding van hierdie titels toon dat die woordeboeke op sommige maniere eenders en op ander ook verskillend is. Die skrywer probeer die maniere aantoon waarop DGC beskou word as 'n meer gevorderde woordeboek wanneer dit met DRC vergelyk word. Alhoewel die argument van die artikel hoofsaaklik gebou is op daardie verskille wat DGC die gevorderdste maak, word ook aandag gegee aan die ooreenkomste tussen die woordeboeke.

Slutelwoorde: ALEX PROJECT, ALRI, DURAMAZWI RECHISHONA, DURAMAZWI GURU RECHISHONA, WOORDEBOEK, SJONA, TREFWOORD, BETEKENIS, EENTALIGE

* This article is a slightly revised version of a paper 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.

1. Introduction

The history of Shona lexicography dates back to the last quarter of the 19th century. Missionaries from different mission stations, strategically located in different parts of the country, were the initiators. Early publications from such stations include small dictionaries and vocabulary lists for dialects of the Shona language spoken in communities surrounding the respective mission stations. The dictionaries published were largely bilingual because they were mainly meant for second-language speakers of Shona. Examples of such dictionaries include *English–Mashona Dictionary* (Hartman 1894), *English–Chiswina Dictionary* (Biehler 1906) and *ChiNdau–English and English–ChiNdau Vocabulary* (Wilder 1915). Although many of these dictionaries were not comprehensive, they were at least adequate for the users targeted. However, they were not very useful to native speakers of Shona because they only provided English lexical equivalents to Shona headwords. The editors did not give explanations that enhance the development of the language they described.

Later, more comprehensive bilingual Shona–English dictionaries, which covered vocabulary from all dialects of Shona were compiled. This was after Doke's (1931) recommendation for the unification of all Shona dialects. Dictionaries such as the *Standard Shona Dictionary* (Hannan 1959) and *Duramazwi* (Dale 1981) were published. These dictionaries are bigger, both in terms of the number of pages covered and the number of headwords defined. Unlike in earlier dictionaries where dialects of Shona were treated as different "languages", headwords and senses for these two dictionaries came from all five dialects of the Shona language, namely Zezuru, Karanga, Manyika, Ndau and Korekore. Today these dictionaries are still being used as reference works for Shona.

Work on monolingual Shona lexicography only started with the launch of the ALLEX Project in September 1992. Through the Project, now transformed into ALRI, the researchers published *Duramazwi ReChiShona* (DRC) in 1996 and *Duramazwi Guru ReChiShona* (DGC) in 2001. As suggested in its title, the second dictionary is a more advanced lexicon compared to the first one. This article discusses the ways in which DGC is considered a more advanced dictionary in comparison with the former. The differences are discussed under the following subheadings: headword selection, sense selection, defining formats and the structure of the dictionaries. The article also describes some of the similarities between the two dictionaries.

2. Differences

2.1 Headword selection

One of the most important stages in the compilation of dictionaries is that of

selecting the words or phrases to be entered as headwords. The decisions to be taken on the type and number of headwords to be included or excluded are influenced by a number of factors, such as the purpose of the dictionary, its intended size as well as its target audience. For example, the issue of the target users is of paramount importance, for it is obvious that a dictionary compiled for primary-school children cannot be expected to be similar to the one intended for students at tertiary institutions. The complexity or simplicity of headwords and other lexicographic information to be included in a dictionary depend on the intended users of the dictionary.

The target groups for DRC and DGC are different. DRC was mainly meant to fulfil the needs of lower secondary school learners. Because of this, it was supposed to be a reference work containing basic Shona vocabulary. The dictionary was also an experiment in monolingual Shona lexicography and was therefore supposed to be relatively small and manageable. On the other hand, DGC was "intended to be a comprehensive reference work, which will serve as a resource for more advanced users, especially those at higher secondary and tertiary education levels" (Chitauro-Mawema 2000: 209). What this entailed was that in addition to all the headwords in DRC, many more had to be selected for DGC. For example, terms from specialised fields, which were not part of DRC, were to be entered in DGC.

Most of the additional headwords for DGC came from the Shona corpus, of which a list of all the items in it was made, and in which the researchers searched for new headwords. Other sources of new headwords included written materials in specialised subject areas. In fact, the selection exercise was comprehensive for it included words from all word categories and also from all spheres of life. Notable additions include phrasal headwords, that is, idioms, proverbs and pithy sayings. Svensen (1993: 108) defines an idiom as "a fixed group of words with a special meaning which is different from the meanings of the individual words". Like idioms, proverbs and pithy sayings are generalising statements whose purpose is to convey certain assertions about life (Svensen 1993: 110). Structurally, they are fixed word combinations that must be shown in full in a dictionary since they mean more than what is implied in their constructions. Because these phrases can only be understood as wholes, they should be presented as such in texts. As a result, they occupy considerable space. This is one of the reasons why they were excluded from DRC. Another reason is that they express meaning in a metaphorical way, and such meanings are captured in long explanations which often need illustrative examples. Zgusta (1971: 153) notes that such long headwords, which also need long explanations, are usually included in comprehensive dictionaries. It is because of the nature of their form and meaning that the title of the dictionary had to be followed by a subtitle: *rine zvirungamutauro* (with idiomatic expressions).

Another category of headwords added in DGC is technical terms. Svensen (1993: 49) notes that technical language arises as a consequence of constant development and specialisation in the fields of science, technology and sociol-

ogy. New concepts are constantly being defined, and in order to exchange information about them, new linguistic expressions have to be found for them. There are many specialised terms used in technical subjects in education, economics, sport, law, medicine and others. However, despite their origin in the terminology of various technical fields, many of these terms make their way into general language and become known to lay-people. Svensen (1993: 49) notes that general and broad terms within a certain area tend to move more readily across from technical to general language than terms representing specific concepts. Not all technical terms were eligible for selection as headwords for DGC. The eligibility of specialised and technical terms was not based on their importance in their respective subject areas, but on their use in general language. Only those that the editors felt were generally or commonly used in the Shona-speaking community were incorporated as headwords. Those that the editors felt were still restricted to their technical fields were excluded because they were considered unfit for a general dictionary such as DGC. Also selected were some international words used in fields such as science and mathematics. Chitauro-Mawema (2000: 212) defines international words as those "technical words which carry specific unchanging and unambiguous senses in the contexts in which they occur and are used internationally". These are terms that are usually borrowed from other languages through the process of acquiring the respective concepts. Terms such as **ajebhura** (algebra), **ikweta** (equator) and **sirabhasi** (syllabus), for example, were thought of as standard terms everywhere and therefore had to be selected as headwords if the dictionary had to be comprehensive in the true sense.

Some slang words were also selected as headwords. These are informal words that are only used by people who know each other well or those who share the same interests. Usually slang words are ephemeral, their use only lasts for a very short time. However, there are some words in this category that tend to stabilise and become part of the normal and conventional vocabulary of a language. It was because of these words that settle in a language that a decision was made sparingly to allow slang words into DGC. Words such as **mushe** (all right), **bhoo** (all right) and **kanjani** (how are you) were entered as headwords because they have been in circulation for a long time and have become part of the Shona vocabulary. However, the problem of dealing with such words is that decisions on whether a particular word has stabilised or not are rather subjective.

A few words that violate the Shona alphabet system were also entered as headwords in DGC. Examples of such words include ***lita** (litre), ***jeli** (jelly), ***loni** (lawn) which have the letter **l** and ***thiyeta** (theatre) which has the cluster **th**. The letter **l** and the cluster **th** are not acceptable in correct Shona spelling. However, these words were included in the dictionary because this is how they are said or pronounced by Shona speakers. It was felt that leaving them out would mean a loss to the language described. However, an asterisk was added to these headwords to show that they are not yet accepted in the writing sys-

tem. For their meanings, users were also referred to corresponding headwords with the letter **r** in place of **l** and **ti** in place of **th**. These headwords are only acceptable in the writing system but do not accurately reflect how they are pronounced by speakers. The headwords with **l** and **th** were entered with the hope that the Shona language committee would consider incorporating these and other such letters in the Shona alphabetic system since they are commonly used. However, by the time of publication of the dictionary, the decision was still pending.

The addition of headwords from various word categories naturally made DGC bigger than DRC. The DGC is more than double the size of DRC, both in terms of the number of entries and the number of pages covered. Whilst DRC occupies 504 pages, DGC consists of 1 280 pages. In terms of the number of headwords, DRC contains about 16 000, whilst DGC has almost 37 000. On the basis of these statistics, DGC has also become the biggest monolingual dictionary in any African language.

2.2 Sense selection

Like the selection of headwords, the selection of senses for dictionary entries is also determined by the type and size of the dictionary being compiled, its purpose as well as its target users. For DRC these factors played a major role in determining the number and kind of senses included in the dictionary. The definitions provided for the words in this dictionary were basic explanations of their meaning. Subsidiary or extended senses were excluded for two main reasons. Firstly, since the dictionary was meant for learners in their early years of secondary education, the explanations were supposed to be simple and were to be those encountered in daily language use. Secondly, the issue of limited space also played a part. This dictionary was not supposed to exceed a prescribed number of pages and to make sure that it was kept within the required size, only an average of two senses per headword was allowed. This was decided after the realisation that most words in Shona, especially verbs, can carry many senses, some being principal and many more being extended, specialised or metaphorical.

The situation was different in DGC since more space was allotted to this dictionary than to DRC. Emphasis shifted from limitation of senses to making the dictionary as comprehensive as possible. In this respect, it was decided that a headword should carry all the senses it has. To illustrate this, the example of the verb **-bata** (touch) can be taken. This verb has a wide range of meanings, most of which are extensions of its basic sense of "touching". Examples of such senses include those of "working, tightening, catching, intoxicating, strengthening" and others that are derived from the basic sense of getting in contact with something. Because all its meanings could not be included in DRC, only four were provided. These were the ones that the dictionary editors felt were basic and would immediately come to anybody's mind when the verb is utter-

ed. However, following the new decision of including in DGC all the senses of a word, the verb **-bata** was provided with a total of 18 senses (14 more than those that were provided for in DRC). Most of the senses added were those that were left out on the grounds that they were extensions of the basic sense(s) of this verb. Like **-bata**, most headwords were provided with more senses in DGC than in DRC. It was for this reason that the idea of adding global definitions to such headwords was introduced. A global definition can be described as a general definition that is put under a headword which has more than one specific sense. It carries the main idea expressed by a word and it takes traits from other definitions provided under that headword. It describes the basic concept to which the headword refers, that which is inherent in every sense provided under it. To illustrate this, the verb **-chaira** which carries three different senses, can be taken as example. A global definition, *kufambisa kuenda mberi* (cause to move forward), was provided to capture the general idea expressed by this verb. All three explanations that follow describe different ideas to which **-chaira** can refer, but all of them have the element of "moving forward".

Another important area which is related to defining and sense selection is that of illustrative examples. Fox (1987: 137) says "the use of examples forms an integral part of the learning of a word". Examples help in reinforcing the meaning of a word, not by acting as a reformulation of the definition, but by showing how the word is actually used in an appropriate context, a typical grammatical structure and together with words that are normally associated with it. They are usually added to illuminate those definitions that are not clear. There were fewer occasions in DRC where examples were found necessary. Besides the need for saving space, the main reason for this had to do with the kinds of senses provided in this dictionary. Since the given meanings were basic, it was felt that users would not have difficulties in understanding the explanations. On the other hand, there often were cases in DGC where examples were considered necessary, because there were a lot of subsidiary, sometimes closely related senses, that were defined in this dictionary. These senses are specialised and/or metaphoric. They usually carry hidden meanings and are also rarely used in everyday interaction. It is because of their nature that it was felt that examples were needed to illustrate the contexts in which they are used. In this respect, the examples were provided so as to aid the dictionary users' understanding of the meanings of the headwords. The examples were also considered useful in showing the ways in which each sense is different from the other(s).

The link between DRC and DGC with the Shona electronic corpus is worth mentioning. DRC has often been described as corpus-aided or corpus-assisted. Very little evidence was drawn from the Shona corpus during headword selection and defining. This was mainly because the corpus, whose compilation started at the same time as that of DRC, was still very small for any meaningful use as a source of dictionary headwords, senses and illustrative examples. On the other hand, DGC can be described as corpus-based. This is because quite a

substantial number of headwords, senses and illustrative examples in this dictionary came from the corpus. For senses, a concordance programme was employed to provide contexts in which words are used in the corpus. A concordance consists of a list of all occurrences of each word in a text. An analysis of the different contexts in which a word is used yields different senses of that word. Because the editors of DRC usually relied on their memory for headwords and definitions, the tendency was to include only those words that were commonly used as well as providing general definitions that they could easily remember. However, a shift from heavy reliance on a few people's memories to the use of the corpus in DGC made it possible for both common and rare headwords and senses to be recalled and included, thus resulting in a more comprehensive product.

2.3 Structure

Different books, including reference works can be arranged in a variety of ways, depending on the kinds of information that they contain. DRC and DGC differ with regard to the way information is arranged in the two dictionaries. Basically the differences emanate from the additions that were made in DGC. DRC is simply divided into the front matter and the main body of the dictionary. Svensen (1993: 16) refers to the front matter of a dictionary as "instructions of use", describing how the dictionary is organised, how it was compiled as well as how it can be used effectively. The main body here refers to the part that contains headwords and their senses, alphabetically ordered from A to Z.

Whilst the main body of DRC consists of just one section, that of DGC is divided into two, that is, *Chikamu I* (Section I) and *Chikamu II* (Section II). *Chikamu I* contains lexical, multiword lexical units and idioms and their senses. Besides the addition of idioms, which are phrasal in nature, this section may be paralleled to the main body of DRC. On the other hand, *Chikamu II* contains other phrasal headwords, that is, proverbs and pithy sayings. These were excluded from the first section of the dictionary for a number of reasons. Firstly, they were excluded because they had to appear in the dictionary as full or complete statements, starting with capital letters and ending with full-stops. It was felt that the inclusion of complete statements in the same section with other entries would yield an unattractive presentation on the page, especially given the fact that the phrases would not be evenly distributed across the dictionary. Proverbs and pithy sayings were also excluded from this section because they are not usually found in smaller dictionaries. Lastly, and arguably most importantly, it was difficult to decide under which headwords respective phrases should be placed. For idioms it was relatively easy since each one was listed under its main verb, because the main verb in an idiom is fairly predictable. However, the reverse is true of proverbs and pithy sayings where there are usually more than one word which can be regarded as the main word.

Unlike DRC, which only has the front matter and the main body, DGC

also has a back matter, that part which comes after the main alphabetical listing of the dictionary. In this section, information which is not needed for the correct use of the dictionary, but which may be useful in other ways, is appended. The appendices provide some practical information that can be utilised by the dictionary user in his/her daily life. The kinds of information appended to DGC include systematic lists of appellations of chiefs in different parts of Zimbabwe and their respective totems, names of African countries and currencies used in these countries, measurements, weights, etc. Whilst information placed under entries in the main body of a dictionary can be regarded as linguistic or as serving linguistic functions, that in the back matter is general, cultural or otherwise.

2.4 Defining formats

Words in a dictionary are not defined in a haphazard manner. Instead, lexicographers have to follow systematic ways of defining which they develop even before they start explaining or describing what the words mean. These systematic ways used have often been referred to as defining formats and have also been described as laid-down principles that "provide guidelines or paths that a definer follows when defining" (Chabata 1995: 2).

Defining formats are developed for each class of words in a dictionary and a number of defining methods or principles can be identified. However, we will only refer to two methods that are relevant to our discussion. These are the traditional method of defining and that of the Collins Birmingham University International Language Database (COBUILD). In short, in the traditional method all the information provided for each headword is contained in the definition and words are defined outside their contexts of use. In fact, statements are made about what the words mean, but very little is said about how they are used (Hanks 1987: 121). This method is usually followed for reasons of saving space. On the other hand, when the COBUILD method is used, words are defined within the contexts of their use. Hanks (1987: 118) notes that the use of this method results in explanations that consist of two parts. The first part represents a departure from the traditional method in that it actually places the word being defined in a typical structure, thus showing its use. For example, the first part of the definition for **banga** (knife) would start with the headword itself, that is, **banga** ... The second part explains the meaning. Because the words being defined form part of their explanations, the practice tends to occupy a lot of space. However, the method is user-friendly.

The defining formats developed for DRC and DGC can be described as a judicious mixture of the traditional method and the COBUILD method. Although a mixture of the two defining methods was adopted for the two dictionaries, a closer look at the formats shows that those for DGC have moved closer to the COBUILD method than those used in DRC. For example, in DRC a principle was adopted that only nouns with at most three syllables were to be

part of the headword's definition(s). In other words, only these nouns were supposed to be defined using the COBUILD method, while those with more than three syllables were not supposed to be included or repeated in the definition. This measure was adopted as a way of saving space. In DRC, space saving was of paramount importance, especially if it is considered that, in Shona, for example, one can coin very long nouns by joining word forms from different word categories. Examples of such nouns include **kadende-mafuta** (small calabash of oil), **chitundu-mutsere-mutsere** (rocket) and **chibaya-mahure** (prickly plant). For such headwords, the traditional method was seen as more suitable since it would help keep definitions short. Although the idea of saving space was still upheld in DGC, it was felt that there was more space allotted to this dictionary in comparison to its predecessor. Because of this, user-friendliness was considered more important than space. This, for example, led to an increase in the number of syllables from three to four for a noun to be repeated in a definition.

It is also important to note that in addition to the defining formats used in DRC, more were created for DGC. The defining formats added are of two kinds. The first is that of formats developed to provide for new categories of headwords such as proverbs and pithy sayings introduced in DGC. The other kind is that of formats developed for categories of headwords which existed in DRC but were intended to augment the already existing ones. For example, new and more explicit formats were developed for nouns. The new formats developed can be regarded as more explicit in giving typical contexts in which particular nouns are used. To illustrate this, the noun **jobhukadhi** (job card) can be taken. This noun has different senses in different contexts. For example, in industry it is a piece of paper on which work intended for each day is written. In garages that repair machines, it is used to refer to a piece of paper on which machine defects as well as the cost of repairing them are recorded. As a way of distinguishing and describing the meanings, the defining format that was used is as follows:

1. *Mumakambani, jobhukadhi* ... (In companies, a job card is ...)
2. *Mumagaraaji, jobhukadhi* ... (In garages, a job card is ...)

When this defining format is applied, one can clearly understand the different, but closely related senses that this word has when used in different contexts.

3. Similarities

So far the focus of this article has been on the differences between DRC and DGC. This is not to suggest that there are no similarities between them. On the contrary, they share a number of similarities which will be looked at in this section.

One of the most outstanding similarities between DRC and DGC is the fact

that they are general monolingual dictionaries. They are described as "general" because they are concerned with the Shona language as it is generally used in the communities where it is spoken. They are also monolingual dictionaries, so far the only ones in Shona. According to Zgusta (1971: 249), in monolingual dictionaries the object of description (that is, the headword) and the descriptive instrument (that is, the meaning or explanation) should be the same. Unlike bilingual dictionaries whose aim is to help in translating from one language into another, the aim of these dictionaries is to describe the Shona language in a way that enhances its development.

Another important point is that these dictionaries were compiled by native speakers of Shona who also happen to be language experts. The dictionaries tend to provide adequate descriptive and/or explanatory senses for headwords. The senses are generally better when compared to the lexical equivalents provided for in the bilingual dictionaries which preceded them. The senses describe phenomena and/or events in a way that helps the user to conceptualise a thing that he/she has not even seen before. This contrasts with dictionaries which were compiled by those we can refer to as 'language tourists', that is, people who were neither mother-tongue speakers nor language experts of Shona. This could be the reason why their dictionaries, for example, those by Hannan (1959) and Dale (1981), leave a lot to be desired in terms of the way they document and also describe the language.

DRC and DGC can also be characterised as synchronic. According to Zgusta (1971: 202), the task of diachronic dictionaries is to deal with the development of the lexicon, whereas the purpose of synchronic dictionaries is to deal with the lexical stock of a language at one stage in its development. Words and their senses in DRC and DGC were collected and presented as they are used or understood in the Shona language community today. There is a field for etymological information on the database for DGC, but it was suppressed during the final stages of editing the dictionary. Although very little historical information was incorporated, this is not to suggest that these dictionaries contain no such information. A few names of people and places that are historically significant, were in fact included as headwords. However, when such headwords were defined, not every detail about them was given; only the basic information was provided. For example, there are two entries for the noun **guruuswa**. The first is defined in general as a forest with thick and long grass. The definition for the second refers to the historically significant place in East Africa where the Shona people once settled on their way from their original abode in West Africa to the present-day Zimbabwe. No information in this definition mentions the socio-economic and political way of life of the Shona people in the *guruuswa* area, information which is usually given in cases where historical detail is prioritised.

Unlike the bilingual Shona–English dictionaries already referred to, in which information on the dialectal sources of headwords and senses is provided, no such details are given in both DRC and DGC. Headwords and senses

are drawn from all areas where Shona is spoken and their sources are not shown in the dictionaries. In some cases, a single headword could have different meanings in different dialectal areas. In these cases, Shona–English bilingual dictionaries would indicate that a specific headword or sense is commonly used in a particular dialectal area. This lexicographic practice, though important in giving such comprehensive details, has the disadvantage of highlighting the differences among varieties of one and the same language. Research carried out by the ALLEX Project in the eastern parts of Zimbabwe prior to the publication of DRC, showed that speakers in this region identify themselves firstly as speakers of particular dialects before being speakers of union Shona. They were very keen to know whether vocabulary from their dialect areas was included in the dictionary. They were delighted to discover that words they use in their daily lives were actually given as headwords.

If the scenario described above is anything to go by, a dictionary that marks dialectal sources of headwords and senses could be described as a catalyst for division among speakers of a language. In fact, it is against this background that the compilers of DRC and DGC decided to exclude such information. It was felt that vocabulary in the dictionaries would be given equal status and treatment. In this way, the dictionaries would become melting-pots in which all Shona words are thrown to result in one product, standard Shona. The dictionaries can therefore be viewed as agents of unification between speakers of different geographical locations in the Shona-speaking communities.

4. Conclusion

In this article, a comparison between DRC and DGC was made. The writer has tried to show the ways in which the latter is considered a more advanced dictionary than the former. DGC is more advanced in two ways, that is, in terms of its size and the presentation of its meanings. The headword and sense selection for this dictionary was more comprehensive for it includes language used in all spheres of life. DGC is also the first dictionary in the history of Shona lexicography to include phrasal headwords such as proverbs, idioms and pithy sayings. The improvement in the quality of presentation of meaning can be accounted for by the use of the Shona corpus as well as the experience the editors gained from having worked on DRC.

The article also discussed a few of the similarities shared by the two dictionaries. That DRC and DGC would have much in common can be surmised from the mere fact that the same team compiled both of them.

References

- Biehler, E. 1906. *English–Chiswina Dictionary with an Outline Chiswina Grammar*. Roermond: J.J. Romen and Sons.

- Biehler, E.** 1950. *A Shona Dictionary with an Outline Shona Grammar*. Cape Town: Longmans/Chishawasha: Jesuit Fathers.
- Chabata, E.** 1995. *Advantages and Disadvantages of Using the Defining Formats Developed for the First Monolingual Shona Dictionary*. Unpublished B.A. Hons. Thesis. Harare: University of Zimbabwe.
- Chimhundu, H. (Ed.)**. 1996. *Duramazwi ReChiShona*. Harare: College Press.
- Chimhundu, H. (Ed.)**. 2001. *Duramazwi Guru ReChiShona*. Harare: College Press.
- Chitauro-Mawema, M.B.** 2000. Challenges Encountered in the Compilation of an Advanced Shona Dictionary. *Lexikos* 10: 209-224.
- Dale, D.** 1981. *Duramazwi: A Basic Shona-English Dictionary*. Gweru: Mambo Press.
- Doke, C.M.** 1931. "Report on the Unification of the Shona Dialects" Carried out under the Auspices of the Government of Southern Rhodesia and the Carnegie Corporation. Hertford: Stephen Austin.
- Fox, G.** 1987. The Case for Examples. Sinclair, J.M. (Ed.). 1987: 137-149.
- Hanks, P.** 1987. Definitions and Explanations. Sinclair, J.M. (Ed.). 1987: 115-136.
- Hannan, M.** 1959. *Standard Shona Dictionary*. London: Macmillan.
- Hartmann, A.M.** 1894. *English-Mashona Dictionary*. Cape Town: Juta.
- Sinclair, J.M. (Ed.)**. 1987. *Looking Up: An Account of the COBUILD Project in Lexical Computing and the Development of the Collins COBUILD English Language Dictionary*. London: Collins.
- Svensen, B.** 1993. *Practical Lexicography: Principles and Methods of Dictionary-Making*. Oxford/New York: Oxford University Press.
- Wilder, G.A.** 1915. *Chindau-English and English-Chindau Vocabulary*. Rhodesia: American Board Mission.
- Zgusta, L.** 1971. *Manual of Lexicography*. The Hague: Mouton.

'n Praktiese benadering tot die samestelling van 'n tweerigtingvakwoordeboek met Afrikaans en Italiaans as taalpaar

F. de Foglio (*defoglio.hum@mail.uovs.ac.za*) en H.J. Lubbe (*fgtlw.rd@mail.uovs.ac.za*), *Eenheid vir Taalfasilitering en -bemagtiging, Universiteit van die Vrystaat, Bloemfontein, Republiek van Suid-Afrika*

Opsomming: Deur die tempo van globalisering op handelsgebied ontstaan daar toenemend 'n behoefte aan tweerigtingvakwoordeboeke. Afrikaans is een van die hooftale in Suid-Afrika, en die landbousektor van die land is oorwegend Afrikaanssprekend. Aangesien een van die outeurs Italiaanssprekend is, by vertaling betrokke is en die leemte aan 'n tweerigtingvakwoordeboek Afrikaans en Italiaans ervaar het, is kriteria bepaal vir die samestelling van so 'n vakwoordeboek op die gebied van die landbou-ekonomie tussen Afrikaans, 'n Germaanse taal, en Italiaans, 'n Romeanse taal. In agtereenvolgende afdelings word die algemene kriteria vir die samestelling van so 'n woordeboek uiteengesit. Daarbenewens word voorafgaande werk wat verrig moet word, bespreek, asook die keuring van materiaal. Ten slotte word, aan die hand van gepaste voorbeelde, die leksikografiese beginsels toegepas. 'n Bylaag wat uit twintig lemmas bestaan, dien as voorbeeld van hoe so 'n vakwoordeboek daar behoort uit te sien.

Sleutelwoorde: AFRIKAANS, ITALIAANS, LEKSIKOGRAFIE, TERMINOGRAFIE, TERMINOLOGIE, TWEERIGTING, TWEETALIG, VAKLEKSIKOGRAFIE, VAKWOORDEBOEK, WOORDEBOEK.

Abstract: A Practical Approach to the Compilation of a Bi-directional Technical Dictionary with Afrikaans and Italian as Language Pair. Due to rapid globalisation in the commercial sphere, there is an ever-increasing need for the compilation of bilingual technical dictionaries. Afrikaans is one of the principal languages in South Africa, and the agricultural sector of the country is predominantly Afrikaans-speaking. Since one of the authors is Italian-speaking, is involved in translation, and has experienced the need for a bi-directional technical dictionary Afrikaans and Italian, criteria were determined for the compiling of such a technical dictionary in the field of agricultural economics between Afrikaans, a Germanic language, and Italian, a Romance language. In subsequent sections, the general criteria for compiling such a dictionary are outlined. In addition, the work that must precede this process is discussed, as is the selection of material. In conclusion, employing appropriate examples, the lexicographical principles are applied. An appendix comprising twenty entries serves as an example of what such a technical dictionary should look like.

Keywords: AFRIKAANS, BI-DIRECTIONAL, BILINGUAL, DICTIONARY, ITALIAN, LEXICOGRAPHY, LEXICOGRAPHY FOR SPECIAL PURPOSES, TECHNICAL DICTIONARY,

TERMINOGRAPHY, TERMINOLOGY.

1. Inleiding

'n Oorheersende kenmerk van die huidige tydsgewrig is globalisering. Handelsgrense word toenemend opgehef, en in- en uitvoer neem toe. Dat Suid-Afrika nie hierdie tendens sal ontsnap nie, is vanselfsprekend. Na die opheffing van sanksies het Suid-Afrikaanse handel met die Europese Unie veelvuldig toegeneem. In hierdie proses speel doeltreffende kommunikasie uiteraard 'n belangrike rol. Dis 'n feit dat die Europese Unie 'n meertalige taalbeleid het, en dat nie aangeneem moet word dat Engels outomaties die voorkeurhandelstaal is nie. Trouens, suksesvolle handel met oorsese lande vereis meer as die gewone kennis van Engels. Kennis van die teikengemeenskap se kultuur word vereis, en sy waardes, manier van onderhandel, behoeftes en begeertes moet ook in ag geneem word. Die Australiese ekonoom McCallum haal die Duitse Minister van Ekonomiese Sake aan: "If you wish to BUY from us, there is no need to speak German, but if you wish to SELL ..." (McCallum 1990: 54). Hy (McCallum 1990: 54) beweer dan ook: "Selling is facilitated by the use of the buyer's language."

Die doel van hierdie bydrae is om praktiese riglyne te verskaf vir die samestelling van 'n tweerigtingvakwoordeboek, spesifiek vir die taalpaar Afrikaans en Italiaans.

Hoewel Afrikaans as 'n relatief "klein" taal gekarakteriseer kan word, is dit een van die hooftale in Suid-Afrika wat deur 14,4% van die bevolking gepraat word, teenoor 8,6% Engels, 22,9% Zulu en 17,9% Xhosa (Statistics South Africa 1998). In teenstelling met die laasgenoemde twee tale beskik Afrikaans oor 'n uitgebreide vakwoordeskat.

Deur sy verbintenis met Nederlands, sowel diachronies as sinchronies, verleen Nederlands toegang, nie net tot één van die lande van die Europese Unie nie, maar tot al sestien (Lo Bianco 1996: 48):

The interesting consideration about the European Union is that to enter one economy, potentially and at least formally and legally, is to enter all 16. They are interlinked. The same trade, tariff and packaging rules apply across the twelve — now sixteen — countries.

Lo Bianco (1996: 48) sê verder dat ander geleenthede vir Afrikaans wag in dele van Suid-Amerika (die Nederlandse Antille en Suriname), as gevolg van hierdie lande se koppeling met die groot Noord-Amerikaanse ekonomieë, en in Indonesië as poort tot die vinnig groeiende ekonomieë in Asië:

These links are not facilitated exclusively through foreign languages or English.

D.E. Barclay van die VSA, 'n gesaghebbende oor kwessies van globalisering,

het tydens 'n besoek aan Suid-Afrika die rol van Afrikaans in die handelswêreld benadruk (Barclay 2000: 6-7):

Afrikaanssprekendes het 'n wanbegrip van die magte van globalisering enersyds en van die voortgaande en groeiende belangrikheid van "plaaslike" tale en kulture andersyds. Weens dít ken hulle gans te veel waarde toe aan die rol van Engels sowel in die globale ekonomie as in Suid-Afrika. [...] Die globale ekonomie staan in die teken van diversiteit, nie uniformiteit nie. 'n Groot deel van die beking van oorsee reis, bestaan in die ontmoeting met vreemde tale. Ek sou net sowel kon tuisbly as ek net Engels wou hoor en lees. Daarom vind ek dit vreemd, trouens dit is 'n verleentheid om in Suid-Afrika te kom en net Engelse wynetikette te kry, of net Engelse toeristebrosjures aan die Kaapse Weskus.

In die lig van die feit dat die Suid-Afrikaanse landbousektor hoofsaaklik Afrikaanssprekend is (Statistics South Africa 1998), lê die voorbeelde wat vir die bespreking byeengebring is, op die gebied van die landbou-ekonomie. Afrikaanssprekende boere is oor die algemeen nie hoofsaaklik met bestaansboerdery gemoeid nie, maar bemark hulle produkte nasionaal en toenemend internasionaal.

As tweede lid van die taalpaar is Italiaans gekies omdat een van die outeurs 'n Italiaanse moedertaalspreker is, by vertaling betrokke is, en die leemte aan 'n tweerigtingvakwoordeboek Afrikaans en Italiaans ervaar het. Sover bekend, is daar geen leksikografiese studie van die verhouding tussen Afrikaans, 'n Germaanse taal, en Italiaans, 'n Romaanse taal nie. Slegs eentalige hulpmiddele bestaan wat dubbele vertaling (Afrikaans na Engels, en Engels na Italiaans) noodsaak. Tweetalige terminologiese databanke en vakwoordeboeke kan die vertaalproses geweldig versnel.

In die volgende afdelings word agtereenvolgens die algemene kriteria vir die samestelling van 'n tweerigtingvakwoordeboek bespreek, voorafgaande werk wat moet verrig word, die keuring van materiaal, en ten slotte word, aan die hand van voorbeelde, die leksikografiese beginsels toegepas. 'n Bylaag wat uit twintig lemmas bestaan, dien as voorbeeld van hoe so 'n vakwoordeboek daar behoort uit te sien.

2. Algemene kriteria vir die samestelling

Op 'n vraag aan die beeldhouer Rodin hoe om 'n beeld te maak, het hy geantwoord: "Neem 'n blok van steen en sny uit wat onnodig is" (Magay 1984: 221). 'n Gelyksoortige tegniek word deur die vakleksikograaf tydens die opstel van 'n vakwoordeboek gevolg. Taal vir Spesiale Doeleindes (TSD) is die soort taal wat in 'n vakwoordeboek opgeneem word, en die res, naamlik Taal vir Algemene Doeleindes (TAD) kan "uitgesny" word. Svensén (1993: 48) stel die TAD gelyk aan die taalgebruik van die gemiddelde moedertaalspreker. Dit gaan nie net oor woorde, uitdrukkings en sintaktiese patrone wat in die aktiewe gebruik van die meeste mense voorkom nie, maar ook oor wat in die passiewe kennis

aanwesig is, m.a.w. die taal wat die meeste gebruikers herken en in bepaalde kommunikasietoestande gebruik. Die vaktaal (TSD), daarenteen, is volgens Alberts (1988: 3) die taal van bepaalde beroepsgroepe of vakdissiplines wat seker kenmerke toon wat dit gedeeltelik van standaardtaal onderskei en daarvan vaktaal maak.

Vaktaal moet onder meer van twee aspekte rekenskap gee, naamlik 'n sosiolinguistiese aspek ("group language") en 'n ónderwerpverbonde aspek ("technical language") (Svensén 1993: 48). Vir Magay (1984: 221) is die samestelling van woordeboeke 'n eindelose reeks van besluite. Die eerste stap sou die beslissing omvat van wat ingesluit of uitgesluit behoort te word. In hierdie proses is drie faktore belangrik:

- die grootte van die woordeboek,
- die doel van die opstelling, en
- die bruikbaarheid daarvan.

Volgens Zgusta (1971: 218-219) behoort 'n omvattende generiese woordeboek van honderd duisend en meer tot ongeveer vier honderd duisend lemmas te bevat as daar ook gespesialiseerde terminologie ingesluit word. 'n Middelslagwoordeboek sal tussen veertig duisend en sestig duisend lemmas bevat, terwyl 'n sakwoordeboek minder as twintig duisend lemmas bevat, en 'n klein woordeboek nie minder as tien duisend nie. Slegs gespesialiseerde woordeboeke (of vakwoordeboeke) kan kleiner as tien duisend terme wees en hulle sal dus gelyk wees aan woordelyste wat aan leksikografiese kriteria voldoen (Zgusta 1971: 220). Die bylae in De Foglio (2001) ressorteer ook hieronder.

Ongeag die grootte van die woordeboek, word die hele leksikon van 'n bepaalde taal benodig by die opstel daarvan. Hoe kleiner die woordeboek, hoe moeiliker is dit om lemmas te selekteer. Kritici en gebruikers (veral vertalers) koppel gehalte van die woordeboek aan dit wat daarin ontbreek (Magay 1984: 220). Dus behoort die gebruikersbehoefes in ag geneem te word by die opstel van 'n woordeboek (vgl. hieroor Gove 1967: 5).

Oor die doel van die opstelling, m.a.w. die potensiële vaktaalgebruiker, moet vooraf helderheid verkry word. Cluver (1992: 40-42) onderskei drie belangrike groepe vaktaalgebruikers en dui aan watter soort definisies benodig word in vakwoordeboeke. Die drie groepe is studente, vakspecialiste en vertalers. Eersgenoemde groep moet verstaan dat begrippe in hiërargieë ingedeel kan word. Die student moet byvoorbeeld presies weet wat die verskil is tussen *atmosfeer* (bowegeskikte term), *mesosfeer* (ondergeskikte term: tussen die bowegeskikte en die newegeskikte term) en *ionosfeer* of *eksosfeer* (newegeskikte term, onderdele van die bowegeskikte term). Hiermee word bedoel dat die student meer ensiklopediese inligting nodig het. Die vakspecialis stel meer belang in die mees resente terme. Hy weet wat met 'n term bedoel word en soek in 'n vakwoordeboek eerder sinonieme en riglyne van die gebruik van die sinonieme as definisies. 'n Ander behoefte van die vakspecialis wat betref 'n vakwoordeboek is verwysingsbronne. Dit gaan nie oor bronne met definisies/omskry-

wings nie, maar eerder oor bronne wat die nuwe tendense in 'n bepaalde vakgebied aandui. Cluver (1992: 42) voeg dan by dat so 'n woordeboek wat die vereistes van die vakspecialis in ag neem, 'n kort lewe het.

Die laaste groep, die vertalers, hou hom besig met tweetalige vakwoordeboeke. Soos dit in die algemene leksikografie die geval is, is hierdie groep ook vir die vakleksikografie die veeleisendste gebruikersgroep. Vertalers benodig definisies, eksakte en erkende terminologie en voorbeelde wat die gebruik in verskillende registers aantoon. Volgens die behoeftes van die vertaler moet 'n vakwoordeboek wat byvoorbeeld die geneeskunde behandel, terme soos *abdomen*, saam met sy verskillende sinonieme (*buik, maag, pens*) bevat, en naas hierdie inligting ook die gebruikregister (Cluver 1992: 42). In die praktyk is die samesteller van 'n gespesialiseerde vakwoordeboek, soos in die geval van die bylae in De Foglio (2001), terselfdertyd die gebruiker. In so 'n geval is die opsteller op homself aangewese vir die identifisering van gebruikersbehoefes. Aanvanklik moet so 'n woordeboek liefs nie in die kommersiële mark aangebied word nie, maar eerder eers aan die beperkte kring vertalers beskikbaar gestel word. Daar behoort 'n wedersydse diensverhouding tussen die leksikografie en die vertaalwese te wees. Die vertaler behoort voortdurend terugvoer en riglyne aan die leksikograaf te verskaf om sodoende die leksikografiese werksaamhede en die gepaardgaande eindproduk so effektief moontlik te hou. Die leksikograaf sou veral aandag aan die vereistes van die vertaler kon gee en volkome afhanklik van hulle riglyne bly. Die vertaler is in der waarheid die grootste gebruiker van tweetalige woordeboeke en veral van vakwoordeboeke (vgl. Gouws 1991: 37 vir die moeilik oplosbare probleem van die bevrediging van die vertaler se behoeftes).

Die doel met die samestelling van 'n tweerigtingvakwoordeboek vir Afrikaans en Italiaans is spesifiek daarop gerig om vertalers behulpsaam te wees om die kommunikasieproses tussen die Suid-Afrikaanse en Italiaanse landbou-sektor te vergemaklik. As die gebruiker verdere inligting op die gebied van meer algemene lemmas of selfs grammatikale inligting benodig, sal hy dit moet gaan naslaan in 'n tweetalige generiese woordeboek.

Die bruikbaarheid van die vakwoordeboek, die derde faktor dus, is veral gerig op die aanwending deur die vertaler wat met die taalkombinasie Afrikaans-Italiaans werksaam is. Sodoende kan die probleem van dubbele vertaling tussen Afrikaans en Italiaans (van Afrikaans na Engels en van Engels na Italiaans, of 'n ander Romaanse taal en omgekeerd) oorkom word. Dit is noodsaaklik om in ag te neem dat die woordeboek in De Foglio (2001) slegs 'n vertaler of vakman kan dien wat beide tale, Afrikaans en Italiaans, magtig is. 'n Gewone gebruiker, die boer of die doeanebeampte wat geen kennis van een van die twee tale het nie, sal nie by die produk baat nie.

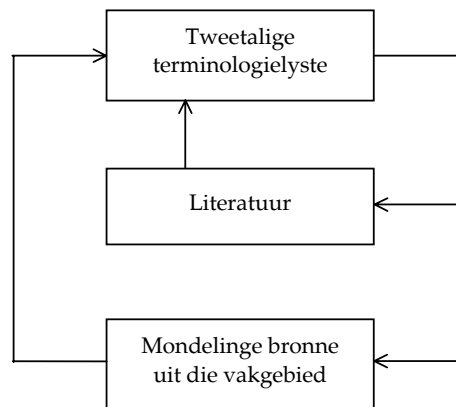
3. Voorafgaande werk

Voordat met die opstel van die vakwoordeboek begin word, moet volledige terminologielyste opgestel word, wat elektronies of in hardekopie geberg kan

word. Die optekening van materiaal kan as verwysingsbron vir toekomstige vertaalopdragte dien, des te meer wanneer die vertaler weet dat daar geen leksikografiese hulpmiddele in dié taalkombinasie beskikbaar is nie. Lemmas en ekwivalente is in die volgende terminologielyste onder die afdeling landbou geklassifiseer en ingedeel:

- landbou-ekonomie,
- landboubemaking,
- landboukrediet,
- handelsterme, en
- landboubeleid.

Die bestaan van sulke selfopgestelde terminologiese versamelings moet tot 'n beter verwerking volgens leksikografiese kriteria bydra. 'n Woordeboek is die vrug van die keuring van opgestelde terminologielyste en die bestaande literatuur soos vaktydskrifte in die betrokke twee tale. Die volgende tabel gee 'n uiteensetting van werk aan 'n vakwoordeboek.



Die terminografiese werk aan 'n tweerigtingvakwoordeboek.

In die geval van die tweerigtingvakwoordeboek vir Afrikaans en Italiaans is terminologielyste uit vorige vertaalwerk en bestaande literatuur (soos aanbeveel deur Bergenholtz en Tarp 1995: 93) gebruik, asook Hartmann (1994: 291-293) se "kontrastiewe tekstologie". Mondelinge terugvoer uit die vakgebied is verder bygevoeg. Hierdie maatreëls het tot die verryking van die korpus en die akkuraatheid van die tweetalige terminologielys bygedra.

Die beskikbare literatuur het hoofsaaklik bestaan uit tweetalige Afrikaanse en Engelse tydskrifte en eentalige Italiaanse tydskrifte oor landbou-ekonomie. Tweetalige ekonomiese vakwoordeboeke (Engels-Italiaans en Afrikaans-Engels) het ook bygedra tot semantiese inligting op die gebied van ekwivalente, tesame met die terugvoering van deskundiges op dié vakgebied (soos aanbeveel deur

Bergenholtz en Tarp 1995: 92; Svensén 1993: 40).

4. Keuring en bewerking van materiaal

Vakwoordeboeke ondergaan 'n strenger keuring van materiaal as meer omvattende algemene tweetalige woordeboeke. Volgens Svensén (1993: 40) is "betroubaarheid" die faktor wat elke leksikograaf in die keuringsproses behoort na te streef. Dit is nie genoeg dat die leksikograaf slegs op sy linguïstiese vermoë staatmaak vir die keuring van materiaal wat in 'n woordeboek opgeneem word nie, maar hy moet ook die betroubaarheid van die terme wat in die praktyk gebruik word, in ag neem. In dié verband beweer Svensén (1993: 40):

The selection for a dictionary must be authentic, that is to say it must include only such linguistic occurrences as actually exist. In order to be sure that a certain linguistic occurrence really is authentic, the lexicographer must find evidence for it in independent sources.

Met onafhanklike bronne word hoofsaaklik bedoel dat mondelinge bronne ook geraadpleeg moet word om 'n bepaalde ekwivalent vas te stel.

Volgens Cluver (1992: 35) is die begrippe agter die vaklemmas meer eksplisiet ingeperk as dié van die lemmas van die algemene taal. Een voordeel van die vakleksikograaf, in vergelyking met die algemene leksikograaf, is dat sy definisies meer eksak en dus kontroleerbaar is. Hierdie faktor versnel ook die keuringsproses van die vakleksikograaf op 'n bepaalde vakgebied.

In die geval van die tweerigtingvakwoordeboek is tekste uit landbou-ekonomiese tydskrifte sorgvuldig "geskandeer" om te bepaal watter terme spesifiek in die taal van die landbou-ekonomie voorkom. Hoewel algemene lemmas soos *bees* of *vrug* wel taamlik algemeen in die landbou-ekonomie voorkom, is hulle nie opgeneem nie omdat veronderstel is dat hulle reeds in algemene tweetalige woordeboeke voorkom.

Woordelyste en woordeboeke oor algemene landbou is geïdentifiseer en die keuring in die afgebakende veld, landbou-ekonomie, is gedoen. Die Italiaans-Engelse ekonomiese woordeboeke was nuttiger as die Afrikaans-Engelse verwante woordeboeke in terme van die reikwydte van ekwivalente. Nietemin was laasgenoemdes onmisbare verwysingsbronne.

In die opstelling van sowel 'n vakwoordeboek as 'n algemene tweetalige woordeboek word primêre én sekondêre bronne gebruik (Svensén 1983: 53-54). Primêre bronne bestaan hoofsaaklik uit mondelinge en geskrewe bronne, terwyl sekondêre bronne woordeboeke insluit. In die leksikografiese werk is van verskillende primêre bronne (tydskrifte, werklike taal van die spesifieke vakgebied en mondelinge gesprekke) en sekondêre bronne (verskillende tweetalige en eentalige woordelyste) gebruik gemaak. Hartmann (1994: 291-293) se tegniek van parallelle tekste is uitvoerig vir die keuring en vasstelling van ekwivalente gebruik.

Tekste verkry uit primêre tweetalige bronne, is met persone, aktief in die

vaktaal, gekontroleer. Uit tweetalige Engelse en Afrikaanse tydskrifte is eerstens Afrikaanse en Engelse ekwivalente verkry. Hulle is as basis gebruik om Italiaanse ekwivalente na te slaan en, indien nodig, ook te vertaal. Die Italiaanse ekwivalente is sorgvuldig met persone wat in die vakgebied werksaam is, gekontroleer. 'n Ekonomiese woordelys van agt honderd lemmas betreffende landbou, handel en die bemerking van landbouprodukte, wat tydens die vertaalpraktyk saamgestel is, is in die woordeboek opgeneem.

Volgens Svensén (1983: 50) bestaan vaktaal nie net uit woorde nie. Verbindings, kollokasies en uitdrukkings behoort ook in 'n vakwoordeboek opgeneem te word. Hy wys daarop dat hierdie aspek nie noodwendig altyd in vakwoordeboeke in ag geneem word nie, omdat hulle gewoonlik slegs as bronne gebruik word om ekwivalente vinnig te bekom. Daar bestaan eintlik geen rede vir die verskillende hantering van tegniese en algemene taal nie.

Vir die keuringswerk is dit veiliger om 'n reeds bestaande, alfabetiese korpus lemmas van 'n bepaalde vakgebied aan te wend. So 'n indeks sou die werk van die leksikograaf aansienlik vergemaklik (Bergenholtz en Tarp 1995: 101). Dikwels is daar geen bestaande lys van die behandelde vakgebied nie en, soos in die geval van hierdie studie, is die leksikograaf op homself aangewese om die keuring te doen uit bepaalde vaktydskrifte, wat per slot van rekening die beste bron van dié gebied se taal behoort te wees.

Omdat geskikte elektroniese infrastrukture ontbreek het tydens die terminografiese werk, is die algehele verwerking hoofsaaklik met behulp van elektroniese ontledingstate uitgevoer, wat onder meer tot die elektroniese alfabetisering van die finale woordelys bygedra het. Bergenholtz en Kaufmann (1997: 107 e.v.) bespreek die rol van rekenaartegnologie by terminografiese werk. 'n Tydelike drietalige woordelys (Afrikaans-Italiaans-Engels) het ontstaan ná akkurate keuring uit literatuur, tydskrifte van landbou en landbou-ekonomie, saam met mondelinge gebruik. Die ontledingstaat het rekenaarmatig gehelp om Afrikaanse en Italiaanse lemmas van die Engelse lemmas af te sonder. Hierdie lemmas is omgekeer volgens die behoeftes van die rigting waarin gewerk is. Grammatikale inligting is nie bygevoeg tydens die terminografiese werk nie en dit het eers in 'n later stadium (dus tydens die leksikografiese verwerking) plaasgevind.

5. Toepassing van leksikografiese beginsels

Uit die terminografiese werksaamhede het 'n terminologie van landbou-ekonomie tussen Afrikaans en Italiaans ontstaan. Die vereiste was om 'n leksikografiese werk ('n woordeboek) tot stand te bring en leksikografiese beginsels is dus toegepas.

5.1 Woordeboekstruktuur

Om 'n oorsigtelike struktuur te verseker, is die volgende kriteria in berekening

gebring:

(a) **Makrostruktuur** As makrostrukturele element verskaf die lemma inligting oor die klem in die vorm van 'n klemteken (**belas'ting**) in die rigting Afrikaans–Italiaans, of in die vorm van 'n onderstreepte klinker (**vendere**) in die rigting Italiaans–Afrikaans.

Op makrostrukturele vlak kan die lemma in toepaslike gevalle 'n simbool (|) bevat wat lettergreepskeiding aandui. Hierdie simbool is belangrik vir die verskaffing van inligting oor die verskillende vorme van die woord. Die morfologie word in die mikrostruktuur aangedui.

In die rigting Afrikaans–Italiaans kan byvoorbeeld die volgende aangetref word:

be | **drag'**, ^s (-drae), ...

addision | **eel'** ^b (-ele), ...

In die rigting Italiaans–Afrikaans is die grammatikale en morfologiese inligting van die lemma deel van die mikrostruktuur:

acces | **so**, *s.m.*, (-si), ...

accaparr | **are**, *w.*, (-ato), ...

agrari | **o(-a)**, *b.*, (-~/-e), ...

acquiren | **te**, *s.m/v.*, (-ti), ...

(b) **Mikrostruktuur** In die rigting Afrikaans–Italiaans word vertaalekwivalente met grammatikale inligting van Italiaans voorsien:

in'vorder, ^w (-ge-), *riscuotere* (*tasse*); **~aar**, ^s (-s), *esattore m*; **~baar**, ^b (-bare), *tassabile, imponibile*; **~ing**, ^s *riscossione v*; **~ing** *op lening*, *rimborso m del credito*; **~ingskoste**, *costo m di riscossione*.

In die meeste vakwoordeboeke in die Germaanse tale is daar min of geen grammatikale inligting ingesluit nie, maar vir sowel Germaanse as Romaanse behoeftes is dit noodsaaklik om grammatikale inligting te verskaf. Grammatikale inligting is soos volg bepaal: in die rigting Afrikaans–Italiaans is die woordsoortlikheid telkens ná die lemma aangedui (^w = werkwoord; ^s = selfstandige naamwoord; ^b = byvoeglike naamwoord; ^{bw} = bywoord). Hierdie inligting word verskaf om vir die behoeftes van die Italiaansspreekende gebruiker wat onderskeidingsinligting omtrent woordsoortlikheid in Afrikaans sal benodig, voorsiening te maak. Genoemde onderskeiding sal oorbodig wees vir die Afrikaansspreekende gebruiker, en dus is die inligting met verkleinde boskrifletters aangetoon. Hierdie werkwyse behoort ook die toegangstruktuur te verbeter.

Ná die lemma word die grammatikale inligting soos volg hanteer: woordsoortlikheid en/of genus, meervoud, verlede tyd, vertaalekwivalente, kollokasies en voorbeeldsinne.

In die rigting Italiaans–Afrikaans word die woordsoortlikheid met 'n ander lettertipe aangetoon (bv. **acquist***o*, *s.m.*, aankoop), waar woordsoortlikheid (*g.m.*) en genus (*s.m.*) 'n belangrike faktor vir die Afrikaanssprekende gebruiker word.¹ Laasgenoemde strategie is eerder gebruik om inligting oor die genus, wat tipies van die Romaanse tale is, aan te dui. In die rigting Italiaans–Afrikaans word daar meer aandag gegee aan die Afrikaanssprekende gebruiker wat meer inligting oor die genus van Italiaans benodig.

Indien nodig is meerfunksionaliteit aangedui. Die verskillende woordsoortlike funksies van 'n lemma met dieselfde vorm (soos *besit*) word met nommers aangedui:

besit', 1 *s*, possesso *m*, proprietà *v*, bene *m*, demanio *m*; ~**nemer**, occupante *m*; titolare *m*; ~**neming**, occupazione *v*; prendita *v* di possesso; ~**reg**, diritto *m* di proprietà, dominion *m*. 2 *w* (~), possedere; occupare ...

(c) **Toegangstruktuur** 'n Effektiewe toegangstruktuur is noodsaaklik vir die gebruik van 'n woordeboek. Dit dien as "soekpad" (Gouws 1996: 19) vir die gebruiker om die benodigde inligting te bekom. In hierdie werk word die inligting op so 'n toeganklike wyse moontlik aangebied. As die gebruiker kollokasies of voorbeeldsinne benodig, sal hy dit in kursief gedruk vind, terwyl samestellings met die lemma in vet gedruk is:

wins, *s* (-te), utile *m*; profitto *m*; reddito *m*; tornaconto *m*; frutto *m*; ~*delende obligasie*, obbligazione *v* con partecipazione; ~**deling**, partecipazione *v* agli utili, compartecipazione *v* agli utili; ~**delingsertifikaat**, certificato *m* di partecipazione agli utili; ~*gewende belegging*, investimento *m* remunerativo; ~**gewendheid**, *s* redditività *v*; ~**grens**, margine *m* di guadagno; ~**marge**, margine di utile, margine di profitto; ~**saldo**, (ONVERDEELDE WINS), profitti *m* indivisi; *deelgeregtigheid in die ~*, ⇔ WINSDELING; *herbelegde ~*; *herinvesteerde ~*, reinvestimento *m* degli utili; *ingehoue ~te*; *teruggehoue ~*, profitti non distribuiti, utili trattenuti; *meevaller~*, utili *m* straordinari; *onuitgekeerde (onverdeelde) ~*, profitti non distribuiti; *papier~*, utile sulla carta; *skoon ~ maak*, chiarificare; guadagnare al netto; *suiwer ~*, profitto netto.

'n Voorbeeld uit die Italiaans–Afrikaanse rigting sien soos volg daaruit:

accredito, *s.m.*, (CREDITO), krediet; kreditering; ~ **bancario**, bankkrediet; *nota di ~*, kredietnota; *somma in ~*, kredietbedrag.

Daar word sover moontlik probeer om die inligting wat in vet gedruk is voor die inligting wat in kursief gedruk is te laat verskyn. Die lemma word in die

mikrostrukturele veld herhaal in die vorm van 'n tilde (~). Hierdie strategie behoort die woordeboek se leesbaarheid te verhoog en die gebruiker in staat te stel om woorde rondom die lemma binne die woordeboekartikel te vind. 'n Interne toegangstruktuur ("inner access structure", vgl. Gouws 1996: 19-20) word ook voorsien in die vorm van die alfabetisering van inligting, in kursief of in vet gedruk.

(d) **Mediostruktuur** Twee tipes verwysingsadresse ("reference addresses", vgl. Gouws en Prinsloo 1998: 20-21) word in die woordeboek gebruik. Albei is eksterne verwysingsadresse ("external reference addresses", vgl. Gouws en Prinsloo 1998: 20). Die funksionaliteit van die woordeboek word hiermee verhoog (Gouws en Prinsloo 1998: 21), en die leesbaarheid daarvan verbeter.

Die eerste tipe verwysingsadres is 'n stelsel van kruisverwysings in klein hoofletters. Hulle word deur 'n verwysingsaanwyser (⇒) voorafgegaan om die inligting wat in die lemma te vinde is, aan te toon (vgl. Gouws en Prinsloo 1998: 19). Voorbeelde hiervan is die volgende:

importeer', ^w (geïm-) ⇒ INVOER ^w.

uit'gifte, ^s (-s) ⇒ EMISSIE.

'n Tweede tipe verwysingsadres dien om sinonieme van die lemma aan te dui wat in ander dele van die woordeboek voorkom. Hierdie metode stel die gebruiker in staat om hierdie sinonieme na te slaan.

obliga'sie, ^s (-s), *obbligazione v*; **~houer**, (SKULDBRIEFHOUER; VERBANDHOUER), *obbligazionista m*.

skuld, ^s (-e), *debito m*; **~akte**, *titolo di debito*; **~brief**, *certificato obbligazionario*; *obbligazione*; **~briefhouer** ⇒ OBLIGASIEHOUER.

verband', ^s (-e), *ipoteca v*, *contratto m ipotecario*; **~akte**, *obbligazione v ipoteca*; **~finansiering**, *finanziamento m ipotecario*; **~houer** ⇒ OBLIGASIEHOUER; **~krediet**, *credito m ipotecario*; **~lening**, *mutuo m ipotecario*, *prestito ipotecario*; **~obligasie** ⇒ VERBANDAKTE; **~rente**, *interesse m ipotecario*; **~sekuriteit**, *garanzia v ipotecaria*; **~skuld**, *debito m ipotecario*; *vruggebruikers~*, *ipoteca usufruttuaria*.

'n Sirkelstruktuur ("circular structure", vgl. Gouws en Prinsloo 1998: 19) word ook aangewend, hoewel dit aanleiding kan gee tot onvergenoegdheid by sekere "ongeduldige" woordeboekgebruikers:

concedente, *s.m.* ⇒ CONCESSORE.

concessore, *s.m.*, (CONCEDENTE), *begiftiger*; *gewer*; *skenker*.

5.2 Hantering van vertaalekwivalente en sinonieme

Die verskillende betekenis van die vertaalekwivalente word met 'n komma-punt (;) geskei en die sinonieme met 'n komma (,) o.a. soos in die *Groot Woordeboek/Major Dictionary* van Eksteen (1997¹⁴). In ander woordeboeke word vertaalekwivalente wat in betekenis verskil soms met nommers aangetoon. Vir die doeleindes van hierdie werk, en ook as gevolg van die gespesialiseerde aard daarvan, is besluit om eerder leestekens te benut aangesien nommers reeds gebruik is om woordsoortlikheid aan te dui. Hierdie strategie behoort weereens die leesbaarheid van die woordeboek te verbeter (d.w.s. 'n beter toegangstruktuur te verskaf).

Sinonieme word sorgvuldig binne die mediostrukturele vlak met kruisverwysings, verwysingadresse en sirkelstrukture aangedui. Dit handel hoofsaaklik oor sinonieme wat, naas lemmas, die vertaalekwivalente verskaf. Volgens die besluit van die opsteller word dit op mediostrukturele vlak hanteer.

5.3 Grammatikale inligting in vakwoordeboeke

Soos ten opsigte van die taalpaar Engels–Italiaans in die geval van Favati (1990), is ook ten opsigte van die taalpaar Afrikaans–Italiaans grammatikale inligting oor Italiaans noodsaaklik vir die Afrikaanssprekendes, en omgekeerd het die Italiaanssprekendes naas die vertaalekwivalent ook semantiese en grammatikale inligting oor Afrikaans nodig.

Uit die vertaalpraktyk is dit duidelik dat blote vertaalekwivalente ontoereikend is weens die onbekendheid van die twee tale onderling. Die probleem word soos volg benader: Waar Italiaans die doeltaal is, word die lemmabewerking naas die vertaalekwivalent ook voorsien van grammatikale inligting, woordsoortlikheid en genus, iets wat Afrikaans, en ook Engels, glad nie het nie. Daarbenewens is die vervoeging van die werkwoordstelsel in Italiaans baie meer flekterend as in Afrikaans ten opsigte van tempus, persoon, getal en kongruensie. Ook vir 'n tweetalige Afrikaans–Italiaanse vertaler of vakman is grammatikale inligting oor die onderskeie tale op die gebied van toegangstruktuur belangrik, want grammatikale inligting speel beslis 'n rol om die vinnige raadpleging van die lemma te bewerkstellig.

Grammatikale inligting kan die vakman meer help as die vertaler wat beide tale en grammatikas beheers. Die inligting sal ook die vertaler in staat stel om vinnig te onderskei tussen die woordsoortlikheid. 'n Voorbeeld hiervan is:

Administrasie, Amministrazione *n.v.*

Genusaanduiding in Italiaans is noodsaaklik vir die Afrikaanssprekende gebruiker wat min kennis van Italiaanse genusverskille het. Die woord *Amministrazione* met 'n *n.v.* (d.w.s. naamwoord/vroulik) wys dadelik aan die gebruiker dat dit 'n vroulike naamwoord is en dat die lidwoord *la* is en nie *il* nie,

m.a.w. skrapping van die klinker *a* (*l'amministrazione* uit *la amministrazione*) het plaasgevind. Vir 'n ervare vertaler, spesifiek vir die Afrikaanssprekende vertaler, kan hierdie grammatikale inligting ook baie nuttig wees al weet hy dat die genus van Italiaans afgelei kan word van die uitsonderlike klinker *e* aan die einde van die woord *amministrazione*, wat nie noodwendig vroulik of manlik is nie.² In teenstelling met Italiaans ken Afrikaans geen genus of vervoeging nie, en dus hoef hierdie inligting nie verstrekkend te word nie. Daar word natuurlik aangeneem dat die Italiaanssprekende gebruiker van so 'n vakwoordeboek (vertaler of tweetalige vakman) reeds weet dat Afrikaans nie sulke grammatikale inligting bevat nie.

6. Slot en aanbevelings

Vir die opstel van 'n tweerigtingvakwoordeboek met Afrikaans en Italiaans as behandelde taalpaar, moet vooraf empiriese navorsing gedoen word om die behoeftes en leemtes van die gebruiker van die betrokke tale te bepaal. Indien die gebruiker hoofsaaklik 'n vertaler of tweetalige vakman is, kan 'n meer gevorderde kennisagtergrond aanvaar word.

Hoewel so 'n vakwoordeboek nie geskik is vir 'n groot publiek of mark nie, kan leksikografiese navorsing wél aangemoedig word tussen Afrikaans en Italiaans, en wel in terme van 'n tweetalige tweerigtingaanleerderwoordeboek wat die behoeftes van die Afrikaanse gebruiker ten opsigte van Italiaans sal dien.

Standaardkriteria van die moderne vertalende leksikografie sal aangepas moet word. 'n Onderliggende databasis in Afrikaans sal saamgestel moet word vir 'n moontlike tweetalige leksikografiese werk met Afrikaans en Italiaans as behandelde taalpaar. Vermoedelik sal die mark en die belangstelling in so 'n woordeboek groter in Suid-Afrika as in Italië wees. Suid-Afrika regverdig die projek aangesien Italiaans as 'n belangrike wêreldtaal beskou word, maar dieselfde kan nie in Italië verwag word nie, aangesien die belangstelling vir Afrikaans nog nie so ontwikkel is nie.

Indien later ook in Italië 'n groter behoefte sal ontstaan, kan dieselfde databasis dien om 'n tweetalige woordeboek saam te stel wat die Italiaanse gebruiker in ag neem.

As voorbeeld hoe so 'n vakwoordeboek daar behoort uit te sien, word twintig lemmas in 'n bylaag aangegee.

Aantekeninge

1. Uit 'n ondersoek na die gebruikersaspek kan daar afgelei word dat die potensiële tweetalige gebruiker eerder Afrikaanssprekend as Italiaanssprekend is. Die mark is potensiëel doeltrefender in Suid-Afrika as in Italië, en dus word meer op die Suid-Afrikaanse gebruiker gefokus.
2. Naamwoorde en byvoeglike naamwoorde in Italiaans dra genus-inligting. As naamwoorde

met 'n *o* eindig, gaan dit oor manlike naamwoorde, as hulle met 'n *a* eindig, is hulle vroulike naamwoorde. Sommige Italiaanse woorde eindig op *e* en dit kan óf na manlike óf na vroulike naamwoorde verwys. In hierdie geval kan slegs woordeboeke die regte inligting oor genus, soos in *amministrazione* (v) en *agente* (m), aan 'n anderstalige persoon gee.

Bibliografie

- Alberts, M. 1988. Die rekenarisering van vakleksikografie. *Suid-Afrikaanse Tydskrif vir Taalkunde* 6 (1): 1-19.
- Barclay, D.E. 2000. Engels: Te veel klem. *Afrikaans Vandag* 7(3): 6-7.
- Bergenholtz, H. en U. Kaufmann. 1997. Terminography and Lexicography. A Critical Survey of Dictionaries from a Single Specialised Field. *Journal of Linguistics* 18: 91-125.
- Bergenholtz, H. en S. Tarp. 1995. *Manual of Specialised Lexicography*. Amsterdam: Benjamins.
- Cluver, A.D. de V. 1992. *Die verskille en ooreenkomste tussen algemene en vakleksikografie*. National Terminology Services 1992: 32-48.
- De Foglio, F. 2001. *Leksikografiese kriteria vir 'n tweerigtingvakwoordeboek: Afrikaans-Italiaans en Italiaans-Afrikaans*. Ongepubliseerde M.A.-skripsie. Bloemfontein: Universiteit van die Vrystaat.
- Eksteen, L.C. (Red.). 1997¹⁴. *Groot Woordeboek / Major Dictionary*. Kaapstad: Pharos.
- Favati, F. 1990². *Dizionario di agricoltura. English-Italian, Italiano-Inglese*. Bologna: Edizioni Agricole.
- Gouws, R.H. 1991. Kommunikatiewe ekwivalensie in vertalende woordeboeke — 'n probleem vir die vertaler. *Suid-Afrikaanse Tydskrif vir Taalkunde* 10(1): 37-42.
- Gouws, R.H. 1996. Bilingual Dictionaries and Communicative Equivalence for a Multilingual Society. *Lexikos* 6: 14-31.
- Gouws, R.H. en D.J. Prinsloo. 1998. Cross-referencing as a Lexicographical Device. *Lexikos* 8: 17-36.
- Gove, P.B. 1967. *The Dictionary's Function*. Gove, P.B. (Red.). 1967. *The Role of the Dictionary*: 5-8. Indianapolis: The Bobbs-Merrill Co.
- Hartmann, R.R.K. 1994. The Use of Parallel Text Corpora in the Generation of Translation Equivalents for Bilingual Lexicography. Martin, W. et al. 1994: 291-297.
- Hartmann, R.R.K. 1984 (Red.). *LEXeter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9-12 September 1983*. Tübingen: Max Niemeyer.
- Lo Bianco, J. 1996. *Language as an Economic Resource*. Language Planning Report 5 (1). Pretoria: DACST.
- Magay, T. 1984. Technical or General: Problems of Vocabulary Selection in a Medium-size Bilingual Dictionary. Hartmann, R.R.K. (Red.). 1984: 221-225.
- Martin, W. et al. (Reds.). 1994. *Euralex '94 Proceedings. Papers Submitted to the 6th EURALEX International Congress on Lexicography in Amsterdam, The Netherlands*. Amsterdam: Free University Press.
- McCallum, A. 1990. Internationalising the Operations of Australian Business with Particular Reference to the Role of Language. *Language is Good Business, Conference Proceedings*. Melbourne: NLLI.
- Statistics South Africa. 1998. *The People of South African Population Census 1996: Census in Brief*. Report No. 03-01-11 (1996). Pretoria: Statistics South Africa.
- Svensén, B. 1993. *Practical Lexicography — Principles and Methods of Dictionary Making*. Oxford: Oxford University Press.
- Zgusta, L. 1971. *Manual of Lexicography*. Den Haag: Mouton.

Bylaag**Afrikaans – Italiaans**

klant, ^s (-e), cliente *m*, compratore *m*; acquirente *m*; ~e, clientela *v*; ~esteun, supporto *m* alla clientela.

kla'ring, ^s compensazione *v*, chiarificazione *v*; pratica *v* di sdoganamento (dogana); ~sooreenkoms, accordo *m* di compensazione.

klein'handel, ^s commercio *m* al minuto; ~aar, ^s (-s/-are), trafficante *m*, rivenditore *m*, commerciante *m* al minuto; ~(s)ketting, catena *v* al dettaglio; ~(s)mark, mercato *m* al dettaglio; ~(s)prys, prezzo *m* al minuto, prezzo al dettaglio; ~(s)prysindeks, indicizzazione *v* del prezzo al minuto; ~(s)verkope, vendite *v* al dettaglio.

klein'kas, ^s piccola cassa, fondo *m* per le piccole spese; ~uitgawes, spese *v* minute.

kleur, ^s (-e), colore *m*, tintura *v* (delle pelli).

kliënt', ^s (-e), cliente *m*; ~ediens, servizio *m* clienti; ~eel', clientela *v*.

klousu'le, ^s (-s), clausola *v*; articolo *m*; paragrafo *m*; *beskermings*~, clausola di salvaguardia {C.E.E.}; *bygevoegde* ~ ⇒ AANHANGSEL; *kontrak*~, clausola di (del) contratto; *meesbegunstigings*~, clausola della nazione più favorita; *omvattende* ~, clausola generale; *straf*~, (STRAFBEPALING), clausola penale.

knel'punt, ^s strozzatura *v*.

koëffisiënt', ^s (-e), (VERHOUDING), rapporto *m*; proporzione *v*.

koel'kamer, ^s (-s), camera *v* frigorifero.

Italiaans – Afrikaans

condutt|ore, *s.m.*, (-ori), (CONDUTTORE DI AZIENDA), plaasbestuurder; boer.

conduzione, *s.v.*, (AMMINISTRAZIONE; DIREZIONE; GESTIONE), bestuur; ~ *di un fondo*, landgoedbestuur; ~ *agricola in affitto*, gehuurde boerdery; ~ *diretta*, direkte bestuur.

confezioname|nto, *s.m.*, (IMBALLAGGIO), verpakking; ~ *della frutta*, vrugteverpakking.

consegnare, *w.*, (-ato), lewer; verskaf; voorsien.

consi|glio, *s.m.*, (-gli), raad; ~ *sulla ricerca agraria*, landbounavorsingsraad.

consor|zio, *s.m.*, (-**zi**), konsortium; unie; trust; ~ **agrario**, landbou-unie; ~ **ban-**
cario, sindikaat; ~ **finanziario**, (CONSORZIO DI EMISSIONE), onderskry-
wingsindikaat; ~ *per investimenti*, effektebelegging; effektetrust.

consul|nte, *s.m/v*, (-**nti**), konsultant, raadgewer; ~ **agrario**, landboukonsul-
tant, landboukundige raadgewer; ~ *di marketing*, (CONSULENTE ADDETTO
ALLA COMMERCIALIZZAZIONE), bemarkingsadviseur.

consumato|re, *s.m.*, (-**ri**), *s.v.*, (UTENTE), verbruiker; ~ **finale**, uiteindelijke ver-
bruiker.

consumatri|ce, *s.v.*, (-**ci**) ⇒ CONSUMATORE.

consum|o, *s.m.*, (-**i**), verbruik; ~ **alimentare**, voedselverbruik; ~ **annuo**,
jaarverbruik, jaarlikse verbruik; ~ *di capitale*, kapitaalvertering; ~ *di der-*
rate, verbruik van voedingstowwe; ~ *di prodotti agricoli*, verbruik van
landbouprodukte; ~ **individuale** ⇒ CONSUMO PRO CAPITE; ~ **interno**,
binnelandse verbruik; ~ **pro capite**, verbruik per capita, verbruik per
hoof, per capita-verbruik; *crisi di ~*, verbruikskrisis; *principali generi di ~*,
stapelartikels; stapelgoedere, stapelprodukte.

Niching as a Macrostructural Procedure*

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

Abstract: Niching is a macrostructural procedure of textual condensation which results in different dictionary articles being clustered into one textblock as subarticles headed by sublemmata. This article offers an identification and critical discussion of different types of niching. Examples from dictionaries are used to examine problems which dictionary users experience as a result of the application of niching. The emphasis is on problems regarding the search route of users via the outer access structure of dictionaries. A distinction is made between single niching and multiple niching. Different aspects of multiple niching receive attention, e.g. the use of remote and exclusively remote multiple niching. The focus is also on an increased degree of textual condensation resulting from these procedures as well as problems in the domain of the access structure. Attention is drawn to the influence of the inclusion of homonym pairs on the application of multiple niching. Problems arising from an inconsistent application of multiple niching are identified and solutions are suggested for these problems in order to enhance the quality of the lexicographic practice. It is shown that a user-driven dictionary would rather opt for the application of procedures of single niching and first level nesting instead of multiple niching.

Keywords: EXCLUSIVELY REMOTE MULTIPLE NICHING, FIRST LEVEL NESTING, HETEROGENEOUS NICHING, HOMOGENEOUS NICHING, HYBRID NICHING, MULTIPLE NICHING, NESTING, NICHING, PARTIAL ARTICLE STRETCH, REMOTE NICHING, SINGLE NICHING, TEXTUAL CONDENSATION

Opsomming: Nisting as 'n makrostrukturele prosedure. Nisting is 'n makrostrukturele teksverdigtingsprosedure waardeur woordeboekartikels in een teksblok as subartikels met sublemmata gegroepeer word. Hierdie artikel bevat 'n beskrywing en kritiese bespreking van verskillende tipes nisting. Aan die hand van woordeboekvoorbeelde word probleme wat gebruikers met die toepassing van verskillende nistingsmetodes kan hê, onder die loep geneem. Daar word veral gefokus op probleme wat nisting kan skep met betrekking tot die soekroete van gebruikers op die eksterne toegangstruktuur van woordeboeke. 'n Onderskeid word gemaak tussen enkelnisting en veelvuldige nisting. Verskillende aspekte van veelvuldige nisting word ondersoek, onder meer die gebruik van verwyderde en eksklusief verwyderde veelvuldige nisting. Die verhoogde graad van teksverdigting wat daardeur teweeggebring word asook die toegangstruktuurprobleme word bespreek. Aandag word ook gegee aan die invloed wat die opname van homoniempare op veelvuldige nisting het. Leemtes en inkonsekwenhede in die toepassing van veelvuldige nisting word bespreek en voorstelle word aan die hand gedoen ter verbetering van die huidige toepassing daarvan in die leksikografiepraktyk. Daar word aangedui dat 'n gebruikersgedrewe woordeboek

* This article is an expanded version of a paper read at the Eleventh International Symposium on Lexicography, University of Copenhagen, Copenhagen, 2-4 May 2002.

eerder van veelvuldige nisting moet afstand doen ten gunste van prosedures van enkelnisting en eerstevlaknesting.

Sleutelwoorde: EERSTEVLAKNESTING, EKSKLUSIEF VERWYDERDE NISTING, ENKELNISTING, GEDEELTELIKE ARTIKELREEKS, HETEROGENE NISTING, HIBRIDIESE NISTING, HOMOGENE NISTING, NESTING, NISTING, TEKSVERDIGTING, VEELVULDIGE NISTING, VERWYDERDE NISTING

1. Introduction

When the macrostructure of a dictionary is observed, the lexical items selected for inclusion as lemmata could be classified according to different sets of criteria, e.g. according to their form as items of the lexicon of the specific language and according to their presentation as macrostructural entries. With regard to their form, a distinction can be made between words, items smaller than words and multiword items, presented as lexical lemmata, sublexical lemmata and multilexical lemmata respectively (cf. Gouws 1991). As far as the presentation of lemmata is concerned, a well-established distinction exists, based on the occurrence of the lemmata as either main lemmata, participating in the vertical ordering of macrostructural elements, or sublemmata, participating in the horizontal ordering of macrostructural elements. Within the category of sublemmata a further distinction is made, i.e. between niched and nested sublemmata (cf. Hausmann and Wiegand 1989). These lemmata function as the guiding elements of niched and nested articles respectively. Procedures of niching and nesting are primarily performed as a space-saving endeavour. The presentation of lemmata in a cluster instead of a linear ordering has to be regarded as a form of textual condensation (cf. Wiegand 1998, Wolski 1989a). All procedures of niching and nesting do not display the same degree of textual condensation and when evaluating this type of macrostructural presentation one has to be aware of the genuine purpose of the specific dictionary and one should focus on the type of dictionary, its target users and the reference skills of those target users.

In this paper various aspects of niching will be discussed with the emphasis on procedures of multiple niching. Some implications of niching for the outer access structure of dictionaries will also receive attention.

2. Niching and nesting

Niched and nested lemmata deviate from the strict application of a linear macrostructural ordering due to the fact that access to the articles with the niched and nested lemmata as guiding elements is only possible when the search route goes via a basis lemma (cf. Wiegand 1989: 390), functioning as nich/nest entrance lemma (cf. Wolski 1989).

Niching (cf. Hausmann and Wiegand 1989: 336) is a strict alphabetical

clustering of lemmata or articles which may or may not be semantically related, whereas nesting is a clustering of lemmata or articles which displays a deviation from the rules of strict alphabetical ordering. According to Wiegand (1989: 391), a nest is a cluster that contains at least one lemma that deviates from the alphabetical ordering. The alphabetical ordering that characterises niching applies on a cluster-internal and cluster-external level (cf. Gouws 2001). The lemmata occurring within the niche are ordered alphabetically and the niche fits into the alphabetical environment framed by the preceding and the following main lemmata.

Wiegand (1989: 391-392) mentions two ways in which nesting exhibits deviation from the strict alphabetical ordering. External deviation implies that the lemmata included in the nest do not fit into the alphabetical environment framed by the preceding and following main lemmata. In this regard nesting typically leads to a clustering of lemmata which alphabetically follows the basis lemma but at least one lemma in this cluster disturbs the alphabetical link with the following main lemma, cf. the following example from the *Verklarende Afrikaanse Woordeboek* (henceforth abbreviated as VAW):

- (1) **ballet'**, **-te**. 1. Kunstige toneeldans. 2. Al die dansers of danseresse saam; ~**dans**; ~**danseres**; ~**liefhebber**; ~**musiek**.
bal'letjie, **-s**. Klein bal; 'n ~ *opgooi*, 'n mening uitspreek om te sien wat die reaksie is.
ballet'korps. Groep danseresse (dansers) wat in 'n ballet saam optree.
ballet'meester, **-s**. Leier van 'n ballet.
balletomaan', **..mane**. Iem. wat op ballet versot is; **..manie'**.

In a strict alphabetical ordering the sublemma *balletliefhebber* should have been ordered between the lemmata *balletkorps* and *balletmeester* and the sublemma *balletmusiek* should have been ordered after the lemma *balletmeester*. In isolation the cluster constituting this type of nest does not differ from an isolated cluster representing a niche because on a cluster-internal level the strict alphabetical ordering is maintained. Gouws (2001: 106) refers to this type of nesting as *first level nesting*.

Internal deviation prevails where the lemmata presented in the cluster do not adhere to a strict alphabetical ordering. This type of nesting (cf. Gouws 2001: 106) is known as *second level nesting*, cf. the following example from *Nasionale Woordeboek*:

- (2) **re'gering (-e, -s)** s.nw. 1. *bestuur, bewind, veral uitoefening van die oppergesag; landsbestuur*. Aan die – kom. 'n Demokratiese –. Die – aanvaar, neerlê. Onder die – van Willem II. 2. *bepaalde kabinet, ministerie, party aan die bewind*. Die – van Hertzog. In die – sit. 'n Maatreël van die –. 3. *owerheid, staat*. **regering-loos**; **regeringsamp**, **-amptenaar**, **-gebou**, **-koste**, **-pos**, **-vorm**; **regeringsaak**, **-stelsel** (by 1); **regeringsbeleid**, **-besluit**, **-blad**, **-hoof**, **-kringe**, **-man**, **-party**, **-tyd** (by 2).

In this example the first nested lemma (*regeringloos*) is a derivation whereas the other are compounds. Morphological differences within the nest motivate the use of a semi-colon to separate the derivation from the compounds. In the first group of compounds (up to *regeringstelsel*) the first sense of the main lemma prevails. This is indicated by the structural marker (*by 1*). In the second group of compounds the second sense of the main lemma prevails. Semantic reasons motivate the use of two distinct groups of lemmata representing the compounds. Within each one of the two groups of nested lemmata an internal alphabetical ordering has been maintained. However, due to morpho-semantic reasons the nest as such does not display an internal alphabetical ordering.

If the main characteristic feature of nesting is a deviation from the strict alphabetical ordering (cf. Wiegand 1989: 391), then first and second level nesting share a strong common ground which clearly distinguishes them from niching. However, Hausmann and Wiegand (1989: 336), argue that nesting stretches the rules of strict alphabetical ordering in order to exhibit morpho-semantic relations between words, and according to Wiegand (1989: 390) a nest-internal deviation from the alphabetical ordering could be regarded as an attempt to separate sublemmata representing derivations and compounds. If these morpho-semantic motivations characterise nesting at best then second level nesting distinguishes itself from both first level nesting and niching. According to Hausmann and Wiegand (1989: 336), the lemmata in a niche may or may not be semantically related. Although the lemmata participating in a procedure of first level nesting usually show a semantic relation with the basis lemma the internal ordering is determined on alphabetical grounds. This then indicates a stronger relation with niching than with second level nesting. A better option would then be to work with first and second level niching where the occurrence of an external deviation from the strict alphabetical ordering represents second level niching.

Although space-saving reasons play an important role in the decision to accommodate lemmata in article clusters, the use of a sinuous lemma file can challenge the ground rules of ordering in a strict alphabetical dictionary. The adherence to morpho-semantic criteria in the ordering of lemmata results in a dictionary displaying a hybrid ordering system.

In this article the existing distinction between niching and nesting will not be investigated or questioned any further. However, the implications of a too strong alphabetical bias will be discussed and attention will also be given to problems arising from a hybrid macrostructural ordering system.

3. Varying degrees of textual condensation¹

One of the characteristic features of a niche is that the lemmata and articles have been clustered into a single textblock, cf. the following example:

- (3) **ballet'**, **-te**. 1. Kunstige toneeldans. 2. Al die dansers of danseresse saam; ~**dans**; ~**danseres**; ~**liefhebber**; ~**musiek**.

A typical textblock is introduced by a guiding element, usually the basis lemma, which is a full lemma functioning in the niche entrance position. A niche displays a lesser degree of textual condensation if the sublemmata maintain their form as full lemmata. However, the degree of textual condensation is often increased by a procedure of phasing out to the left. A textblock containing a cluster of niched articles can be attached to the article of a main lemma but instead of having a sequence of full sublemmata in the niche, a procedure of phasing out to the left, followed by a further procedure of textual condensation, leads to more complex textual condensation. The mutual first component of the sublemmata is phased out to the left and stranded in the niche entrance position. There it does not function as a treatment unit but only as a guiding element to lead the user to the partial lemmata remaining in the niche. In the following example from the *Elektroniese Verklarende Handwoordeboek van die Afrikaanse Taal* (henceforth abbreviated as EIHAT), the sublemmata *kleinkasboek*, *kleinkasboekhouer* and *kleinkashouer* are the targets of a process of phasing out to the left. This results in the lemma external part *kleinkas-* positioned in the niche entrance and the partial lemmata *~boek*, *~boekhouer* and *~houer* completing the niche (cf. Wolski 1989):

- (4) **klein'kas** (*handel*) Geld in hande vir daaglikse klein kontantuitgawes. **kleinkas:** *~boek, ~boekhouer, ~houer.*

A higher degree of textual condensation can occur when the phasing out to the left goes beyond the borders of the attached cluster, resulting in the basis lemma functioning in a *remote niche entrance* to combine with each one of the partial lemmata in the niche individually as treatment units. This is illustrated by the following example from *New Words/Nuwe Woorde*:

- (5) **heat** *n.*, (*Br. sl.*) polisieondersoek; (*sl.:* *pistool, rewolwer, skietding*) yster, twa, rollie, 7 mm; *the ~*, (*sl.:* *die polisie*) die pote/honde/boere; *put the ~ on s.o.* die wêreld vir iem. warm maak, druk op iem. uitoefen; *take the ~ for s.t.*, (*Am. sl.*) deurloop (*of* onder skoot/vuur kom) oor iets. *~ death* hittedood. *~ engine* hittemasjien, warmte-enjin. *~ flash* hitteflits (*by kernontploffing*). *~proof* hittevas, *≧bestand*. *~ pump* hittepomp. *~-seeking* *a.* (*attr.*) hittegeleide (*missiel*); hitesensitiewe, warmtegevoelige (*kamera*). *~-sensitive* hitesensitief, warmtegevoelig (*materiaal, papier, selle, ens.*). *~ shield* (*teg.*) hiteskerm.

In such a textblock, the retrieval of information demands a reconstruction of lemma signs and this reconstruction process exceeds the boundaries of the single article. Although it is more difficult to achieve a successful dictionary consultation procedure in a niche with partial lemmata, the user still has the advantage of a full lemma in the niche entrance position, albeit that the niched articles are separated from this niche entrance position. Still, the search route of the user from the basis lemma to the niched lemmata does not have to exceed the boundaries of a single textblock.

A similar situation is seen when the cluster of unexplained compounds is attached to the end of the article of a full lemma. This niche presents a sinuous file of partial lemmata without either a full lemma or a lemma part representing the first component of the compounds in a niche entrance position in close proximity. Reconstructing the niched lemmata demands procedures exceeding article boundaries, cf. the following example from VAW²:

- (6) **haard, -e.** Vuurmaakplek, es, herd; tuiste; oorsprong; *die eie ~ is goud waard*, die eie huis is die beste plek op aarde; ~kleedjie; ~kool; ~mat; ~plaat; ~steen; ~yster.

Yet again the user is assisted by the fact that the niche has been attached to an article with a full lemma as guiding element so that the search route to the lemmata in the niche can still proceed via the full lemma as basis lemma, albeit in a remote niche entrance position. An entry like *haard* in example (6) has a dual function. It is the lemma sign and treatment unit of an article but it also functions as lemma external part in a remote niche entrance position.

Textual condensation of a high degree also prevails where a textblock containing a niched cluster which has undergone phasing out to the left is not joined to the article of a full lemma but the lemma part in the niche entrance position is accommodated within the vertical ordering of the dictionary, cf. the following example from *Tweetalige Woordeboek/Bilingual Dictionary* (henceforth abbreviated as TW):

- (7) *kin'ders*: ~aard child nature. ~aftrek rebate (deduction) for children, tax abatement for children. **kin'deragtig**, **kinderag'tig** *≠e* childish, babyish, infantile; petty, foolish, silly, puerile. **kin'deragtigheid**, **kinderag'tigheid** childishness, infantilism; foolishness, silliness, puerility, puerilism. ~arbeid child labour. ~arts children's specialist, p(a)ediatrist, p(a)ediatrician. ~baring child-bearing. ~bed child's bed, cot, crib; childbed. ~bedjie bassinet, cot, crib, child's bed. ~bedkoors childbed fever, puerperal fever. ~beskerming child (life) protection; *vereniging vir* ~ child life protection society, children's aid society. ~bessie *Halleria elliptica*. ~beul bully (to children). ~bewaarploas, *≠plek* crèche, day-nursery; nursery school. ...

In the users' guidelines text of this dictionary, it is explained that the italics of the entry *kin'ders* indicate that this entry follows a related previous lemma (*kind*). This adds to the degree of textual condensation and complicates the procedures of reconstruction and deconstruction the user has to apply in order to retrieve the needed information.

Due to orthographic conventions textblocks containing this kind of niched cluster could display additional textual condensation by presenting an extended lemma-external niche entrance (cf. Wiegand 2002). In the following example from the *De Gruyter Wörterbuch Deutsch als Fremdsprache* (henceforth abbreviated as DGWDF), the entries *Eigen/eigen* function as extended lemma-

external niche entrance.

- (8) **Eigen/eigen** ['..] | **-sinn, der** (o.Pl) 'hartnäckiges Beharren auf seinem Tun, seiner Absicht, Meinung, verbunden mit einer Unzugänglichkeit für vernünftige Argumente'; ↗ FELD I.2.1: *sturer ~; aus ~ an einer Meinung festhalten*; vgl. *Trotz* (1.1) ❖ ↗ Sinn; **-sinnig** <Adj.; Steig. reg.> 'mit, aus Eigensinn' /vorw. auf Personen bez./; ↗ FELD I.2.3: *ein ~es Kind; ~ auf etw. beharren; unser Kind war früher sehr ~* ❖ ↗ Sinn ...

Neither of these lemma parts in the niche entrance can combine with both the partial lemmata included in the cluster. This type of phasing out to the left confronts the user, in this dictionary a learner of German, with extreme problems to achieve a successful dictionary consultation procedure. These problems are increased by the fact that the DGWDF does not employ the system of hyphenated entries to indicate that the element functioning in the extended niche entrance is only a lemma part. Only the entries presented as partial lemmata in the lemma niche have a hyphenated status.

These problems have been accounted for in a slightly more satisfactory way in TW. Although there is orthographic variation with regard to the first stem of the compounds functioning as niched lemmata, the default form is presented in the niche entrance and the deviation of this default model is indicated at the relevant occurrences. The lemma parts in the niche entrance position are presented as post-hyphenated entries (cf. Gouws 1989: 216), with the partial lemmata of the niche presented as pre-hyphenated entries.³ The following example illustrates the TW approach:

- (9) **blou** *se, n.* (shade of) blue, **blou** *~, se seer ste, a.* blue; cyanotic; ~ **aniline** aniline blue; ~ **asbes**, *vid.* **blou-asbes**; ~ **baba** b. baby; ~ **bloed** b. blood; *die ~ dam* the sea; *geen ~e duit werd nie* not worth a fig (a brass farthing); *hy het geen ~e duit nie* he hasn't got a bean; ~ *van die koue* blue with cold; ~ **maan**, *son* blue moon, sun; ~ **Maandag** black Monday; *elke ~ Maandag* every now and then; *sy het 'n ~ Maandag by ons gekuier* she stayed with us some short time; *die B~ Nyl* the B. Nile; *iem. 'n ~ oog slaan* give s.o. a black eye; ~ **pas** walking ticket (papers, orders); *iem. die ~ pas gee* send s.o. packing; ~ **plekke**, (*also*) bruises; *'n ~ skeen kry* be refused (*by a girl*); *die B~ Trein* the B. Train; ~ **walvis**, *vid.* **blouwalvis**; ~ **wimpel** b. ribbon (*of the Atlantic*); ~ **winde**, *vid.* **blouwinde**; *hy kan praat tot hy ~ word* he can talk till he is blue in the face. ~-**aap**, ~-**apie** vervet (monkey), blue monkey. **blou'agtig** *se, vid.* **blouerig**. ~-**apiestuipe**: *dit gee my die ~* it gives me the creeps. ~-**asbes** crocidolite, b. asbestos. ~**baadjie** convict, prisoner with indeterminate sentence. **B~baard** Bluebeard. ~**beer** b. or glacier bear. ~**bek** *n.* tribulosis (*in sheep*). ...

Although an entry like *B~baard* in the penultimate line of this example is a highly condensed macrostructural item, the niche as a whole is subjected to a lesser degree of textual condensation. This does, however, raise another question in terms of the success of the niching. One of the decisions a lexicographer

has to make as part of the planning of a dictionary and which has to be accounted for in the instruction book for the specific dictionary regards the access alphabet. Clear decisions should be taken to ensure a proper ordering of lemmata with orthographic differences, not only in terms of the alphabet letters used but also in terms of other orthographic conventions, e.g. the use of hyphens, capital letters, the umlaut, the circumflex, etc. (cf. Bergenholtz 1990: 22). Where the orthography of a word pair differs only in terms of the use of a capital versus a small letter the macrostructural ordering of these words as lemma signs should be influenced by this difference. The access alphabet of the specific dictionary should determine which form takes precedence. From a user-perspective, it adds to the textual condensation if the dictionary does not make provision for this kind of ordering. An arbitrary ordering or an ordering which merely ignores the difference is not acceptable. In this regard, the system prevailing in both the DGWDF and the TW is unacceptable. The successful utilisation of a procedure of niching demands that the first component, i.e. the component presented as lemma part in the lemma external niche entrance position, of all lemmata clustered into a niche must be fully identical in terms of orthography. Differences based on e.g. small versus capital letters and other orthographic conventions should be accounted for in the discussion of the access alphabet of the dictionary, and word pairs displaying such differences may not be included in the same niche.

4. Single and multiple niching

It has already been indicated that the pertinent difference between first level nesting and niching lies on the level of the adherence to or deviation from a consistent application of strict alphabetical ordering. In both these types of macrostructural ordering, the strict alphabetical principle is maintained on a cluster-internal level. Dictionaries in which sinuous lemma files prevail often limit the clustering procedures to a combination of niching and first level nesting. Second level nesting can be regarded as not only a higher degree of textual condensation but also as a more sophisticated lexicographic procedure which demands more comprehensive dictionary-using skills.

The decision to use a macrostructural ordering system which relies on sinuous lemma files should not be taken without proper consideration and an awareness of the needs and reference skills of the intended target user. However, once this decision has been taken, the dictionary plan should make provision for the best possible implementation of the system of clustering. Space saving has to be regarded as one of the most imperative reasons for the use of sinuous lemma files. This leads to the grouping of articles into single textblocks. Finding one's way in a cluster demands a cluster-internal ordering system which can ensure the user easy access to the desired guiding element of an article accommodated within a cluster. Niching implies the bringing together in one textblock of various articles (cf. Wiegand 1989: 388).

In a dictionary, the focus on the user and his/her ability to retrieve the needed information from the presented data should never be subordinate to space-saving procedures. Lemma clustering should be done in a way which does not confront the user with an unrealistic challenge to reach the desired lemma sign. In this regard, the type of articles brought together in a niche or nest should co-determine the way in which the dictionary deals with its sinuous lemma files. Where the textblock presents clustered lemmata which function as the guiding elements of articles with a restricted treatment, e.g. a clustering of unexplained compounds and derivations, the textblock structure could be of a very simple nature, characterised by the application of strict horizontal ordering, cf. the following example from the VAW:

- (10) **hand'doek, -e.** Absorberende doek waaraan 'n mens jou hande en gesig afvee na dit gewas is; *die ~ ingooi*, die stryd gewonne gee; **~goed; ~linne; ~rak; ~roller; ~stof.**

A knowledge of the alphabet and a bit of lexicographic endurance will allow the user to reach the desired lemma without too many problems. Especially if the way in which lemmata are ordered in a niche is explained in the users' guidelines of the dictionary, the system of niching does not impair the outer access structure of the dictionary. The clustering in example (10) is an illustration of *single niching*, i.e. where the clustering of complex words with a mutual first component is restricted to a single niche attached to the article introduced by the basis lemma. The application of single niching is not confined to a clustering of unexplained compounds and derivations. Articles with a more comprehensive treatment can also be clustered by means of single niching.

Niching is a result of various textblocks containing full lemmata with their accompanying articles being condensed into one textblock, albeit not one article. Procedures of horizontal ordering, e.g. in examples (4), (5), (6) and (10), lead to a non-vertical ordering of the textblock-internal lemma file. This procedure is also known as textual grouping (cf. Bergenholtz, Tarp and Wiegand 1999: 1815).

Grouping and the application of a strict horizontal ordering can be an access impediment in a textblock where the clustered lemmata are guiding elements of articles with a more comprehensive treatment. In the planning phase of such a dictionary, the micro-architecture should be devised to allow a textblock structure which enhances rapid access to the clustered lemmata. This could probably lead to a textblock-internal ordering which eschews the principle of grouping in favour of a presentation of listing (cf. Hausmann and Wiegand 1989: 336), with each niched article commencing on a new line. Yet again, the specific nature of the data-presentation in these niched articles should determine the decision to apply a procedure of grouping or listing. However, whether grouping or listing applies does not change the status of the cluster as belonging to one textblock. Niching should ideally result in all the textblocks that are candidates for participation in the textual condensation process di-

rected at a specific basis lemma combining into one textblock, in which a strict alphabetical ordering guides the user to the desired lemma sign. From a user-perspective, such a niche can be accessed in an unproblematic way.

In his discussion of niching in the DGWDF, Wiegand 2002 indicates the problems a user experiences when the results of niching in a given dictionary stretch over numerous columns and even pages. He rightly concludes that this is the wrong position for the application of space-saving procedures. When these procedures of clustering in the DGWDF are observed, a different picture is exposed from the applications of niching so far described. In this dictionary, the macrostructure is characterised by niching but it happens quite often that there is only a relation of distant addressing (cf. Louw and Gouws 1996: 98) between a niche and the relevant basis lemma.

In the DGWDF, all articles selected for participation in textual condensation, by means of clustering directed at one basis lemma, are not necessarily accommodated in the same niche but frequently in a series of niches separated by articles not participating in the relevant clustering (cf. Stark 2001: 21). This is illustrated by the presentation of lemmata with *Haus/haus* as first component:

- (11) **Haus** [haus], **das**; ~es, Häuser ['hɔɪzɐ]; vgl. *Häuschen* 1.1. 'ein- od. mehrstöckiges Gebäude, das zum Wohnen dient'; ↗ FELD I.15.1: *ein hohes, modernes, baufälliges, neues, altes* ~; *ein* ~ *bauen, einrichten, beziehen, sanieren, verkaufen, umbauen, renovieren, abreißen*; *ein* ~ *besitzen*; *das* ~ *steht unter Denkmalschutz*; *von* ~ *zu* ~ *gehen* 1.2. (o.Pl.; vorw. o.Art.; + Präp.) /meint den Bereich, in dem jmd. (ständig) wohnt/: *jmdn. aus dem* ~ *jagen*; ...
- Haus/haus** [..] | **-arbeit, die** 1. (o.Pl.) 'im Haushalt (1) anfallende Arbeiten wie Waschen, Kochen u.Ä.': *nach dem Dienst ist noch allerlei* ~ *zu erledigen*; *sie schafft die* ~ *nicht mehr* 2. '(umfangreiche) schriftliche Hausaufgabe': *eine* ~ *in Biologie, Geschichte schreiben müssen* ❖ ↗ *Arbeit*; **-aufgabe, die** (oft im Pl.) 'dem Schüler vom Lehrer aufgetragene Aufgabe, die zu Hause erledigt werden soll': *~n aufgeben, aufhaben*; *die* ~ *n erledigen, machen* ❖ ↗ *gegeben*; **-backen** (Adj.; Steig. reg., ungebr.) 'bieder, reizlos und etwas langweilig wirkend' /auf Personen, Denkweise bez./: *jmd. hat* ~ *e Ansichten*; *seine* ~ *e Kleidung*; *sich* ~ *kleiden* ...
- Häuschen** ['hɔɪʃçən], **das**; ~s, ~; vgl. *Haus* 'kleines Haus': *wir haben ein* ~ *am See* ❖ ↗ **Haus** * umg. /jmd./ **aus dem** ~ **geraten** ...
- hausen** [houzŋ] (reg. Vb.; hat) 1. /jmd./ *irgendwo* ~ 1.1. 'unter schlechten Bedingungen irgendwo dürftig wohnen'; ↗ FELD V.2.2: *in Baracken* ~ (*müssen*) 1.2. ...
- Haus/haus** [haus..] | **-frau, die** 'den eigenen Haushalt führende (und nicht berufstätige) Frau': *sie is (nur)* ~, *ist* ~ *und Mutter* ❖ ↗ *Frau*; **-halt, der**; ~s/auch ~es, ~e 1.1. ...
- Haus halten** (er hält Haus), hielt Haus, hat Haus gehalten /jmd./ *mit etw.* ~ 'mit etw. sparsam umgehen' ...
- Haus/haus** | **-herr, der** veraltend 'der Familienvater (als Gastgeber)': *der* ~ *begrüßte die Gäste, schenkte den Wein ein* ❖ ↗ *Herr*; **-hoch** ...
- häuslich** ['hɔɪs.] (Adj.) 1. (o. Steig.; nur attr.) 'das Zuhause, die Familie, den Haushalt (1.2) betreffend': *die* ~ *en Verhältnisse von jmdm. (nicht) kennen* ...

Haus [haus..] | **-nummer, die** 'eine der laufenden Nummern, mit denen die Häuser einer Straße bezeichnet sind': *bei der Adresse Straße und ~ angeben* ❖ ↗ Nummer; **-schuh, der** (vorw. Pl.) 'bequemer, aus weichem Material gefertigter Schuh, der nur in der Wohnung getragen wird': *~e an-, ausziehen* ❖ ↗ Schuh; **-seggen, der** * umg. scherzh. **bei jmdm. hängt der ~ schief** ('in jmds. Ehe gibt es aus einem bestimmten Anlass (längeren) Streit, herrscht gereizte Stimmung'); **-suchung** [zuç..], **die**; ~, ~en 'polizeiliche Durchsuchung einer Wohnung, eines Hauses zur Aufklärung einer Straftat': *bei jmdm. eine ~ durchführen; die ~ hat nichts ergeben* ❖ ↗ suchen; **-tier, das** 'vom Menschen bes. zur wirtschaftlichen Nutzung gehaltenes Tier'; ↗ FELD II.3.1: *die Katze ist ein ~; Kühe, Pferde, Schafe, Hunde sind ~e* ❖ ↗ Tier; **-wirt, der** 'jmd., der Eigentümer eines Hauses ist, in diesem Haus wohnt und die anderen Wohnungen des Hauses vermietet hat': *der ~ lässt unsere Wohnung sanieren, hat die Miete erhöht* ❖ ↗ Wirt

In this partial article stretch in example (11), *Haus* is presented as a main lemma and the guiding element of an article displaying a full treatment. The next guiding element presented in the vertical ordering of the macrostructure of the DGWDF is the extended lemma-external niche entrance *Haus/haus*. These lemma parts have been phased out to the left and introduce a cluster in which the partial lemmata *-arbeit, -aufgabe* and *-backen* head the niched articles. This cluster of grouped lemmata is followed by two vertically ordered main lemmata, i.e. *Häuschen* and *hausen*, preceding a next cluster with *Haus/haus* as phased out lemma parts in the entrance position of a new niche, which contains the sublemmata *-frau, -halt* and *-halten*. This niche is followed consecutively by the main lemma *Haus halten*, by another niche featuring *Haus/haus* in the entrance position and the partial lemmata *-herr* and *-hoch* in the cluster, before *häuslich*, as a new main lemma, interrupts the niching. The treatment of sublemmata with *Haus* as first component is concluded with the presentation of *Haus* as lemma-external niche entrance and the cluster of articles headed by the partial lemmata *-nummer, -schuh, -seggen, -suchung, -tier* and *-wirt*.

The partial article stretch which starts with the textblock introduced by the lemma sign *Haus* and ends with the final article, headed by the grouped partial lemma sign *-wirt*, in the condensed text block introduced by the lemma-external niche entrance *Haus* is constituted by different textblocks. As a partial article stretch it is therefore not an example of textual grouping (cf. Bergenholtz, Tarp and Wiegand 1999: 1815) neither is it a functional article stretch (cf. Wiegand 2002).

This partial article stretch contains various niched articles with either *Haus* or *haus* as first component. These niched articles display a semantic and a morphological relation with both the basis lemma, i.e. the word *Haus*, and the other main lemmata in this particular article stretch, i.e. the complex forms *Häuschen, hausen, Haus halten* and *häuslich*. The reasons why these complex forms have been presented as main and not as niched lemmata are not made known to the user. A consistent application of a traditional system of clustering could have rendered all the niched articles and the main lemmata in this partial article

stretch into one text block — ignoring for the moment the orthographic differences regarding the initial capital/small letter of the first component. Instead, the lexicographer has opted not to condense all the lemmata with *Haus/haus* as first component into one cluster but rather to present some of them as main lemmata and some grouped into different clusters. Such hybrid niching does not enhance the quality of the outer access structure.

In this partial article stretch from the DGWDF, the lexicographer repeatedly applies the procedure of niching to render a presentation of *multiple niching*, i.e. niching in one partial article stretch, characterised by the repetitive occurrence of niches with the same lemma part operating in the niche-external entrance position. The partial article stretch in example (11) contains no less than four niches resulting from a procedure of multiple niching.

The biscopal bilingual dictionary *Groot Woordeboek/Major Dictionary* (henceforth abbreviated as GW), with Afrikaans and English as treated language pair, utilises a system of niching comparable to that of the DGWDF. In the GW, the application of single niching which leads to a clustering of articles grouped into one text block with the basis lemma functioning, through a textual condensation process of phasing out to the left, as a remote niche entrance can be regarded as the default pattern, cf. the following example:

- (12) **pup'pet**, speelpop, handpop, draadpop, marionet; strooipop; werktuig, speelbal, figurant; ~ **government**, marionetteregering; ~-**man**, marionetspeler; ~-**play**, poppespel; ~-**ry**, poppespel; skynvertoon; poppekastery; ~-**show**, poppekas, poppespel, marionet(te)spel; ~ **state**, vasalstaat, popstaat; ministaat; ~-**valve**, stootklep.

This clustering adheres to the criteria for single niching. On both the internal and the external level a strict alphabetical ordering is maintained. The grouping of articles gives the user access to the niched lemmata via the main lemma which functions as basis and as niche entrance lemma. This niching does not confront the experienced user with too many problems when retrieving the desired information from the GW. However, this is not the only way in which niching is performed in the GW. This dictionary frequently indulges in the application of procedures of multiple niching, motivated by an exaggerated adherence to the alphabetical approach. By doing so it stretches the limits of article clustering to domains where the target user struggles to find lemmata included in the article niches.

5. Multiple niching in partial article stretches which include semantically related main lemmata

As is the case in the DGWDF's treatment of niched articles with *Haus/haus* as first component, the GW also displays partial article stretches which include repetitive niching of complex words with the same first component as well as

main lemmata heading non-niched articles. Quite often all the lemmata included in such partial article stretches are semantically related because the main lemmata and the sublemmata share mutual components. In these partial article stretches, the niches resulting from a procedure of multiple niching are not interrupted by unrelated main lemmata, cf. the following example:

- (13) **soft**, (n) papbroek, stumper; (a, adv) sag (saf); mal; ... **~ball**, sagtebal; **~boiled**, sag gekook; **~brimmed hat**, bolhoed(jie), dotjie; **~ cover**, slap band; ... ; **~ener**, versagter, leniger; versagmiddel.
soft'ening, (n) verweking, versagting; ontharding (water); (a) versagtend; *~ of the BRAIN*, harsingverweking; *~ UP*, makmakery.
soft: **~ furnishings**, gordynstowwe; meubelstowwe; **~ goods**, wolstowwe, ... ; **~ sore**, sagte (veneriese) sweer; **~spoken**, vriendelik, sag.
soft spot, teer plek, swak; *attack someone's ~*, iem. in sy swak tas; *have a ~ for women*, 'n swak hê vir vroue.
soft: **~ steel**, weekstaal; **~ talk**, mooipraatjies; **~ touch**, ligte aanslag; ... ; **~ wa-ter**, sagte water; **~wood**, greinhout; naaldhout; jonghout; **~y**, (**..ties**), papbroek, stumper, sul; teerhartige, goeierd.

This partial article stretch is the target of textual condensation by means of, among other methods, niching. However, instead of using the main lemma *soft* as a basis lemma giving access to all the niched articles, the lexicographer has opted for a different approach. The micro-architecture of the articles in the GW allows that an article niche can be attached to the article of a main lemma to form one textblock, e.g.:

- (14) **soft**, (n) papbroek, stumper; (a, adv) sag (saf); mal; ... **~ball**, sagtebal; **~boiled**, sag gekook; **~brimmed hat**, bolhoed(jie), dotjie; **~ cover**, slap band; ... ; **~ener**, versagter, leniger; versagmiddel.

The macrostructural elements in this niche (starting with **~ball**) are not presented as full lemmata but only as partial lemmata. These pre-hyphenated entries are not the treatment units. Each partial lemma contains the second component of the relevant treatment unit and this component is preceded by a hyphen functioning as place-keeping symbol, substituting the element given in the remote niche entrance, here the lemma part *soft*. The full treatment units can only be identified through a process of textual reconstruction via the basis lemma as search route. In this regard the structure of this niche and the procedures to reconstruct the treatment units do not differ from those applicable in clusters resulting from single niching, e.g. example (12), the cluster introduced by *puppet* as basis lemma.

The given partial article stretch in example (13) contains two other niches in which complex forms with *soft* as first component have been clustered:

- (15) **soft**: **~ furnishings**, gordynstowwe; meubelstowwe; **~ goods**, wolstowwe, ... ; **~ sore**, sagte (veneriese) sweer; **~spoken**, vriendelik, sag.

soft spot, teer plek, swak; *attack someone's ~*, iem. in sy swak tas; *have a ~ for women*, 'n swak hê vir vroue.
soft: ~ **steel**, weekstaal; ~ **talk**, mooipraatjies; ~ **touch**, ligte aanslag; ... ; ~ **water**, sagte water; ~-**wood**, greinhout; naaldhout; jonghout; ~**y**, (**..ties**), papbroek, stumper, sul; teerhartige, goeierd.

These niches constitute separate textblocks and they are not attached to the article or textblock of a basis lemma. The articles included in these niches participated in the process of textual condensation that have clustered certain complex forms into the textblock introduced by the basis lemma. However, the alphabetic bias prevailing in this dictionary has ensured the establishment of additional niches, separated from the textblock of the basis lemma by other textblocks featuring main lemmata and their treatment. The niches in example (15) are the result of multiple niching and more specifically a process of *remote multiple niching*, i.e. the formation of niches in textblocks separated from the textblock which accommodates the basis lemma by one or more textblocks headed by their own main lemmata. The further a niche is removed from the basis lemma the more difficult it is for the user to perform a successful dictionary consultation procedure in terms of rapidly accessing the textblock which contains the relevant niched lemmata.

From both a space-saving perspective, the dominant reason for lemma clustering, and a semantic perspective there is no reason why *softening* and *softspot* should have been given main lemma status. A consistent application of textual condensation by means of lemma clustering could have accommodated these two lemmata along with all those included in the three niches presented in this partial article stretch in one amalgamated textblock. The inclusion of *softening* and *softspot* as main lemmata could have been motivated by their usage frequency. This, however, is not of importance in the present discussion. If, for whatever reason, their inclusion as main lemmata is justifiable, the question remains how the other complex words with *soft* as first component should have been ordered as sublemmata. They could easily have been included in one textblock attached to the main lemma *soft*, functioning as basis lemma and lemma part in lemma-external remote cluster entrance. An internal alphabetical ordering could easily have been maintained in this cluster.

In the vertical ordering of lemmata the main lemma *soft* would then have been followed by the main lemmata *softening* and *softspot*. By doing so the lexicographer would have deviated from a strict alphabetical ordering because the cluster of lemmata does not adhere to an external alphabetical ordering. According to a strict alphabetical ordering, both the lemmata *softening* and *softspot* should have preceded some of the lemmata in the cluster. This presentation would have been an example of first level nesting.

Niching, by definition, does not take cognisance of morpho-semantic relations but only adheres to a strict alphabetical ordering. When niching leads to the clustering of all the sublemmata with a mutual first component into one text block, the users of the dictionary get accustomed to a system where one main lemma gives access to all the sublemmata relevant to the basis lemma. Al-

though this kind of textual condensation, especially in a niche with many sub-articles, can create some retrieval difficulties for the user, the systematic application of this type of clustering can lead to a relative easy access to the desired lemma signs. Grouping in an article cluster leads to a sinuous lemma file where each article does not start on a new line. In a niched ordering where the cluster contains many articles, the application of a system of listing or non-grouping where the micro-architecture allows each sublemma to be positioned at the beginning of a new line, enhances user-friendliness and an easier access to the lemma signs.

Whether a system of grouped or listed sublemmata is employed the essence of niching should be regarded as the clustering of lemmata into a single text block. Multiple niching confronts the user with a more problematic information retrieval challenge due to the more complicated external access route. The basis of multiple niching is an ardent alphabetic bias. This bias often prevails at the cost of user-friendliness. Multiple niching leads to the formation of more than one article cluster. In an application of first level nesting the same sublemmata would have been clustered into one textblock, attached to the article of one basis lemma. However, due to multiple niching they are accommodated in different niches, constituting different textblocks. These niched textblocks are interrupted by other textblocks presenting non-clustered main lemmata. From an access-structure perspective, multiple niching increases the difficulties users experience in their attempt to reach a desired lemma sign. This applies even if the main lemmata heading the interrupting textblocks are semantically related to the niched lemmata. The search route is no longer directed only at one cluster but the user has to go through the first cluster and then has to cross textblock barriers, negotiate the articles of main lemmata to which no clusters are attached and only then move into further niched clusters in the hope of finding the desired lemma.

The application of multiple niching leads to additional problems for the dictionary user. When a dictionary employs procedures of multiple niching it necessarily also has to use single niching to establish the default niches. Such a system relies on the use of different types of niches. In the default niche, the article of the lemma *puppet* (cf. example (12)), the user can find all the subarticles directed at one basis lemma accommodated in one textblock. Contrary to this, the occurrence of multiple niching spreads these subarticles over various textblocks. This hybrid method of niching adds to the problems the typical target user of a dictionary experiences during a dictionary consultation procedure. The combination of single niching and first level nesting will ensure that a user can consistently find all the subarticles clustered into one textblock.

6. Multiple niching in partial article stretches which include semantically unrelated main lemmata

Niched articles resulting from the application of multiple niching are more easily accessible in a partial article stretch where all the lemmata are semantically related (cf. example (13)). When the niches are interrupted by lemmata

which do not display a semantic relation with the basis lemma of the first in the series of multiple niches, access to the subarticles participating in a procedure of multiple niching is further impeded, cf. the treatment of complex words with *sea* as first component in the GW:

- (16) **sea**, see; deining, golf, branding; oorvloed; menigte; AT ~, op see; ...; *a ~ of TROUBLES*, 'n oseaan van leed; ~ **acorn**, eendemossel; ~-**action**, seegeveg; ~ **aerodrome**, ~ **airport**, ...; ~-**kale**, seekool; ~-**king**, seekoning; ~-**kit**, seemansuitrusting.
sea¹, (n) seeleeu, rob; (v) robbe vang.
sea², (n) seël; stempel; beseëling; stempelafdruk; ykmerk; lak; *the GREAT S-*, die Rykseël; ...
sea: ~ **lace**, seewier; ~ **lamprey**, seeprik; ...; ~ **lawyer**, vitterige persoon; korrelkop; ~-**legs**, seebene; *find (get) one's ~-legs*, jou seebene kry, leer om jou maklik op die dek te beweeg.
sea'er¹, robbejagter; robbeskip.
sea'er², verseëlaar; deklaag.
sea'-letter, seebrief.
sea'-level, seevlak, seespieël; *ABOVE ~*, bokant die seespieël (seevlak); *AT ~*, by die seevlak (seespieël).
sea'-fishery, robbevissery, robbevangs.
Seal Island, Seal Island.
sea: ~ **life**, seelewe; ~-**line**, horison, kim; kus; lang vislyn.
sea'ing¹, robbejag.
sea'ing², bekragtiging, bevestiging; digting; ~ **agent**, digtingsmiddel; ~ **coat**, deklaag; afdiglaag; ~ **ring**, digtingsring; ~ **tree**, kiaal (*Pterocarpus angolensis*).
sea'ing vessel, robbevaarder.
sea'ing-wax, lak.
sea: ~-**lion**, seeleeu; **S~ Lord**, lid van die admiraliteit in Engeland.
sea'-ring, seëling.
sea: ~-**rookery**, robbekolonie; ~-**skin**, robbevel.
seam, (n) soom, naat; aar, laag (kole); rimpel; ...
sea: ~-**man**, seeman, matroos; seevaarder, varengesel; ~-**manship**, seemanskap, stuurmanskap; ~-**mark**, seebaken; ~-**men's institute**, seemanshuis; ~-**mew**, seemeeu; ~ **mile**, seemyl; ~ **mist**, seenewel; seemis.
sea'mer, naatbouler.
seam: ~ **bowler**, naatbouler; ~-**bowling**, naatboul; ~-**bowling attack**, naat(boul)-aanval; ~-**lap**, naatrand; ~-**less**, sonder naat, naatloos.
sea' monster, seemonster.
seam: ~-**presser**, parsyster; ~-**stress**, (-es), naaister; ~-**weld**, naatsweislas; ...; ~-**y**, vol nate; ...
sé'ance, sitting, séance.
sea: ~ **needle**, geep; ~-**nymph**, seenimf; ~-**onion**, seeajuin; ...; ~-**quake**, seebewing.
sear¹, (n) tuimelaar (geweer).
sear², (v) brand, verskroei, uitbrand, toeskroei, digskroei; verhard; ~ *a wound*, 'n wond toeskroei; (a) droog, dor.
search, (n) ondersoek; soekery, soek; ...; ~-**able**, te ondersoek; ~-**er**, soeker; deur-

soeker, huissoeker; visenteeryster; botterboor; visenteermes; sondeernaald, peilstif; **~ing**, (n) ondersoek; deursoeking; (a) deurdringend, ondersoekend, deurvorsend, grondig; noukeurig, diepgaande; ... ; **~-warrant**, visenteerbrieff, lasbrief vir huissoeking.

sea' reed, helmgras.

sear'ing-iron, brandyster.

sea: **~-risk**, seerisiko; **~ rover**, seeskuimer, roofskip; swerwer op see; **~-salt**, seesout; **~-sand**, seesand; ...; **~sick**, seesiek; **~sickness**, seesiekte.

sea'side, strand; seekant; *AT the ~*, aan die strand; *GO to the ~*, strand toe gaan; **~ bungalow**, strandhuis; **~ cottage**, strandhuisie; **~ resort**, strandplek, strandoord; **~ wear**, stranddrag.

sea: **~ slug**, seeslak; tripang, seekomkommer; **~ snail**, seeslak; **~ snake**, seeslang.

sea'son, (n) seisoen, jaargety; geskikte tyd; tydperk; bronstyd; ... ; **~ MORTAR**, kalk insuur; **~ oneself to hard WORK**, jou aan harde werk gewoon maak; **~able**, geskik, geleë, tydig; aktueel; passend by die seisoen.

sea'sonal, seisoen(s)_≠, jaargetye_≠, seisoenaal, binneseisoens; **~ greetings**, Kers_≠ en Nuwejaarsgroete; **~ valve**, weerklep, seisoensklep; **~ industry**, seisoenbedryf; **~ variation**, jaargetyverskil; **~ worker**, seisoenwerker.

sea'soned, gekruie, gesout, gepeper; beleë, droog (hout); **~ food**, gekruide kos; **~ iron**, getemperde yster; **~ soldiers**, geharde soldate; **~ traveller**, deurwinterde reisiger; **~ wood**, gedroogde (beleë) hout.

sea'son: **~ing**, voorbereiding, klaarmaak; kruiesous, smaakmiddel, kruisel; geursel; ontvarsing; kruidery; droging (hout); **~less**, sonder jaargety; **~-ticket**, seisoenkaartjie.

sea' spray, bruiswater, seeskuim, brandersproei.

seat, (n) sitplek; stoel; gestoelte; sitting; setel; toneel; buiteplaas; ...

sea' tangle, seewier.

seat: **~back**, rugleuning; **~belt**, sitplekgordel; veiligheidsgordel.

sea term, seemansterm; vlootterm; **~ ~inology**, vlootterminologie, seeterme, seeterminologie.

seat'ing, sitplek; onderstuk; fundasie; **~ accommodation**, sitplek; plaasruimte.

sea: **~-toad**, seeduiwel; **~-tossed**, deur die see geslinger; **~ training**, seeopleiding; **~ trout**, seeforel; **~ trumpet**, trompetskulp; **~ turtle**, seeskilpad; ...; **~worthy**, seewaardig; **~-wrack**, seegras.

The process of textual condensation to cluster complex words with *sea* as first component leads to multiple niching. The given partial article stretch, example (16), contains the article with *sea* as main lemma with a niche attached to this article and no less than eight niches with *sea* as lemma-external lemma part in the niche entrance position, constituting a procedure of remote multiple niching. A number of main lemmata included in this partial article stretch show a semantic relatedness with *sea*. Besides the *sea*- and *sea*-related lemmata this partial article stretch also contains a variety of lemmata which are semantically unrelated to the main lemma *sea*. This partial article stretch also contains other examples of niching and multiple niching. The ordering of all the lemmata in this partial article stretch is only determined by the application of a strict alphabetical principle.

This strict alphabetical ordering is, however, not user-friendly (cf. Stark 2001: 21). The user looking for, say, *seasick* in the GW goes to the article of the lemma *sea* and the attached niche where the desired form is not found. The following textblocks in the partial article stretch contain lemmata unrelated to *sea* and the user easily deduces that *seasick* has not been included in the macrostructure of this dictionary. If the user does look a bit further he/she may notice some other niches with *sea*-words and he/she could eventually, with some luck and many problems, reach the lemma *seasick*.

One real problem experienced in struggling through a partial article stretch like this is that the user never knows for sure whether a specific niche, attached to the article of a basis lemma, is the only niche accommodating complex words with the form represented by the basis lemma as first component or whether there are some other relevant alphabetically ordered niches further down the article stretch. Having reached an article cluster which presents one of the multiple niches for a given first component, the user is still at a lost and does not know whether any other niches prevail. This enhances user-frustration.

The clustering of *sea*-lemmata could have looked like this:

- (17) **sea**, see; deining, golf, branding; oorvloed; menigte; AT ~, op see; ...; *a ~ of TROUBLES*, 'n oseaan van leed; ~ **acorn**, eendemossel; ~-**action**, seegeveg; ~ **aerodrome**, ~ **airport**, ...; ~-**kale**, seekool; ~-**king**, seekoning; ~-**kit**, seemansuitrusting; ~ **lace**, seewier; ~ **lamprey**, seeprik; ... ; ~ **life**, seelewe; ~-**line**, horison, kim; kus; ... ; ~-**lion**, ... ; ~-**man**, seeman, matroos; ... ; ~ **monster** ... ; ~ **needle**, geep; ... ; ~ **reed**, helmgras; ~-**side**, strand; seekant; AT *the ~*, aan die strand; *GO to the ~*, strand toe gaan; ~-**side bungalow**, strandhuis; ~-**side cottage**, strandhuisie; ~ **slug**, seeslak; ... ; ~ **spray**, bruiswater, seeskuim ... ; ~ **term**, seemansterm; ... ; ~-**toad**, seeduiwel; ~-**tossed**, deur die see geslinger; ... ; ~-**worthy**, seewaardig; ~-**wrack**, seegras.

In the partial article stretch from the GW, example (16), similar changes could also have been made to the niching of words with *seal*¹, *seam* and *season*. In all these clusters the use of a lesser degree of textual condensation regarding the lemma sign, i.e. refraining from place-keeping symbols and the use of partial lemmata, as well as the use of a system of listing could have added to the user-friendliness of this dictionary. This will, however, not be discussed in the present article.

Although the proposed clustering of *sea*-words would have resulted in a lengthy textblock with a relative high degree of textual condensation there would only have been one clear search route in the outer access structure guiding the user to the relevant lemma sign.

The suggested textblock for the treatment of *sea*-words deviates from the external alphabetical ordering and represents an application of first level nesting. As Gouws (2001: 108) has indicated, multiple niching confronts the dictionary with many more problems than either first level nesting or single niching.

Both these procedures are much more user-friendly than multiple niching. Especially when multiple niching results in a partial article stretch where the niches are interrupted by main lemmata showing no semantic relation to the basis lemma or the lemmata participating in the clustering, users can easily lose their way on the search route of the outer access structure. For this reason the lexicographer preparing a user-driven dictionary will do well to avoid multiple niching.

7. Exclusively remote multiple niching

The following partial article stretch comes from the GW:

- (18) **salt**, (n) sout; silt; seerob, ... ; *as ~ as LOT'S wife*, so sout soos brem; *a ~ RIB of mutton*, 'n soutribbetjie.
sal'tant, springend (*her.*); dansend.
salta'tion, dans; dansery; springery; sprong.
sal'tatory, springend.
salt: ~-**bag**, soutsak; ~-**box**, soutvaatjie, soubak; ~-**briquette**, soubriket; ~-**bush**, vaalbos, soubos; ... ; ~-**fish**, soutvis; ~-**herring**, southaring.
sal'tigrade, (n) springspinnekop; (a) springend, spring_z.
salt: ~-**iness**, souterigheid, siltigheid; ~-**ing**, insouting; ... ; ~ **monopoly**, soutmonopolie; ~-**ness**, southeid; ~-**pan**, soutpan.

The textblock in which the article of the lemma sign *salt* appears has no niche attached to it. However, the given partial article stretch contains two niches accommodating complex words with *salt* as first component. Both these niches function as separate textblocks introduced by a lemma part in a lemma-external niche entrance position. This is yet another example of multiple niching and more specifically remote multiple niching. In partial article stretches like the one presented as example (13), remote multiple niching complemented niching attached to the article carrying the basis lemma. Contrary to this, example (18) illustrates the result of *exclusively remote multiple niching*, i.e. when none of the niches participating in a given procedure of multiple niching are attached to the article of the basis lemma.

Being familiar with niching and the use of the default niched article in the GW, the user starts the search route at the basis lemma and then works towards the attached niche. Having found no niche attached to the article of the lemma sign *salt*, the user can easily deduce that a word like *saltmonopoly* has not been included in the dictionary. Exclusively remote multiple niching can be regarded as an additional impediment in the outer access structure of a dictionary.

8. Homonymy and multiple niching

In dictionaries, homonyms are usually entered as separate macrostructural entries identified as such by superscript numbers immediately preceding or

following the lemma sign, cf. the following example from the EIHat:

- (19) **grif**¹ b.nw. en bw. (*deftig*) Vlug, geredelik, vaardig; grifweg: *Hy is nie juis grif van begrip nie. Grif op 'n vraag antwoord. Iets grif erken.*
grif² ww. (gegrif) **1** Met 'n stif kras, skryf, aanbring: *Letters in 'n klip grif. Gegrifte tekens.* **2** (*fig.*) Blywend, onuitwisbaar aanbring asof met 'n stif gekras: *Dit staan in my geheue gegrif.*

In a dictionary where procedures of niching prevail, each member of a homonym pair which is included as a vertically ordered main lemma could, where applicable, function as basis lemma for a niche with the niche attached to the same textblock, or even as lemma part in lemma-external niche entrance.

Afrikaans has three lexical items *bak* and they should be entered as three separate treatment units in a dictionary, cf. the following example from the EIHat:

- (20) **bak**¹ □ s.nw. (-ke) **1** (Oop) kis waarin iets bewaar kan word: *'n Bak vir meel, hout, steenkool.* **2** Deel van wa, kar, motor, ens. wat op die onderstel rus. **3** Houer vir vloeistof: *'n Bak met water. Drink-, suip-, doopbak.* **4** Hoeveelheid wat in 'n bak gaan: *'n Bak water, meel.* **5** Hol kant van 'n kromming: *Die plate staan met 'n bak.* UITDR.: *In die bak raak*, agter raak. *Iemand in die bak sit* (nie alg.), hom oortref. □ b.nw. en bw. Soos 'n bak: *Sy bene, ore staan bak. Die hande bak hou.*
bak² ww. (bakkende; gebak) **1** Gaarmaak deur hitte: *Eiers in die pan bak. Gebakte eiers.* **2** Gaar word; warm kry: *In die son sit en bak.* **3** Hard (laat) word deur hitte: *Stene bak.* **4** Hitte afgee: *Die son bak op die stoep. In die bakkende son.* ▽ *Die son bak blink* (M.M. Walters). UITDR.: *Mooi broodjies bak* — sien onder BROOD. *Bak en brou*, knoei. **bak**: ~dag, ~hoender, ~huis, ~kis, ~mengsel, ~skottel, ~trog.
bak³ b.nw. en bw. (*geselst.*) Piekfyn, uitstekend: *'n Bak kèrel. Dit gaan vandag net bak.* Vgl. BAKGAT.
ba·ka·tel' (-le) **1** Nietigheid: *Waarom al die lawaai — die hele saak is net 'n bakatel.* **2** Sien BAGATEL. [F. *bagatelle* kleinigheid]

The article of the second lemma sign *bak* has an attached article cluster. This cluster features a strict internal alphabetical ordering of lemmata but introduces a deviation from the strict external alphabetical ordering. The next main lemma, i.e. *bak*³ representing the last member of the homonym paradigm, and the following main lemma (*bakatel*) should, in an application of strict alphabetical ordering, have been ordered prior to all the lemmata included in the article cluster. The fact that clustering in the EIHat is not restricted to niching but that allowance is also made for first level nesting leads to a situation where a niche could be attached to the article of the lemma representing any member of a homonym paradigm. The user is given the advantage of finding a clustered lemma by progressing via a basis lemma which displays a semantic relation with the first stem of the complex lexical item represented by the clustered lemma. This is user-friendly lexicography.

Where only niching is employed instead of both niching and first level nesting, clustering may not be performed in the articles of the non-final members of a homonym paradigm because it would lead to a deviation from the external alphabetical ordering. In a dictionary like the GW, the articles of these non-final members of a homonym paradigm are consistently excluded from niching procedures, cf. the following example:

- (21) **roei**¹, (s) (-e), tail of a comet; lattice; mullion.
roei², (w) (ge-), row; pull (oars); ~**bank**, rowing-bench; thwart; ~**boot**, rowing-boat; ~**dol**, rowlock; ~**er**, (-s), rower; oarsman, waterman; sculler; oar; ~**klamp**, rowlock; ~**klub**, boating (rowing) club; ~**kuns**, oarsmanship; watermanship; ~**mik**, rowlock; ~**pen**, thole; ~**plank**, surf board; ~**plankry**, surfing; ~**riem**, oar; ~**spaan**, (.spane), oar, scull; paddle; ~**sport**, rowing.
roei'ster, comet.

The alphabetical bias allows the niche to be attached to the article of the lemma sign *roei*² but no niche may be attached to the article of the lemma sign *roei*¹. Instead, a separate main lemma *roeister* is included. This article could have been attached to that of the lemma sign *roei*¹ if a procedure of first level nesting had been permitted.

When only niching instead of both niching and first level nesting as a form of external textual condensation (cf. Wiegand 1998) is utilised, the adherence to a strict alphabetical ordering compels the lexicographer to employ procedures of multiple niching to lemmatise and treat complex lexical items with a first component which is a stem variant of a lexical item functioning in a homonymous relation with another lexical item, with both these lexical items included as main lemmata. This is illustrated by the next example from the GW:

- (22) **roes**¹, (s) rust, blight, brand (in corn); *hy IS 'n ~*, he is a wastrel; *'n ~ in iem. se SAK*, a drain on one's pocket; (w) (ge-), rust, corrode; get rusty; *ou LIEFDE ~ nie*, true love never grows old; *my REKENKUNDE is al 'n bietjie ge~*, my arithmetic is a bit rusty.
roes², (s) intoxication bout, drunken fit; ecstasy, frenzy; flush (of victory); *sy ~ UITSLAAP*, sleep off the effects of one's debauch; *in die ~ van VRYHEID*, in the intoxication of liberty.
roes: ~**agtig**, (-e), rusty; ~**bruin**, rust-brown; ~**dopluis**, rusty scale.
roe'semoes, confusion, disorder, tumult, bustle; *see also geroesemoes*.
roe'serig, (-e), rusty; ~**heid**, rustiness.
roes: ~**hoop**, scrap-heap; ~**kleur**, rust colour; ~**kleurig**, (-e), rust-coloured, foxy; ~**laag**, patina; ~**oplosmiddel**, rust solvent; ~**rooi**, rust-red; ~**terig**, (-e), rusty; ~**terigheid**, rustiness; ~**vlek**, rust stain; ~**vry**, rust-proof, stainless; ~**werend**, (-e), anti-rust, anti-corrosion, rust-proof.

Both members of the homonym paradigm *roes* have been included as lemmata. The lexical item represented by the second lemma *roes* does not function as a stem in complex lexical items and consequently the article of this lemma has no

niche attached to it. Contrary to this, the lexical item represented by the first lemma *roes* frequently occurs as first stem of complex lexical items. For space-saving reasons the lexicographer wants to cluster these complex forms into a single textblock. Ideally a procedure of first level nesting would have attached this cluster to the article of the first lemma *roes*. This is prevented by adherence to a strict alphabetical ordering, and consequently procedures of remote niching and multiple niching have to be performed.

Lemmata representing homonyms are clearly identified to assist users in their endeavour to access the article of the proper source-language item. The application of multiple niching often impedes a clear and unambiguous access to the correct source-language form because the lemma part in lemma-external niche entrance cannot be linked with the appropriate main lemma without venturing into the rest of the niche. On the other hand, the consistent application of a strict alphabetical ordering does function as a prominent guideline for the user on the search route towards the needed source-language form.

9. Semantically motivated multiple niching

The GW adheres to a strict alphabetical ordering and this results in multiple niching. As already indicated, the strict alphabetical ordering leaves the article(s) representing the non-final member(s) of a homonym paradigm stranded when niches need to be attached to existing articles. A consistent adherence to the strict alphabetical principle should not allow semantic reasons to influence the ordering of lemmata in remote niches. The following example gives evidence to the contrary:

- (23) **kant¹**, (s) lace; *GEKLOSTE* ~, bobbin lace; *NAGEMAAKTE* ~, imitation lace; *met* ~ *VERSIER*, trim with lace.
kant², (-e), side, edge, brink, border, margin; limb (of a fold); flank (of army); face; hand; ...
kant'aantekening, marginal note; gloss.
kantari'de, (-s), Spanish fly, cantharides.
 ...
kant: ~*tussensetsel*, lace insertion; ~*verkoopster*, lace woman; ~*verkoper*, lace man.
kant: ~*voorspeler*, wing forward, flanker; ~*vyl*, cant, barette-file.
kant: ~*werk*, lace-work, lace; ~*werker*, lace-maker; lace man; ~*werkster*, (female) lace-maker; ~*winkel*, lace shop.

In this partial article stretch, incompletely presented here, the two lemmata *kant* represent two homonyms. The remote niches indicate an interesting aspect of the clustering of lemmata. The first and the third remote niche both contain lemmata which relate semantically to the first lemma *kant* whereas the lemmata in the second remote niche relate to the second lemma *kant*. If semantic reasons had not played a role these three niches could have been clustered into one

textblock:

- (24) **kant**: ~**tussensetsel**, lace insertion; ~**verkoopster**, lace woman; ~**verkoper**, lace man; ~**voorspeler**, wing forward, flanker; ~**vyl**, cant, barette-file; ~**werk**, lace-work, lace; ~**werker**, lace-maker; lace man; ~**werkster**, (female) lace-maker; ~**winkel**, lace shop.

The consistent application of a system like the one used in the GW can assist the user to access the needed data in a predictable way. An inconsistent application leads to user-frustration because it defies the important lexicographic criterion of predictability, cf. the following example:

- (25) **heel**¹, (w) (**ge-**), heal, cure; granulate (wound).
heel², (w) (**ge-**), receive (stolen goods).
heel³, (b) (**hele**), whole, entire, complete, unbroken, undamaged; sound; ...
 ...
heel: ~**konfyt**, conserve; whole preserve; ~**koring**, whole wheat; ~**krag**, healing (curative) power; ~**krag'tig**, (-e), healing, curative; ~**kunde**, surgery; ~**kun'dig**, (-e), surgical; ~**kun'dige**, (-s), surgeon; ~**kuns**, art of healing; ~**maak**, (-ge-), mend; ~**maker**, (-s), mender.

The niche contains lemmata which relate to the first lemma *heel*, e.g. *heelkrag*, *heelkragtig* and *heelkuns*, whereas others relate to the second, e.g. *heelkonfyt* and *heelkoring*. Such a mixed niche, and this dictionary contains quite a number of them, diminishes the potential success of the target user to gain rapid access to the desired lemma sign. The consistent application of a well-explained system is necessary to ensure the success of a dictionary as an authoritative reference source.

10. Heterogeneous niching

Due to its strong adherence to a strict alphabetical ordering, the application of niching can lead users astray when unrelated words are forced into a single niche, cf. the following example from the *Reader's Digest Afrikaans-Engelse Woordeboek / English-Afrikaans Dictionary*:

- (26) **mira'kel** ...
mirasi'dium *≈s* miracidium. **mirbaan'olie** mirbane oil, nitrobenzene. **Mir'jam** Miriam. **mirliton'** *≈s* mirliton, kazoo (gazoo). **Mirmidoon'** *≈one* Myrmidon.
mir're myrrh.
mirt ...

The lemma *mirasidium* is positioned in the default vertical ordering of the macrostructure. It is preceded by the vertically ordered main lemma *mirakel*. On macrostructural level, the lemma *mirasidium* is followed by the lemmata *mir-*

baanolie, *Mirjam*, *mirliton*, *Mirmidoon* and *mirre*. These lemmata display no semantic relations with the lemma *mirasidium*, neither are they complex words with *mirasidium* as one of the components. However, the lexicographer has opted to order them in a sinuous lemma file, resembling a niche with *mirasidium* in the apparent niche entrance position. This sinuous file is followed by *mirt*, the next vertically ordered lemma.

A typical application of single niching leads to the clustering of lemmata with a mutual first component. This can be regarded as the default type of single niching. The occurrence of a mutual component allows the phasing out to the left of that lemma part. Due to the presence of a mutual component in all the complex forms clustered into the niche, this can be regarded as *homogeneous niching*. In the given article cluster from the *Reader's Digest Afrikaans-Engelse Woordeboek / English-Afrikaans Dictionary*, the niched lemmata cannot participate in a process of phasing out to the left because not all of them are complex words and they have no mutual morpheme. This can be regarded as an example of *heterogeneous niching*.

To the user heterogeneous niching and the consequential horizontal ordering of the lemmata *mirbaanolie*, *Mirjam*, *mirliton*, *Mirmidoon* and *mirre* could be extremely confusing. When looking for a specific lemma, say *mirliton*, the user will not see the vertically ordered lemma *mirasidium* as a guiding element leading him to the needed lemma. Being familiar with the way in which single homogeneous niching is done in this dictionary, the user will interpret the vertically ordered textblocks

mira'kel ...
mirasi'dium
mirt ...

as an indication that *mirliton* has not been included in the macrostructure of this dictionary.

Unless the system of heterogeneous niching is explained comprehensively in the users' guidelines text, the application of such a niching procedure will be an impediment in the outer access structure of a dictionary. Procedures of textual condensation which cluster different dictionary articles into a single textblock should limit the participating lemmata to complex items which can be the target of a process of phasing out to the left due to the repetitive occurrence of a mutual morpheme in all the lemmata.

11. In conclusion

Niching is a lexicographic procedure which can help the lexicographer to render a more condensed macrostructure. The space-saving motivation for niching may never be at the cost of the users in their attempt to gain access to the data presented in the dictionary. Multiple niching can be regarded as a more complex form of textual condensation and it increases the level of difficulty of dic-

tionary-consultation procedures. This is especially true of procedures of exclusively remote multiple niching and multiple niching where the interrupting textblocks do not have main lemmata which are semantically related to the basis lemma or the sublemmata. Where multiple niching is not applied in a consistent way, e.g. where a basis lemma represents a member of a homonym paradigm and the underlying homonymy is not maintained in the niches, users struggle to access the articles from which they need to retrieve information. This is also the case where heterogeneous niching and hybrid niching procedures prevail.

Lexicographers should continue to evaluate the contribution of different niching procedures. A more general use of a more diverse presentation of article clustering which includes the use of procedures of first level nesting should also be investigated.

Endnotes

1. In this article, the term *textual condensation* is used to include both *Textverdichtung* and *Textkomprimierung* (cf. Wiegand 1998: 31).
2. In the electronic version of this dictionary, the textual condensation has been decreased with the niched lemmata presented in their full form.
3. It is important to distinguish these pre- and post-hyphenated *entries* from what Gouws (1989: 215) calls pre- and post-hyphenated *lemmata*. A pre- or post-hyphenated *lemma* is a main lemma presented in the vertical ordering of lemmata to introduce a stem or affix as treatment unit. Pre- and post-hyphenated *entries*, as used in this article, refer to lemma parts and partial lemmata participating in a procedure of lemma clustering to present sublemmata contained in a niche or a nest.

Bibliography

Dictionaries

- Bosman, D.B. et al. (Eds.). 1984⁸. *Tweetalige Woordeboek / Bilingual Dictionary*. Cape Town: Tafelberg.
- De Villiers, M. et al. (Eds.). 1988⁷. *Nasionale Woordeboek*. Cape Town: Nasou.
- Du Plessis, Madeleine. 1999. *New Words / Nuwe Woorte*. Cape Town: Pharos.
- Eksteen, L.C. et al. (Eds.). 1997¹⁴. *Groot Woordeboek / Major Dictionary*. Cape Town: Pharos.
- Eksteen, L.C. et al. (Eds.). 1992⁸. *Verklarende Afrikaanse Woordeboek*. Pretoria: J.L. van Schaik.
- Grobbelaar, P. (Ed.). 1987. *Reader's Digest Afrikaans-Engelse Woordeboek / English-Afrikaans Dictionary*. Cape Town: Readers' Digest Association South Africa.
- Kempcke, G. et al. (Eds.). 2000. *De Gruyter Wörterbuch Deutsch als Fremdsprache*. Berlin: Walter de Gruyter.
- Odendal, F.F. and R.H. Gouws (Eds.). 1998. *Elektroniese Verklarende Handwoordeboek van die Afrikaanse Taal*. Midrand: Perskor.

Other literature

- Bergenholtz, H.** 1990. Lexikographische Instruktionen für ein zweisprachiges Wörterbuch. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 43: 19-37.
- Bergenholtz, H., S. Tarp and H.E. Wiegand.** 1999. Datendistributionsstrukturen, Makro- und Mikrostrukturen in neueren Fachwörterbüchern. Hoffman, L., H. Kalverkämper and H.E. Wiegand (Eds.). *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.
- Gouws, R.H.** 1989. *Leksikografie.* Cape Town: Academica.
- Gouws, R.H.** 1991. Toward a Lexicon-based Lexicography. *Dictionaries* 13: 75-90.
- Gouws, R.H.** 2001. The Use of an Improved Access Structure in Dictionaries. *Lexikos* 11: 101-111.
- Hausmann, F.J. and H.E. Wiegand.** 1989. Component Parts and Structures of General Monolingual Dictionaries: A Survey. Hausmann, F.J. et al. (Eds.). 1989-1991: 328-360.
- Hausmann, F.J. et al. (Eds.).** 1989-1991. *Wörterbücher. Ein internationales Handbuch zur Lexikographie/Dictionaries. An International Encyclopedia of Lexicography /Dictionnaires. Encyclopédie internationale de lexicographie.* Berlin: De Gruyter.
- Louw, P.A. and R.H. Gouws.** 1996. Lemmatiese en nielemmatiese adressering in Afrikaanse vertalende woordeboeke. *S.A. Tydskrif vir Taalkunde* 14(3): 92-100.
- Stark, M.** 2001. *Die toegangstruktuur in verklarende aanleerderwoordeboeke.* Unpublished M.A. thesis. Stellenbosch: University of Stellenbosch.
- Wiegand, H.E.** 1989. Aspekte der Makrostruktur im allgemeinen einsprachigen Wörterbuch: Alphabetische Anordnungsformen und ihre Probleme. Hausmann, F.J. et al. (Eds.). 1989-1991: 371-409.
- Wiegand, H.E.** 1998. Lexikographische Textverdichtung. Entwurf zu einer vollständigen Konzeption. Zettersten, A., V.H. Pedersen and J.E. Mogensen (Eds.). 1998. *Symposium on Lexicography VIII:* 1-35. Tübingen: Max Niemeyer.
- Wiegand, H.E.** 2002. Über textuelle Strukturen der Wörterbuchartikel und Artikelnischen im "de Gruyter Wörterbuch Deutsch als Fremdsprache". Zugleich ein Beitrag zur Weiterentwicklung einer Theorie der Wörterbuchform. Wiegand, H.E. (Ed.). 2002. *Perspektiven der pädagogischen Lexikographie des Deutschen II:* 497-595. Tübingen: Max Niemeyer.
- Wolski, Werner.** 1989. Das Lemma und die verschiedenen Lemmatypen. Hausmann, F.J. et al. (Eds.). 1989-1991: 360-371.
- Wolski, Werner.** 1989a. Formen der Textverdichtung im allgemeinen einsprachigen Wörterbuch. Hausmann, F.J. et al. (Eds.). 1989-1991: 956-967.

The Ndebele Language Corpus: A Review of Some Factors Influencing the Content of the Corpus*

Samukele Hadebe, *Institutt for Nordistikk og Litteraturvitenskap, Seksjon for Leksikografi, University of Oslo, Oslo, Norway (samukeleh@yahoo.co.uk)*

Abstract: The Ndebele language corpus described here is that compiled by the ALLEX Project (now ALRI) at the University of Zimbabwe. It is intended to reflect as much as possible the Ndebele language as spoken in Zimbabwe. The Ndebele language corpus was built in order to provide much-needed material for the study of the Ndebele language with a special focus on dictionary-making and research. Like most corpora, the Ndebele language corpus may in future be used for other purposes not thought of at the time of its inception. It has been designed to meet generally acceptable standards so that it can be adaptable to various possible uses by various researchers. The article wants to outline the building process of the Ndebele language corpus with special emphasis on the challenges that faced compilers, and possible solutions. It is assumed that some of these challenges might not be peculiar to Ndebele alone but could also affect related African languages in a more or less similar situation. The main focus of the discussion will be the composition of the Ndebele language corpus, i.e. the type of texts that constitute the corpus. The corpus is composed of published texts, unpublished texts and oral material gathered from Ndebele-speaking districts of Zimbabwe. It will be argued that the use of the corpus and its reliability for research depends among other factors on its contents. It will also be shown that the contents of a corpus depend on a number of factors, some of which include sociolinguistic, political and economic considerations. These considerations have implications on both the content and quality of published and oral texts that constitute the Ndebele language corpus.

Keywords: CORPUS, ORAL MATERIALS, CODE-MIXING, CODE-SWITCHING, MOTHER-TONGUE, NDEBELE

Opsomming: Die Ndebeletaalkorpus: 'n Oorsig van sommige faktore wat die inhoud van die korpus beïnvloed. Die Ndebeletaalkorpus wat hier beskryf word, is dié saamgestel deur die ALLEX Project (tans ALRI) by die Universiteit van Zimbabwe. Dit is bedoel om soveel moontlik te weerspieël van die Ndebeletaal soos in Zimbabwe gepraat. Die Ndebeletaalkorpus is opgebou om veelbenodigde materiaal te verskaf vir die studie van die Ndebeletaal, met spesiale fokus op woordeboeksamestelling en navorsing. Soos die meeste korpora, kan die Ndebeletaalkorpus in die toekoms gebruik word vir ander doeleindes waaraan nie by tye van sy ontstaan gedink is nie. Dit is ontwerp om aan algemeen aanvaarde standaarde te voldoen sodat

* This article is a slightly revised version of a paper 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.

dit aanpasbaar kan wees vir verskillende moontlike gebruike deur verskillende navorsers. Die artikel wil die bouproses van die Ndebeletaalkorpus skets met spesiale klem op die uitdagings wat die samestellers ondervind het, en moontlike oplossings. Dit word aanvaar dat sommige van hierdie uitdagings nie eie aan Ndebele alleen mag wees nie, maar ook verwante Afrikatale in 'n min of meer soortgelyke situasie mag raak. Die hoofokus van die bespreking sal op die samestelling van die Ndebeletaalkorpus wees, d.w.s. die soort tekste wat die korpus uitmaak. Die korpus is saamgestel uit gepubliseerde tekste, ongepubliseerde tekste en mondelinge materiaal versamel in Ndebelesprekende distrikte van Zimbabwe. Daar sal geredeneer word dat die gebruik van die korpus en sy betroubaarheid vir navorsing op onder andere sy inhoud berus. Daar sal ook getoon word dat die inhoud van die korpus op 'n aantal faktore berus, sommige waarvan sosiolinguistiese, politieke en ekonomiese oorwegings insluit. Hierdie oorwegings het implikasies vir beide die inhoud en gehalte van gepubliseerde en mondelinge tekste wat die Ndebeletaalkorpus uitmaak.

Slutelwoorde: KORPUS, MONDELINGE MATERIAAL, KODEVERMENGING, KODE-OMSKAKELING, MOEDERTAAL, NDEBELE

Introduction

The Ndebele language corpus described here is that compiled by the ALLEX Project (now ALRI) at the University of Zimbabwe. It reflects or is intended to reflect as much as possible the Ndebele language as spoken in Zimbabwe. The Ndebele language corpus was built in order to provide much-needed material for the study of the Ndebele language with a special focus on dictionary-making and research. As would later be demonstrated in this article, the composition of texts and their conversion to machine-readable documents reflect the underlying focus of the main objective, which is lexicography. Like most corpora, the Ndebele language corpus may in future be used for other purposes not thought of at the time of its initial compilation. The main focus of this article is the content of the Ndebele language corpus, i.e. the type of material that constitutes the corpus. The corpus is composed of published texts, unpublished texts and oral material gathered from Ndebele-speaking districts of Zimbabwe. It will be argued that the use of the corpus and its reliability for research depend among other things on its contents. The contents of a corpus depend on a number of factors, some of which include sociolinguistic, political and economic considerations. These considerations have implications on both the content and quality of published and oral texts that constitute the Ndebele language corpus.

Background: The Ndebele Language

Language policy factors have a bearing on the content of the corpus of the Ndebele language. This is so because of the status of Ndebele in Zimbabwe. Ndebele, together with Shona, are the recognised national languages of Zimbabwe while English enjoys the almost exclusive monopoly as language of admin-

istration and medium of instruction in schools. Ndebele is therefore confined to informal domains while official business is done mainly in English. For this reason and other related factors, Ndebele has not developed a vocabulary for other spheres of activity. For instance, there are no Ndebele books for subjects such as history, geography, science or mathematics. In short, the Ndebele language lacks published material in and about the language. Although Ndebele is taught up to university level in Zimbabwe, this has not led to as many advantages as would be expected, that is, in terms of research on and publications in the language. One factor that has hindered this otherwise normal development is that instead of teaching Ndebele as spoken in Zimbabwe, Zulu was taught. For this reason one big question that will always evade compilers is whether to include Zulu texts in the Ndebele language corpus. Secondly, the continued use of English as language of instruction and official language of administration has denied the Ndebele language the opportunity to develop vocabulary and terminology in fields such as agriculture, commerce, law, science, etc. With this background it is only natural that creative works would dominate the corpus.

Gathering of Oral Material

Areas where Oral Material Was Collected

Ideally oral material had to be gathered in all the areas where there are mother-tongue speakers of Ndebele. This would have given the desired representative sample of spoken Ndebele from all geographical areas. However, not all areas were as well covered as researchers would have wanted. Firstly, such an endeavour was impracticable financially, considering the cost involved in such an undertaking. Secondly, mother-tongue speakers of Ndebele are not confined to Ndebele-speaking districts and towns only, but some pockets are scattered in other non-Ndebele-speaking districts. Locating all these communities would not only have been time-consuming and costly but was also felt to be unnecessary. The areas of focus were therefore the Ndebele-speaking districts, which are mainly in the provinces of Matabeleland North, Matabeleland South, and the Midlands. These three provinces constitute nearly half the size of the country geographically although accounting for probably one fifth of the country's population. This implies that human settlements are far apart and very scattered, resulting in high cost in travelling through the districts. This also became a factor in reducing potential areas for oral material collection.

The research was also to serve as a sociolinguistic survey of the language map of the country. Until that of Hachipola (1998), no prior comprehensive survey of the language situation was available. The districts which are commonly described as Ndebele-speaking areas are also populated by speakers of the so-called minority languages: Kalanga, Venda, Tonga, Nambya, Sotho, and in the Midlands districts consist of both Ndebele and Shona speakers. There were

debates as to whether it was worth collecting data from areas where other languages are also spoken. There were concerns that the type of Ndebele spoken by these people who also speak other languages, was likely to be heavily influenced by these languages and therefore not appropriate for the envisaged dictionary. This argument posed another problem of how to distinguish between acceptable Ndebele and unacceptable varieties. There were fears that it would be politically wrong to exclude other people deliberately because they were speakers of other languages, as all children in these districts learn Ndebele in any case. So there were arguments that all varieties of Ndebele should be gathered as this would reflect the linguistic reality at ground level. Although it was eventually agreed that oral material should only be collected from mother-tongue speakers of Ndebele, this was impossible in practice. However, to minimise the influence of other language groups, the majority of student research assistants were deployed only in those areas where only Ndebele was the community language. For instance, the Beitbridge district was not covered because of its predominantly Venda population, and only one research assistant was deployed in Binga, which is a Tonga territory. The table below shows the rough estimate of Ndebele speakers in Beitbridge as extracted from Hachipola (1998: 32):

Areas of Language Mixture in the Beitbridge District

Area	Dominant Community	Other Communities
1. Tshipise	Venda	Shangani
2. Tshitulipasi	Venda	Shangani
3. Tshikwalakwala	Venda	Shangani
4. Dendele	Venda	Sotho
5. Maramane	Venda	Sotho
6. Shashe	Sotho	Venda
7. Malibeng	Venda	Sotho
8. Makombe	Venda	Pfumbi
9. Siyoka 2	Venda	Ndebele

As becomes clear from this table, it would have been a costly venture to collect oral material in Beitbridge owing to the paucity of mother-tongue speakers of Ndebele. Ndebele is the language taught in schools and used in the public domain in such areas as Beitbridge but the users are not first-language speakers of Ndebele and their type of Ndebele was considered not of the desired standard for a monolingual dictionary for learners.

The other determining factor in the choice of areas to be covered was the availability of student research assistants in certain areas. According to the regulations of the University of Zimbabwe on remuneration for student research assistants, there is no allowance for transport and accommodation. It therefore meant that student research assistants should come from those areas where research was to be conducted. As a result some areas could not be covered

because there were no available students from these areas. Students researching in their home areas had the advantage that it was easier for them to conduct interviews among communities with which they are familiar.

Competence of Interviewers

The student research assistants were largely drawn from undergraduates who had done the course on translation and lexicography. However, some had just taken Ndebele language courses in their first year at college. None had prior experience of research in this field, but their performance was considered satisfactory by the corpus compilers and most of them fulfilled the targets that had been set. They had undergone a crash course on the basics of fieldwork that included training in the use of audio-recorders and transcribing recorded material.

All the student research assistants were fluent mother-tongue speakers of Ndebele. Of the twenty-six research assistants, eleven were female and the rest male. Their ages ranged between twenty and twenty-four years. In terms of their academic ability as well as their proficiency in Ndebele the group was competent enough to assume the task. These are some of the key issues that have a bearing on the quality and reliability of the results of the research. The fact that the student research assistants were working in their home districts had an added advantage in that they knew most of the people as well as their potentialities to provide certain information. Similarly it was easier to approach potential informants by people who already knew them. However, because these interviewers were almost all of the same age group, there are topics they seemingly handled very well but in some cases their youth was a limitation when considering the cultural orientation of the Ndebele society. Topics related to everyday events, which dominated the interviews, were handled well but specialised topics such as aspects of religion or sexuality were not satisfactorily treated.

The reason for this is that in Ndebele society it is considered improper to discuss certain topics with young people. In the same way sexuality cannot be discussed by opposite sexes, which meant that male interviewers could not ask certain questions to female informants and vice versa. The other limitation was the student research assistants' own lack of knowledge about certain topics so that they could not pose suitable questions to elicit more information from the informants. The researchers and compilers of the Ndebele language corpus had foreseen some of these inadequacies in the student research assistants. Therefore, the students had been given notebooks that were to be used as diaries throughout the whole fieldwork period. In these they had to give detailed descriptions of their daily contacts and work within the community as well as their own evaluation of informants they met. It is here, also, that they had to note down potential informants they could not interview or those they felt had more information but could not disclose it to them because of their age or sex.

Another way in which these limitations were noted was when the researchers went through the transcribed texts as well as through the tapes. In some instances it became clear that more could have been obtained had the interviewer been knowledgeable enough to lead the discussion fruitfully.

All these limitations had been foreseen and ways of overcoming them prepared. It was originally planned that the researchers would make follow-up interviews in those areas where it was felt follow-ups were needed. An inventory of potential informants was compiled but unfortunately so far no follow-up interviews could be conducted to fill the gaps left by the student research assistants. As there already was a lot of oral material to be processed, no immediate follow-ups were considered. There was also no money available to pursue further research. However, the potential informants are known and recorded for possible future interviews.

Method of Collecting Oral Material

Most of the oral material was collected by means of structured and unstructured interviews. Each of the student research assistants was responsible for determining whether to use a structured or an unstructured interview. Some began as structured but flowed into more or less unstructured discussions. Guidelines had been given for typical structured interviews on specific topics on which they were required to gather material, for instance, topics related to Ndebele marriage customs, child care or cattle farming. Students were given the discretion to choose between structured and unstructured interviews depending on what they thought best in prevailing circumstances. Apart from aiming at creating a word-bank for the Ndebele language, it was also envisaged that the material would be useful for oral history and cultural studies as well as for various language studies other than lexicography. Although the primary aim was an oral corpus for dictionary-making, its other possible uses were not forgotten. The student research assistants themselves had no prior knowledge of a corpus or dictionary-making based on a corpus, which could lead them to the assumption that detailed oral material is required to obtain the meanings of words. However, although their assumptions were not always entirely correct, they succeeded in collecting a rich variety of oral material.

While interviews were the most prevalent in the oral material collection, there were a few cases where recordings of dialogues or other discussions were made. For instance, there were recordings of songs, either at social functions such as weddings, or in churches and in schools. Church services were also recorded as well as classroom sessions in both primary and secondary schools. All these were done only after prior permission was sought from the authorities concerned. In some classroom recordings the teachers involved did the recordings by themselves so as to avoid the presence of a stranger, that is the student research assistant, in the class. One student research assistant managed to record a traditional court session while some recorded normal conversations

in workplaces. While this type of recordings are valuable for yielding real-life situations, they have, however, some limitations, the most conspicuous being the problem of identifying the particular speaker in terms of name, age, occupation and gender. All oral interviews are marked with these details for record purposes.

As already mentioned, all student research assistants were given notebooks that they had to use like diaries to record in detail all research experiences. They would also write down the names of trees, grass, birds and plants found in their areas of research. These notebooks are therefore a rich collection of oral material, especially as far as the names of birds, animals, trees and the like are concerned. Family praise names were sometimes similarly obtained. Audio-recording informants could not easily have yielded this kind of valuable information. Although the bulk of the oral material the student research assistants collected was through audio-recordings, the notes they made in their notebooks have proved very useful.

Written Texts

Renouf (Sinclair 1987: 2) makes the following observation:

When constructing a text corpus, one seeks to make a selection of data which is in some sense representative, providing an authoritative body of linguistic evidence which can support generalisations and against which hypotheses can be tested.

As it describes an ideal situation, this observation holds true for any language corpus. However, for languages with a relatively short and recent literary history such as Ndebele, it is not always practical to have a representative selection. As Renouf (Sinclair 1987: 2) states further, a selection is possible where there is a range or variety from which a representative sample can be drawn:

The first step towards achieving this aim is to define the whole of which the corpus is to be a sample.

For Ndebele, with a very small number of published books whether it be fiction or non-fiction, the whole implies all publications in the language. The long-term objective is to include all published texts in the Ndebele language corpus. The little that has been published represents a neat selection of material used for educational purposes. Apart from religious texts, most publications in Ndebele, both fiction and non-fiction, are in fact targeted at schools. A number of factors account for this. One reason is that the cost of producing and publishing books in Zimbabwe is relatively high and in order to offset these costs there is also a need for a ready market for the books. In a country where there is not yet a reading culture, only schools offer that ready market, and publishers would publish only those works that could be used in schools. However, the Ndebele

language corpus in its current state does not reflect the long-term ideal nor is it likely to do so in the near future. It is a sample of what has so far been published in Ndebele and this sample cannot be described as representative until qualified.

It should be explained why certain texts were excluded from the Ndebele language corpus. The early written works in the Ndebele language may be categorised as falling between 1852 and 1950. The first date marks the first publications in Ndebele by the London Missionary Society, while the latter date marks significant departures from the early Ndebele orthography. Publications spanning this period, few as they may be, are very important in the history of Ndebele but had to be excluded. These are in the old Ndebele orthography, which few people can read today and unless these are rewritten in the current orthography (which is very unlikely) they cannot be included in the corpus. Some of the symbols used would even pose problems for the scanner to detect. Therefore, all texts in the old orthography, which include scripture texts and Ndebele language newspapers and leaflets, have been deliberately excluded owing to the orthography used in them.

The earliest publication of fictional work in Ndebele dates from 1957. It should be noted that the Ndebele language corpus is largely composed of novels. However, a number of novels originally planned to be included in the text corpus were later excluded, some temporarily. Most of the books published in the sixties and seventies were on cheap quality paper. It is difficult and time-consuming to scan works on this kind of paper, most of these works also having been printed in a small font size. If such texts had to be scanned, it means that the time of proofreading them is almost the same as that of typing them. However, the compilers of the Ndebele language corpus had a time frame and target to meet and apart from corpus building, they were also compiling a dictionary, which had to be completed within a given time and target date. Under these time constraints the compilers preferred to scan and proofread those texts that only took the minimum time.

The majority of books included in the corpus were therefore published within the last twenty years, that is, between 1979 and 1999. As already mentioned, the bulk of these are creative works, especially narratives. No poetry collections or anthologies have been included and there are no immediate plans to do so. Poetic language is not popular in general corpus work and for lexicographic purposes it would be less useful. One drama text has been included and so far also one textbook. More textbooks will be included as the corpus keeps growing. Scanning, proofreading and tagging textbooks are more demanding than doing the same with novels, for instance. For this reason it seems the compilers have postponed the inclusion of textbooks, which they will have to do eventually, if they keep to their original plan.

As the Ndebele language corpus consists mostly of novels, the selection criteria for this category must be discussed. Firstly, there were efforts to bring about a balance between male and female writers. There are more published

male than female Ndebele writers, so an effort to include a representative sample of novels by women was made. Another selection criterion was the popularity of the works. Writers who are considered popular had their works included. Writers usually become popular when their works are either prescribed in schools or broadcast in the media. Two such leading Ndebele writers are a male, Ndabezinhle Sigogo, and a female, Barbara Makhalisa. All their works other than drama and poetry have been included. Some works were included on the basis of the richness of their language. Corpus compilers who are competent literary critics made these judgements. Novels were further chosen according to the themes they handle, for instance, attempts were made to have a representative sample of war novels, love and marriage themes, witchcraft themes, and historical novels. Some themes dominate, partly because of the colonial policy prescribing certain themes for writers.

Composition of the Corpus

The corpus consists of both oral and written texts, all transcribed and converted into machine-readable texts. The oral material can be subdivided into oral interviews, oral recordings (of classroom lessons, church sermons, court sessions, etc.) and radio and television recordings. The written texts include publications and manuscripts. Within the category of manuscripts are unpublished dissertations and some selected documents and manuscripts. The published texts are divided into novels, drama and textbooks. There are other materials that have been included such as newspaper articles and advertisements.

Texts in the Ndebele Corpus

Type of Texts	Quantity in %
Written Material	80
Publications	
Novels	
Drama	
Textbooks	
Manuscripts	
Unpublished dissertations	
Unpublished documents	
Oral Material	18
Oral interviews	
Oral recordings	
Radio/Television recordings	
Other Materials	2

(The percentages given are estimates; as the corpus keeps growing its composition is not static.)

The composition of the Ndebele language corpus reflects the history of publishing in Zimbabwe, especially that of the indigenous languages. The case of Ndebele is further complicated by the reliance on Zulu literature for the teaching of Ndebele. As the majority of publications are biased towards school textbooks and novels, these texts dominate the Ndebele corpus. Some efforts were made to include types of texts other than school textbooks and novels. One way of offsetting this imbalance was to include what has been categorised as manuscripts. These are mostly unpublished dissertations and other documents and reports. The dissertations were collected from Ndebele departments in the various teacher-training colleges. As they are research papers, they contain some form of formal academic language. For instance, some dissertations are on Ndebele grammar while others are on teaching methods. Dissertations on literary criticism of Ndebele were also sampled. These papers have a potential of yielding language that is not ordinarily found in novels. However, the major limitation of these manuscripts is that as unpublished works they remain private and personal, and the language they contain is not standardised.

Similarly, as far as the oral material is concerned, it was felt that more was needed than the data collected through interviews throughout the Ndebele-speaking districts. The oral interviews were complemented by recordings of programmes from radio and television stations. The advantage of these is that compilers would have listened to or seen the programmes and so could choose whether to include them or not. Such material could also be systematically chosen to find the desired types of material. The disadvantage, however, was that compilers could not obtain any previous recordings, as the stations of the Zimbabwean Broadcasting Corporation destroy all tapes about two weeks after having broadcast them. Therefore, material was to be limited to that broadcast during the collection of the corpus material. One other disadvantage of radio and television material is that it lacks adequate biographical details of informants in terms of age, sex, occupation and educational background. Such information is essential for various research purposes and all oral interviews therefore have such details marked.

What has been labelled as other materials include various types of language, those that can be found in advertisements, posters or letters. One other notable kind of material in this category is the unfinished Ndebele dictionary that was supposed to be published by Longmans Zimbabwe.

Conversion of Texts

The coming of computers into language study has helped to address the question of corpus accessibility to other researchers. However, before the corpus can be shared by many researchers, it must be made machine-readable. The ALEX Project corpora (both Ndebele and Shona) use the Standard Generalised Mark-up Language (SGML). This is in line with the international choice and preference of this method. "Because of its power, flexibility and independ-

ence of particular software systems," says Kennedy (1998: 82), "the Standard Generalised Mark-up Language (SGML) has become increasingly accepted outside the publishing industry as the standard way of encoding texts."

In addition to the SGML, the text encoding initiative guidelines were also followed. Although these are not standardised, they are flexible and adaptable to the compilers' needs. According to Kennedy (1998: 83) "the TEI Guidelines were designed to apply to any texts regardless of the language, the date of production or the genre". The use of these internationally used mark-up techniques makes the Ndebele corpus accessible to most international users and it can be rated as user-friendly and up-to-date.

As mentioned previously, the compilers tagged the texts to suit their immediate lexicographic needs. As "the corpus compiler has flexibility as to how much detail is marked-up for any particular corpus" (Kennedy 1998: 84), there is room for additional tags depending on the needs of the researcher. For instance, most of the Ndebele oral corpus has tags giving the biographical details of informants such as age, sex, education and occupation. There are also details on the header about the district where the material was gathered.

Implications of the Content of the Corpus

Collecting oral material was not easy for the compilers of the Ndebele language corpus. Apart from the large financial resources that were expended on the activity, there was also the problem of who should be interviewed. The majority of those who are counted as Ndebele speakers today are in fact mother-tongue speakers of the so-called minority languages such as Venda, Kalanga, Nambya, Sotho and Tonga. An ideal Ndebele language corpus should be the language of mother-tongue speakers of Ndebele; however, in reality, the Ndebele language is spoken by people whose first language is not Ndebele. The question of choosing from which districts to collect oral material was therefore problematic as it brought a number of sociolinguistic and political factors into play. Even if one were strictly to isolate mother-tongue speakers of Ndebele (if that is ever possible) there is still the problem of English that seems to be a characteristic of most Ndebele speakers, especially the middle-aged and most urban dwellers. There is a lot of code-mixing and code-switching and outright use of English words. This becomes problematic in transcribing the tapes as it gives orthographic problems. Decisions had to be made in certain cases whether to write some words in their English spelling or give them a Ndebele version. Notwithstanding the above-mentioned problems, the ALLEX Project of the University of Zimbabwe began to compile the Ndebele language corpus basing it on the language as currently spoken by mother-tongue speakers.

Conclusion

It has been shown that the content of a corpus depends on a number of factors

that include sociolinguistic, political and economic considerations. It has also been shown how these factors have a bearing on both the content and quality of published and oral texts that constitute the Ndebele language corpus. As this corpus grows year by year, the present limitations would be addressed gradually. For instance, as the present state of the corpus is predominantly creative work, a deliberate effort would be made to include more textbooks so as to create a balance. The changes in the status of Ndebele as a language in Zimbabwe might significantly influence writing and publishing in the Ndebele language and thus influence the content of the corpus.

References

- Hachipola, J.S.** 1998. *A Survey of the Minority Languages of Zimbabwe*. Harare: UZP.
- Kennedy, G.** 1998. *An Introduction to Corpus Linguistics*. London: Longman.
- Sinclair, J.M. (Ed.)**. 1987. *Looking Up: An Account of the COBUILD Project in Lexical Computing and the Development of the Collins COBUILD English Language Dictionary*. London: Collins.

Business Metaphors in a Bilingual Business Lexicon*

Li Lan (*egtilan@polyu.edu.hk*) and Grahame T. Bilbow
(*eggrabil@polyu.edu.hk*), *Department of English, Hong Kong Polytechnic University, Hong Kong*

Abstract: General purpose dictionaries benefit users at large in many ways, but the definitions and examples might not satisfy the diverse needs of different professional users. This is especially true of metaphors. The article discusses the treatment of business metaphors in the PolyU Business Lexicon derived from the trilingual PolyU Business Corpus (PUBC). During the process the concordances are grouped by senses, and then separated according to their literal and metaphorical meanings, which in turn lead to the decisions of sense order, word meaning and translation equivalents. Since different cultures have different 'bags' of metaphors, and metaphorical meanings also vary in different registers, the focus is primarily on the differences between Chinese and English in terms of culture, psychology, language and how such differences can be translated and presented in a corpus-based business lexicon with a minimum loss of their original connotations. Cultural transformations, such as direct translation, image substitution, explanatory notes and abandonment of the figure of speech, are suggested to bridge interlanguage metaphorical gaps.

Keywords: METAPHOR, DEFINITION, TRANSLATION, CULTURAL DIFFERENCE

Opsomming: Sakemetafore in 'n tweetalige sakewoordeboek. Woordeboeke vir algemene doeleindes bevoordeel gewone gebruikers op baie maniere, maar die definisies en voorbeelde mag dalk nie die uiteenlopende behoeftes van verskillende professionele gebruikers bevredig nie. Dit is veral waar van metafore. Die artikel bespreek die behandeling van sakemetafore in die PolyU Business Lexicon gebaseer op die drietalige PolyU Business Corpus (PUBC). Gedurende die proses word die konkordansies deur betekenis gegroepeer, en dan geskei volgens hul letterlike en metaforiese betekenis, wat vervolgens lei tot die besluite oor betekenisorde, woordbetekenis, en vertaalsekwivalente. Aangesien verskillende kulture verskillende "sakke" metafore het, en metaforiese betekenis ook varieer in verskillende registers, is die fokus primêr op die verskille tussen Sjinees en Engels in terme van kultuur, psigologie, taal en hoe sulke verskille vertaal en aangebied kan word in 'n korpusgebaseerde sakewoordeboek met 'n minimum verlies aan hul oorspronklike konnotasies. Kulturele transformasies, soos direkte vertaling, beeldvervanging, verklarende aantekeninge en prysgewing van die stylfiguur, word voorgestel vir die oorbrugging van die tussentalige metaforiese gapings.

Sleutelwoorde: METAFOR, DEFINISIE, VERTALING, KULTURELE VERSKIL

* This article is an adapted version of a paper presented at the International Conference on Translation and Bilingual Dictionaries at the Chinese University of Hong Kong, Hong Kong, 23–24 February 2002.

Metaphor is an important linguistic expression of analogy, and the study of metaphor helps us understand the social systems and structures that shape languages. Lakoff pointed out that metaphor is not just a part of novel poetics but very much a part of everyday, ordinary language. He also argued that metaphor is not merely a matter of language but a matter of thought; consequently, it is an inseparable part of the manner in which the world is ordinarily conceptualized (Lakoff 1993²: 46). In the course of creating the PolyU Business Lexicon, which is based on the PolyU Business Corpus (PUBC) (see Li and Bilbow 2001), it was noticed that the metaphorical use of words has a high frequency of occurrence but their meanings are hard to be found in reference works. Take, for example, the following two sentences:

1. The Dow Industrials Average, which jumped 132 points the past three trading days, was higher in early US trade, enabling the FTSE to *flirt with* the 4,300-point level for the first time in two weeks.
2. Big companies such as stockbroker Merrill Lynch and retailer Wal-Mart Stores are beginning to *flirt with* the concept.

The definition of *flirt* in the *Cambridge International Dictionary of English* (CIDE) is '1. to behave as if sexually attracted to someone, although not seriously ; 2. (flirt with) to consider, although not seriously'. These meanings from a general dictionary do not fit into the above sentences, and a specialized dictionary, *Longman Dictionary of Business English* does not include the word *flirt*. It was assumed that general dictionaries cannot embody all the specific meanings of a word, and specialized ones might rule them out as general English words. The assumption was borne out by looking up a number of metaphorically used verbs in PUBC. The following is an example from the *English-Chinese Dictionary*, one of the most authoritative bilingual dictionaries published in China:

flirt / flo:t / I ● vt. 1. 轻快地摆动, 挥动 (尾巴, 扇子等): ~ a fan 扇扇子 / The bird ~ed its tail. 鸟儿轻快地摆尾巴 2. (罕) 急扔, 急弹 ● vi. 1. 调情, 卖俏 (with): The boys stood on the corner ~ing with the passing girls. 那些青年站在角落里同路过的姑娘们调情。2. 不认真地考虑 (或对待) (with): I've been ~ing with the idea of leaving my job. 我脑子里一直浮现着辞职的念头。/ The bullfighters are ~ing with death. 斗牛士们在玩命儿。3. 摇摇晃晃地 (或不稳地) 动, 跳, 急冲, 飘动 II n. 1. 调情者, 卖俏者 2. 急动, 急扔, 急弹, 猛掷: (尾巴等的) 快速摆动, 飘动 || **flirt-er** n. / **flirt-ing-ly** ad. [<旧词 flert, flurt < ? OFr fleureter to touch lightly, 原义: move from flower to flower < fleur flower]

Semi-technical words like *flirt* are often polysemic and metaphorical. The sense order of such words in a general dictionary is often arranged according to the

frequency of occurrences, therefore its meaning in a specific register might be buried in the huge amount of information, and its metaphorical meaning can be even more difficult to find. To fill the gap between general dictionaries and specialized dictionaries, it was decided to embody both highly specific terms and semi-technical words in the word list of the Business Lexicon and to record the meanings that emerge from the business corpus only. The word *flirt* has little sexual connotation in the PUBC corpus, therefore its definition was settled as 'to experiment with'.

Analysis of business metaphors

Business texts are a rich source of metaphorical language. There has been a renewed upsurge of interest in describing metaphor in general, and in understanding the metaphorical models, paradigms or images that shape and construct business knowledge, practices and structures. In *The Language of Metaphors*, Goatly (1997: 46-53) groups metaphorical patterns in the English lexicon as follows:

- general reifying: including create, destroy, transform, transfer, handle, possess, impact, space, place, proximity, dimension, shape, parts, perception and seeing.
- specific reifying: physical properties can figure as abstract qualities.
- animizing and personifying: abstract entities are represented not simply as concrete, but as animate and human.
- materializing abstract process: internal mental processes are metaphorically represented as perceptual processes.

All these patterns exist in the PolyU Business Corpus. The metaphors are based on certain image-schemata which recur in our everyday experience, e.g. containers, paths, balance, up and down, part and whole, front and back. However, the nature of business leads us to map business metaphors into several conceptual or master metaphors: business as a war, as a game, as a machine, as a human being and as a plant. Each category has a graded structure, with some prototypical category members being more central than others. It is also noted that the process of word-formation involves metaphor, and metaphors operate as nouns, verbs, adjectives, adverbs or as longer idiomatic phrases (Baldick 1990: 134).

Meaning

'Dictionaries are certainly the cemeteries and the mortuaries, definitely the dormitories, and generally the resting place for the populations of metaphors' (Goatly 1997: 31). Although inactive metaphors have become lexicalized, acquired a new conventional semantic meaning and found their way into diction-

aries, active metaphors are highly dependent on inferential pragmatic principles connected with language use and users in contexts. Active metaphors are more difficult to define and translate.

Metaphorical interpretation in business texts demands experience of language either through the process of metaphorical transfer, or through the narrowing or extending of senses. Look at a group of metaphors referring to *water*:

Verb: pour, flood, flow, dry up, sank, plug, swim, dive.

Noun: source, drain, fluid, liquidity.

With the help of the information offered by the business corpus, these metaphors can be lexicalized with exact and unambiguous explanations relating to business contexts. For instance:

pour: supply in large amounts or quantities

dry up: use up

sank / dive: fall steeply

plug: connect

liquidity: the state of having cash or goods that can easily be sold for money

However, the degree of metaphoricality of the above words is lower than that of the words *curb*, *cushion*, *shrug* and *mount*, which require more, what Andersen calls, 'metaphor-competence'. The strategy of understanding such metaphors includes the linguistic manifestation of the metaphor, its identification and interpretation, and its differences from and similarities to other linguistic signs (Andersen 2000: 58). For instance, metaphors indicating trends, or up-down metaphors, are highly relevant to the field of business. Normally *up* is better and *down* is worse. However, *up* is not always better. When the metaphors *mount* and *escalate* are used, they describe costs, debts, inflation or unemployment. *Up* is definitely worse in these cases and the definition and translation should reflect this.

Sense order

In the PUBC, the majority of warfare items are used metaphorically and only a very small portion is used literally. Therefore, in the creation of the lexicon, it was attempted to draw users' attention directly to the metaphorical senses of words in definitions and in translation equivalents. For instance, of the 97 occurrences of the verb *attack*, only a few actually describe military or physical action (shown in bold in the concordance below).

al Financial Regulation" to	attack	the problem. The "Impossib
ficult for the speculators to	attack	the Hong Kong dollar. 3.
get points for speculators to	attack .	The present practice provid

l bank, speculators would	<u>attack</u>	the currency with a view to b
losses in Thailand, anxiety	<u>attacks.</u>	But is it fair to draw paral
esire to do it. He also	<u>attacked</u>	present accounting standards
e reduction is expected to be	<u>attacked</u>	by World Bank critics, who ar
1,000 each. It has been	<u>attacked</u>	by a range of professional an
d result in the insured being	<u>attacked.</u>	As that was what actually ha
to believe that the militia	<u>attacked</u>	and killed hundreds of Britis
boys."Apple has relentlessly	<u>attacked</u>	C.H. Tung for watering down
rtunity, currency speculators	<u>attacked</u>	the HK dollar, leading to a
inggit. As the Thai baht was	<u>attacked</u>	again in early July, the Bank
cially Hong Kong dollar, were	<u>attacked,</u>	and there was a significant

The lexicon entry for the verb *attack* is therefore as follows:

attack vt.

1. to attempt to harm someone or something
2. to criticize strongly and negatively
3. to take aggressive physical action against someone or something

It is interesting to note how polysemous words find their way into the business lexicon. According to the concept of common core vocabulary, a word has a central meaning and a series of radiations. Radiations from the core are multi-dimensional, stretching out in all directions. Consequently, there are differences in terms of the degree of 'closeness to the core' of a group of polysemes. When compiling a general language dictionary, the editor may try to cover as many dimensions as possible. In a specific domain, in contrast, word radiation tends to be mono-dimensional; therefore, when working on a specific dictionary, one direction might meet the criteria for inclusion.

The core meaning of the word *strategy*, for example, is 'the art of planning in advance the movements of armies or forces in war' (*Oxford Advanced Learner's Dictionary*), and its extended meaning is 'a detailed plan for achieving success in situations such as war, politics, business, industry or sport, or the skill of planning for such situations' (CIDE). In the Business Lexicon, only the extended meaning of *strategy* is included.

Metaphorical use in business discourse also tends to limit the inclusion of synonyms and antonyms in the Business Lexicon. For example, there is only limited metaphorical aptness for words such as *assault* and *strike*, which are normally considered synonyms of the word *attack*, to collocate with the words *currency*, *market*, *accounting standard*, etc.

Picture in another language

When metaphors occur in different cultures, the cognitive prototype may be the same, may be partially the same or may be totally different. Mühlhäusler

notes that natural metaphor tends to be culture-dependent and universally understood, in contrast to culturally invented neutral metaphor. As such, metaphor may provide important insights into language universals: 'Developmentally very early clusters of meaning (archetypal metaphor) may reflect universals in the semantics of human languages' (Mühlhäusler 1983: 19).

It is obvious that many discussions on metaphor translation focus on the conviction that 'metaphor is text' (Snell-Hornby 1988). But in compiling a dictionary or a lexicon, 'one-word metaphor' has to be dealt with. Such translation is not the simple task of converting thoughts and words into the words of a different language or discipline, and then changing them back again into similar concepts represented by different, but equivalent, words. Sherblom argues that translation is transformative in nature rather than simply translative, and is a larger, more encompassing, dynamic, communication process: 'As a transformative business communication phenomenon, translation becomes a more complex process engaged in by whole communities and cultures' (Sherblom 1998: 75). As the word list of the PolyU Business Lexicon is generated from the English corpus as base, interest lies mainly in identifying how English metaphor translates into Chinese or Japanese, not in identifying metaphorical equivalents across all three languages.

Three strategies have been adopted when translating business metaphors in the PolyU Business Lexicon:

- direct translation,
- image substitution, and
- abandonment of the figure of speech.

Direct translation

Because all human beings have the same basic perceptual apparatus and share many similar experiences, it is inevitable that there are some strong similarities in metaphorical structure across languages. The first and most satisfying procedure for translating metaphor is to reproduce the same image in the target language, provided it has comparable frequency and currency in the appropriate TL register (Newmark 1988: 108). Direct translation is used when both source language and target language have the same or similar cognition of metaphors.

Human metaphors have many similarities across languages. For example, organizations have *heads* in both English and Chinese. A *handshake* and 握手言和 both mean 'the reaching of an agreement'; *to shoulder a loss or a responsibility* has the Chinese equivalent 肩負. English and Chinese-speaking business people both suffer from *headaches*, 頭痛, and *fevers*, 過熱. The metaphorical verb *to slam* also has a direct equivalent in Chinese meaning 'to criticize severely'.

Many warfare metaphors in business are also translated directly into Chinese.

<i>bidding war</i>	奪標戰
<i>mortgage war</i>	按揭戰
<i>price war</i>	削價戰
<i>mortgage rate war</i>	按揭減息戰
<i>MPF war</i>	強積金戰

Image substitution

Another common procedure for translating metaphor is replacement of the image with another established TL image. *Parent* in English can be regarded as gender-free. The Chinese term for *parent company*, however, is 母公司 (mother company), not 父母公司 (father and mother company) or 雙親公司 (parents company). *Infant companies* are different from a term common in Chinese, 子公司 (literally *son company* or *subsidiary*, having no equivalent in English), as the following example illustrates.

Conversely, venture capital can play a different role in funding risky *infant* companies.

Infant, as used in *infant company*, *infant insurance industry*, etc., has no equivalent in Chinese. The definition of *infant company* then takes on an extended meaning of 'newly-established company or business'.

Chinese has a metaphorical use of the verb *to swallow* which means 'to tolerate something that is unpleasant'. For example, the expression 'swallow a bitter pill/bitter wine' refers to tolerating a situation rather than swallowing a specific object or liquid. Chinese does not, however, permit the expressions *to swallow high costs*, *interest rates* or *risks*. In this case, the translation has to have its extended meaning: 吸收, 忍受.

Partial coverage occurs in some cases. For example, the verb *to ignite* has the same extended meaning in Chinese but it most often occurs in political texts, and only rarely in business texts. In English, however, the verb *to ignite* appears quite commonly, as in the following example:

Fears for further interest rate rises *were ignited* on Thursday when slightly higher-than-expected US gross domestic product data for the first quarter indicated rising inflation.

Abandonment of the image

According to the *New Oxford English Dictionary*, *lobby* refers to any of several large halls in the Houses of Parliament in which MPs may meet members of the public. The verb sense (originally US) derives from the practice of frequenting the house of a legislature to influence its members into supporting a cause. In the PUBC English corpus, for example, the following is found:

Why firms *lobby*, they argue, is due to government intervention which has an effect on their cash flows.

This sense of the word *lobby* is, however, absent in Chinese culture. The Chinese translation in dictionaries is 大廳, 休息廳 (entrance hall), which has neither a political sense nor a verbal function. Its verbal figurative speech has, therefore, to abandon the image of *lobby* and is translated as 游說, 拉票.

Shrug off is a body gesture that is common among Westerners to show indifference or dismissal. It is not, however, a common gesture among Chinese. The expression to *show indifference* in Chinese is 不屑一顧 (not worth one glance), relating to eye movement. From the following business context, we have opted to use the expression 漠視, 擺脫 in the lexicon.

UK equity market was able to	<u>shrug</u>	off some of the uncertainty
19 per cent this week alone,	<u>shrugged</u>	off early selling pressure to
the next millennium must be	<u>shrugged</u>	off with a laugh.
ers said Hong Kong stocks had	<u>shrugged</u>	off a weak Wall Street overnight
per cent firmer. Investors	<u>shrugged</u>	off the news from New York, s
a ebbed, the Hong Kong market	<u>shrugged</u>	off higher closes on Wall Stree
ofitable in 1998. Mr Wang	<u>shrugged</u>	off worries that the ethylene

The following is a concordance of the word *foot* from the PUBC English corpus:

y, don't get caught with your	<u>foot</u>	in your mouth - visit our Ja
said. Foreign firms eye	<u>foot</u>	in door of lucrative pension
re-based consultancy to get a	<u>foot</u>	in doors now closed to fore
uing it would force them to	<u>foot</u>	the insurance bill for their
s and generally would have to	<u>foot,</u>	most of the bill for any even,
in a battle over who should	<u>foot</u>	a multi-million dollar bill f

The metaphorical use of the word *foot* is shared in both languages, but its metaphorical extensions in each language reveal striking differences. English expressions such as *to have one's feet on the ground* and *to put somebody on his feet* find equivalences in Chinese: 腳踏實地 and 使某人自立. But Chinese does not allow *to foot a bill* or *to put one's foot in one's mouth*. The translation equivalents of these expressions show no trace of *foot* whatsoever.

Another example of the abandonment of an image occurs with the word *nerve*.

itory have evolved into a	<u>nerve</u>	centre, acting as a control r
g out this week and there are	<u>nerves</u>	the US [market] may see a bi
bond market was depressed by	<u>nerves</u>	ahead of the release later o
tion-related data that calmed	<u>nerve</u> s	the US Federal Reserve would
summit'' with Beijing calmed	<u>nerve</u> s,	but brokers expect market v
600 that had held for months.	<u>Nerve</u> s	over dollar weakness and t

ncial stocks, reflecting more	<u>nerve</u> s	about interest rates, were m
ocks as investors overcame	<u>nerve</u> s	about the presidential elect
rkets racked by interest rate	<u>nerve</u> s,	Hong Kong blue chips are
e translation. "Do I have the	<u>nerve</u>	to pursue an investigation
you have finally raised the	<u>nerve</u>	to invest in Russian stocks,
5,550. 'To steady traders	<u>nerve</u> s	the FTSE needs at the very l
l as domestic factors had un	<u>nerve</u> d	regional markets. Tokyo sl

In both English and Chinese, use of the word *nerve* goes beyond the word's biological sense, extending to central control of emotional functions. The following table compares metaphorical use of the word *nerve* in each language.

English Metaphor	Meaning	Chinese Metaphorical Equivalent
nerve centre	central control system	神經中樞
nerve-racking	causing stress or anxiety	神經緊張
nervous breakdown	mental illness resulting from severe stress	神經崩潰
to have/raise the nerve to	to develop the courage to	—
to calm nerves	to reduce the tension	放鬆/松弛 神經
to reflect nerves	to show worries	—
to overcome nerves	to become less worried	—
to steady nerves	to soothe	—

Conclusions

Metaphors are particularly prevalent in business texts both in English and in Chinese, and metaphorical meaning often takes pride of place over literal meaning. This tendency has to be taken into consideration when deciding the definition and sense order of entries in a business lexicon.

Every culture has its own 'bags' of metaphors and their meanings also vary in register. Differences lie in individual metaphors used in each language, rather than in conceptual structure. Since the process of communication is a creative experience, metaphors invariably embody cultural values and analogies specific to the contexts from which they spring. Even where similar images are used, for example, there are significant differences in metaphorical collocation and usage between one language and another. It has been shown that different strategies for cultural transformation, such as direct translation, image substitution and abandonment of the figure of speech can be used to bridge interlanguage metaphorical gaps.

References

Adam, J.H. (Ed.). 1989². *Longman Dictionary of Business English*. Essex: Longman.

- Andersen, M.S.** 2000. Metaphor Matters. *Herms: Journal of Linguistics* 24: 57-80.
- Baldick, C.** 1990. *The Concise Oxford Dictionary of Literary Terms*. Oxford: Oxford University Press.
- Crowther, J. (Ed.)** 1995⁴. *Oxford Advanced Learner's Dictionary*. Oxford: Oxford University Press.
- Gusun, L.U. (Ed.)** 1993. *The English-Chinese Dictionary*. Unabridged. Shangai: Shangai Translation Press.
- Goatly, A.** 1997. *The Language of Metaphors*. London: Routledge.
- Lakoff, G.** 1993. The Contemporary Theory of Metaphor. Ortony, A. (Ed.). 1993². *Metaphor and Thought*: 202-251. New York: Cambridge University Press.
- Li, Lan and G.T. Bilbow.** 2001. From a Business Corpus to a Business Lexicon. *Lexikos* 11: 209-221.
- Mühlhäusler, P.** 1983. *Towards an Explanatory Theory of Metaphor*. Trier: Linguistic Agency.
- Newmark, P.** 1998. *More Paragraphs on Translation*. Clevedon/Philadelphia: Multilingual Matters.
- Pearsall, J. (Ed.)** 1998. *The New Oxford Dictionary of English*. Oxford: Clarendon Press.
- Procter, P. (Ed.)** 1995. *Cambridge International Dictionary of English*. Cambridge: Cambridge University Press.
- Sherblom, J.C.** 1998. Translation Business Communication by Building on Forman's Translation Metaphor. *The Journal of Business Communication* 35(1): 74-86.
- Snell-Hornby, M.** 1988. *Translation Studies: An Integrated Approach*. Amsterdam/Philadelphia: John Benjamins.

Synopsis Articles in the Planning of a Trilingual Dictionary: Yilumbu–French–English

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

Abstract: A distinction is often drawn between **single articles** and **synopsis articles**. A single article is the so-called **default article**. It does not deviate from the traditional microstructural approach of the dictionary because it presents the minimum data for each lemma sign treated, while a synopsis article gives additional data for each treated lemma. The classical conception of the synopsis article originates with Bergenholtz, Tarp and Wiegand (1999). In their view, synopsis articles which were devised for language for special purposes (LSP) dictionaries have to make provision for the treatment of lemma signs functioning as superordinate terms (general lemmata or topics) as well as their hyponyms. The aim of this article is to present a new lexicographic approach to the study of synopsis articles against the background of a planned trilingual dictionary. It will be shown that synopsis articles have a more general use than its restricted application in LSP dictionaries. In fact, the focus will not be on the treatment of technical or scientific vocabulary but rather on cultural terms. This article also raises among other things, questions about the purpose of the planned dictionary, its target users, different types of microstructures as well as different data categories to be included in synopsis articles.

Keywords: SYNOPSIS ARTICLE, SINGLE ARTICLE, DIFFERENT TYPES OF MICROSTRUCTURES, LEMMA SIGN, LINGUISTIC AND EXTRA-LINGUISTIC ASPECTS, LSP DICTIONARIES, TARGET GROUP, DATA CATEGORIES, CULTURE-BOUND LEXICAL ITEMS, SPECIAL-FIELD TERMS

Abstract: Articles synopsis dans la planification d'un dictionnaire trilingue: yilumbu–français–anglais. Une distinction est souvent faite entre **articles simples** et **articles synopsis**. L'article simple est encore connu sous la dénomination d'**article par défaut**. Il ne dévie pas de l'approche microstructurale traditionnelle du dictionnaire parce qu'il présente le minimum pour chaque lemme traité, alors qu'un article synopsis offre des données additives pour chaque lemme traité. La conception classique d'article synopsis est de Bergenholtz, Tarp et Wiegand (1999). De leur point de vue pour les articles synopsis qui ont été conçus pour des dictionnaires de langue de spécialité (LSP), des dispositions doivent être prises pour le traitement de signes-lemmes fonctionnant comme termes superordonnés (lemmes généraux ou thèmes) ainsi que leurs hyponymes. L'objectif de cet article est de présenter une approche lexicographique nouvelle pour l'étude des articles synopsis avec en toile de fond un dictionnaire trilingue en proposition. Il sera montré que les articles synopsis ont un usage plus large que leur application limitée aux dictionnaires LSP. En effet, le centre d'intérêt portera non pas sur le traitement des termes du vocabulaire

technique ou scientifique mais plutôt sur des termes culturels. Cet article soulève également entre autres choses des interrogations quant à la visée du dictionnaire proposé, ses utilisateurs cibles, les différents types de microstructures ainsi que les différentes catégories de données à inclure pour chaque article synopsis.

Mots-clés: ARTICLE SYNOPSIS, ARTICLE SIMPLE, DIFFÉRENTS TYPES DE MICROSTRUCTURES, SIGNE-LEMME, ASPECTS LINGUISTIQUES ET EXTRA-LINGUISTIQUES, DICTIONNAIRES LSP, GROUPE CIBLE, CATÉGORIES DE DONNÉES, ITEMS LEXICAUX CENTRÉS SUR LA CULTURE, TERMES TECHNIQUES

Introduction

According to Hausmann and Wiegand (1989: 340), the microstructure of a dictionary article is the total set of linearly ordered information items following the lemma sign. As far as this point is concerned, a distinction is drawn between **single articles** and **synopsis articles**. A single article is the so-called **default article**. It does not deviate from the traditional microstructural approach of the dictionary because it presents the minimum data for each lemma sign treated, while a synopsis article gives additional data for each treated lemma sign. Moreover, some synopsis articles will focus on linguistic aspects but others will go further including extra-linguistic aspects.

The classical conception of synopsis articles originates with Bergenholtz, Tarp and Wiegand (1999). According to their view, synopsis articles which were devised for language for special purposes (LSP) dictionaries, have to make provision for the treatment of lemma signs functioning as superordinate terms (general lemmata or topics) as well as their hyponyms. In this article an innovative lexicographic approach to the study of synopsis articles is proposed. Before discussing this proposition however, the purposes, nature and typology of the planned trilingual dictionary, Yilumbu–French–English, will be offered.

Guthrie (1953) classifies Yilumbu as B. 44, which falls in the same language group as Ghisira (B. 41), Yisangu (B. 42) and Yipunu (B. 43). It is a Bantu speech form spoken in Gabon and in its neighbouring countries, the Republic of the Congo and the Democratic Republic of the Congo.

In Gabon, Yilumbu speakers mostly reside in two of the nine Gabonese provinces, namely Nyanga (especially in Mayumba) and Ogooué-Maritime (especially in Gamba and Setté-Cama). This geographical distribution has led to the existence of two major dialects: Yilumbu *yí ghângu* (the variety of the Nyanga province) and Yilumbu *yí menaáne* (the variety of the Ogooué-Maritime province)¹.

Although the main focus of this article is on the microstructure of the planned dictionary, a variety of lexicographic data regarding the addressing structure, the access structure, the mediostructure, the data distribution structure as well as some aspects of the user perspective will necessarily come into play.

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

It is a well-attested fact that the **genuine purpose** (cf. Wiegand 1999: 299) of a particular dictionary should be identified prior to the compilation phase and provision to fulfil the real needs and reference skills of the intended target users should be made. At this stage the genuine purpose of the planned dictionary can be described as follows: On account of the decoding and encoding functions, the planned dictionary should assist native- and foreign-language speakers on an equal basis in retrieving relevant information as quickly as possible. More specifically, the planned dictionary should fulfil two main functions. These are: (a) A dictionary for the Balumbu helping them to understand Yilumbu texts. (b) A dictionary for the Balumbu helping them to produce Yilumbu texts. Moreover, the planned dictionary will be a **typological hybrid** (cf. Gouws 1999: 39) in the sense that it will have features of both translation and monolingual dictionaries. As far as monolingual features are concerned, the planned dictionary will provide users with a brief paraphrase of the meaning of the lemma sign in the source language (Yilumbu). This paraphrase of meaning can be expanded for the compilation of a comprehensive version of the dictionary. With regard to translation features, the planned dictionary will present translation equivalents for lemmata where available. The low-density orientated microstructure will give the work the features of a pedagogical dictionary. In other words, by presenting the data as explicitly as possible the planned dictionary could be used by high school pupils, students and learners. With regard to the interaction between users' needs and dictionary typology, the planned dictionary will be a typological hybrid in the sense that it will also include special-field terms. Finally, the inclusion of data regarding the proto-Bantu reconstruction of the genuine Yilumbu lemma signs as well as the source language items for Yilumbu loan words will give the planned dictionary all the features of both historical and etymological dictionaries.

2. The synopsis article: A new conception

As stated above, synopsis articles were devised for dictionary articles displaying a variety of lexicographic data that could not fit into the so-called single article. Given the fact that a certain category of lexical items require more data than the traditional treatment to be allocated to the average lemma sign, Bergenholtz, Tarp and Wiegand (1999) argue in favour of a category of articles, which will focus on extra-linguistic aspects, the traditional encyclopedic approach. According to Gouws (1999: 41), typical items to be treated in synopsis articles are scientific words used in general conversation between expert and lay person, e.g. medical and juridical terminology.

Language for special purposes, although very important, may not be regarded as the only field of application for synopsis articles. In other words, it is believed that synopsis articles have a more general use than its restricted

application in LSP dictionaries. In fact, the focus will not be on the treatment of technical or scientific vocabulary but rather on cultural terms. More specifically, priority will be given to the treatment of lexical items with a high degree of cultural information. One of the representatives of these lexical items is the so-called "culture-bound lexical items". According to Tomaszczyk (1984: 289), "culture-bound lexical units include items which represent objects, ideas and other phenomena that are truly unique to a given speech community. In the pursuit of this definition it should be emphasised that culture-specificity is merely a matter of degree". This is particularly relevant because the culture-bound lexical items in question have to be regarded as cross-cultural lexemes in the African languages² in general and the Gabonese languages in particular, although they, to a greater or lesser degree, throw light on some aspects of the typical socio-cultural background of the Yilumbu lexicon. With regard to macrostructural selection the following criteria will apply: without necessarily working with the idea of superordinate versus hyponym as in Bergenholtz, Tarp and Wiegand (1999), lexical items of the same genus and other scientific terms as well as cultural terms will be included. The dictionary conceptualisation plan (cf. Prinsloo and Gouws 2000: 153) should therefore provide a data distribution structure which warrants a differentiated approach with regard to the structure of the dictionary articles. The point here is threefold. Instead of using inserted inner texts, some articles (especially those presenting a lemma sign representing a high degree of cultural information) will have, according to the specific case: (a) a cultural or encyclopedic section presenting texts in the three languages of the dictionary (Yilumbu–French–English), (b) a cultural or encyclopedic section presenting texts in only the target languages of the dictionary (French–English), and (c) a cultural or encyclopedic section presenting excerpts from relevant literature dealing with the lexical items treated. The foregoing also applies to dictionary articles dealing with a general discussion of special-field terms: the classical conception of synopsis articles.

This wider application is important, because dictionary articles do not have to be treated alike. The concept of synopsis articles is used here as the venue or landing spot³ for the treatment of ethnographical terms. This will also give the planned dictionary features 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". In addition, these synopsis articles may be qualified because articles with the encyclopedic section presented in the three languages of the dictionary may differ from one with a bilingual presentation or even with one merely containing a comment on pragmatics following the linguistic presentation. This difference in presentation may be motivated on account of culture specificity as well as the subject knowledge of the users. This discrepancy in the treatment should also be explained in the user's guide. User proficiency will furthermore determine the nature and extent of the lexicographic presentation. For example, in the case of surrogate terms⁴, speakers of English and French may need more

cultural information than mother-tongue speakers of Yilumbu because for the latter a certain degree of familiarity with the treated surrogate terms will be assumed. These metalexigraphic foundations have to be entered in the front matter of the specific dictionary. Moreover, according to Gouws (1999: 42), "the back matter could also contain a text with a list of all the items which are treatment units in the central list's synopsis articles". As will be shown later in this article, the microstructural treatment in synopsis articles is open to a variety of classes. Therefore, if not planned carefully, the microstructural arrangement of the dictionary articles dealing with these lexical items may create communicative problems for the target users of the specific dictionary. It may be worthwhile to devote some space to this important aspect of the dictionary conceptualisation phase.

3. The target group

Any dictionary project should clearly identify its target users prior to the compilation phase. The target users of the dictionary under discussion range, on the one hand, from senior high school pupils to academics who have Yilumbu as first language and a relative good command of French or English, and, on the other hand, pupils and scholars who want to improve or learn Yilumbu as second language. That at this stage no dictionary for special-field terms exists in Yilumbu (in fact, there is no Yilumbu dictionary at all) is a motivation for the inclusion of technical vocabulary. With regard to the question of how many languages for special purposes should be included, the intention is that the planned dictionary should include special-field lemmata from among others astrology, agriculture and traditional pharmacopoeia. These three special fields are not arbitrarily motivated but is based on empirical observation. In fact, people use medical, astrological and agricultural terminologies on a daily basis. With regard to the extent to which technical vocabulary should be treated, it should be made clear that the data presented are not intended for specialists but for laypersons. Obviously, there is a limit to what can be included in a dictionary. Therefore the questions to be addressed are: What can be included? What should be left out, and why? For example, as far as the treatment of plant names is concerned, it has been decided to leave out data such as the scientific description (height and diameter, arrangement and texture of the foliage and leaves, etc.), the distribution and the cultivation requirements of the plant dealt with. This data will be interesting to the knowledgeable reader but of little use to the uninitiated or non-specialist. As far as this data is concerned, the specialist can always be referred to LSP dictionaries and botanical journals. Contrary to this approach, a simplified presentation of the description, habitat and uses of the plant dealt with may be useful to users. Given their relevance as data types, the scientific name as well as the genus of the plants has to be included. However, the metalanguage should be kept as simple as possible. This aspect receives attention once again in a subsequent section of this article.

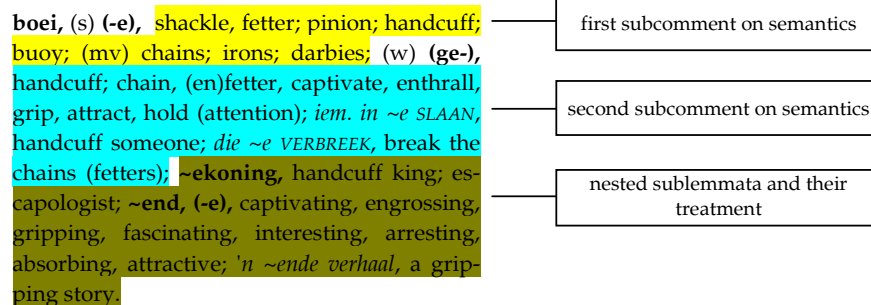
In addition, a pictorial illustration may be more accurate than a complete scientific description of the plant name lemmatised. This chiefly depends on the space the publisher is willing to allocate to a trilingual dictionary of this kind that obviously has space constraints. The ideal for this kind of work will be to have a final product combining dictionary and encyclopaedia.

4. Different types of microstructures

Hausmann and Wiegand (1989: 354) distinguish three types of microstructures, namely: integrated microstructures, partially or semi-integrated microstructures and unintegrated microstructures. According to their approach (Hausmann and Wiegand 1989: 346), a dictionary article displays an unintegrated microstructure when the items indicating semantic data and the items giving idioms and the competence examples are presented in two different subsections, the first subcomment on semantics and the second and further subcomments on semantics respectively. Along the line of Hausmann and Wiegand (1989) but rather focusing on the concept of addressing, Gouws (1999: 47) proposed a reformulation of the unintegrated microstructure:

An unintegrated microstructure displays a distant addressing between a co-text entry and the relevant paraphrase of meaning/translation equivalent. In the case of a lemma sign representing a polysemous lexical item, a bilingual dictionary will give all the translation equivalents and then present the co-text entries. Contrary to the above-mentioned microstructure, an integrated microstructure displays a system of direct addressing between a paraphrase of meaning/translation equivalent and its co-text entry/entries. Each paraphrase of meaning/translation equivalent is immediately followed by the co-text entry illustrating the typical usage of the lexical item in question.

This is clear from Text examples 1 and 2.



Text example 1 (from *Groot Woordeboek*, 1986¹³: 97).⁵

Compilers of the *Groot Woordeboek* have interpreted the lexical item **boei** as a lemma sign with more than one part of speech function. In fact, it can function as a noun and as a verb. In the treatment of the article of the lemma **boei**, the user is provided with a list of translation equivalents (first subcomment on semantics) followed by the item giving the form of the lemma sign functioning as a verb (w) (**ge-**). The latter is then followed by a list of translation equivalents as well as by collocations (*iem. in ~e SLAAN* and *die ~e VERBREEK*), illustrating the use of **boei** functioning as a noun. One may speak of these collocations as a distant addressing because they do not belong to this second subcomment on semantics. They should have been presented under the first subcomment on semantics, before the item giving the form of the lemma sign functioning as a verb. The distant addressing relation existing between co-text entries and their relevant translation equivalents is taken a bit further in the treatment of the sublemmata **boei-ekoning** and **boei-end**. After a mere listing of translation equivalents without any indication of the co-text in which they are generally used, the user only gets the entry '*n ~ende verhaal*.

ukuthi qàbà fkz/ideo: 1 ukugalela into esisingxaza kodwa ungayithi whuu koko uyithi thontsi; ukuqabaza, ukuthontsiza:	1 pour a small quantity of liquid; trickle, dribble, drip (tr):	1 ('n vloeistof) laat drup; 'n paar druppels daarvan uitgooi:
<i>yithi qaba intwana yeyeza kule komityi: ...</i>	<i>trickle a small quantity of medicine into this cup; ...</i>	<i>gooi 'n paar druppels medisyne in die koppie; ...</i>

Text example 2 (from the *Greater Dictionary of Xhosa*, 1989: 1)

The dictionary article above is user-friendly because the compilers have made the typographical presentation of the lexicographic text in such a way that it is clearly perceived by the user. Each paraphrase of meaning is immediately followed by the appropriate illustrative example.

Gouws (1999: 47) describes a semi-integrated microstructure as follows:

A semi-integrated microstructure is a hybrid form displaying features of both an integrated and an unintegrated microstructure and is typically used in more comprehensive dictionaries where lengthier articles with a variety of data types and search zones occur. The article structure has two distinct sections to deal with the two microstructural types, cf. Wiegand (1996). In the first article component, a single co-text entry is added to each entry given as a paraphrase of meaning/translation equivalent. A separate article component is presented to accommodate additional co-text entries. In this unintegrated component the relation between each co-text entry and the relevant paraphrase of meaning/translation equivalent entry is explicated by means of a clear and unambiguous cross-reference entry marking the co-text entry as addressed at a specific paraphrase of meaning/translation equivalent entry.

This point is illustrated by Text example 3 which is a proposed article structure for the treatment of the lemma **fumu**.

fumu (+fumu) [fùmù/(tsí)fùmù] <i>n. cl. 9/10 < *kúmy</i>		
1 ♦ Mútu wó avátúmi. ♦	1 (F) Chef, propriétaire (non au sens légal du terme mais plutôt au sens d'avoir un privilège sur ou des responsabilités vis-à-vis de quelqu'un ou quelque chose).	1 (E) Chief, owner ⁶ (not in the legal sense of the word but rather in the sense of having privilege over or responsibility for something or someone).
Σ Fumu anáwiítsa.	(F) Le chef est arrivé.	(E) The chief has arrived.
2 mwána fumu (aussi/also (<i>apoc.</i>) mwá fumu/Fumu) Mwána úbura sá mwána my-óoghú wó mwána muvhíga.	2 (F) Personne de sang noble par opposition à l'esclave.	2 (E) Person of noble blood as opposed to a slave.
Σ Yá ábanganga fumu ífumba.	(F) Il était le chef de la famille (par le sang).	(E) He was the head of the family (by blood).
3 Mútu wó aváwúkígha bátu.	3 (<i>par ext.</i>) (F) Rassembleur, unificateur.	3 (<i>by ext.</i>) (E) Assembler, unifier.
Σ Mí fumu iwála myóoghú. Mo ífumba isána búrandu.	(F) J'ouvre les bras en signe de rassemblement. Au sein de la famille je n'ai pas de parti pris.	(E) I open my arms as a sign of assembling. In the bosom of the family I am not biased.
4 Fumu (<i>o ingánga</i>) Mwána Nzâmbi.	4 (<i>chris.</i>) (F) Le Seigneur.	4 (<i>Chris.</i>) (E) The Lord.
Σ Á! Fumu!	(F) Ah! Seigneur!	(E) Oh! Lord!
<p>■ (<i>cf. 1</i>) fumu bísalu (F) patron (E) manager, boss; (<i>Ghâng.</i>) fumu bwála, (<i>Men.</i>) fumu dímбу (F) chef du village (E) chief of the village (⇒ shééfu bwála, shééfu dímбу, díkaata di bwála, ívhunda i dímбу); fumu bulôngu (F) président, chef de l'État (E) president, head of state (⇒ táata kwabi, ívhunda bulôngu); fumu dyâmbu (F) propriétaire de la palabre (E) owner of the problem resulting from the discussion; fumu íbôngu (F) maître du village des fantômes (E) master of the village of ghosts (⇒ íbôngu); fumu ífumba (F) chef de famille (E) head of the family; fumu kându (F) chef tribal, chef de clan (E) tribal chief, head of a clan (⇒ fumu ífumba, divhâta); fumu ndônga (F) leader, personne qui est à la tête (E) leader, person who is at the head; fumú nyimbi (F) sorcier en chef (E) chief wizard (⇒ fumu íbôngu); fumú tera (F) propriétaire du terrain (E) owner of the land; (<i>cf. 2</i>) fumu (ngúlu) Nyânga (bikútu) (F) le noble par excellence (E) noble <i>par excellence</i>; (<i>cf. 4</i>) (Nzâmbi) Fumu (F) Dieu, Jésus-Christ (⇒ Yesu) (E) God, Jesus Christ. ■</p>		
Σ (<i>Ghâng.</i>) Yabé fumu bwála anáwiítsa.	(F) (<i>cf. 1</i>) Lui-même le propriétaire est arrivé.	(E) (<i>cf. 1</i>) The owner himself has arrived.
Σ Uvárómbe úweela vângi ngébé fumu dyâmbu usána dupátaaghu dumwéégha. Disásíla.	(F) Tu veux te marier mais toi-même (la personne concernée) tu n'as pas un seul franc (en poche). Ça ne se fait pas!	(E) You want to get married but you yourself (the person concerned) do not have a single penny (in your pocket). It should not be done!
Σ (<i>Men.</i>) Uyi mbíre! Ágho á fumu.	(F) (<i>cf. 2</i>) Il y a l'aigle! Il est le roi (de tous les oiseaux).	(E) (<i>cf. 2</i>) There is the eagle! He is king (of all the birds).
<p>▼ (<i>Ghâng.</i>) Divhâta mu utúba tí: fumu kându. Fumu kându bí? Yá fumu yífumba. Mambú mwáli, mambú matátu tí: Monanganu yábé fumu kându, divhâta bé díng! Yândi yá vátúmi. Yândi yá díkaata. Yândi yá múnkúruntu. Ké Yilúmbu ukénápota nána. Vali díkaate, vali fumu kândwe, vali divhâte, vali dí fumu yibôngwe!</p>		

Text example 3

In the treatment of this article a variety of data types have been presented in four search zones that can be clearly perceived by the user and which display a

semi-integrated microstructure. The first search area presents the lemma sign followed by its stem, the phonetic transcription item, the part of speech indicator, the item giving the class number as well as the proto-Bantu reconstruction.⁷ The second search area starting with the Arabic numerals presents a system of direct addressing between the paraphrases of meaning/translation equivalents and their co-text entries. In the third search zone introduced by the symbol ■, the lexicographer deals with a paradigm of compounds with **fumu** as first component. Apart from the compounds **fumu (ngúlu) Nyânga** and **(Nzambi) Fumu**, groupings of sublemmata display a strictly alphabetical ordering. Although the lexicographer assists users by clearly indicating which sense of **fumu** applies to specific lemma clusters, this deviation on alphabetical level is not motivated on semantic or morphological grounds. Therefore this sinuous lemma file may be regarded as an illustration of a **second level niching** (cf. Gouws 2002: 136). In the fourth section of the article introduced by the symbol Σ||, one finds a paradigm of co-text examples displaying an unintegrated microstructure due to the distant addressing prevailing between these lexicographic data and their relevant paraphrases of meaning/translation equivalents that are to be found in section two of the article. In order to anticipate any problem in the retrieval of the information sought by the user, the latter is once again helped by an entry clearly indicating which sense of **fumu** applies to a specific co-text example. In the fifth search area the treatment proposed is of an ethnographical nature.

5. Different data categories to be included

Wiegand (1978 as quoted by Smit 1993: 143-144) takes the view that all parts of the dictionary article, such as the meaning explanations, the items indicating word class, etymology and spelling, can be regarded as answers to specific user questions. These questions that may be anticipated by the lexicographer also determine the dictionary type.

In a dictionary article (cf. Hausmann and Wiegand 1989: 353), the user is usually provided with both macro- and microstructural data. The first data types are referred to as the **comment on form** (e.g., the lemma sign and the item giving the part of speech) whereas the second data categories are known as the **comment on semantics** (the paraphrase of meaning/translation equivalent and illustrative material in the form of citations, co-text examples or contextual data). All these data types will be included in synopsis articles.

Below follows more detailed information regarding the inclusion of a few data categories, e.g. paraphrase of meaning/translation equivalent, citation and etymology. In the case of the planned dictionary which is a typological hybrid, it can be assumed that the user will need some meaning descriptions in Yilumbu. The short explanation of the meaning is deliberately written in ordinary language. However, this does not imply that the data presented in the meaning descriptions have to be superficial. As mentioned above, the para-

phrase of meaning is meant in the first place to improve the abilities of Yilumbu speakers in their own language. The traditional way in explanatory dictionaries is to present meaning explanations without translation equivalents. Such an approach is normal if one considers that monolingual dictionaries are primarily aimed at the description of meaning. However, in bilingual or multilingual dictionaries, although paraphrases of meaning may be included, the priority is not the description of meaning but the co-ordination of source and target language forms. The translation of the paraphrase of meaning and other texts given in the target languages has the genuine purpose (as defined by Wiegand 1999) of giving the adequate translation equivalent to each source-language form. In addition, it may serve as a useful aid in the acquisition of the languages in question by the learners of both French and English and will help to shift the focus from the source language to the target languages (cf. Gouws 1996: 158).

In all existing bilingual dictionaries of the Gabonese languages, coverage is strong on traditional crafts (e.g. fishing, hunting, weaving, pottery, wood-carving, etc.) (cf. Nyangone Assam and Mavoungou 2000). However, citations are hardly ever used and references to relevant literature are seldom found. Citations and references to relevant literature should be included where necessary. These references will provide researchers with additional information when needed. Provision is not made for the inclusion of etymological items explaining from a historical point of view why and how a lexical item has become what it is. That would fall within the scope of a historical dictionary. The inclusion of diachronic data in a synchronic dictionary rather results from the lexicographer's willingness to take both contemporary language and items from former language stages into account.

When the different data categories to be included are discussed, it is also important to comment on their distribution within the dictionary. This is the concern of the data distribution structure which basically determines where the different data categories have to be included. Some data types will make it into the central list, whereas others will be included in the different outer texts of the dictionary (i.e. idioms and proverbs will be included in the back matter section). In addition to the foregoing, provision should also be made with regard to the different search zones within the dictionary articles. It is the task of the data distribution structure to determine the article internal presentation and the different search zones within the dictionary articles. For example, if three data categories have to be included, then the lexicographer has to make provision for three article slots or search zones in the dictionary article.

— Proposed article structures for the treatment of special-field terms

The dictionary article as shown in Text example 4 has a limited data type. It is clear that the focus is not on linguistic data, e.g. etymology, translation equivalent, etc. but on the field (pharmacopoeia and, to a lesser degree, ethnol-

ogy). After the items giving the proto-Bantu form of the lemma sign as well as the paraphrase of the meaning, the user is provided with target language entries that are introduced by **structural markers**, e.g. ⟨E⟩ (for English) and ⟨F⟩ (for French). Co-text entries in the source language are introduced by the symbol ∑||. The English usage example is followed by a text introduced by another structural indicator, namely the black upside down triangle (▼) used to mark a specific article or search zone in which the user is provided with an additional text or lexicographic description of an encyclopedic or extra-linguistic nature.

mufúma, mi (+ fuma) [mùfùmà] <i>n. cl. 3/4 < *-fúma or -fúma</i>		
◆ Mwíri múneni o báyisi. ◆	⟨F⟩ Grand arbre tutélaire des génies; ⟨T⟩ fromager, kapokier; faux cotonnier; arbre à ouate (Ceiba pentandra).	⟨E⟩ Big tree of tutelary spirits; ⟨T⟩ kapok tree (Ceiba pentandra).
∑ Mufúma avábúra mikòndu mívhema.	⟨F⟩ Le fromager produit une soie végétal blanche.	⟨E⟩ The kapok tree produces a white vegetal silk.
∑ Mufúma, mwíri mi bíl- ongu.	⟨F⟩ Le fromager est un arbre médicinaal.	⟨E⟩ The kapok tree is a medicinal plant.
<p>▼ (Men.) Umábúru mwána, mwána muyéeyi, ubúkulu mufúma unéengi mú tseengi. Ubwísyanga mwána, mwána akúrúgha. Kòndé kyeeri ágho múru. Ukeba mamba gho múru, múru ukúrúghu ⟨F⟩ Si tu mets au monde un enfant chétif, prends des écorces du fromager que tu déposes dans une cuvette. Fais prendre au bébé un bain et il grossira. Toutefois le danger (de ce bain) se situe au niveau de la tête. Il faut soigneusement éviter de mouiller la tête (de l'enfant) avec l'eau (du bain) de peur que celle-ci grossisse de façon disproportionnée ⟨E⟩ If you give birth to a tiny child, take bark of the kapok tree and put it in a basin. Bath the baby and he/she will grow bigger. However, the danger (with this bath) lies at the level of the head. Great care should be taken to avoid pouring water onto the (child's) head lest the latter becomes disproportionately big ⇒ quot/cit.</p> <p>Quot/cit ⟨F⟩ D'après RAPONDA et SILLANS (1961: 106) le tronc (du fromager) sert parfois à faire des pirogues. Le kapok est vendu dans le commerce. Chez les indigènes, il sert à garnir les coussins et les matelas ou bien il est filé pour en faire des sacs de voyage. L'écorce des jeunes arbres, débarassée des épines, est usitée pour faire des cloisons de cases. La décoction de l'écorce est employée comme vomitif ou en lavement. Les feuilles sont émollientes ou calment les névralgies. On tire de l'huile de ses graines. Ce végétal géant est considéré par les Noirs comme un arbre sacré. Lorsqu'on voit sur l'emplacement des anciens villages deux pieds de <i>Ceiba</i> côte à côte, c'est l'indice que là, autrefois, sont nés deux enfants jumeaux. On le plante aussi comme arbre principal du fétiche-protecteur ou sur les tombes. C'est au pied de ces arbres que l'on dépose les offrandes faites aux mânes des ancêtres ou aux génies tutélaires.</p> <p>⟨E⟩ According to RAPONDA-WALKER and SILLANS (1961: 106) the trunk (of the kapok tree) is sometimes used to make canoes. The kapok is sold on the market. Among the indigenous population, it is used to stuff cushions and mattresses, or else it is spun to make travelling-bags. The bark of the young trees, with the thorns removed, is used to make partitions for huts. A decoction of the bark is used as an emetic or for enemas. The leaves can be used as an emollient or to soothe neuralgia. Oil can be obtained from its seeds. This giant plant is regarded by the Blacks as a sacred tree. When two kapok trees, standing side by side, are seen on the site of ancient villages it is an indication that twins were born there long ago. It is also the tree that is mainly planted as protecting fetish and on graves. It is at the foot of these trees that offerings to ancestral spirits or to tutelary spirits are left.</p>		

Text example 4

díbogha (+ bogha) [díbòǵà] <i>n. cl. 5</i> < *-búga or -búka		
◆ Mwíri múdíídi ivávéggha (tsi)mbânzi mu úyaaba búkulu bu bulôngu. ◆	⟨F⟩ Arbuste dont les râpures d'écorces fraîches, la poudre de racines séchées ou encore la décoction des racines est donnée aux initiés afin qu'ils acquièrent une connaissance du monde; ⟨T⟩ iboga (Tabernanthe iboga).	⟨E⟩ Shrubby tree of which fresh raspings of the bark, powder of the dried roots or else a decoction of the roots is given to initiates in order that they acquire a knowledge of the world; ⟨T⟩ iboga (Tabernanthe iboga).
Σ Bamúvegha díbogha.	⟨F⟩ On lui a donné l'iboga.	⟨E⟩ He was given iboga.
Σ Anághólu na díbogha.	⟨F⟩ L'iboga le fait soûler.	⟨E⟩ The iboga has intoxicated him.
<p>▼ ⟨F⟩ Arbrisseau des sous-bois de la forêt, l'ibogha fait partie des produits réputés de la pharmacopée traditionnelle gabonaise. Il possède un double usage, médical et magique. Les râpures d'écorce ainsi que les racines de l'arbrisseau se consomment comme fortifiant, aphrodisiaque ou encore comme coupe-faim. À faible dose les racines combattent efficacement les coliques. Les vertus magiques de l'ibogha sont connues des populations locales depuis longtemps. Mais ce n'est que dans ce qu'il convient d'appeler la religion du Bwiti (⇒ Bwiti) que l'ibogha est placé en très haute estime. En effet, c'est l'arbre 'sacré' des adeptes de cette religion. ⇒ quot/cit.</p> <p>Quot/cit D'après RAPONDA et SILLANS (1961: 90), d'un point de vue scientifique "(...) l'iboga est utilisé comme stimulant neuro-musculaire (dépressions et asthénies physiques et intellectuelles); antitoxique (convalescences des maladies infectieuses, intoxications). (...) c'est surtout dans les pratiques fétichistes qu'ils (les indigènes) en font usage (de l'ibogha). C'est en effet la plante magique par excellence des adeptes du Bouiti. Elle sert principalement pour la cérémonie rituelle d'initiation à cette société secrète. L'absorption des râpures d'écorce ou du bois de la racine détermine une sorte d'ébriété, d'hébétude, de torpeur dans les facultés intellectuelles. À doses massives, l'iboga fait perdre la raison, provoque des hallucinations et parfois la mort. L'état de léthargie dû à l'usage immodéré de l'iboga dure 4 à 5 jours pendant lesquels le patient ne prend aucune nourriture."</p> <p>"(Les adeptes du Bwiti) font également usage d'un breuvage magique confectionné avec les râpures d'écorce de Tabernanthe iboga. Cette mixture les plonge dans un état d'ébriété pouvant aller jusqu'à annihiler momentanément toute notion du monde extérieur. Au cours de cette cérémonie religieuse qui est le Bouiti, on ne se contentait pas d'absorber un tel «léthé»; on y a pendant longtemps procédé à des agapes d'une nature particulière d'où les végétariens devraient probablement être exclus (...) Allons, ne rions donc pas des pratiques de sorcellerie des Noirs d'Afrique car il est bien rare en effet que l'homme, même évolué, néglige le Merveilleux! Nous avons bien aussi nos cartomanciennes, tous les dispensateurs d'horoscope, les voyants, sans compter les sorcières du Moyen Age ..." RAPONDA-WALKER et SILLANS (1961: 32).</p> <p>⟨E⟩ Shrubby tree of the undergrowth of the forest, the ibogha forms part of the well-known products of the traditional Gabonese pharmacopoeia. It has a double use, medical and magical. Raspings of the bark as well as the roots of the tree are taken as stimulant, aphrodisiac and hunger-depressant. A small dosage of the roots is an effective remedy for colic. The magical properties of the ibogha have been known to the local community for a long time. But it is only in the so-called Bwiti religion (⇒ Bwiti) that the ibogha is held in very high esteem. In fact, it is the 'sacred' tree of the adepts of this religion.</p>		

Text example 5

These typographical structural elements form part of the rapid access structure of the dictionary (cf. Hausmann and Wiegand 1989: 329). The use of the upside down triangle enables a user who is not interested in the grammatical aspect of the lemma signs **mufúma** and **díbogha**, immediately to jump to the upside-

down triangle. So the dictionary user does not have to read everything until he/she comes to the information on plant uses. From this introductory text, the user is referred to relevant literature in which the lemma sign is discussed at length. According to the approach of Wiegand (1996), the lemma sign **mufúma** or **díbogha** is the reference position. The reference entry \Rightarrow quot/cit. (see quotation/citation) consists of a reference marker (\Rightarrow) and an entry indicating the reference address (quot/cit). Given the fact that this mediostructural procedure does not exceed the boundaries of the article, it can be referred to as an internal reference address (cf. Gouws and Prinsloo 1998: 20-22). In addition, both Text examples 5 and 6 have a limited data type.

ibũngu ¹ (+ bungu) [ibũngù] <i>n. cl. 7 < *-bũngu or -bũngu</i>		
◆ Bwála bu mátenɡu°. (Bwála a búna bwála bú nyimbí. Bátu bootsu basábúmóna. Bwála bu bátu báfu kála.) ◆	⟨F⟩ Village des revenants. (Ce village est occulte. Par conséquent il est invisible pour le profane. C'est le village des morts.)	⟨E⟩ Village of ghosts. (This village is regarded as being occult. It is therefore invisible to the layperson. It is the village of the dead.)
Σ Anámúbóka anámútúula mo ibũngu yandi.	⟨F⟩ Il l'a tué et l'a mis dans son village des revenants.	⟨E⟩ He killed him and put him in his village of ghosts.
fumu ibũngu Fumu nyimbi.	⟨F⟩ Le maître du village des fantômes.	⟨E⟩ The master of the village of ghosts.
Σ (Mangúmba.) Fumu ibũngu áká?	⟨F⟩ Où est la personne responsable de la veillée (mortuaire)?	⟨E⟩ Where is the person in charge of the (funeral) wake?
Σ Mɔna fumu ibũngwa!	⟨F⟩ Voici la personne responsable de la veillée!	⟨E⟩ Here is the person in charge of the wake!
<p>▼⟨F⟩ Le maître du village des revenant utilise les esprits de ceux et celles dont il/elle a tué et asservi pour accéder à une position sociale élevée. Cette quête d'un meilleur statut social se traduit souvent par une ascension fulgurante. Ces esprits sont également utilisés comme gardes du corps et hommes de main. Le propriétaire d'un village de fantômes jouit de la protection surnaturelle des esprits dont il/elle a le contrôle. Ces esprits passent pour commettre des assassinats ou crimes rituels au sein des populations, faisant ainsi passer en leur maître la force vitale des gens qu'ils mettent à mort. Mais les privilèges dont le maître d'un village de fantômes jouit ont un prix. Ce dernier pour se concilier ces esprits, il se doit de leur offrir annuellement un sacrifice humain. En se refusant à honorer les termes de ce contrat, le propriétaire des dits-esprits s'expose à leur colère impitoyable.</p> <p>⟨E⟩ The master of a village of ghosts uses the spirits of the people he/she has killed to achieve a high social position. This quest for a better social status often results in a quick ascent. These spirits are also used as bodyguards and killers. The owner of a village of ghosts enjoys supernatural protection from the spirits he/she has under control. These spirits are believed to commit assassinations or ritual crimes amongst the population, in this way causing the vital force of the people they kill to be passed on to their master. However, the privileges enjoyed by a master of a village of ghosts bears a price. The latter has annually to provide these spirits with a human sacrifice. By refusing to respect this contract, the owner of these spirits exposes him-/herself to their merciless anger.</p>		

Text example 6

The knowledge, beliefs, and practices of the Yilumbu society are extensively reflected in these articles. However, in the case of the article of the lemma **díbogha** in particular, reservations can be made by target users with regard to

an unbalanced presentation of data. It appears that in the encyclopedic section presenting excerpts from relevant literature dealing with the treated lexical item, the focus is on dictionary users who are familiar with French. Although this is detrimental to the English user, this discrepancy may be motivated on the grounds of user proficiency. Given the fact that the planned dictionary is primarily aimed at meeting the needs of Balumbu mother-tongue speakers with a relative good command of French, these target users should get the primary focus.

In the treatment of the article in Text example 6, *bwâla bu mâtengu*[°] (village of ghosts) is the real paraphrase of the meaning of the lemma **ibûngu**. The sentences *Bwâla a bûná bwâla bú nyimbí*, *Bátu bootsu basâbûmóna*, and *Bwâla bu bátu báfú kála* are explanatory additions and marked as such by the use of parentheses. In addition to the foregoing, the symbol [°] after the lexical item *mâtengu* (ghosts) means that this word will be explained somewhere else in the planned dictionary. What follows after the English co-text example is of an ethnographical nature. This encyclopedic entry in both French and English gives the user more information regarding the cultural context of the lemma **ibûngu**.⁸

An implicit presentation of culture-bound lexical items may be acceptable for those who might have a sound knowledge of their mother tongue and native culture, but for those having a passive knowledge of Yilumbu this may prove to be highly problematic. That is why the lexicographer must make the data maximally explicit in the lexicographic text. As learners are also dealt with in this research, the lexicographer has to be well aware of the needs of these learners. The planned dictionary has to be user-orientated.

With regard to the dictionary article given as Text example 7, a lot of examples of sublemmatic addressing can be found, in particular in the section dealing with a niched paradigm of compounds with **ngwîsi** as first component, introduced by the symbol ■. The ordering within this sinuous lemma file is strictly alphabetic and very user-friendly. In fact, each compound is followed by an explanation in the target languages so that the user gets a better idea of the meaning. As long as the explanation of the meaning continues, there is also a transfer of focus from the source to the target language. The purpose of the use of non-typographical markers in the case of the entries **coupure de cordes**, **levée de terre** and **port de deuil** which have been made "visual" to the dictionary-user by the use of bold characters, and indicated by (frGab), is that these lexicographic units are the so-called culture-bound elements in Gabonese French.

In rural areas, the corpse of a person who has passed away can be kept for two days on the maximum. By the third day it has to be buried. If the person who passed away was a married man, the tradition requires his widow(s) to sleep on leaves of the banana-tree (*Musa paradisiaca*) for four days. On the fifth day a ceremony is organised in order to allow the widow(s) to sleep on a mat on the ground. Two weeks after the burial, a ceremony called *mangumba* is arranged in order to pacify the spirit of the deceased. During this ceremony, held at night, songs and dances are performed. In the morning, the cutting of

the pall that people have around their wrists as sign of mourning takes place. This is referred to as the "coupure de cordes" in Gabonese French. Then members of the family, the children of the deceased man and his widow(s) have to undergo purificatory rites. According to custom, an uncleanness caused by death is believed to rest on widows, in particular, after the death of the husband. This uncleanness has to be washed away by a ritual bath before the members of the family can wear the mourning of the deceased (referred to as "port de deuil" in Gabonese French). This ceremony also serves as a rising from the earth which is referred to as "levée de terre" in Gabonese French. For this part of the funeral rite, the widow's bed (the sleeping mat) is put upside down (**uwúdígha tãngi**) which literally means that the widow is now allowed to sleep on a bed. Finally, after a year, a celebration is held officially to end the mourning period ("retrait de deuil" in Gabonese French).

ngwísi (+ gwísi) [ngwísi] (aussi/also ngúsi) n. cl. 9/10 < *-guj̥si		
◆ Muyínu o dúfu ná gho báyisi. ◆	(F) Dance ordinaire ou cérémonie à l'honneur des génies tutélaires.	(E) Customary dance or ceremony for tutelary spirits. ⇒ ngwédi .
Σ (Ghâng.) Yoonu batsísi ngwísi i mfúúmbi Mákaya. (Men.) Yoonu batsívanga ngwísi i mfúúmbi Mákaya.	(F) Hier, ils ont organisé la cérémonie (mortuaire) de feu Mákaya.	(E) Yesterday they organised the late Makaya's (funeral) ceremony.
Σ (Ghâng.) Ngwísi be yína utsíwítsíla pwééla bátu. (Men.) Ngwísi be yína iyétsíla bátu ba bĩnga.	(F) Beaucoup de personnes ont assisté à cette cérémonie (mortuaire).	(E) A lot of people attended this (funeral) ceremony.
<p>■ ngwísi dúfu est une veillée mortuaire/ <i>is a funeral wake</i>; ngwísi i Mábaantsi rite de passage statutaire qui permet aux jeunes filles de passer de l'âge de puberté à l'âge adulte. Le rite a été pratiqué à l'origine par les femmes. Aujourd'hui, les Mabaantsi est une société secrète initiatique mixte/ <i>is a rite of passage that allows girls to pass from puberty to become adults. The ritual was originally practised by women. Today the Mabaantsi is a mixed secret initiation society</i>; (Men.) ngwísi i mĩtsi (⇒ (Ghâng.) ngwísi misĩnga) est une cérémonie funéraire marquée par la (frGab) coupure de cordes provenant du drap funéraire ou mortuaire que les gens portent autour des poignets en signe de deuil. Cette cérémonie est également organisée en guise de (frGab) levée de terre et elle annonce éventuellement le (frGab) port de deuil/ <i>is a funeral ceremony marked by the cutting of the pall people bear around their wrists as sign of mourning. This ceremony also serves as a rising from the ground and it eventually announces the bearing of the mourning</i>; ngwísi Mughulu est une société secrète et initiatique qui viendrait des Bisira/ <i>is a secret initiation society believed to have come from the Bisira</i>; ngwísi Mulóghu est une société secrète et initiatique. Lorsqu'une personne de sexe féminin est frappée par le Mulóghu, elle tombe ordinairement dans une espèce de transe (⇒ utútúgha), un genre de folie qui est supposée causée par l'esprit des ancêtres. Seul le père (au sens que ce terme a dans les langues gabonaises) peut apaiser la colère de l'esprit des morts. C'est pour cette raison que les gens font généralement appel à lui pour un rituel de délivrance consistant simplement à des paroles de bénédiction/ <i>is a secret initiation society. When a female person is struck by the Mulóghu, she usually falls in a kind of trance (⇒ utútúgha), a sort of madness assumed to be caused by the ancestral spirits. Only the father (in the sense that this word has in the Gabonese languages) can appease the anger of the spirits of the dead. It is for this reason that people generally call upon him to perform a ritual of deliverance that merely consists of words of blessing.</i></p>		

Text example 7

6. On the metalanguage used in the proposed article structures

— The problem

If French or English is chosen as the metalanguage (e.g. the language used for the texts of definitions as well as notes in the form of labels) of the planned dictionary, the latter will be less informative for Yilumbu speakers than their French and English counterparts whose language will predominantly be used to reflect on the meaning of lexical items as well as the culture that underlies them.

— Decision taken

The solution would lie in the combination of a few standardised sets of abbreviations in the form of labels and the use of Yilumbu on the metalinguistic level. With regard to French and English, the standardised sets of abbreviations (in the form of labels) used in these two proposed articles mainly include the use of the part of speech label *n.* (*nom/noun*). Given the fact that the primary target user group of the planned dictionary is the Balumbu mother-tongue speakers, the metalanguage for the paraphrases of meaning, the so-called diatopic or regional labels such as (*Ghâng.*) and (*Men.*), sphere of usage labels such as (*Mangúmba*), and encyclopedic information in synopsis articles addressed at mother-tongue speakers should be in Yilumbu. This is important and can be motivated by what follows. With regard to dictionary typology, Zgusta (1989: 71) has mentioned the role of a subcategory of dictionaries instrumental in the revival of languages. Given the fact that nowadays the Gabonese in general and the Balumbu in particular live in a society where all aspects of life have become increasingly westernised, the cultural dimension must be said to be one of the areas that really need to be examined in more detail. Thus all attempts towards preventing the indigenous cultures from disappearing (both at state or individual level or by means of a dictionary, for example) may be referred to as **culture revival**.

7. On the access structure of the proposed article structure

— The problem

Dictionaries are known to present great access problems to users. That is why one of the challenges facing lexicographers is to devise a user-friendly access structure enabling inexperienced dictionary users to maximally benefit from using a particular dictionary.

— Decisions taken

Given the fact that dictionary users usually experience great difficulties in

understanding the boundaries of a given entry, the definition in particular, a decision has been taken to make use of the so-called ordering devices to mark the boundaries of each simple entry. For example, all the data occurring between the following structural indicators (◆ ◆) represents the paraphrase of meaning of the treated lexical item. On the same principles, co-text examples are introduced by the structural indicator (Σ||). In addition and with reference to Wiegand's concept of **microarchitecture**, each piece of data type starts on a new line. Such a foundation is likely to greatly improve the access structure of the planned dictionary because the user perceives each data type clearly. To be effective, these metalexigraphic procedures must be reflected on in the compulsory users' guide of a particular dictionary. In present-day lexicography more and more proposals are made towards introducing a component "dictionary using skills" at school's level that will equip users with the necessary expertise to use dictionaries more efficiently as sources of reference.

Conclusion

Devising the microstructure of a dictionary is one of the greatest responsibilities of the lexicographer, especially when it comes to the structure of a specific category of articles, namely synopsis articles as opposed to single articles. The metalexigraphical discussion above shows that synopsis articles have a more general use than its restricted application in LSP dictionaries. In accordance with the data distribution structure of the planned dictionary, synopsis articles should not only include scientific terminology but also cultural lexical items. In the traditional approach, synopsis articles are characterised by a low density of data categories because data types such as the items giving the translation equivalent, the etymology or the usage example are not so important. The focus is usually on extra-linguistic aspects.

The lexicographic treatment presented in synopsis articles usually results in lengthier articles with a variety of data types and search zones that could impede the successful retrieval of the information users are seeking. Therefore there should be a system of contextual entries to assist users in the most effective way (cf. Gouws 1999: 47). In addition, mediostructural representations, and access and addressing procedures should also come into play. As far as the addressing structure is concerned, the distant addressing prevailing between co-text entries and paraphrases of meaning/translation equivalents in unintegrated microstructures should lead the lexicographer to use an integrated microstructure in synopsis articles. However, a semi-integrated microstructure could be used in synopsis articles of which the lemma sign is of a polysemous nature. Because of a relation of distant addressing between a co-text entry and the relevant paraphrase of meaning/translation equivalent in the second section of a semi-integrated microstructure, the lexicographer will still have to provide users with contextual data enabling them to disambiguate the different senses presented in the article. Moreover, by providing users with a list (in the

back matter section) of all the items which are treatment units in the central list's synopsis articles, the particular dictionary will be polyfunctional and polyaccessible.

Abbreviations, Symbols and Labels Used in the Text Examples

- ⟨F⟩: French
- ⟨E⟩: English
- ⟨T⟩: Translation equivalent(s)/traduction(s)
- 1, 2, 3, 4, ... These are markers of polysemy.
- ◆◆ These mark the boundaries of the definition.
- n.*: noun
- cl.*: class
- <: comes from
- *: proto-Bantu reconstruction
- Σ||: This introduces examples.
- ▼: This marks a specific article slot in which extra-linguistic data are given.
- ⇒: Reference marker
- °: When given after a particular lexical item, this symbol indicates that the lexical item in question is explained somewhere else in the specific dictionary.
- : This introduces a sublemma.
- Quot/cit**: Quotation/citation
- (*apoc.*): apocope
- (*by ext.*): by extension
- (*chrét.*): religion chrétienne
- (*Chris.*): Christian religion
- (*frGab*): Gabonese French
- (*Ghâng.*): Yilumbu yi ghângu
- (*Mangúmba*): Dance held at funerals
- (*Men.*): Yilumbu yí menááne
- (*o íngána*): Christian religion (lit.: "in church")
- (*par ext.*): par extention

Acknowledgement

Comments, discussions and suggestions from Prof. Sven Tarp, Prof. R.H. Gouws and Dr J.C.M.D. du Plessis on an earlier version of this article are greatly appreciated.

Endnotes

1. In this article I use the official Gabonese spelling and not the traditional semiphonetic spell-

ing employed by Guthrie, Jacquot, Kwenzi-Mikala and others. The underlined e, o and n represent [e], [o] 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 it by a, as in Make.

The accent signs on vowels indicate tones (for a description, see Afane Otsaga 2002: 76-79).

2. It must be kept in mind that the term "African languages", used here for the sake of sensitivity with respect to the term "Bantu languages" in the South African context, has a wider application than languages belonging to the "Bantu language family", such as the Gabonese languages.
3. This information comes from a class lecture of Prof. R.H. Gouws.
4. Some scholars in the field of lexical semantics (cf. Lehrer 1974 and Lyons 1977), have drawn the attention to the existence of lexical gaps. A distinction is usually made between linguistic and referential gaps. According to Gouws (1996: 25-26), a linguistic gap exists where in two different speech communities the speakers of a language X do not have a word for an object, which only exists in language Y and vice versa. A referential gap occurs where a lexical item from language X does not have a translation equivalent in language Y simply because the speakers of that language are not familiar with the referent in question. In this case, lexicographers quite often use a paraphrase of meaning or a loan word as a surrogate equivalent. The lack of lexical items in e.g. English or French to offer as translation equivalents for the Yilumbu "mupééyi" (wild mangoes that are pounded without being previously dried) and "ngwâmba" (absence of meat in the homes) represent a referential gap because both speakers of French and English are not familiar with the referents of these lexical items.
5. The use of the inserted inner text and other typographical features is not from Kritzinger et al. but from the author.
6. Cf. Naden 1996: 87.
7. The proto-Bantu reconstructions in Text examples 4-7 come from Nsuka-Nkutsi (1980).
8. I gladly acknowledge information obtained from Guy-Roger Mihindou as far as illustrative examples given in the treatment of the articles of lemmata **ibûngu** and **ngwîsi** are concerned.

Bibliography

- Afane Otsaga, Thierry.** 2002. Les tons dans les dictionnaires de langues gabonaises: situations et perspectives. *Lexikos* 12: 75-89.
- Bergenholtz, H., S. Tarp and H.E. Wiegand.** (1999) Datendistributionsstrukturen, Makro- und Mikrostrukturen in neueren Fachwörterbüchern. Hoffmann, L., H. Kalverkämper, H.E. Wiegand (Eds.), together with Ch. Galinski and W. Hüllen. *Fachsprachen. Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft / Languages for Special Purposes. An International Handbook of Special-Language and Terminology Research.* Berlin/New York: De Gruyter.
- 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 Lexicography Units.* Research report submitted to the Pan South African Language Board.
- Gouws, R.H.** 2002. Niching as a Macrostructural Procedure. *Lexikos* 12: 133-158.
- Gouws, R.H. and D.J. Prinsloo.** 1998. Cross-Referencing as a Lexicographic Device. *Lexikos* 8: 17-36.
- 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.
- Kritzinger, M.S.B. et al. (Eds.).** 1986¹³. *Groot Woordeboek. Afrikaans–Engels/Engels–Afrikaans*. Pretoria: Van Schaik.
- Lehrer, A.** 1974. *Semantic Fields and Lexical Structures*. Amsterdam/London: North-Holland Publishing Company.
- Lyons, J.** 1977. *Semantics. Volume 1*. Cambridge: Cambridge University Press.
- Naden, Tony.** 1996. Ancestor Non-worship in Mamprali. *Lexikos* 6: 71–103.
- Nsuka-Nkutsi, F.** 1980. Quelques reflexes du proto-Bantu en Punu. Nsuka-Nkutsi, F. (Éd.). 1980. *Éléments de description du Punu*: 129–179. Lyons: PUL.
- Nyangone Assam, B. and P.A. Mavoungou.** 2000. Lexicography in Gabon: A Survey. *Lexikos* 10: 252–274.
- Prinsloo, D. J. and R.H. Gouws.** 2000. The Use of Examples in Polyfunctional Dictionaries. *Lexikos* 10: 138–156.
- Raponda-Walker, A. and R. Sillans.** 1961. *Les plantes utiles du Gabon: essai d'inventaire et de concordance des noms vernaculaires et scientifiques des plantes spontanées et introduites du Gabon*. Paris: Lechevalier.
- 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.
- Tomaszczyk, J.** 1983. The Culture-Bound Element in Bilingual Dictionaries. Hartmann, R.R.R. (Ed.). *LEXeter '83 Proceedings*: 289–297. Tübingen: Max Niemeyer.
- Wiegand, H.E.** 1978. Lexikographische Praxis — vom Standpunkt unterschiedlicher Wörterbuchtypen und Wörterbuchkonzepte. *Zeitschrift für germanistische Linguistik* 6: 326–330.
- Wiegand, H.E.** 1996. Über die Mediostrukturen bei gedruckten Wörterbüchern. Zettersten, A. and H. Pedersen (Eds.). *Symposium on Lexicography VII*: 11–43. Tübingen: Max Niemeyer.
- Wiegand, H.E.** 1999. *Semantics and Lexicography. Selected Studies (1976–1996)*. Edited by Anntje Immken and Werner Wolski. Tübingen: Max Niemeyer.
- Zgusta, L.** 1987. Translational Equivalence in a Bilingual Dictionary. *Dictionaries* 9: 1–43.
- Zgusta, L.** 1989. The Role of Dictionaries in the Genesis and Development of the Standard. 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*: 70–77. Berlin: Walter de Gruyter.

The Cause of Urban Slang and its Effect on the Development of the Northern Sotho Lexicon*

V.M. Mojela, *School of Languages and Communication Studies, University of the North, Polokwane, Republic of South Africa (mojelav@unorth.ac.za)*

Abstract: Urban slang terminology is extensively used today by most indigenous communities when speaking various South African indigenous languages. This is usually the case with informal conversations where the type of language used is also regarded to be informal. There are many instances where slang terms are used, consciously or unconsciously, in the form of code-switching with the indigenous languages in formal speech or informal conversations. These urban slang terms are regarded by most of the South African indigenous communities as prestige terms, especially by the young generation. Just as English and Afrikaans terminologies which had tremendous influence on the vocabularies of the indigenous languages, urban slang terminology has great influence on the daily usage of the South African indigenous languages. Despite this pressure, Northern Sotho and other South African indigenous languages do not have officially recognized adoptives which are derived from slang terminology. This is due to the fact that most language authorities do not want to associate their standard languages with urban slang because of various reasons which include, *inter alia*, the conservative idea of associating urban slang with crime and juvenile delinquency. Despite the negative attitudes of most of the conservative language authorities who regard the loan words of urban slang origin to be "contaminating" the South African indigenous languages (such as Northern Sotho), this slang terminology is every day becoming more popular and seems to have come to stay rather than be removed from the language of the people. A considerable fraction of the vocabulary of Northern Sotho, for instance, is made up of words which can be regarded as "loan words" derived from urban slang, even though formally we are still refusing to accept them as part of the lexicon of official Northern Sotho.

Keywords: EUPHEMISM, FLYTAAL, INFERIOR DIALECT, INFERIOR LANGUAGE, MULTILINGUALISM, NORTHERN SOTHO LEXICON, PRESTIGE LANGUAGE, PRESTIGE TERMINOLOGY, SECRET TERMINOLOGY, TSOTSITAAL, URBAN SLANG, UNDERWORLD

Opsomming: Die oorsprong van stedelike sleng en die uitwerking daarvan op die ontwikkeling van die Noord-Sotholeksikon. Stedelike slengterminologie word tans wyd gebruik deur die meeste inheemse gemeenskappe wanneer hulle verskillende inheemse Suid-Afrikaanse tale praat. Dit is gewoonlik die geval met informele gesprekke waar die

* This article is a slightly adapted and extended 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.

gebruikte tipe taal ook as informeel beskou word. Daar is baie gevalle waar slengterme, bewustelik of onbewustelik, gebruik word in die vorm van kodewisseling met die inheemse tale in formele spraak of informele gesprekke. Hierdie stedelike slengterme word as prestigeterme deur die meeste van die inheemse Suid-Afrikaanse gemeenskappe, veral deur die jong generasie beskou. Net soos Engelse en Afrikaanse terminologieë wat groot invloed op die woordeskat van die inheemse tale gehad het, het stedelike slengterminologie 'n groot invloed op die daaglikse gebruik van die inheemse Suid-Afrikaanse tale. Ten spyte van hierdie druk het Noord-Sotho en ander inheemse Suid-Afrikaanse tale nie amptelik erkende leenwoorde wat van slengterminologie afgelei is nie. Dit is die gevolg van die feit dat die meeste taaloutoriteite nie hul standaardtale geassosieer wil hê met stedelike sleng nie, vanweë verskillende redes wat, onder andere, die konserwatiewe idee van die assosiasie van stedelike sleng met misdaad en jeugmisdaad insluit. Ten spyte van die negatiewe houdings van die meeste konserwatiewe taaloutoriteite wat die leenwoorde van stedelikeslengoorsprong beskou as "kontaminerend" vir die inheemse Suid-Afrikaanse tale (soos Noord-Sotho), word hierdie slengterminologie elke dag meer populêr en lyk dit of dit blywend is eerder as verwyderbaar uit die taal van die mense. 'n Aansienlike deel van die woordeskat van Noord-Sotho, byvoorbeeld, bestaan uit woorde wat beskou kan word as "leenwoorde" verkry uit stedelike sleng, alhoewel ons nog formeel weier om hulle as deel van die leksikon van offisiële Noord-Sotho te aanvaar.

Slutelwoorde: EUFEMISME, FLAAITAAL, ONDERGESKIKTE DIALEK, ONDERGESKIKTE TAAL, MEERTALIGHEID, NOORD-SOTHOLEKSIKON, PRESTIGETAAL, PRESTIGETERMINOLOGIE, GEHEIME TERMINOLOGIE, TSOTSITAAL, STEDELIKE SLENG, ONDERWÊRELD

1. Introduction

This paper focuses attention on the reasons why urban slang terminology is not only commonly used among the Northern Sotho communities, but is every day becoming more popular, especially among the young generation. The second and main objective of this paper is to highlight some of the major consequences resulting from the use of urban slang terminology in the development of the vocabulary in Northern Sotho. Even though there are many works which have already been published dealing with slang, very few of these deal with the effects of slang on the vocabulary of the South African indigenous languages, especially the use of slang in Northern Sotho.

The few scholars who contributed research works in this field include people such as De Klerk (1991), Bailey (1985), Coertze (1969) and Mfusi (1992). Mfusi gives an excellent analysis of urban slang spoken in the area of Soweto and Johannesburg from the Zulu point of view. He explains the status, the origin and the importance of urban slang among the urban Zulu-speaking communities, e.g. the youth, tsotsis, etc., but does not show how the vocabulary of other African languages, such as Northern Sotho, are also affected. Coertze investigates the impact of the use of urban slang in the area around Pretoria, especially in the township of Atteridgeville. This research is based on the tra-

ditional township slang, e.g. "flytaal" (as he calls it), which is dominated by Afrikaans, in the area where Tswana is predominant. Just like that of Mfusi, Coertze's research does not investigate the effect of the modern urban slang terminology which is used today in the form of code-mixing with the indigenous languages such as Northern Sotho, but deals with the terminology of the original so-called "tsotsitaal" or "flytaal" which was dominant in the early fifties and sixties in the PWV (Pretoria, Witwatersrand and Vereeniging) area. Other writers, such as De Klerk (1991) and Bailey (1985), researched on slang pertaining to English usage in general.

Some of the major factors which influence the development of urban slang among the Northern Sotho people include the following: multilingualism, prestige and recognition, brevity and conciseness, euphemism and disguise. These factors are usually associated with a desire by the South African indigenous communities to acquire a knowledge of the two status languages, i.e. English and Afrikaans. This desire for knowledge of English and Afrikaans is usually coupled with prestige as well as factors such as secrecy and euphemism, which some scholars, such as Mfusi (1992: 57), refer to as "lexical avoidance".

2.1 Multilingualism

Slang develops in a contact situation where two or more languages, usually of different status, are spoken, i.e. in a multilingual society. According to Mansour (1980: 250), multilingualism covers two distinct phenomena, i.e. (a) the co-existence and close contact of several languages within a given community where, owing to the nature of community life, a special pattern of language use needs to be adopted; and (b) the co-existence of several ethnolinguistic entities within the boundaries of a state, with each entity occupying a relatively well-defined territory. In situations such as these, most speakers of the so-called "inferior languages" have a desire to know the prestige languages, such as English and Afrikaans, and start imitating them using the little knowledge of the vocabulary and pronunciation they have already acquired, which in most instances amounts to a mixture of vocabulary and structures of the indigenous languages and those of the prestige languages in order to forge understanding.

The urban community in the metropolitan areas of South Africa is a good example of a multilingual society. Besides the existence of the two "prestige languages" which used to be official languages in the Republic of South Africa, i.e. Afrikaans and English, the other indigenous languages such as Zulu, Xhosa, Tswana, Southern Sotho, Venda, Tsonga and Ndebele exist in the same proximity, and this resulted in the creation of a "mixed society" which required a "mixed variety" (such as slang) to act as *lingua franca* to simplify communication between several peoples of different linguistic backgrounds. Mfusi (1992: 53) confirms this as follows with regard to the situation in Soweto:

In a heterogeneous urban community such as Soweto's, the majority of the

population speaks several languages. In almost all cases the knowledge of English and/or Afrikaans is indispensable for the acquisition of any type of employment. Actually, historical conditions in South Africa interrupted a natural extension of the functional range of local languages. With the institutional pattern, imposed by the colonial powers, arose the need to employ European languages not only in the role of languages of wider communication, but also as official languages.

As a result of multilingualism in the metropolitan areas of South Africa, especially the PWV, the people from various linguistic backgrounds felt a need to communicate in a sort of *lingua franca* variety which is understood by most people without prior formal instruction. As a "register" incorporating vocabularies from different languages, slang seemed to be an appropriate *lingua franca* for the urban indigenous communities, especially the youth. In most cases slang uses most of the known terminology from different languages in a form of metaphor which makes it easily understood without prior instruction in these "languages".

2.2 Prestige and Recognition

Prestige seems to be among some of the most important reasons for the development of slang among the youth. For instance, the development of *tsotsitaal* among the urban youth may be ascribed to their desire for communication in the prestige languages. Lack of sufficient knowledge of these languages resulted in the youth mixing the vocabulary and structures of these languages with their own vernaculars, which ultimately led to the development of a "language" which is neither English nor Afrikaans, nor any of the indigenous languages, but a different variety of mixed languages commonly known as "*tsotsitaal*" or "*flytaal*". In theory, slang varieties such as *tsotsitaal* seem to be of a low standard, especially when considering facts such as the following: (a) their lack of orthographies, (b) their habit of continuous and rapid change of terminology, (c) their association with juveniles, as well as (d) their association with the criminal elements in society. Nevertheless, in spite of all these factors, the prevailing situation is one in which these varieties are highly regarded by most of the young generation in our country, and for them the ability to speak *tsotsitaal* does not only signify a sense of belonging to a group of *clevas* (clever people), but also a realisation of having a high social status. The prestige of urban slang is explained by Mfusi (1992: 60) as follows with reference to Soweto Zulu slang:

Soweto Zulu slang holds a special position of prestige which derives from its use in music and political circles. Actually, it has become quite prestigious for the youth to use this code as it identifies them as Sowetan. People from Soweto are regarded with awe by Blacks from other parts of South Africa, and are also considered as trend-setters in fashion and also in the liberation struggle.

The prestige of flytaal is derived from its association with the metropolitan areas in South Africa since it originated in these urban environments. Most urban youth use urban slang terminology in their conversation in order to distinguish themselves from the rural youth, whom they regard as *moggos* (morons). In turn, the rural youth also want to be associated with the urban life by learning to use urban slang terminology in their communication so that they may not be regarded as *moggos*. As such, a knowledge of the so-called "lingo" (which is another synonym for "flytaal") is enough to elevate one to a status of *clevas*, *majitas*, *bras*, *motsiengs* or *mncas*, which are some of the tsotsi names for "clever people" or "tsotsis". To speak a refined standard vernacular without the inclusion of slang terms is today regarded by the youth as a sign of stupidity, and as such, a person who does not seem to have a knowledge of their lingo is referred to as *barri*, *spaza*, *moggo* or *kgashu* (which are terms for "stupid" or "a stupid person"). Code-mixing with urban slang terminology is today used in South Africa by many rural youths speaking different indigenous languages. In this way, the use of urban slang terminology helps one to be recognized as belonging to the "advanced" group, and that is why every teenager (rural or urban) strives to know this type of terminology as a means of identification or association with the urban environment. With regard to how they learn this terminology, De Klerk (1991: 81) says the following:

No one teaches these words, they must be picked up by careful observation, and used carefully and coolly. Asking about what they mean would be tantamount to admitting failure as a teenager.

The Northern Sotho youth can in this regard be divided into two categories, i.e. those who practically stay or belong in the metropolitan areas, and those who reside in the rural areas in Leboa (the North). Almost all of the Northern Sotho youth who live in the PWV speak a vernacular (Northern Sotho) which, besides terminology from Afrikaans and English, is rich in urban slang vocabulary. A considerable number of the young generation whose lives are still entirely restricted to the rural areas still use a refined standard Northern Sotho which is free of slang terminology, especially the female youth, while the majority of the rural youth, especially the male youth, try to imitate the urban youth in order to be recognized as belonging to a group of a "sophisticated" young generation.

2.3 Disguise and Euphemism

Another reason which leads to the use of slang is to disguise or conceal information which the speakers do not want other people to know. For this reason, the speakers sometimes use a special vocabulary which is solely intelligible to them. With regard to South African urban slang, the need to disguise and conceal information became necessary, especially with tsotsitaal, after the tsotsis "turned to robbery, mugging and other violent crimes" (Coplan 1985: 162). To-

day, tsotsis are associated with criminal activities such as theft, robbery, rape and drug-dealing, which are punishable by law, and therefore, secrecy is one of the major objectives when using their "language" to avoid imprisonment. Disguise is usually a primary reason why in most cases slang terminology changes all the time, or in some cases there are many terms referring to one concept. This is due to the fact that a secret term is only secret for some time before it is known by other people, which obviously forces the users of the term to find another secret term for the same concept to keep the secrecy. This is mostly the case with all the illegal or unlawful activities which might lead to the speakers being arrested if the concealed information should reach the authorities. For instance, the following terms either refer to marijuana or are marijuana-related concepts: *maruana*, *mingus*, *gantja*, *the holy herb*, *dabas*, *giggle sticks*, *greens*, *majat*, *tarries*, *zol*, *boppa*, *moshwang*, *fasa-nkhola*, etc. Anderson and Trudgill (1990: 79) say the following in this regard:

It has been said that one function of the language of thieves and drug addicts is to keep the content of their conversations secret — outsiders should not understand what is being said. This is sometimes called anti-language. Since, however, most of the words they use are not at all hard to understand, this claim is dubious. However, there are other ways to keep outsiders outside. A member of the narcotics police has informed us that the language of drug addicts changes rapidly, which makes it very hard for the police to train informers for infiltration into these groups. It is easy to learn the slang words, but it is hard to keep up to date and use and combine words correctly. In this way it is easy for the group members to tell who is a true member of the group.

Besides disguise, slang is also used in the form of euphemism. This is usually the case when the speaker prefers to use slang terms to avoid using standard indigenous lexical items which may not only be associated with the harsh and unpleasant reality of the meaning of the terms but also, according to Stern (1931: 331), "have numerous undesirable associations with the offensive things, and mean exactly what they mean".

As such, most of the words relating to things such as the embarrassing consequences of overindulgence in drinking, and the embarrassing and indecent indigenous terms relating to sexual organs and sexual activities are replaced by decent slang terms. This amelioration of indecent terminology resulted in the creation of many lexical items originating from slang, which are used by most Northern Sotho people. Even though most of such words are still not yet adopted as loan words in the Northern Sotho vocabulary, most people prefer to use them instead of their counterparts which are regarded as Northern Sotho standard terms. For instance, people prefer to use the term *spotso* instead of *bjaleng*, which has an additional connotation of "stupefying drunkenness" in it. The term *tsheri* is more prestigious and less offensive than the standard Northern Sotho words *nyatse*, *motlabo* or *mokaola* which all mean "girlfriend" and which are not as decent as the former. This is the case with well-

known slang terms such as *jola*, *pholaka*, *skebereshe*, *maruana*, etc. as is demonstrated in the following sentences:

Mosadi yo ke skebereshe instead of *Mosadi yo ke kwababane/nkwahla/sefefe*
(This lady is morally weak)

Tate o pholakile instead of *Tate o tagilwe*
(Father is half-drunk)

Simon o jola le Queen instead of *Simon o nyatselane/kaolane le Queen*
(Simon has a love affair with Queen)

Ditsotsi di tsuba maruana instead of *Ditsotsi di tsuba patše*
(Tsotsis smoke dagga)

Tate o ile spotsong instead of *Tate o ile bjalleng*
(Father went to a shebeen)

3. The Effect of Slang on the Vocabulary of Northern Sotho

The use of urban slang has both negative and positive consequences on the development of the vocabulary of standard languages. The negative influence is associated with the so-called "contamination" of the vocabulary of standard languages. The "purity" of the language (Northern Sotho) is destroyed by the use of many words from foreign nonstandardized "registers", i.e. urban slang such as *flytaal*, which, it is feared, might reduce our language to the status of a mere street or gangster language. For this reason, we try to protect the vocabularies of our languages by discouraging the use of these "registers" by our people, and all the time encouraging formal use of the standard language. In his article, Swanepoel (1978) advances the "purist" idea of protecting our standard languages by criticizing the role of urban slang as it is used in the urban communities. He (Swanepoel 1978: 8) says:

Many readers of this article will immediately remark: "o yes that is the so called 'tsotsi' language of some youngsters in our towns". Who of us will not recognize these common terms?: "notch", "cherry", "groove", "mca", "smack", "dribble", "bra", etc. My answer to these people is that this type of language is used so commonly today that it has found its way right into the compositions and other written work. I want to stress another important point: The use of this type of language does not indicate any form of language ability, on the contrary this displays poor taste and exactly the opposite of what the user had in mind.

Unfortunately, we are fighting a losing battle since the use of slang terminology within our languages grows at an alarming pace. This type of terminology is used not only by the youth in high schools and tertiary institutions, but also by some of the teachers, academics, doctors, language specialists and many people whom we regard as professionals and custodians of the "purity" of the stan-

dard languages. These people think of standard language only when they want to be formal to suit formal situations. When conversing with friends outside these formal situations (conversations which take almost 70% of their talking time), they switch over to a variety which is dominated by slang terminology. In this way, slang terminology today has rendered most standard words into disuse, and these standard terms become decorations in our dictionaries, while in practice they are not used. The following slang terms are examples in this regard:

baiza (*paeza*) instead of *gakanega* (to be puzzled/to be confused), e.g.
O a paeza ge a bona basetsana instead of *O a gakanega ge a bona basetsana*
(He is easily confused when he sees girls)

jojo instead of *tsogo-le-kobong* (kickback), e.g.
O amogetše jojo instead of *O amogetše tsogo-le-kobong*
(He has received a kickback)

tšheri instead of *nyatse* (girlfriend/mistress), e.g.
Ke tšheri ya ka instead of *Ke nyatse ya ka*
(She is my girlfriend)

diza instead of *go fa dipute* (give presents/money to a girlfriend/fiancée), e.g.
Tšhelete ya gagwe e fetšwa ke go diza instead of *Tšhelete ya gagwe e fetšwa ke go fana ka dipute*
(He wastes his money by giving it to girlfriends)

This shows that the positive consequence of urban slang can be said to be related to the increase in the vocabulary of languages. Even though the "purists" are concerned about the status of the standard languages, in the South African indigenous languages such as Northern Sotho, slang has contributed more to the expansion of their lexicon than the damage it has done. Even though most of the terms which are derived from slang are still excluded from our orthographies, the fact that these words are used more than their indigenous counterparts means that they are here to stay, and it will not be possible for the "purists" to remove them from the language of the people. The sooner lexicographers start recording these terms either in the standard dictionaries of the indigenous languages or in the specialized dictionaries dealing with loan words from urban slang, the better for the positive development of these languages.

4. Conclusion

It is obvious that our indigenous languages do need the terminology which is made available by urban slang to enrich their own vocabulary. This need is

demonstrated by the number of slang terms our people are using and the number of standard terms they are neglecting in their daily informal and formal communications. If some of these slang terms can be adopted as loan words in the vocabulary of the indigenous languages, such as Northern Sotho, these languages will be enriched by the new terms which are favoured and used by the people, thereby bridging the gap between the written and the spoken language. Of equal importance is the compilation of special dictionaries for these slang terms which will remain as treasuries for the future of our languages. This means that there is still a lot of research awaiting lexicographers of the South African indigenous languages with regard to the structuring of the slang lexicon in these languages.

Glossary

The following miniglossary shows a few examples of the slang terms commonly used in Northern Sotho by most of the urban and rural youth:

<i>Baizane</i>	A type of white-coloured fowl
<i>Bari (Baru)</i>	Stupid
<i>Blasa</i>	Play music/Drink liquor
<i>Blind (Blaend)</i>	Bad/Too much
<i>Bloma</i>	Sit/Stay
<i>Bombai</i>	An imitation (which is not up to standard)
<i>Bra</i>	Friend/ <i>Tsotsi</i> /Brother
<i>Daesa</i>	Sell
<i>Des</i>	Nice/Beautiful
<i>Dladleni</i>	Home
<i>Fasa</i>	Propose love to a girl/To be with a girlfriend
<i>Flopa</i>	Fail
<i>Jampas (Jumpers)</i>	At night/In the evening
<i>Jiwish</i>	Attire/Clothes
<i>Josi</i>	Johannesburg
<i>Juba</i>	Boyfriend
<i>Juleiti</i>	Work/Employment
<i>Khawada</i>	Disappoint/Terrorize
<i>Kuza</i>	Money
<i>Laiti</i>	Young boy
<i>Lani</i>	White man/lady
<i>Lethaima (Timer)</i>	Father
<i>Magegamba</i>	Money
<i>Mazuma (Mazumba)</i>	Money
<i>Ncanda</i>	Take
<i>Ngamola</i>	Tycoon/White man
<i>Skele</i>	School

<i>Smeka</i>	Money
<i>Smoko</i>	Trouble
<i>Spane</i>	Work/Employment
<i>Squeza</i>	Sister-/Brother-in-law
<i>Squila</i>	Refuse
<i>Stalala</i>	Johannesburg
<i>Topi</i>	Father
<i>Vaya</i>	Go/Leave
<i>Zak</i>	Money
<i>Zwakala</i>	Come

References

- Anderson, L. and P. Trudgill.** 1990. *Bad Language*. Oxford/Cambridge Mass.: Basil Blackwell.
- Bailey, R.** 1985. South African English Slang: Form, Function and Origins. *South African Journal of Linguistics* 3(1): 1-42.
- Coertze, R.D.** 1969. *Atteridgeville, 'n Stedelike Bantoeewoonbuurt*. Pretoria: University of Pretoria.
- Coplan, D.B.** 1985. *In Township Tonight!: South Africa's Black City Music and Theatre*. Johannesburg: Ravan Press.
- De Klerk, V.** 1991. What's the Current Slang? *English Usage in Southern Africa* 22: 68-82.
- Mansour, G.** 1980. The Dynamics of Multilingualism: The Case of Senegal. *Journal of Multilingual and Multicultural Development* 1(4): 273-293.
- Mfusi, M.J.H.** 1992. Soweto Zulu Slang: A Sociolinguistic Study of an Urban Vernacular in Soweto. *English Usage in Southern Africa* 23: 39-83.
- Mojela, V.M.** 1991. *Semantic Changes Accompanying Loan-words in the Northern Sotho Lexicon*. Unpublished M.A. Thesis. Pretoria: Vista University.
- Mojela, V.M.** 1999. *Prestige Terminology and its Consequences in the Development of Northern Sotho Vocabulary*. Unpublished D.Litt. Thesis. Pretoria: University of South Africa.
- Stern, G.** 1931. *Meaning and Change of Meaning*. Gothenburg: Goteborgs Hogskolas Arrskrift.
- Swanepoel, J.J.** 1978. Urban Slang in Compositions. *Educamus*, December 1978.

Le corpus lexicographique dans les langues à tradition orale: le cas du dialecte fang-mekè*

Yolande Nzang-Bié, *Département des Sciences du Langage, et Groupe de Recherches en Langues et Cultures Orales (GRELACO), Université Omar Bongo, Libreville, Gabon (yolnzang@yahoo.fr)*

Résumé: Les corpus sont à la base de la plupart des recherches en linguistique et particulièrement lexicographique. La compilation d'un corpus est une activité spécialisée dont dépend le résultat de la recherche en question. Le sujet de cet article est la compilation du corpus lexicographique dans les langues à tradition orale, et exige une démarche différente de celle ayant une longue tradition écrite. De ce fait, ces dernières disposent d'une importante documentation pouvant servir comme base pour de nombreux sujets de recherche. L'auteur propose comme approche une analyse qui permettrait de mieux rendre compte des spécificités lexicales et sémantiques des langues à tradition orale.

Par le truchement de la production orale libre, l'auteur base ses hypothèses de recherche sur une expérience en dialecte fang-mekè, une variante linguistique localisée au Gabon. Les résultats permettent de mettre l'accent sur deux données essentielles du processus de compilation dans les langues à tradition orale: les informateurs et la représentativité du corpus. Cette dernière, qui doit s'exprimer à travers des champs lexicaux diversifiés mais également équilibrés, permettrait d'élaborer des dictionnaires dans lesquels les locuteurs, qui en sont les premiers utilisateurs, doivent se reconnaître.

Mots-clés: CORPUS, LEXICOGRAPHIE, LANGUES À TRADITION ORALE, LANGUES À TRADITION ÉCRITE, INFORMATEURS, EXHAUSTIVITÉ, REPRÉSENTATIVITÉ, CHAMPS LEXICAUX, ORALITÉ, ÉCRITURE, MÉTHODE, DIALECTE FANG-MEKÈ, CORPUS ÉQUILIBRÉ.

Abstract: The Lexicographic Corpus in Languages with an Oral Tradition: The Case of the Dialect Fang-Mekè. Corpora form the basis of most linguistic and especially lexicographic research. The compilation of a corpus is a specialised activity on which depends the result of the research to be undertaken. The subject of this article is the compilation of a lexicographic corpus in languages with an oral tradition which requires a different approach from those having a long written tradition. Because of this fact, the latter have an important documentation which can serve as basis for many subjects of research. The author proposes an analytic

* Cet article a été réalisé pendant que l'auteur participait à un programme d'échange pour l'année académique 2000-2001 au Département d'afrikaans et de néerlandais à l'Université de Stellenbosch en Afrique du Sud. L'article doit beaucoup aux discussions et aux enseignements que l'auteur a eus avec professeur R.H. Gouws dont les points de vue dans le domaine lexicographique lui ont été bénéfiques et enrichissants.

approach which will allow better to account for lexical and semantic particularities of languages with an oral tradition.

Through the method of free oral production, the author bases her hypotheses of research on an experiment in the Fang-Mekè dialect, a language variant of Gabon. The results make it possible to stress two essential factors in the compilation process of languages with an oral tradition: the informants and the representativeness of the corpus. The latter which must encompass varied but also balanced lexical fields, would allow the planning of dictionaries in which the speakers who are the first users, can recognize themselves.

Keywords: CORPUS, LEXICOGRAPHY, LANGUAGES WITH AN ORAL TRADITION, LANGUAGES WITH A WRITTEN TRADITION, SPEAKERS, EXHAUSTIVENESS, REPRESENTATIVENESS, LEXICAL FIELDS, ORALITY, WRITING, METHOD, DIALECT FANG-MEKÈ, BALANCED CORPUS.

1. Introduction

Dans le domaine de la recherche linguistique, il existe une littérature abondante et variée sur la thématique du corpus et son importance capitale pour la plupart des recherches linguistiques. Aussi, notre travail n'a pas pour objectif de théoriser sur ce sujet où beaucoup a déjà été dit. Bien que l'on constate une évolution rapide des méthodes pour le perfectionnement de ce domaine, je pense comme Kennedy (1998) que la notion de ce que constitue un corpus valide reste un sujet à controverse.

La lexicographie, plus que les autres domaines des sciences du langage, n'échappe pas à ce principe. Ma réflexion a pour principal objectif de montrer le caractère primordial, essentiel, représentatif, qualitatif et quantitatif du corpus lexicographique, et davantage pour la spécificité du corpus lexicographique dans les langues à tradition orale où il existe, manifestement, peu de documents écrits pouvant constituer une base de données primaires pour l'élaboration des dictionnaires. En effet, de nombreuses références sur cette question sont basées sur des langues ayant une longue tradition écrite et bénéficiant de ce fait d'un fond documentaire divers, pouvant servir de référence à la recherche dans différents domaines de la linguistique, et particulièrement à la compilation des corpus lexicographiques. Il me semble que les langues à tradition orale, qui sont pour le grand nombre des langues africaines, devraient être abordées selon une approche relativement différente de celles à tradition écrite. Ma contribution se veut plus pratique que théorique. Elle me permet de partager mon expérience du corpus lexicographique dans les langues à tradition orale, à travers le dialecte fang-mekè, une variante linguistique localisée au Gabon, avec d'autres chercheurs que cette question intéresse. Toutefois, avant d'examiner les données du corpus lexicographique dans les langues à tradition orale, il est nécessaire de rappeler quelques principes inhérents au corpus linguistique en général et lexicographique en particulier. Aussi, ma réflexion va s'articuler sur les points suivants:

- le corpus lexicographique et ses objectifs;
- le corpus dans les langues à tradition orale;
- l'expérience du dialecte mekè; et
- la planification rationnelle du corpus lexicographique dans les langues à tradition orale.

2. Le corpus lexicographique et ses objectifs

En linguistique, de manière générale, le corpus est présenté comme la première étape dans la plupart des projets de recherche. Il constitue la source essentielle d'informations. De nombreux chercheurs le présente comme un ensemble de textes écrits ou transcrits qui sert de point de départ à la plupart des analyses ou descriptions dans une langue donnée. Cette approche est aussi celle de Dubois Charlier (1997: 312), pour qui le corpus est "une base de données textuelles réunissant des textes écrits et oraux de provenances variées et de registres divers: livres, journaux, débats etc.", ou encore celle de Francis (1992: 17) "a collection of texts assumed to be representative of a given language, dialect, or other subset of a language, to be used for linguistic analysis". Il constitue pour Kennedy (1998: 4) "an empirical basis not only for identifying the elements and structural patterns which make up the systems we use in a language, but also for mapping out our use of these systems".

En principe, le corpus a une fonction "représentative". Il doit par conséquent être représentatif du répertoire oral, grammatical et écrit maîtrisé par les locuteurs natifs performants dans leurs divers usages de la langue, et permettre de rendre compte de la totalité des caractéristiques d'une langue donnée. À cet effet, le corpus doit être l'objectif d'une étude exhaustive et non sélective; c'est ce que Quirk (1992: 458) désigne par "the vital principle of total accountability". Par ailleurs, le corpus doit aller au-delà des missions ou buts qui lui sont initialement assignés, comme le dit justement Svartvik (1992: 9): "corpora are not simply language samples that may provide useful illustrative examples but a theoretical resource."

En lexicographie, le corpus ne déroge pas aux principes observés plus haut, mais il tient sa particularité du caractère de l'exhaustivité, ce qu'on n'observe pas, par exemple, dans la description linguistique. Cette dernière utilise des corpus questionnaires¹ qui remplissent parfaitement les fonctions qui leurs sont assignées, même si l'on relève pour des descriptions plus exhaustives des insuffisances. Celles-ci sont supplées par diverses productions orales ou écrites libres qui ont pour but de vérifier la fonctionnalité des règles élaborées en contexte naturel et variable.

La lexicographie forme la structure de base d'une langue. Par ce fait, les données qui proviennent de la langue doivent refléter une représentativité et une exhaustivité sans faille, et l'on insistera jamais assez sur ce principe de

représentativité et d'exhaustivité qui caractérise le corpus lexicographique. C'est pour cette raison qu'un tel corpus doit prendre en compte les différents domaines d'application dont fait usage la langue dans ses multiples actes de parole, le résultat étant tributaire des données compilées. Cette représentativité du corpus permet de découvrir continuellement comment les langues se comportent dans toutes leurs diversités. La phase de compilation du corpus qui est l'acquisition des données matérielles pour tout projet de dictionnaire, doit être, comme l'écrit Gouws (2001) "a highly skilled activity". Le corpus lexicographique doit être en mesure de fournir en théorie tous les emplois possibles avec leur fréquence, leur dispersion, leur contexte (Dubois Charlier 1997: 312).

Il est manifeste que le corpus n'est pas une fin en soi, mais il est une source évidente de données dans un processus d'élaboration d'un dictionnaire. Toutefois, l'on ne doit pas perdre de vue que la compilation d'un corpus dépend du type de dictionnaire que l'on veut élaborer. À mon avis, tout projet de compilation d'un corpus lexicographique doit pouvoir répondre à quelques interrogations, comme

- à qui est adressé le projet du dictionnaire;
- quel type de dictionnaire veut-on élaborer; et
- quels sont les objectifs visés.

En effet, il me semble erroné qu'un projet de dictionnaire axé sur la thématique des techniques de chasse, par exemple, s'étende démesurément sur des domaines de la langue qui n'ont rien à voir de près ou de loin avec le sujet en question. Aussi, le terme de la "représentativité" doit être clairement cerné, et ne doit pas prêter à confusion. Je suis d'avis qu'un corpus, quel que soit le but visé, ne doit pas être sélectif dans le sens d'être limité à un simple questionnaire. C'est pour cette raison que j'adhère à l'approche qui stipule de "faire des provisions" dans la planification de tout corpus lexicographique. J'entends par "faire des provisions", le principe de la collecte exhaustive des données qui permettrait, par la suite, d'extraire du corpus les données dont on a besoin pour la thématique concernée. Cette approche que je tiens du professeur R.H. Gouws présente l'avantage que le corpus puisse être utilisé pour d'autres sujets. C'est dans ce contexte que l'on peut parler du caractère multifonctionnel d'un corpus.

3. Le corpus lexicographique dans les langues à tradition orale

De l'avis de nombreux chercheurs, le corpus lexicographique doit s'orienter sur deux axes parallèles: les sources écrites et les sources orales. Le constat que l'on est amené à faire, c'est qu'une place prioritaire est accordée aux sources écrites. Cette approche est celle qu'on retrouve dans les propos de Kennedy (1998: 7): "In this case of corpus based research, the evidence is derived directly from

texts"; mais aussi d'autres sources tel que le Bureau du WAT (1999: 4), qui écrit: "The most common source of language material is the written text. Because words are available in their full context, the texts are readily accessible and can be utilized immediately. Written sources include the following: published sources such as books, journals and magazines, newspapers, reports and proceedings; and unpublished sources (ephemera) such as advertising material, financial reports, pamphlets, minutes and letters. Included in the published written sources are other dictionaries and word lists (technical as well as general) which the editor may consult during the compilation process and which may also serve to check whether all important words and terms have been included"; et plus loin encore, on peut lire: "it is advisable to start a material collection with written materials as it is more easily collected and processed". Cette approche qui s'applique parfaitement aux langues ayant une longue tradition écrite, ne s'adapte pas, à mon avis, aux langues à tradition orale pour lesquelles l'essentiel de la communication dans ses divers usages est centré sur l'oralité et où, dans la plupart des cas, il n'existe pas ou très peu de données écrites. Bon nombre de langues du Gabon pour lesquelles il n'existe aucune documentation écrite peuvent l'illustrer et dans les cas où il y aurait des traces, celles-ci sont limitées à quelques fascicules dont la graphie est souvent douteuse et ne reflète pas la situation de la prononciation réelle. Il me semble que les sources écrites dont disposent actuellement certaines langues à tradition orale, sans pour autant être négligées, doivent faire l'objet de vérifications par de nouvelles enquêtes, en vue de leur actualisation, afin d'aboutir à un travail qualitatif et référentiel.

Une autre catégorie de sources orales vers laquelle l'on devrait également s'orienter, sont de nombreuses bandes sonores entreposées dans des musées et les institutions de communication (radios, télévisions). Celles-ci sont de nombreuses sources d'informations dans des domaines divers, mais elles sont généralement inexploitées par manque de transcriptions. Ce problème de transcription du matériel oral disponible mais non-exploité, représente l'une des difficultés que l'on rencontre dans la compilation du corpus dans les langues à tradition orale.

Aussi, fort de ce qui précède, j'encourage fortement pour les langues à tradition orale l'approche inverse plus adaptée et rationnelle qui considérerait comme

- les sources primaires, les données orales (qu'elles soient à compiler ou à transcrire); et
- les sources secondaires, les données écrites.

Dans la première catégorie de données, la langue est fonctionnelle, dynamique et pleinement efficace. Cette approche permet de saisir la langue dans son contexte naturel et de capter les domaines d'usage qui constituent l'activité culturelle du peuple, ce qui est l'objet de notre étude. Dans cette activité de compi-

lation du corpus, il est question de la compétence aussi bien que de la performance des locuteurs. Les sources écrites, qui sont également un aspect capital, doivent suppléer les sources orales. Elles contiennent indéniablement de nombreuses informations. C'est dans ce contexte bi-directionnel que les principes de représentativité, de diversité et d'exhaustivité tiennent toute leur importance.

L'approche du questionnaire comme élément de base pour la compilation d'un corpus lexicographique a perdu son importance primordiale. Limité par ce fait, un questionnaire, aussi exhaustif soit-il, ne peut prendre en compte l'essentiel des données d'un domaine spécifique, tout en restant subjectif du fait qu'un questionnaire représentatif pour un chercheur, ne l'est pas forcément pour un autre. Toutefois, je pense que le "corpus questionnaire" peut servir de point de départ dans le contexte d'un corpus lexicographique thématique ou dans un domaine spécialisé qui exige une terminologie spécifique. Mais il doit, nécessairement être renforcé par ce que j'ai appelé "les données orales à production libre". Elles présentent l'avantage d'être produites de manière spontanée et dans leur contexte naturel. Le questionnaire peut également servir de point d'appui dans la recherche des mots dont les définitions ou les portées sémantiques ne figurent pas explicitement dans le corpus global.

De tout temps, la linguistique a fonctionné à l'aide de "corpus questionnaire". Si cette approche reste encore valable pour certaines disciplines de la linguistique, et de la lexicographie des langues à tradition orale, la compilation des données est un processus dynamique et continu. Toutefois, elle doit prendre en compte l'inquiétude formulée par le WAT (1999: 6), qui est aussi mienne: "be aware of an overabundance of language material or inadequate filtering methods, which may slow down the process of manuscript making."

4. L'expérience dans le dialecte fang-mekè

Mon expérience sur le corpus en dialecte fang-mekè est à l'origine de cette réflexion sur le corpus lexicographique dans les langues à tradition orale. Elle a été motivée par une formation en lexicographie à l'Université de Stellenbosch en Afrique du Sud. C'est en prévision de cette formation que j'ai procédé à une enquête libre auprès d'un groupe de locuteurs natifs fang, du dialecte mekè, sans trop lui donner un objectif précis.

Pour la localisation du groupe-cible, l'on retiendra brièvement que le dialecte mekè est l'une des composantes du groupe linguistique fang². Il est localisé, selon diverses sources (entre autres Mba-Nkoghe 1981, Voltz 1990, Ondo-Mebiame 1992) à Libreville et Cocobeach dans la province de l'Estuaire, mais aussi à Mitzic dans la province du Woleu-Ntem. Démographiquement, je ne dispose pas de données fiables sur ce groupe de locuteurs.

Sur le plan documentaire, le dialecte mekè présente très peu de documents écrits. Actuellement, hormis quelques descriptions linguistiques non exhaustives et des fascicules laissés par des missionnaires, le document le plus impor-

tant écrit en dialecte mekè est la traduction du nouveau testament biblique (1962). Cet ouvrage contient également des cantiques, des prières et des oraisons en dialecte mekè. Cette source écrite est d'une grande valeur quand on sait que la Bible est considérée depuis le 18^e siècle comme l'une des premières pièces significatives dans la recherche des corpus (Kennedy 1998).

L'expérience menée sur le dialecte mekè a été faite à base d'un matériel très simple, à savoir un enregistreur et des cassettes audio, ce qui est une des techniques recommandées (Kennedy 1998: 20).

4.1 La méthode utilisée

J'ai procédé à une enquête à "production libre" auprès d'un groupe de locuteurs natifs performants, ayant reçu leur éducation dans la langue maternelle et n'utilisant le français qu'en cas de besoin spécifique. Les entretiens se sont déroulés de manière libre et collégiale entre les différents locuteurs. Chacun était libre de développer sa pensée, mais il pouvait à tout moment être interrompu par un autre locuteur pour compléter l'information, expliciter les propos du prédécesseur et même les rectifier, ajouter ou introduire de nouvelles données. Cette méthode a fait que j'ai assisté à un échange actif et fructueux. L'un des constats que j'ai eu à faire est la spontanéité dans les réactions des différents locuteurs.

Après avoir expliqué à mes interlocuteurs quelles étaient mes attentes, j'ai introduit mon enquête sur la thématique de l'histoire des Fang. À la suite de ce thème, d'autres se sont spontanément développés sans que cela ait été planifié et ce dans une logique déconcertante. C'est ainsi que l'on pouvait passer, par exemple, du thème du mariage à celui de la mort ou encore de la nature, sans qu'il y ait rupture dans le développement des idées chez mes différents interlocuteurs. Tout se présentait comme si tous les thèmes abordés "inconsciemment" étaient le processus logique d'une même chaîne et dont on passait, sans rupture, d'un maillon à un autre.

Tout au long des échanges, j'ai mené les débats de sorte que les différents interlocuteurs mais aussi l'enquêteur pouvaient, sans dérouter l'intervenant, poser des questions pour d'amples informations. En exemple, les locuteurs hommes utilisaient énormément de proverbes pour illustrer leurs propos. Aussi, ai-je à un moment interrompu mon locuteur pour avoir des explications sur le proverbe. J'ai eu droit à une longue définition du proverbe qui a été suivi par une série d'autres proverbes, énoncés et explicités par mes locuteurs. De fait, les explications sur un point pouvaient conduire à un autre thème et les interruptions servaient à faire des remarques sur des points précis.

À propos de ces interruptions, j'ai relevé la fréquence des expressions telles que: "je peux ajouter à ce qu'il vient de dire que ...; c'est aussi ce qu'on appelle en fang³; ce que tu dis n'est pas complet ...; tu as omis de préciser que ...; ton explication n'est pas juste ...; tel type de proverbe peut aussi illustrer ton propos ...; ce mot n'est pas de notre dialecte nous l'avons pris chez ...; est ce que quelqu'un à quelque chose à ajouter ...; si je me trompe que untel ou untel me

reprenne ...; je ne pense pas tout connaître sur le sujet, mais je peux dire que ..." et bien d'autres encore.

J'ai noté que mes locuteurs n'étaient pas omniscients. Dans certains domaines spécifiques, les informateurs me renvoyaient vers quelqu'un de plus compétent avec des expressions telles que "je ne connais pas grand chose dans ce domaine, mais tu peux aller voir untel qui peut très bien t'en parler; cependant je peux dire que ...", ce qui me laisse supposer que tout locuteur compétent et performant n'est jamais totalement ignorant des concepts relatifs à son milieu.

Il est significatif de souligner que pendant les entretiens, les locuteurs pouvaient se lever librement, vaquer à quelques occupations pas très prenantes et revenir rejoindre le groupe sans qu'il y ait rupture dans le déroulement des entretiens. Cette méthode présente ainsi l'avantage de ne pas être contraignante, aussi bien pour les enquêteurs que pour les informateurs.

À la fin du premier entretien, j'ai constaté que mes informateurs, réticents au départ, avaient été très heureux de me fournir ces informations. Ils donnaient l'impression d'avoir contribué à quelque chose de constructif et pérenne pour leur langue. Aussi, étaient-ils entièrement disposés pour la suite de l'enquête.

L'expérience que je viens de relater n'est certainement pas celle d'autres chercheurs qui se sont intéressés à cette problématique, mais je pense que si elle est planifiée, cette méthode de travail dans la compilation du corpus lexicographique dans les langues à tradition orale, devrait mener à des résultats probants.

4.2 Les résultats

Pour une période assez courte de quatre jours, j'ai pu enregistrer quatre cassettes audio d'une durée de 1h30mn chacune. L'objectif de ces enregistrements n'est pas de faire le traitement lexicographique⁴ en vue d'établir des listes de fréquence, de concordance, de fréquence d'un niveau de langue, moins encore l'élaboration du dictionnaire. Il vise essentiellement les divers champs lexicaux ou domaines de recherche que ma méthode et mon expérience m'ont permis de mettre en évidence, et qui constituent, à mon avis, un échantillon des données lexicales pour beaucoup de langues africaines à tradition orale.

Les champs lexicaux qui ont pu être extraits du corpus sont représentés dans le tableau 1.

Tableau 1: Champs lexicaux et sous-thèmes du corpus

Thèmes représentant des termes superordonnés	Hyponymes ou sous-thèmes
<ul style="list-style-type: none"> -- histoire -- les récits⁵ 	contes, proverbes, récits épiques, devinettes

<ul style="list-style-type: none"> -- le mariage -- éducations des enfants (jeunes filles et garçons) -- la palabre -- la mort -- la procréation -- l'adultère -- le veuvage -- la réinsertion sociale -- la sémantique des noms propres -- le maquillage (extrait les termes relatifs à la beauté) -- le décompte du temps -- la destinée -- la circoncision -- les travaux champêtres -- rituels initiatiques -- la pêche -- la chasse -- les constructions -- l'occupation de l'espace -- les astres -- l'artisanat -- l'éducation sexuelle -- le corps humain -- les aliments -- l'écosystème -- les couleurs 	<p>le choix du conjoint, les présentations, les fiançailles, la dot, mariage par rapt, mariage par consensus</p> <p>l'annonce, l'enterrement, le deuil, la levée de terre, le port du deuil, le retrait de deuil, les funérailles</p> <p>la grossesse, l'accouchement, les matrones, le bain rituel</p> <p>heures, jours, semaines, mois, saisons, années</p> <p>les techniques, les instruments, la pêche par les hommes, la pêche par les femmes, les périodes</p> <p>les techniques, les instruments, les types, les périodes, les techniques de pièges</p> <p>les techniques, les matériaux, les types</p>
---	--

J'ai noté, et cela est peut-être une particularité de mon groupe d'informateurs, qu'ils évitaient subtilement de parler des questions relatives aux sexes ou aux organes génitaux⁶, même lorsque les échanges devaient les y conduire naturellement. J'ai personnellement évité d'insister sur ce point pour ne pas heurter la sensibilité des informateurs. Je pense que si la tendance se confirme dans d'autres expériences, cet aspect, dans l'enquête, peut faire l'objet d'un traitement particulier à définir par les enquêteurs.

Cet échantillon de corpus qui n'est ni exhaustif, ni représentatif est, à mon avis, une ébauche, une esquisse de ce qui pourrait constituer un corpus lexicographique des langues à tradition orale. Il a pu faire ressortir quelques thèmes saillants, ce qui dénote que menée sur une échelle représentative, cette méthode peut conduire à une vue globale et non moins précise des champs lexicaux récurrents et fonctionnels permettant d'aboutir à un corpus équilibré.

5. Pour une planification du corpus lexicographique rationnelle

L'activité de compilation d'un corpus lexicographique, dans les langues à tradition orale, doit prendre en compte deux données essentielles: les informateurs et la représentativité du corpus équilibré.

5.1 Les informateurs

Les informateurs locuteurs natifs constituent le pilier central pour toute compilation d'un corpus lexicographique. Dans le cadre d'un corpus lexicographique, le nombre d'informateurs doit varier en fonction de l'ampleur du projet à mener. Dans tous les cas, de par son ampleur et sa diversité, aucun corpus lexicographique, aussi restreint soit-il, ne peut être compilé par un seul informateur ou par un seul groupe d'informateurs. Le nombre doit aller croissant selon les besoins du projet mené. En effet, l'univers linguistique est extraordinairement vaste et complexe, et aucun humain n'a la prétention de le maîtriser. À mon avis, et d'accord avec Chafe (1992: 80) "it is impossible for any individual to understand more than a tiny segment of it. Different individuals understand different tiny segments — some perhaps slightly more than others — but no human understanding can go very far. Understanding may be enlarged when it is distributed over groups of individuals". Cette approche de Chafe plaide en faveur de la diversité dans le choix des informateurs.

Le dictionnaire reflète la compétence, la dynamique linguistique d'une société donnée. Il véhicule l'activité culturelle d'un peuple et doit, par conséquent, refléter les différents domaines d'activité dont fait usage une société donnée. En effet, toute élaboration d'un dictionnaire doit être adaptée aux besoins de la société qui en est le premier utilisateur. Celle-ci doit se reconnaître dans le dictionnaire constitué à son usage. À ce titre, le corpus, à travers ses informateurs, doit être représentatif de la langue dans ses différents domaines comme un tout. Toutefois, l'on doit garder en esprit qu'aussi exhaustives qu'elles soient, les données d'un corpus ne seront jamais qu'un minuscule échantillon de toute l'activité linguistique reçue ou produite par les locuteurs d'une langue donnée. Ce qui est essentiel, à mon avis, c'est que cet échantillon soit représentatif de l'activité humaine à travers ses locuteurs. Je partage à cet effet la réflexion de Sinclair (1991: 20) qui écrit que "corpora of finite size were inherently deficient because any corpus is such a tiny sample of language".

Les informateurs performants locuteurs natifs de la langue vers laquelle le

projet est mené doivent être majoritairement des notables, des personnes d'âges mûrs, des spécialistes de domaines spécifiques des deux sexes usant le moins possible une langue importée, à l'instar du français; car comme nous l'avons déjà dit, l'activité de compilation d'un corpus est une question de performance, mais aussi de compétence.

Les informateurs doivent provenir principalement de deux milieux: le milieu rural et le milieu urbain. Cette répartition, source d'enrichissement pour le corpus, est importante, car elle permet de prendre en compte les différents milieux naturels dans lesquels vivent quotidiennement les informateurs, et qui sont susceptibles d'influencer, d'une manière ou d'une autre, leurs performances linguistiques. Aussi, le choix de locuteurs dans un univers limité peut fausser la représentativité du corpus et déséquilibrer les données.

Les informateurs doivent travailler dans un cadre naturel, non contraignant et le plus souvent mis en situation de conversations, d'entretiens, à travers lesquels les différents locuteurs s'expriment librement. L'une des approches que je propose est la participation des enquêteurs à des cérémonies ou des rencontres en milieu et en contexte naturels. Mais on peut aussi organiser, comme c'est généralement le cas, des séances "artificielles"⁷ de rencontres. On devrait éviter, le plus possible, de faire débiter sans motivation des listes de mots aux informateurs, ce qui par expérience finit par lasser ceux-ci qui peuvent aller jusqu'à demander l'arrêt de l'enquête. Toutefois l'on doit veiller à éviter des situations cacophoniques où l'on ne sait plus qui dit quoi, de même qu'on doit veiller à créer un climat de confiance d'une part entre les enquêteurs et les informateurs, et d'autre part entre les différents informateurs. Cette atmosphère est nécessaire au bon déroulement de la compilation du corpus.

5.2 La représentativité du corpus équilibré

Ce travail n'a pas la prétention d'aborder la grande question du contenu du corpus lexicographique en ce qui concerne le nombre de mots, les différents types de mots que doit contenir un corpus ou encore le nombre de mots d'un exemple. De nombreuses références disponibles en font état, et il ressort qu'un bon corpus se chiffre en dizaines de millions de mots (Kennedy 1998). Je retiens, par exemple, que Sinclair (1991: 20) a suggéré que "10–20 million words might constitute a useful small general corpus".

Ce travail se propose de mettre l'accent sur la représentativité du corpus lexicographique, en termes de diversité de champs lexicaux dans les langues à tradition orale.

Les champs lexicaux sur lesquelles doivent se baser une compilation de corpus dans une société donnée ne sont pas facilement décelables. Cette tâche est très difficile quand on sait que, même sous le contrôle de toute une communauté, il est difficile, voire impossible de recenser les différents domaines d'emploi du langage. Kennedy (1998: 72) a proposé le tableau ci-après pour les catégories de textes oraux ("categories of spoken texts").

Tableau 2: Catégorie des textes oraux (selon Kennedy 1998)

— Monologue
— formal
- written to be read aloud
- prepared but unscripted public speech
— less formal
- academic lectures
- commentaries on public occasions
- sports commentaries
- demonstrations
— Dialogue
— face-to-face dialogue
— public discussion (e.g. on radio)
— business or professional transactions, e.g.
- client and professional
- workplace interaction between colleagues
- commercial (sales) transaction
— informal
- within family
- between friends
— telephone dialogue
- between interlocutors known to each other
- between interlocutors unknown to each other
— structured interaction
- interview
formal, e.g. arts programme
less formal, e.g. with survivors and witnesses to events such as accidents
- debate
- committee meeting

Cette liste permet de voir que bien de sujets sont facilement applicables aux langues à tradition orale. Toutefois à cette liste, j'ajouterai les sujets suivants contenus dans le tableau 3.

Tableau 3: Catégories des données orales (en complément du tableau 2, selon l'auteur)

— Monologue
— émission radio ou télévisée
- présentation des nouvelles, spot publicitaires, annonces publiques, annonces culturelles, jeux éducatifs
- musique
- informations documentaires musique
- films ⁸
- théâtre ⁹
— Dialogue
— production libre (permet d'accéder à divers thèmes) ¹⁰

- participation à diverses cérémonies
 - rituelles
 - funéraires (deuil, levée de terre, veuvage, retrait de deuil, etc.)
 - folkloriques
 - initiatiques
 - religieuses
 - règlement de palabre
 - tribunal traditionnel
 - maquillage (termes relatifs à la beauté)
 - soins de santé
 - purification
 - divination, voyance
 - mariage
 - fêtes
 - pêche, chasse, récolte, agriculture
 - naissances
 - installation de notables
 - transactions de types divers
 - politiques
 - économiques
 - culturelles (production de différents genres littéraires, jeux collectifs)
 - détente

Cette liste a une fonction purement indicative et référentielle, car comme l'écrit Kennedy (1998: 72) "taken over a whole community, it is simply impossible to get a balance which represent actual use of the language". Les dictionnaires de langues africaines doivent refléter des réalités africaines. Aussi, cette liste fournit des pistes de recherches pouvant aider dans la compilation des thèmes relatifs au corpus lexicographique, en mettant l'accent sur des facettes courantes des langues à tradition orale.

J'ai insisté, tout au long de ce travail, sur le défi de la représentativité du corpus lexicographique. Un autre défi est celui de la notion de corpus équilibré. Cette notion n'est pas considérée dans le texte, tenant compte des nombreuses sources, et la proportion des données équitables entre les sources écrites et les sources orales ("60% spoken texts and 40% written texts", Kennedy 1998: 62). Elle est retenue selon la répartition équitaine des données à l'intérieur d'un même sujet qu'il soit spécifique ou général. Le but du corpus est d'être fonctionnel. Aussi, un projet de compilation d'un dictionnaire général doit prendre en compte, non seulement tous les domaines possibles repérables dans la langue, mais aussi des aspects identifiables à l'intérieur de chaque domaine. Lorsque le projet est spécifique, les différents aspects, même ceux qui paraissent anodins, doivent être exploités. Cela permet d'aboutir à un corpus fiable.

6. Conclusion

J'ai examiné dans ce travail la problématique du corpus dans les langues à tradition orale pour lesquelles l'essentiel de la communication est basé sur la pro-

duction orale. Je reste consciente que ce type d'enquête est dur et coûteux, du fait qu'il nécessite des moyens matériels, financiers et surtout d'énormes ressources humaines.

L'activité de l'élaboration d'un corpus lexicographique n'est pas une activité nouvelle, mais il semble aujourd'hui encore que les langues à tradition orale n'ont pas réellement trouvé leur place dans les banques de données existant sur ce sujet. Aussi, dans ce siècle de mondialisation, il est plus qu'imprévisible que les lexicographes et particulièrement les lexicographes africains donnent aux langues à tradition orale une place référentielle dans le traitement du langage par le truchement de la compilation des corpus lexicographiques. Pour que cette tâche soit effective, je pense comme Gouws (2001) que les unités lexicographiques doivent prendre les dispositions suffisantes dans leur planification pour cet aspect important de leur entreprise lexicographique. Ceci ne revient pas à jeter un discrédit sur ce qui a déjà été fait, mais d'améliorer les données disponibles afin de constituer des bases de données lexicales et thématiques considérables pour les langues à tradition orale. Il ne s'agit pas de "réinventer la roue", mais d'attirer l'attention de nombreux lexicographes africains sur une problématique réelle. Botha (2000) parle d'une coopération franche entre les lexicographes, mais aussi entre les lexicographes et les utilisateurs des dictionnaires afin d'éliminer les obstacles entre les utilisateurs et les concepteurs. J'ai proposé des méthodes de recherche qui peuvent amener à mieux aborder ce problème. L'accent a été mis sur des points prioritaires dans la compilation du corpus dans les langues à tradition orale.

Ces quelques pages ont posé une réflexion dans la perspective du développement rationnel et fonctionnel des langues à tradition orale par l'usage de cet outil indispensable qu'est le dictionnaire. Je pense que ces langues ont besoin de compiler, pour l'instant, des données fonctionnelles menant à l'élaboration des dictionnaires, mais permettant de constituer des bases de données qui aideraient à produire, à long terme, des travaux aussi importants que les dictionnaires monolingues, des encyclopédies et bien d'autres ouvrages de référence dans des domaines divers. Aussi, cette réflexion de Botha (2000) pourrait trouver toute sa signification: "the much talked about African Renaissance will, through lexicographic co-operation, be characterised by emergence of dictionaries for all African languages."

Notes

1. Je désigne par corpus questionnaire, des listes de questions établies pour l'élaboration des systèmes phonologiques, morphologiques et syntaxiques. L'on parle ainsi du *Questionnaire d'enquête linguistique* de J. Greenberg et W.E. Welmers (traduit par Jean Doneux), Université de Dakar, 1967.
2. Le groupe linguistique fang est classé chez Guthrie (1948), où il porte le sigle A75. Il est constitué, en plus du mekè, de cinq autres dialectes que sont: l'atsi, le dzaman, le mveny, le ntumu et l'okak.

3. Mes locuteurs parlent invariablement le fang (qui est la désignation du groupe linguistique) ou le parler mekè (nkobe wa meke (e)ne: notre parler meke d'ici).
4. Je consacre cet autre aspect à une recherche qui est en cours et qui permettra de faire paraître, avec les données disponibles élargies, un dictionnaire parallèle du dialecte mekè.
5. Les divers récits sont constitués de thèmes divers. J'ai relevé ceux relatifs à la jalousie, la cupidité, la haine, la méchanceté, la sorcellerie, l'amour, l'amitié, la confiance, la violence, l'autorité, le vampirisme, la palabre, la faune, la flore, les interdits, la nature, le respect, la patience, la ruse, l'adresse, la force, l'épreuve, l'initiation, les classes d'âge, les jeux, la notion du temps, la tradition.
6. Chez le peuple Fang, les interdictions relatives aux sexes sont culturelles et relativement taboues en public.
7. Je désigne par séances artificielles, des séances d'enregistrements qui ne se déroulent pas dans un contexte naturel.
8. et 9. sont classés dans cette catégorie du fait que l'enquêteur n'a pas accès aux informateurs.
10. Un aperçu de cet aspect à été donné à la section 4.

Bibliographie

- Atkins, B.T.S. 1994. A Corpus-based Dictionary. Grundy, Marie-Hélène Corréard-Valerie (Éd.). 1994: xix-xxxii.
- Atkins, Sue et Jeremy Clear. 1992. Corpus Design Criteria. *Literary and Linguistic Computing* 7: 1-16.
- Botha, Willem. 2000. We All Stand Together, Don't We? — African Renaissance through Dictionaries. Heid, Ulrich et al. (Éds.). 2000: 163-168.
- Bureau of the WAT. 1999. Language Material Collection and the Compilation of a Data Base. Notes de cours non publiées.
- Chafe, Wallace. 1992. The Importance of Corpus Linguistics to Understanding the Nature of Language. Svartvik, Jan (Éd.). 1992: 79-96.
- Clear, Jeremy. 1996. Technical Implications of Multilingual Corpus Lexicography. *International Journal of Lexicography* 9(3): 265-273.
- Crowdy, Stephen. 1993. Spoken Corpus Design. *Literary and Linguistic Computing* 8: 259-265.
- Dubois Charlier, Françoise. 1997. Compte rendu de l'Oxford-Hachette French Dictionary. *International Journal of Lexicography* 10(4): 311-329.
- Francis, W. Nelson. 1992. Language Corpora B.C. Svartvik, Jan (Éd.). 1992: 17-32.
- Gouws, R.H. 2001. Lexicographic Training: Approaches and Topics. Emejulu, James Duplessis (Éd.). 2001. *Éléments de lexicographie gabonaise. Tome I: 58-94*. New York: Jimacs-Hillman.
- Guthrie, Malcolm. 1948. *The Classification of the Bantu Languages*. Londres: Oxford University Press.
- Grundy, Marie-Hélène Corréard-Valerie (Éd.). 1994. *Oxford-Hachette French Dictionary*. Oxford: Hachette Oxford University Press.
- Heid, Ulrich et al. (Éds.). 2000. *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000. Stuttgart, Germany, August 8th-12th 2000*. Stuttgart: Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.
- Kennedy, Graeme. 1998. *An Introduction to Corpus Linguistics*. Londres: Longman.
- Koček, Jan, Marie Koprivova et Vera Schmiedtova. 2000. The Czech National Corpus. Heid, Ulrich et al. (Éds.). 2000: 127-132.

- Mba-Nkoghe, J.** 1981. *Phonologie et classes nominales en fang (Langue bantoue de la zone A)*. Thèse 3^e cycle. Paris: Université Sorbonne Nouvelle.
- McEnery, Tony et Andrew Wilson.** 1996. *Corpus Linguistics*. Édimbourg: Edinburg University Press.
- Ondo-Mebiame, P.** 1992. *De la phonologie à la morphologie du fang parlé à Aboumezok (langue Bantu A. 78)*. 2 volumes. Thèse de doctorat. Tervuren: Université Libre de Bruxelles.
- Quirk, Randolph.** 1992. On Corpus Principles and Design. Svartvik, Jan (Éd.). 1992: 463-469.
- Svartvik, Jan.** 1992. Corpus Linguistics Comes of Age. Svartvik, Jan (Éd.). 1992: 7-13.
- Svartvik, Jan (Éd.).** 1992: *Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82*. Stockholm: Mouton de Gruyter.
- Sinclair, John.** 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Voltz, M.** 1990. Fanj (Ntumu). *Revue Gabonaise des Sciences de l'Homme* 2: 143-173.
- Zgusta, Ladislav.** 1971. *Manual of Lexicography*. La Haye: Mouton.

Logistiek: Die oorsprong, konseptuele ontwikkeling en betekenis daarvan in hedendaagse gebruik

Wessel J. Pienaar, *Departement Logistiek, Universiteit van Stellenbosch, Stellenbosch, Republiek van Suid-Afrika (wpienaar@sun.ac.za)*

Opsomming: Hierdie artikel beskryf die oorsprong van die woord *logistiek*, stip die ontwikkeling daarvan uit en verklaar die hedendaagse betekenis daarvan. Die ontwikkelingsgang van die begrip word aangedui vanaf die oorsprong daarvan in Grieks, met inagneming van die opeenvolgende opname daarvan in Laat-Latyn, Frans en Nederlands, tot by die huidige gebruik daarvan in Afrikaans. Die vestiging van die Franse *logistique* as militêre konsep gedurende die Napoleontiese oorloë, en die uitbreiding daarvan tot 'n bedryfsbegrip ná afloop van die Tweede Wêreldoorlog word verduidelik. Die ontwikkelingsgang sedertdien van die betekenis van die konsep in bedryfsverband word bespreek. Eietydse omskrywings van *logistiek*, betreffende die gebruik daarvan in die bedryfswêreld en in die omgangstaal, word verskaf. Vervolgens bied die artikel 'n eietydse definisie van *logistiek* in bedryfsverband. Ten slotte word aangetoon hoe doeltreffende bedryfslogistiek waarde toevoeg.

Slutelwoorde: BEDRYFSLOGISTIEK, BESITNUT, BESTUUR, DIENS, DOELMATIGHEID, DOELTREFFENDHEID, GOEDERE, GRONDSTOF, INLIGTING, KLIËNTEDIENS, KLIËNTE-TEVREDENHEID, LOGISTIEK, MILITÊRE LOGISTIEK, NUT, PLEKNUT, PRODUK, TYDNUT, VORMNUT, WAARDE

Abstract: Logistics: Its Origin, Conceptual Evolution and Meaning in Contemporary Usage. This article traces the origin of the word *logistics*, outlines its conceptual evolution, and explains its meaning in contemporary usage. The contextual development of the concept is traced from its origin in Greek, along with its successive adoption in Late Latin, French and Dutch, up to its contemporary usage in Afrikaans. The establishment of the French *logistique* as military concept during the Napoleonic wars, and its later extension to commercial usage subsequent to the Second World War are described. The development since then of the meaning of the concept in a commercial context is detailed. Contemporary descriptions of *logistics* relating to its meaning in a commercial sense and in common parlance are discussed. Thereafter the article offers a contemporary definition of *logistics* in a commercial context. Finally it is indicated how effective business logistics adds value.

Keywords: BUSINESS LOGISTICS, CUSTOMER SATISFACTION, CUSTOMER SERVICE, EFFECTIVENESS, EFFICIENCY, FORM UTILITY, GOODS, INFORMATION, LOGISTICS, MANAGEMENT, MILITARY LOGISTICS, PLACE UTILITY, POSSESSION UTILITY, PRODUCT, RESOURCE, SERVICE, TIME UTILITY, UTILITY, VALUE

1. Inleiding

Die woord *logistiek* word toenemend in die algemene omgangstaal en in die bedryfslewe gebruik. In die omgangstaal verwys *logistiek* gewoonlik na die poging om 'n ingewikkelde gebeurtenis vlot te laat verloop. In 'n bedryfskonteks verwys dit na die proses om grondstowwe vanaf hul oorsprong te verplaas, die omvorming daarvan te ondersteun en die voltooide produkte op die regte plek en tyd vir verbruik beskikbaar te stel.

Die doel van hierdie artikel is om die oorsprong van die woord *logistiek* na te speur, die konseptuele ontwikkeling daarvan uit te stip en die betekenis daarvan in hedendaagse gebruik te verklaar.

2. Evolusie van die begrip *logistiek*

Die woord *logistiek* is afgelei van die Griekse byvoeglike naamwoord *logistikos*, wat beteken "geskool in berekening", "met verwysing na rekenkunde" of "wat gemoeid is met rede". Die byvoeglike naamwoord is afgelei van die werkwoord *logizomai*, wat volgens die konteks beteken "om te bereken", "om te redeneer" of "om te dink". Die werkwoord is gegrond op die woord *logos*, wat vertaal word as *rede*, *woord* of *gesprek*. Die Grieke het die lede van 'n mediese skool wat rede en mediese praktyk gekombineer het, *logistikoi* genoem.

Die Griekse *logistikos* het die Europese hooftale deur die Laat-Latynse *logisticus* bereik, wat *logistique* in Frans geword het. Die selfstandige naamwoord *logistique* het sedert 1546 begin verskyn. In *Discours du songe de Poliphile* (Gesprek oor Polifilus se droom), wat in 1593 gepubliseer is, word *rede* verteenwoordig deur 'n persoon genaamd *Logistique* (hy wat redeneer) (Boucher 1998: 1). In 1611 is die selfstandige naamwoord *logistique* met verwysing na die vier prosesse van rekenkunde en algebra (optel, aftrek, maal en deel) gebruik. In 1765 is die byvoeglike naamwoord *logistique* wat "met verwysing na berekening" beteken, van hierdie naamwoord afgelei. Hierdie twee betekenis het grotendeels in onbruik geraak (Dictionnaires Le Robert 1976: 1004). (Dieselfde het gebeur met *logarithmes logistiques* (logistiese logaritmes), waar die logaritme van 3 600 as gelyk aan nul beskou is ten einde sekere berekening wat met sterrekundige waarnemings verband hou, te vereenvoudig.)

Die gebruik van die woord *logistique* vir die vermoë om die inkwartiering, vervoer, voorsiening en onderhoud van troepe tydens 'n militêre veldtog te kombineer en koördineer deur beredenering gegrond op berekening, word toegeskryf aan Antoine Henri, Baron van Jomini. Jomini was 'n militêre teoretikus wat as brigadier-generaal in Napoleon se leër gedien het. Hy het die woord in 1836 in hierdie konteks bekend gestel in sy hoogs invloedryke werk *Précis de l'art de la guerre* (Opsomming van die krygskuns) (Parkinson 1977: 90; Macksey en Woodhouse 1991: 198), soos dit in 1840 deur die Franse Akademie erken is (Dictionnaires Le Robert 1996: 1299).

Die Franse *logistique*, met die betekenis soos deur Jomini gebruik, het

sedert 1855 as *logistiek* in Nederlands begin verskyn (Van Veen en Van der Sijs 1997: 518). In 'n laat negentiende-eeuse Nederlandse werk *Handboek van Vreemde Woorden, Uitdrukkingen enz.* (Baale en Baale s.j.: 629) word *logistiek* omskryf as: "vr., letterrekening = Algebra, ook onderdeel der krijgskunde, dat den tijd en de plaats welke voor de uitvoering eener tactische beweging noodig zijn, leert berekenen." Van Dale se *Groot Woordenboek der Nederlandse Taal* (1984¹¹: 1596) se verwoording van *logistiek* onderskei drie punte in Jomini se geheelbeeld van die begrip: "1. (thans w.g.) wetenskap van de deductie, geformaliseerde of mathematische logica; 2. (milit.) alle voorbereidingen en handelingen die nodig zijn om de troepen op de meest doeltreffende wijze van goederen en voorraden te voorzien en onder de gunstigste omstandigheden te doen strijden; — ook deze voorziening self; 3. beheersingsproses van goederenbewegingen."

Die woord *logistiek* het ná afloop van die Tweede Wêreldoorlog in Suid-Afrika begin posvat, waarskynlik as gevolg van die intensiewe logistieke poging wat daardie oorlog geveg het en Suid-Afrika se betrokkenheid daarby. Die vroegste gebruik van *logistiek* in Afrikaans soos aangehaal in die *Woordeboek van die Afrikaanse Taal* (WAT) (1994: 348) lui soos volg: "Die maneuvres is bedoel as 'n toets vir die leër se taktiek en logistiek" (*Die Burger*, 16 April 1949: 3). (Die Tweede Wêreldoorlog het van 1939 tot 1945 geduur.)

Die *Verklarende Handwoordeboek van die Afrikaanse Taal* (Odendal 1994³: 625) se verklaring van *logistiek* is ooglopend gegrond op punte 1 en 2 van Van Dale se verklaring daarvan: "1. (*mil.*) Wetenskap wat hom besighou met beplanning, ten uitvoerbringing, berekening van veral militêre aksies, voorsiening van voorrade op die doeltreffendste wyse sodat troepe onder gunstige omstandighede kan veg; berekening i.v.m. die verplasing en inkwartiering van troepe. 2. Wetenskap van die deduksie." Die WAT (1994: 348) verklaar *logistiek* soos volg: "1.a. (krygskunde) Gesamentlike voorbereidings en handelinge gerig op die beweging en onderhoud van krygsmagte, asook die wetenskap onderliggend hieraan. b. Alle aspekte rakende die beplanning en totstandbrenging van 'n bep. onderneming. 2. (filosofie) Simboliese logika." (Punt 1b se gegewe dui eksplisiet daarop dat *logistiek* ook 'n niemilitêre betekenis het.)

Hoewel laasgenoemde twee omskrywings in woordeboeke van 1994 verskyn, verwoord hulle gevestigde opvattinge van logistiek. Plowman (1964: 1) het reeds opgemerk (skrywer se vertaling):

Logistiek beteken, in die militêre wetenskap, die voorbereiding, hantering en organisering van personeel, ook die relevante materiaal, fasiliteite en ander faktore. *Logistiek* beteken dus die beredenering, veral wiskundige ontleding en sintese, van die ingewikkelde en onderling verwante probleme om personeel, voorrade en huisvesting te koördineer. Militêre *logistiek* is 'n oorwegende faktor ten einde dit vir 'n leër moontlik te maak om opmarse te loods, te veg en veldslae te wen. Aldus gedefinieer behels *militêre logistiek* die onderling verwante en hoofsaaklik wiskundige prosedures, en die gevolglik gekoördineerde besluite.

3. Voorkoms van logistiek in 'n bedryfskonteks

Die verwoesting wat tydens die Tweede Wêreldoorlog in Wes-Europa aangerig is, het daartoe gelei dat Jomini se nalatenskap van militêre logistiek later as *bedryfslogistiek* bekend geword het. Daardie oorlog het groot gedeeltes van Wes-Europa se infrastruktuur en fisiese kapitaalvoorraad kreupel gelaat. Ook was groot gebiede plaasgrond wat tydens die oorlog die rol van slagvelde vervul het, nie geredelik vir landboudoeleindes beskikbaar nie.

Ná afloop van die vyandelikhede in 1945 het die ekonomiese heropbou van Wes-Europa begin. Deskundiges vanuit die militêre logistieke veld, wie se toewyding nie meer vir oorlogvoering benodig was nie, is in hierdie heropbou en herlewingsproses benut. Die lande betrokke in hierdie streekspoging was België, Wes-Duitsland, Frankryk, Italië, Luxemburg en Nederland.

Die herstellpoging was intensief en gekoördineerd. Dit is as noodsaaklik geag dat primêre produksie, sekondêre vervaardiging, sowel as tersiêre dienslewering spoedig op dreef moes kom ten einde ekonomiese selfbeskikking te herwin. Die herstel van, en selfs verbetering op vorige standaarde van plaaslike en tussenstreekse toeganklikheid en mobiliteit het voorrang gekry, aangesien dit duidelik was dat Wes-Europa se herlewing ten beste gedien sou word deur samewerking, streekspesialisasie, intrastreekse grondstofverkryging en die skep van 'n gemeenskapsmark. Die katalisator waardeur dit bereik kon word, was die inwerkingstelling van 'n doeltreffende logistieke stelsel, geskoei op die militêre model. Binne die streek het logistieke voorsieningskettings ontstaan asof daar geen internasionale grense was nie.

Bogenoemde ses lande het die Europese Gemeenskap (EG) as 'n streeksorganisasie in 1958 met die Ooreenkoms van Rome gestig. Die ooreenkoms het voorsiening gemaak vir die geleidelike opheffing van doeanehettings en ander handelsbepelings binne die EG-gebied, die skep van gemeenskaplike eksterne tariewe vir ander lande, die geleidelike invoer van 'n gemeenskaplike landboubeleid, asook waarborge vir die vrye beweging van arbeid, kapitaal en grondstowwe (Hinkelman 2000⁴: 75). Teen hierdie tyd het gekoördineerde logistieke praktyk reeds posgevat in Wes-Europa. Die strategie om logistieke voorsieningskettings tot stand te bring wat primêre produksiepunte binne spesialiseringstreke op gekoördineerde wyse met plekke van verbruik en finale aanwending verbind het, is deeglik begryp.

Die Franse Akademie het die gebruik van *logistiek* in bedryfsverband (naas die militêre betekenis) in 1960 erken en het dit soos volg gedefinieer: "Ensemble de moyens et de méthodes concernant l'organisation d'un service, d'une entreprise et spécialement les flux de matières avant, pendant et après une production" (Alle wyses en metodes van die organisasie van 'n diens, 'n onderneming en veral die vloei van materiale voor, tydens en ná produksie) (Dictionnaires Le Robert 1996: 1299).

Die eerste handboek buite Europa wat die voordele van gekoördineerde bedryfslogistieke bestuur voorgelê het, het in 1961 verskyn. Dit verduidelik

gedeeltelik waarom 'n algemeen aanvaarde definisie van *bedryfslogistiek* toe nog in 'n vormingsproses was (Smykay, Bowersox en Mossman 1961). (Ofskoon hierdie werk primêr gerig is op fisiese distribusiebestuur, dit wil sê op die vloei van produkte vanaf vervaardiging tot by die plek van verbruik of aanwending, het dit ook na die vloei tussen die grondstofbron en produkvervaardiging verwys, wat kennelik op die Franse uitleg van *bedryfslogistiek* gegrond is.)

K.M. Ruppenthal, Direkteur van die First Annual Business Logistics Forum wat in 1962 by die Stanforduniversiteit plaasgevind het, het die eerste omvattende oorsig van *bedryfslogistiek* gepubliseer. Hy (Ruppenthal 1963: xiv) skryf soos volg (skrywer se vertaling):

Geen knap militêre aanvoerder sal daarvan droom om 'n offensief te loods indien hy nie redelik verseker is van logistieke steun nie. Die verantwoordelike generaal moet weet dat voorrade in voldoende hoeveelheid gelewer sal word en dat dit beskikbaar sal wees waar en wanneer dit benodig word. Hierdie belangrike kenmerk van beskikbaarheid is lank reeds bekend as *logistiek* — die fisiese verplasing van materiaal en voorraad van die verskaffingsbron tot by die plek van uiteindelige verbruik.

Ook in die bedryfslêwe is die logistieke funksie die noodsaaklike skakel — in hierdie geval tussen produksie en bemarking. Bedryfslogistiek is gemoeid met die inwaartse beweging van materiaal en voorraad en die uitwaartse beweging van voltooide produkte. Die doel daarvan is die aflewering van die voltooide produkte, soos deur die bemarkingsdepartement versoek, by die punt waar dit benodig word, wanneer dit benodig word, asook op die mees ekonomiese wyse.

Dit is dus voor die hand liggend dat bedryfslogistiek meer as bloot die invul van vragbriewe, keuse van karweiers en die monitering van goederebewegings behels. Dit het ook met 'n aantal ander sake te make wat inbreuk maak op doelmatige beweging en die koste daarvan:

1. Aspekte van aanlegvestiging soos dit verband hou met die beskikbaarheid, koste en algemene doelmatigheid van vervoer.
2. Pakhuiskwessies in verband met vervoer.
3. Keuse van die beste vervoermodus vir spesifieke karweiwerk.
4. Kwessies van vraggeld en tariefvasstelling.
5. Optimale besendingsgroottes, met deeglike inagneming van die produksie- en bemarkingsafdelings se behoeftes.
6. Voorraadbeleid, soos dit met vervoer verband hou.
7. Verpakkingsoorwegings, soos dit met vervoer verband hou.
8. Kwessies rakende beroepsvervoer, bruikhuur van voertuie en eie vervoer.
9. Wetgewing, regulering en ander beleidsaspekte wat vervoeraangeleenthede raak.

10. Aanwending van ekonomiese en statistiese tegnieke vir vervoerprobleme.
11. Proefneming, navorsing en ontwikkeling afgestem op die sentrale logistieke funksie.

Dit sal niemand verbaas dat hierdie lys feitlik elke aspek van die sakeonderneming beliggaam nie. Geen hedendaagse sakeonderneming kan bekostig om in netjiese kompartemente verdeel te word nie; in der waarheid is alle departemente verwant. Dit is hierdie verwantskap tussen die departemente van hedendaagse sakeondernemings wat die terrein van die bedryfslogistiek bepaal.

Om bedryfslogistiek te ignoreer, om die ondeelbare aard van die hedendaagse sakeonderneming mis te kyk, is om ondergang te bewerkstellig — aldus Ruppenthal. Ná Ruppenthal se weergawe van bedryfslogistiek in 1962 het omskrywings van *logistiek* in bedryfsverband nog 'n kwarteeu lank op die vloei van goedere gefokus. Dit was eers gedurende die laat jare tagtig dat daar in die literatuur ook na die belangrikheid van inligting- en dienstevloei verwys is. In 1986 het die Council of Logistics Management (CLM), die grootste logistieke beroepsliggaam in die VSA, *logistiek* gedefinieer as (skrywer se vertaling) "die proses van beplanning, implementering en beheer van die doelmatige, kostedoeltreffende vloei en opberging van grondstowwe, halfvoltooide produkte, voltooide produkte en verwante inligting vanaf die plek van oorsprong tot die plek van verbruik met die doel om aan kliëntevereistes te voldoen" (CLM 1986: 2). In 1987 is die heersende opvatting van bedryfslogistiek deur Ballou (1987²) so weergegee: "Bedryfslogistiek dek alle beweging en opbergingsaktiwiteite wat produk vloei vanaf 'n plek van grondstofverkryging tot by die plek van finale verbruik fasiliteer, sowel as die inligtingvloei wat die produk in beweging stel, met die doel om voldoende vlakke van kliëntediens teen billike koste te voorsien." (Let daarop dat *produkte goedere of dienste* kan wees.)

Routledge se *Concise Dictionary of Management* (Statt 1991: 81) erken die feit dat die term *logistiek* posgevat het in die Engelse sakewoordeskate deur dit te omskryf as "a term that was originally used in the military to describe the organizing and moving of troops and equipment". Dit sê voorts: "It is now often applied to any detailed planning process in an organization which entails the distribution or redistribution of resources." In ooreenstemming hiermee en samehangend met die Franse Akademie se verklaring van die begrip (*Dictionnaires Le Robert* 1996: 1299) omskryf die *Oxford Dictionary for the Business World* (1993: 485) logistiek as "1 organization of (orig. military) services and supplies. 2 control of the movement of materials in a factory, from the arrival of raw materials to the packaging of the product, known as *materials management*, together with *distribution management* (or *marketing logistics*), which includes the storage of goods and their distribution to distributors and consumers. 3 organization of any complex operation". In 1995 gee die *Cambridge International Dictionary of English* (1995: 837) gelyke erkenning aan die term

logistiek in militêre en sakeverband deur dit te verklaar as "the careful organization of a complicated military or business activity so that it happens in a successful and effective way".

In 1991 het die CLM erkenning verleen aan die belangrikheid van dienste-vloei en sy definisie van *logistiek* (skrywer se vertaling) gewysig na: "die proses van beplanning, implementering en beheer van die doelmatige, doeltreffende vloei en opberging van goedere, dienste en verwante inligting vanaf die plek van oorsprong tot by die plek van verbruik met die doel om aan kliëntevereistes te voldoen." Die CLM (1991: 4) beklemtoon dat hierdie definisie inwaartse, uitwaartse, interne en eksterne verplasing, sowel as goedereretoere ter wille van omgewingsdoeleindes insluit.

In 1998 het die CLM hierdie definisie soos volg herbewoord (skrywer se vertaling): "Logistiek is daardie deel van die voorsieningskettingproses wat gemoeid is met die beplanning, implementering, en beheer van die doelmatige, doeltreffende vloei en opberging van goedere, dienste en verwante inligting vanaf die plek van oorsprong tot by die plek van verbruik met die doel om aan kliëntevereistes te voldoen" (CLM 1998). Hierdie definisie verander nie die betekenis van die konsep soos verwoord in die CLM se 1991-definisie nie — die CLM voeg bloot by dat logistiek as 'n onderdeel van die voorsieningsketting-proses beskou word (Lambert, Cooper en Pagh 1998: 3). Met die bekendmaking van hierdie definisie vermeld die CLM (1998; 1999: 4) ongelukkig nie wat hy met die begrip *voorsieningskettingproses* bedoel nie.

Die CLM se 1991-definisie van *bedryfslogistiek* geniet wye internasionale aanvaarding en dit word gewoonlik wêreldwyd as die ankerdefinisie aangehaal in hedendaagse bedryfslogistieke literatuur. Opmerkings oor hierdie definisie is gepas:

Eerstens, soos die definisie aandui, sluit *bedryfslogistieke bestuur* die handeling in wat die vloei van goedere, dienste en inligting tussen die plek van oorsprong en plek van verbruik of aanwending (a) voorberei (*beplan*), (b) organiseer (*implementeer*) en (c) uitvoer (*beheer*) (Bowersox en Closs 1996: 2572; Ballou 1999: 23). *Logistieke voorbereiding of beplanning* sluit die keuse in van (a) vestigingsplekke (insluitende tipe, aantal, ligging en kapasiteit), (b) distribusiepartye (insluitende groothandelaars, kleinhandelaars en derdeparty-diensverskaffers) en (c) karweiers (insluitende hulpvervoer teenoor beroepsvervoer en modus) wat nodig is om die diensvlak te lewer wat kliënte verlang, ten einde die ondernemingsdoelwitte te verwesenlik. *Logistieke organisasie of implementering* sluit in (a) die toewysing en plasing van hulpbronne asook (b) die skedule-ring en uitvoering van besendings en aktiwiteite op 'n doelmatige wyse in die nastreef van die onderneming se doelwitte. *Logistieke uitvoer of beheer* omsluit monitering en hersiening van prestasie (soos diensgehalte, uitgawes, produktiwiteit en benutting van bates) ten einde toe te sien dat (a) die behoeftes van kliënte doeltreffend onderskraag word, (b) die onderneming se bronne doelmatig toegewys (ontplooi) word, en (c) regstellend opgetree word sodra prestasie nie in pas met doelwitbereiking is nie.

Tweedens, aangesien logistieke bestuur gerig is op die beplanning, organisering en uitvoer van die logistieke proses, omsluit dit strategiese, taktiese en operasionele ondernemingsaktiwiteite. Op grond van Hax en Candea (1984) se bespreking van produksievoorraadstelsels, kan logistieke bedrywighede in die volgende drie vlakke verdeel word:

- Die *strategiese vlak* behels beplanning wat 'n langdurige invloed op die onderneming het, soos die soort, aantal, ligging, uitleg en kapasiteit van fasiliteite, sowel as die duursame toerusting wat nodig is vir die vloeï van produkte en inligting deur die logistieke netwerk.
- Die *taktiese vlak* sluit organisatoriese aspekte in wat met tussenposes van 'n maand tot 'n jaar bygewerk word. Hieronder ressorteer ook verkrygings- en vervaardigingskedules, voorraadbeleid, vervoerpraktyk en die reëlmaat waarvolgens kliënte besoek word.
- Die *operasionele vlak* verwys na die uitvoering van dag-tot-dag-pligte, soos roetebepaling, die skedulering van voertuie en bemanning, asook bestelprosessering.

Derdens sluit logistiek die vloeï van goedere tussen die plek van oorsprong en die plek van verbruik of aanwending in. Hierdie bestek omsluit die verplasing van grondstowwe en halfvoltooide produkte vanaf primêre produsente en verkopers na vervaardiging, en die verplasing van voltooide produkte na distribusiefasiliteite en verder tot by kliënte (CLM 1986: 2). Die verskaffingsketting se oorsprong is daár waar daar nie primêre voorsieners is nie. Alle voorsieners aan die plek van oorsprong is slegs ondersteunende leweransiers. Die plek van verbruik of aanwending is daár waar daar nie verdere waarde toegevoeg word nie — die produk word hier verbruik of aangewend sonder om as primêre inset vir 'n volgende voorsieningsketting te dien (Porter 1984: 3).

Vierdens sluit die definisie die vloeï in van goedere en dienste binne sowel die vervaardiging- as die dienstesektor (Smith et al. 1991: xvii). Die dienstesektor (of die tersiêre sektor in ekonomieseterminologie) sluit in: handel, elektrisiteitsvoorsiening, vervoer, kommunikasie, onderwys en finansiële, persoonlike en openbare dienste.

Vyfdens sluit die definisie, bykomend tot die verplasing van goedere en dienste, ook inligtingvloeï in (Bowersox en Closs 1996: 2572). Kliënte verwag gewoonlik dat produkte (d.i. goedere en dienste) en verwante inligting gelyktydig beskikbaar moet wees; bestel- en vragbriewe, handleidings, waarborg- en instandhoudingsinligting en fakture word byvoorbeeld gewoonlik saam met afgelewerde goedere benodig. So vind daar ook dikwels transaksie- en in transito-kommunikasie tussen 'n kliënt en 'n logistieke diensverskaffer plaas om te verseker dat 'n doeltreffende diens verskaf word.

Sedens is die doelwit met logistieke bestuur om oor die hele stelsel heen doelmstig en doeltreffend te wees. Die oogmerk is om koste op gekoördineerde

wyse stelselwyd te minimeer, vanaf die plasing van fasiliteite, vervoer en distribusie, tot die opberging van grondstowwe, halfvoltooide en voltooide goedere, en verpakkingskoste, telkens gedagtig aan die diensdoeltreffendheid wat kliënte vereis. Die klem is nie bloot op die goedkoopste of vinnigste vervoer of minimum voorraadvlakke nie, maar eerder op 'n geïntegreerde en gekoördineerde stelselbenadering tot die logistieke proses (Bramel en Simchi-Levi 1997: 3).

Hoewel die CLM se definisie van logistiek die begrippe *doelmatigheid* en *doeltreffendheid* insluit, verwys dit nie na die versoening van die twee elemente nie. 'n Poging om koste stelselwyd te minimeer, terwyl daar ook aan kliënte se vereistes voldoen moet word, is 'n lofwaardige oogmerk, maar dit behoort binne die konteks van 'n gemeenskaplike ideaal gehanteer te word. Vanuit die onderneming se oogpunt is dié ideaal die maksimering van sy welvaart oor die lang termyn.

Logies beskou, moet die minimering van koste dus slegs daarop gerig wees om kosteverspillings uit te skakel, en nie kostebesnoeiing ten koste van groter wins nie. 'n Vermindering van logistieke koste is slegs sinvol wanneer die prysgawe van wins kleiner is as die kostebesnoeiing. Insgelyks sal die aangaan van bykomende logistieke koste, byvoorbeeld 'n lugvragbesending in plaas van 'n besending met 'n goedkoper en stadiger vervoermodus, maar wat nie gesteun word deur 'n vermindering van voorraaddrakoste of ander logistieke kostebesparings nie, slegs geregverdig wees indien die netto inkomste as gevolg daarvan styg.

Die optimale vlak van logistieke uitgawe kom voor wanneer grenskoste (d.i. die koste van die laaste uitseteenheid) gelyk is aan grensinkomste (d.i. die inkomste wat die laaste uitseteenheid realiseer). Om hierdie aspek in ag te neem, kan *bedryfslogistiek* gedefinieer word as "die proses van beplanning, organisering en uitvoering van die doelmatige, doeltreffende vloei en opberging van goedere, dienste en verwante inligting, vanaf die plek van oorsprong tot by die plek van verbruik of aanwending, op so 'n wyse dat daar optimaal aan kliëntevereistes voldoen word met die doel om die welvaart van die onderneming te help maksimeer".

4. Waardetoevoeging deur logistiek

Vier tipes nut kan waarde tot 'n produk toevoeg, naamlik: *vormnut*, *pleknut*, *tydnut* en *besitnut*. *Vormnut* word geskep deur vervaardigingsaktiwiteite, *pleknut* en *tydnut* deur logistieke bedrywighede, en *besitnut* deur bemarkingsaksies. *Vormnut* ontstaan wanneer grondstowwe in 'n produksie- en/of vervaardigingsproses saamgevoeg word om 'n voltooide produk te vorm waarvoor daar 'n vraag is. Deur middel van vervoer skep logistiek *pleknut* deur goedere te verplaas van plekke waar hulle in onbenutbare vorm voorkom of waar daar 'n oorskot is na plekke waar hulle in benutbare vorm verwerk word of waar daar 'n vraag daarna of 'n behoefte daaraan voorkom. Logistiek skep *tydnut* deur

goedere op te berg en dit later by 'n plek van aanvraag af te lewer op 'n tyd soos deur 'n kliënt verlang. Ofskoon *besitnut* deur produkbevestiging geskep word, word dit deur logistiek ondersteun deurdat plek- en tydnut voorvereistes is om kliënte beskikkingsmag oor produkte te gee.

Elke logistieke aktiwiteit van 'n onderneming kan daartoe bydra dat 'n kliënt die regte produk en inligting op die aangewese plek en tyd in die vereiste toestand en hoeveelheid teen 'n aanvaarbare koste of prys ontvang. Kliëntediens is die gemeenskaplike dryfveer wat al die logistieke aktiwiteite doelgerig saambind. Die oogmerk met 'n kliëntgerigte benadering is om die kliënt meer bereidwillig te maak om vir die onderneming se produkte te betaal en só die omset te stimuleer. Kliëntetevredenheid kom voor wanneer die onderneming se totale nutskepping (waardetoevoeging) sodanig is dat die kliënt geen behoefte ervaar om die produk op alternatiewe wyse te bekom nie. Optimale kliëntediens word gelewer wanneer netto inkomste per eenheid (dit is totale inkomste per eenheid minus totale koste per eenheid), vermenigvuldig met die totale aantal eenhede verkoop, 'n maksimum waarde aanneem.

5. Gevolgtrekking

Die woord *logistiek* is afgelei van die Griekse byvoeglike naamwoord *logistikos*, wat beteken "geskool in berekening", "met verwysing na rekenkunde" of "wat gemoeid is met rede". Die Griekse *logistikos* het die Europese hooftale deur die Laa-Latynse *logisticus* bereik wat *logistique* in Frans geword het. Die selfstandige naamwoord *logistique* het sedert 1546 begin verskyn. Die bekendstelling van die woord *logistique* vir die vermoë om die inkwartiering, vervoer, voorsiening en onderhoud van troepe tydens 'n militêre veldtog te kombineer en koördineer deur beredenering gegrond op berekeninge, is in 1840 deur die Franse Akademie erken.

Ná afloop van die Tweede Wêreldoorlog het die betekenis wat aan die begrip *logistiek* geheg is, uitgebrei na velde buite militêre verband. Sedert die vroeë jare sestig word die begrip ook in 'n bedryfskonteks gebruik. In hierdie konteks kan *logistiek* beskryf word as die proses van beplanning, organisering en uitvoering van die doelmatige, doeltreffende vloei en opberging van goedere, dienste en verwante inligting, vanaf die plek van oorsprong tot by die plek van verbruik of aanwending, op so 'n wyse dat daar optimaal aan kliëntevereistes voldoen word met die doel om die welvaart van die onderneming te help maksimeer.

In gewone omgangstaal is *logistiek* die gedetailleerde beplanning, organisering en uitvoering van 'n ingewikkelde operasie wat die toewysing en verspreiding van middele behels.

Ongeag of die woord in 'n militêre of bedryfsverband, of in alledaagse situasies gebruik word, is die doel van logistiek om toe te sien dat die regte produkte (d.i. goedere en dienste) en inligting op die aangewese plek en tyd in die vereiste toestand en hoeveelheid teen 'n aanvaarbare koste of prys beskikbaar gestel word.

Verwysings

Woordeboeke en ensiklopedieë

- Baale, L.M. en C.H. Baale.** s.j. *Handboek van Vreemde Woorden, Uitdrukkingen enz.* Zutphen: W.T. Thieme.
- Bowersox, D.J. en D.J. Closs.** 1996. Logistics. *International Encyclopedia of Business and Management. Volume 3.* Londen: Routledge.
- Buro van die WAT.** 1994. *Woordeboek van die Afrikaanse Taal. Deel 9.* Stellenbosch.
- Cambridge University Press.** 1995. *Cambridge International Dictionary of English.* Cambridge.
- Dictionnaires Le Robert.** 1976. *Petit Robert: Dictionnaire Alphabétique et Analogique de la Langue Française.* Parys.
- Dictionnaires Le Robert.** 1996. *Petit Robert: Dictionnaire Alphabétique et Analogique de la Langue Française.* Parys.
- Hinkelman, E.G.** 2000⁴. *Dictionary of International Trade: Handbook of the Global Trade Community.* Novato, Kalifornië: World Trade Press.
- Macksey, K. en W. Woodhouse.** 1991. *The Penguin Encyclopedia of Modern Warfare: 1850 to the Present Day.* Londen: Viking.
- Odendal, F.F. (Red.).** 1994³. *Verklarende Handwoordeboek van die Afrikaanse Taal.* Midrand: Perskor.
- Oxford University Press.** 1993. *Oxford Dictionary for the Business World.* Oxford.
- Parkinson, R.** 1977. *Encyclopedia of Modern War.* Londen: Routledge.
- Statt, D.A.** 1991. *The Concise Dictionary of Management.* Londen: Routledge.
- Van Dale Lexicografie.** 1984¹¹. *Van Dale Groot Woordenboek der Nederlandse Taal. Tweede deel J-R.* Utrecht/Antwerpen.
- Van Veen, P.A.F. en N. van der Sijs.** 1997. *Etymologisch Woordenboek: De herkomst van onze woorden.* Utrecht/Antwerpen: Van Dale Lexicografie.

Ander bronne

- Ballou, R.H.** 1987². *Basic Business Logistics, Transportation, Materials Management, and Physical Distribution.* Englewood Cliffs, N.J.: Prentice-Hall.
- Ballou, R.H.** 1999⁴. *Business Logistics Management.* Englewood Cliffs, N.J.: Prentice-Hall.
- Boucher, C.** 1998. Logistique. Private faksimilekommunikasie ontvang van professor Claude Boucher, Universiteit van Sherbrooke, Kanada, oor die etimologie van die woord *logistique* in die Franse taal. (Faksimileekopie in besit van die skrywer.) Montreal.
- Bramel, J. en D. Simchi-Levi.** 1997. *The Logic of Logistics: Theory, Algorithms, and Applications for Logistics Management.* New York: Springer.
- Council of Logistics Management.** 1986. *What It's All About.* Oak Brook, Illinois.
- Council of Logistics Management.** 1991. *What It's All About.* Oak Brook, Illinois.
- Council of Logistics Management.** 1998. www.CLM1.org.
- Council of Logistics Management.** 1999. *What It's All About.* Oak Brook, Illinois.
- Hax, A.C. en D. Candea.** 1984. *Production and Inventory Management.* Englewood Cliffs, N.J.: Prentice-Hall.

- Lambert, D.M., M.C. Cooper en J.D. Pagh.** 1998. Supply Chain Management: Implementation Issues and Research Opportunities. *The International Journal of Logistics Management* 9(2): 1-18.
- Plowman, E.G.** 1964. *Lectures on Elements of Business Logistics*. Palo Alto, Kalifornië: Graduate School of Business, Stanforduniversiteit.
- Porter, M.E.** 1984. *Competitive Advantage — Creating and Sustaining Superior Performance*. New York: The Free Press.
- Ruppenthal, K.M. (Red.)**. 1963. *New Dimensions in Business Logistics. Proceedings of the First Annual Business Logistics Forum, 1962*. Palo Alto, Kalifornië: Graduate School of Business, Stanford-universiteit.
- Smith, P.A., J. Barry, J.L. Cavinato, J.J. Coyle, S.J. Dunn en W. Grenoble.** 1991. *Logistics in Service Industries*. Oak Brook, Illinois: Council of Logistics Management.
- Smykay, E.W., D.J. Bowersox en F.H. Mossman.** 1961. *Physical Distribution Management: Logistics Problems of the Firm*. New York: Macmillan.

Equivalence in Bilingual Lexicography: Criticism and Suggestions*

Herbert Ernst Wiegand, *Germanistisches Seminar, University of Heidelberg, Heidelberg, Germany (wiegand.h.e.oberurff@t-online.de)*

Equivalence in difference is the cardinal problem of language and the pivotal concern of linguistics (Roman Jakobson)

Abstract: A reminder of general problems in the formation of terminology, as illustrated by the German *Äquivalenz* (Eng. *equivalence*) and *äquivalent* (Eng. *equivalent*), is followed by a critical discussion of the concept of equivalence in contrastive lexicology. It is shown that especially the concept of partial equivalence is contradictory in its different manifestations. Consequently attempts are made to give a more precise indication of the concept of equivalence in the metalexigraphy, with regard to the domain of the nominal lexicon. The problems of especially the metalexigraphic concept of partial equivalence as well as that of divergence are fundamentally expounded. In conclusion the direction is indicated to find more appropriate metalexigraphic versions of the concept of equivalence.

Keywords: EQUIVALENCE, LEXICOGRAPHIC EQUIVALENT, PARTIAL EQUIVALENCE, CONGRUENCE, DIVERGENCE, CONVERGENCE, POLYDIVERGENCE, SYNTAGM-EQUIVALENCE, ZERO EQUIVALENCE, CORRESPONDENCE

Abstrakt: Äquivalenz in der zweisprachigen Lexikographie: Kritik und Vorschläge. Nachdem an allgemeine Probleme der Begriffsbildung am Beispiel von dt. *Äquivalenz* und dt. *äquivalent* erinnert wurde, wird zunächst auf Äquivalenzbegriffe in der kontrastiven Lexikologie kritisch eingegangen. Es wird gezeigt, dass insbesondere der Begriff der partiellen Äquivalenz in seinen verschiedenen Ausprägungen widersprüchlich ist. Sodann werden Präzisierungen zu den Äquivalenzbegriffen in der Metalexikographie versucht, die sich auf den Bereich der Nennlexik beziehen. Insbesondere der metalexikographische Begriff der partiellen Äquivalenz sowie der der Divergenz werden grundsätzlich problematisiert. In welche Richtung man gehen kann, um angemessenere metalexikographische Fassungen des Äquivalenzbegriffs zu finden, wird abschließend angedeutet.

Stichwörter: ÄQUIVALENZ, LEXIKOGRAPHISCHES ÄQUIVALENT, PARTIELLE ÄQUIVALENZ, KONGRUENZ, DIVERGENZ, KONVERGENZ, POLYDIVERGENZ, SYNTAGMEN-ÄQUIVALENZ, NULLÄQUIVALENZ, KORRESPONDENZ

* Translation of a German paper delivered at the international symposium Deutsche und Bulgaren im Gespräch in Sofia, April 2001. The original paper will be published in the symposium proceedings. Due to the interdisciplinary nature of the symposium, I have refrained from many linguistic abstractions. Some questions which I left aside in the present article will be treated in Wiegand 2002. I thank Maria Smit for the translation.

1. Aspects of the problem of equivalence: a sketch

The problem of equivalence lies in a fact about which there exists interdisciplinary consensus: the lexical-semantic structures of the lexicon of a particular language are language-specific and therefore partly unique. This implies that the lexical-semantic structures of two (or more) languages are not isomorphic. The non-isomorphism (Zgusta 1971: 294) within the area of the lexicon (which of course also applies to the area of grammar) forms the pre-theoretical, observable empirical circumstances, the study of which led to discipline-specific manifestations of the problem of equivalence. Accordingly, the problem in question here has a long scientific history. Knowledge of this history and its most important basic outlines is a useful presupposition by which one can structure the problem of equivalence. My — by all means limited because of annoying language barriers — knowledge of this history of the problem enables me to distinguish the following dimensions:

- (1) a dimension of the problem, in which general problems of scientific conceptualization are embedded;
- (2) a contrastive lexicological dimension of the problem; and
- (3) a genuine metalexicographical dimension of the problem.

The last-mentioned has two sub-dimensions, namely:

- (3)(a) one which relates to the subject of the dictionary (in the sense of Wiegand 1998: 302), and
- (3)(b) one which relates to the dictionary form.

In what follows, I shall briefly give attention to the specific dimensions. According to my theme, the focus will be on the metalexicographical aspect. I have to leave out the sections on dictionary form touched on by the theme, hence specifically the questions in connection with the particular dictionary type and therefore also the presentation of equivalent items and items indicating differences in equivalence determined by the functions of the dictionary, as well as the question of addressing (cf. in this regard, for example, Meyer and Wiegand 2000, and Wiegand 2000, 2001, 2002).

1.1 General problems of concept building: a reminder

Problems of concept building and conceptual vagueness can have different causes. A typical constellation that can easily lead to inaccurate ways of thinking and argumentation exists when language expressions — in our case the nouns *equivalence* and *equivalent* as well as the adjective *equivalent* — are used in both general language and in various scientific disciplines, and in addition to this, also in practical professional fields belonging to the scientific disciplines.

Both the nouns *equivalence* and *equivalent* as well as the adjective *equivalent* are used — apart from their use in general language — in, amongst others, the following disciplines: in the theory of science, logic, mathematics, physics, jurisprudence, ethics, translation theory, in dictionary research as well as in the

practical field of lexicography. Thus one easily finds illustrations for *equivalent* in bilingual dictionaries of which the circle of addressees consists of educated potential users. For example, in the user instructions of the *Wörterbuch Deutsch–Sanskrit* (Mylius 1988: 10), there is a paragraph headed "Das Sanskrit-Äquivalent" (The Sanskrit Equivalent). The use of the particular words in question is not at all uniform in the mentioned fields. Especially diverse and relatively unclear is the use of *equivalent* and *translation equivalent* in translation theory (cf., for example, Kade 1973, 1975, Wilss 1977: 156 ff. and Koller 2001: 159 ff.).

Although the use of the three expressions in question is semantically really different, there are however some common traits. These can be found in the origin of the generally used loan word *equivalent*. Middle Latin *aequivalentia* belongs to the Latin *aequus* meaning "same" and to the Latin *valere* meaning "having value". An appropriate general correspondent expression for *equivalent* would therefore be "having the same value" (cf., for example, GIWDS 1995, s.v. *Äquivalenz*). Accordingly, the adjective *equivalent* would mean, in general usage, the same as "having the same value", and equivalents are in accordance various items which have the same value with regard to at least one other item.

"Having the same value" must clearly be distinguished from identicalness. Two items are — roughly speaking — identical when they correspond with regard to all their distinctive features. They have, on the other hand, the same value when they have the same purpose on the basis of at least one identical feature in the context of a thought or action. Identical items have the same value with regard to a similar context of thought or action; the opposite is, however, not applicable. In my view, almost all subject-specific concepts of equivalence have something to do with "having the same value" in this general sense. The given similarity of concept building in the various subject fields does however not automatically lead to an appropriate understanding of this particular theory-specific concept of equivalence in question. One can rather only reach, on the basis of its use in general language, a more or less vague pre-conception. If somebody, for example, states: an expression in language A is equivalent to an expression in language B when they have the same meaning, then such a proposal is, even though not untrue and quite understandable from the viewpoint of the general language use of *equivalent*, so hopelessly vague from a linguistic point of view that one can hardly do anything with it in the scientific sense. A concept of equivalence specifically for dictionary research should not be built counter-intuitively from its use in general language, but should be conceived more precisely and should also be differentiated from the concepts of equivalence from neighbouring disciplines, especially those of contrastive linguistics and translation theory.

1.2 A brief look at contrastive lexicology

The terms in question only have a marginal role in lexicology, when lexicological investigations are only concerned with one language. In this way, one will, for example, refer to lexical synonyms within the designative lexicon (such as *Orange* and *Apfelsine*) which are extensionally equivalent, meaning that they

have exactly an equal number of denotations. In addition, one will also debate whether the relationship of lexical synonyms can be valid as equivalent relationships in the area of scientific theory (cf., for example, Fischer 1973).

The concepts of equivalence, on the other hand, have a crucial role in contrastive or confrontational lexicology. There are also different lexicological manifestations of the problem of equivalence. Concepts of equivalence in contrastive lexicology also had an effect on metalexicographical publications (cf., for example, Karl 1982). In what follows, I can only give a limited critical perspective on a few variants of the concept of equivalence.

Contrastive lexicology is seen as a partial discipline focusing on *langue*. Its concepts of equivalence accordingly focus on the language system, but are mostly relatively vague. There are several reasons for this. The designative lexicon has as its basis a polysemous concept of language signs. Noun items in the lexicon can therefore be n times polysemous (with $n \geq 2$). When contrasting a source language noun item with a corresponding target language item, the denotative relationship is usually taken as basis for the comparison. Accordingly, equivalence, which is then usually called *semantic equivalence* (cf., for example, Karl 1982: 34), is present in a polysemous noun item precisely when, *firstly*, the number of sememes in the source language are equal to those in the target language and, *secondly*, their denotation correspond in m pairs of sememes with regard to a source and target language sememe in each pair.

In Figure 1, this type of equivalence is illustrated.

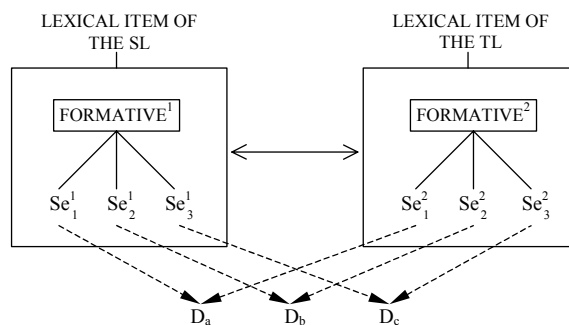


Figure 1: System-related semantic equivalence in threefold polysemous lexical items ($m = 3$). Abbreviations: SL = source language; TL = target language; Se = sememe; D = denoted item; Notational convention: "—" means *is a sememe in*; "↔" means *is semantically equivalent to*; "x -----> y" means *x has as denoted item y*.

The above-mentioned concept of equivalence (using the structuralist terms of its advocates) has various weak points. I can only briefly go into two of them. Firstly, the relationship of semantic equivalence is defined with sets of which the elements are not semantic units at all, but bilateral linguistic signs; the second weak point exists because the point of departure is a concept of equivalence based on semantic units, namely, sememes. Sememes have to "correspond" with reference to the denotation; this means, however: they have to be

denotatively equivalent. In contrast to semantic equivalence, one therefore also refers here to *sememic equivalence* (cf., for example, Karl 1982: 35). Besides, it is not always clear at all whether, for sememic equivalence to be present, one only needs an equal structure of denotative semantic markers, or whether this also applies to non-denotative semantic markers.

Already on a systemic level, one consequently works here with two different, but related, concepts of equivalence, which must necessarily lead to complications. Accordingly, the concept *partial equivalence* already causes considerable confusion in many publications, a fact that I cannot demonstrate here with citations. For, *in the first instance*, one talks about *partial equivalence* when the polysemy structure of a source language lexical item does not correspond with that of the target language on the systemic level (cf., for example, Sternemann 1983: 43f). This case is illustrated in Figure 2 as a first case example.

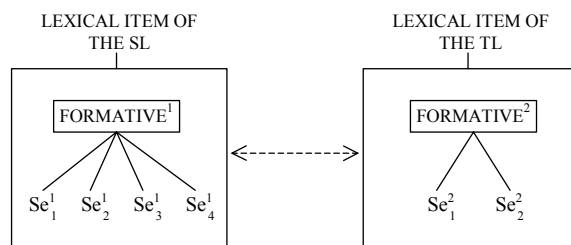


Figure 2: System-related partial semantic equivalence (Case 1a). Notational convention: "<----->" means *is partially semantically equivalent with*.

A second case example (1b) is illustrated in Figure 3.

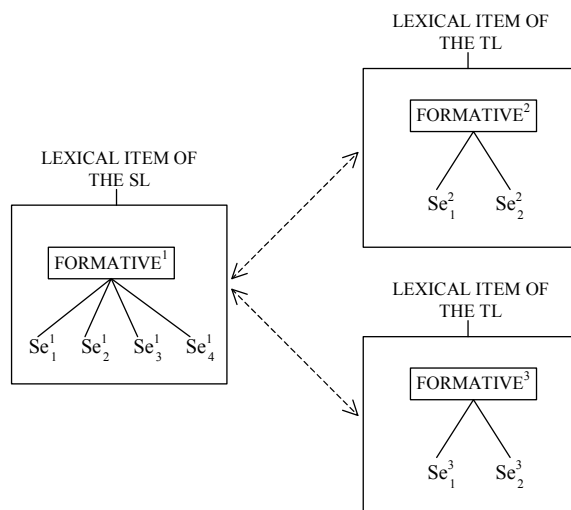


Figure 3: System-related partial semantic equivalence (Case 1b).

In the second place, one uses the expression *partial equivalence* when there is either a hyperonym void or a hyponym void in one of the contrasting partner languages. For example, in Russian, there is no word which corresponds to the German word *Kirschbaum*. In Russian, only matches for the German *Süßkirschenbaum* and *Sauerkirschenbaum* have been lexicalised. *Kirschbaum* is then a partial equivalent to both the Russian *čerešnja* and *višnja*. The reason for this type of partial equivalence on the systemic level is based incorrectly on factors that have to do with *parole*, because one argues that a hyperonym expression could refer to the same referential object as one of the hyponym expressions in the text (cf., for example, Sternemann 1983: 44). The second type of partial equivalence is illustrated in Figure 4.

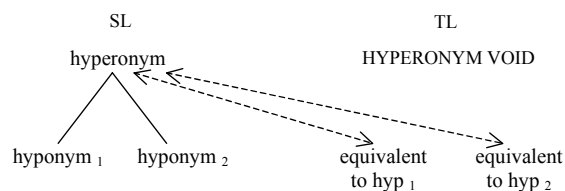


Figure 4: System-related partial semantic equivalence (Case 2).

In the third place, one uses the expression *partial equivalence* when the set of classes of items referring to a collective noun in the source language (for example, German *Vieh*) either do not totally correspond with a collective noun in the target language, or, if there is correspondence of at least one class, it only occurs in one set in both languages (cf., for example, Vietze 1981: 78f). In German, for example, pigs also belong to *Vieh*, but not in Mongolian. On the other hand, in Mongolian, camels, amongst others, also belong to the class denoted by *Vieh* in German. An illustration for case 3 is seen in Figure 5.

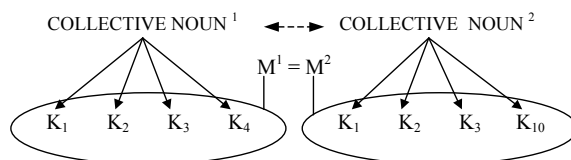


Figure 5: System-related partial semantic equivalence (Case 3). Abbreviations: C = Class; S = Set; Notational conventions: "→" means *denotes*; "=" means *is equal to*.

It is not possible to remain on the topic of contrastive lexicology here any longer. It is clear, however, that concepts not treated here, such as *mono-equivalence*, *poly-equivalence*, *non-equivalence* (cf., for example, Karl 1982: 34ff.), *approximate*, *facultative*, *communicative* and *total equivalence* (cf., for example, Kade 1968: 1973; Wotjak 1982: 113ff.) as well as numerous other postulated types of equivalence (cf., for example, Wilss 1977: 156ff; Scholze-Stubenrecht 1995: passim),

cannot be conceived sufficiently exact if they are conceptually related to concepts of equivalence such as the ones criticized above. The same applies for terms such as, amongst others, *congruence*, *divergence*, *poly-divergence* and *convergence* (cf., for example, Vietze 1981: 78f; Hausmann 1977: 54f.; Rettig 1985: 98ff; Gouws 1996) as well as for terms for the most diverse types of equivalence, such as *quasi-equivalent*, *assisting equivalent* (Karl 1982: 40f) and numerous others (cf. altogether, for example, also Koller 2001: 159ff).

1.3 Describing concepts of equivalence in dictionary research more precisely

In the past three decades, many important insights were gained on equivalence in bilingual lexicography in numerous publications. I mention only a few of these publications: Atkins 1996; Baunebjerg Hansen 1990; Duval 1991; Gouws 1996, 2000, 2002; Hartmann 1994; Hausmann 1977, 1988, 1995, 1997; Jarošová 2000; Meyer and Wiegand 2000; Petkov 2001; Scholze-Stubenrecht 1995; Sinclair 1996; Tognini-Bonelli 1996; Werner 1999; Wiegand 2000; Zgusta 1971, 1984. The insights are concerned, amongst others, with the following aspects, into which I cannot go into detail:

- evaluation of equivalents,
- the microstructural presentation of equivalent items,
- addressing equivalent items,
- discriminating between several presented equivalents by means of "items discriminating between equivalents",
- distinguishing between types of equivalents and types of equivalence, and
- distinguishing between various functions of equivalent items depending on the dictionary type and dictionary function.

However, it is remarkable that the concept of equivalence — in contrast to contrastive lexicology and translation theory — has hardly been discussed in meta-lexicographical publications. In my view, there exist, in the meantime, grave differences of opinion which have led to a whole range of misjudgments about the features of equivalent relationships in bilingual lexicography. In what follows, I will try to verify these critical statements. Accordingly, I will now concentrate on the concept of equivalence, but can only single out some aspects.

It should in the first instance be immediately clear: the *langue*-related concept of equivalence of contrastive lexicology is inappropriate for bilingual lexicography, because bilingual dictionaries are not conceptualised as aids for contrastive studies of language systems (even though some advocates of contrastive lexicology use them in this way). They are rather meant, in the first place, as a means to understand and produce foreign texts and to make translations in both directions. Because the systemic level can also play a part in dictionary research and lexicography, I suggest that one speaks, with reference to Koller (2001: 216ff), of *correspondence* instead of *equivalence* when one deals with lan-

guage systems. It would of course also be necessary precisely to determine the concept of correspondence. In the theory of bilingual dictionaries, *equivalence* would then be a term reserved for *parole* phenomena. Another possibility would be permanently to distinguish between *systemic equivalence* and *parole equivalence*. The *parole*-relatedness brings a metalexicographical concept of equivalence close to most of the concepts in translation theory, so that here one has to draw a clear distinction. The concepts of equivalence in translation theory thoroughly differ. They have, however, the following in common: they refer to whole texts and their translations. Without taking proverbs into account, this is not at all the case in lexicography. Here one deals with the equivalence of meaning-bearing units below the level of sentences. It therefore also deals with the equivalence of word formation devices, words, free syntagmas, and with equivalence of various items that consist of several words which do not form sentences, especially idiomatic expressions and collocations. One should already point out here that a metalexicographical concept of equivalence should not be reduced to lexical items. Otherwise, several cases cannot be taken into account, for example, when a source language item is equivalent to a non-lexicalised target language item. Thus the French *espacer* is equivalent to the German *Zwischenraum lassen zwischen*. The German equivalent is, however, not lexicalised (cf. in this regard Rettig 1995: 93ff versus Hausmann 1977: 53ff).

Up to now, it has only very roughly been outlined which language items from the dictionary contents can feature in the pre- and post-domain of a bilingual equivalence relationship. The following should of course be very clear: independent of which concept of meaning is used in detail to interpret language expressions below the level of sentences, each polysemous item can only belong to an equivalence relationship, based on *parole* conditions, with one of its meanings.

In what follows, I limit my consideration to noun items. Bilingual equivalence is a relationship between a source and target language item which is present when the threefold predicate *x is equivalent to y with regard to z* is true; "z" is the variable for the criterion of equivalence. All features of noun items can in principle function as equivalence criteria. For instance, the following statement is true:

- (1) German *Scheiße* is equivalent to British *arse* with reference to the pragmatic label "vulgar".

Even though both the expressions *Scheiße* and *arse* are obviously in a relationship of equivalence, they are just as obviously not equivalents in a lexicographical context. This is because the essential requirement for lexicographical equivalents to be present in the area of nouns consists of the source and target language item denoting the same object in usual texts (in the sense of Wiegand 1996a). This means: they have to be *referentially-semantically equivalent*, in short: *semantically equivalent*. To form the concept *lexicographically equivalent*, one therefore has to evaluate the possible equivalence relationships: the most important one being semantic equivalence. One can accordingly only speak of *lex-*

icographical equivalence when semantic equivalence occurs.

Two items which are semantically equivalent can also be in other equivalence relationships.

- (2) The German *Arsch* and British *arse* are semantically equivalent.

However, it is also true that

- (3) *Arsch* and *arse* are equivalent with reference to the label "vulgar".

Both expressions therefore have two equivalence relationships. If one speaks of an equivalence relationship of which the equivalence criterion is a dimension of pragmatic labels, of pragmatic equivalence, then *Arsch* and *arse* are semantically as well as pragmatically equivalent, in short: semantically-pragmatically equivalent. If one has at one's disposal a system of labelling, then one can distinguish several pragmatic equivalence relationships. If one dimension of labelling, for example, is "style", one can state:

- (4) *Arsch* and *arse* are semantically-stylistically (or semantically and stylistically) equivalent.

Two expressions which are semantically equivalent, can also be *pragmatically* non-equivalent. This is the case when one of the expressions is pragmatically labelled and the other not. In this case, one can refer to *quasi-equivalents*, with reference to Karl (1982: 40f). One can then state: a pair of quasi-equivalents is to a lesser extent equivalent than a pair of semantic-pragmatic equivalents. In bilingual lexicography, one therefore strives for semantic-pragmatic equivalence.

To introduce some necessary further distinctions, we can subsequently look at a dictionary article from Neubert and Gröger 1991³.

bed [bed] 1. *s* Bett *n* (< and board *Jur* Bett *n* u. Tisch *m*; ~ and breakfast *Übernachtung f* mit Frühstück; double ~ Doppelbett *n*; single ~ Einzelbett *n*, *Übernachtung f* für eine Person; to be brought to ~ of niederkommen mit; to go to ~ ins Bett gehen; to keep/ take to one's ~ das Bett hüten, im Bett bleiben müssen; to make the ~ das Bett machen; ~ of roses leichtes od unbeschwertes Leben; ~ of thorns Schmerzenslager *n*) | Bett *n*, Matratze *f* (feather ~ Unterbett *n*) | (Tier) Lager *n* | Lager *n* (aus Stroh u. ä.) | (Fluß-) Bett *n* | *Tech* Unterlage *f*, Bettung *f*, Fundament *n* | *Eisenb* Unterbau *m*, Schotterbett *n* | *Arch* Untermauerung *f* | *Geol* Lager *n*, Schicht *f* | *Bergb* Flöz *n* | (Blumen-) Beet *n*; 2. (~ded, '-ded) *vt* (jmdn.) ins Bett legen, zu Bett bringen | betten (*auch übertr*) | *Tech* betten, einlegen, einmörteln, festlegen (in in); ~ down (Pferd) in den Stall bringen, mit Streu versorgen | *Gartenb* (Blumen u. ä.) einpflanzen, in Beete pflanzen; ~ out *Gartenb* (Pflanzen) auspflanzen; *vi auch* ~ down sich schlafen legen, sich niederlegen | zusammen schlafen (with mit) | (Tier) lagern

Figure 6: Dictionary article 1 (da₁) from Neubert and Gröger 1991³.

Only the first part of the article in which the noun has been treated lexicographically is of interest here.

In dictionary article da_1 , the German equivalent item "Bett" is addressed to the English item giving the form of the lemma sign *bed*. Both expressions are pragmatically non-labelled. This is also a type of pragmatic equivalence. The lexicographical statement therefore is: the standard language noun *bed* used in its usual sense is semantically-pragmatically equivalent with the standard language noun *Bett* used in its usual sense. This means: if one forms a usual text with *bed* in English (that is, a text which is in accordance with habitual contexts of designation for *bed*), the German word *Bett* can be used in the German translation. The English *bed* could then — as we can furthermore conclude from dictionary article da_1 — occur in the following standard language co-texts (= stands for *is semantically-pragmatically equivalent with*):

bed and breakfast = *Übernachtung mit Frühstück*
single bed = *Übernachtung für eine Person*
to be brought to bed of = *niederkommen mit*
bed of roses = *leichtes od. unbeschwertes Leben*
bed of thorns = *Schmerzenslager*

In addition to these standard language co-texts mentioned in the dictionary, there are others which are domain-specific and which are referred to in the dictionary by means of other items. One either presents specific classes of referential objects, for example "(Tier) Lager" (this means it is *Lager* in German when the English *bed* refers to the "bed" of animals), or one presents specific subject fields, for example "*Bergb Flöz*".

The article da_1 is consequently constructed in such a way that the general case of semantic-pragmatic equivalence is dealt with in the first instance, then followed by the specific cases.

Let us now look at the co-text item "bed and breakfast" from da_1 . Of course, this item was presented because the word equivalence of *bed* and *Bett* does not occur. Rather, an equivalence of syntagmas is present. The English syntagma *bed and breakfast* and the German syntagma *Übernachtung mit Frühstück*, in which the word *Bett* obviously does not occur, are pragmatically-*semantically* equivalent.

In both cases discussed above, the elements in the pre- and post-domain of the equivalence relationship are on the same hierarchical level. They can, however, also be on different hierarchical levels. For example, the German compound *Schwarzmarkt* is semantically-pragmatically equivalent to the English syntagma *black market*. This is a case of *word-syntagma equivalence*. And the French *marché noir* is also semantically-pragmatically equivalent to the German *Schwarzmarkt*; here also, *syntagma-word equivalence* is present.

When no equivalent is given in the target language, one refers to *non-equivalence*. One may ask here: exactly when does non-equivalence occur? When considering this question, one is faced with the fact that each language item of a particular language always has a paraphrase in this language which

can be more detailed or less detailed. If there is no semantic equivalent for an item in another language, one can translate its paraphrase in the language lacking an equivalent. Accordingly, another question arises: is this translation then an equivalent? Let us look at an example. For the German *Amtsgericht*, there is no equivalent in French, neither on the word level nor on the syntagmatic level. If one wants adequately to explain to a French person what *Amtsgericht* means, one has to form a longer French text, which one can also consider a translation of a corresponding German text. In my view, it makes no sense to let such a text pass as equivalent in dictionary research. One should rather state the following definition criterion for equivalents: Only items below the sentence level which can be used in target language sentences can feature as equivalents of the target language in lexicography, to which can be added — as already mentioned — that, for equivalent syntagmas in the target language, no stipulations regarding lexicalisation should be formulated, so that equivalence can also be free from the limitations imposed by the lexicon. In my opinion, one should either give up the distinction between so-called "translational equivalents" and "explanatory equivalents" or, on the other hand, define it more precisely, so that one can only refer to "explanatory equivalents" when target language syntagmatic equivalents in sentences are applicable. Furthermore, it should be clear that non-lexicalised equivalents should only be presented when no lexicalised ones can be found. Non-equivalence is therefore present when no word or syntagma which is at least semantically equivalent can be found in the target language. Non-equivalence should explicitly be marked in the dictionary article (for example, by means of "0") (cf., for example, Wiegand 1996: 228ff in this regard).

At present, since Hausmann (1977: 54f), one refers to *divergence* and *convergence* in dictionary research. Examples of divergence can be found in Figure 7.

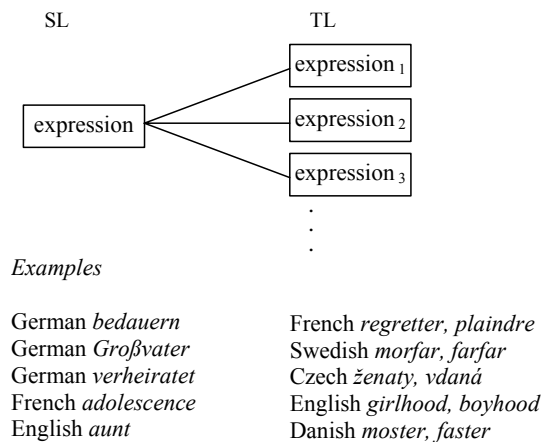


Figure 7: Divergence; Notational convention: "———" means *corresponds*.

In Figure 8 one finds examples of convergence; all examples are from Hausmann (1977), Koller (2001: 228ff) and Kromann et al. (1991).

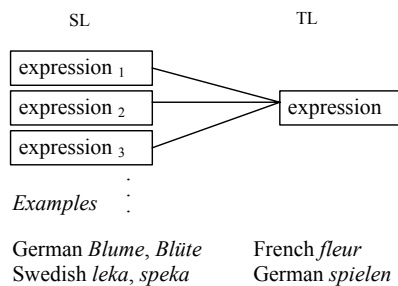


Figure 8: Convergence.

In Hausmann (1977: 54), divergence and convergence belong to the "basic types of equivalence relationships". One can interpret this in such a way that, for example, in the case of divergence, one can speak of *several* equivalence relationships. Hausmann (1977: 55) formulates this as follows: "Two or more equivalents in the target language corresponds to a source language word: *bedauern* = *regretter*, *plaindre*." Divergence and convergence are not to be understood extensionally in Hausmann. It is not concerned with relations representing sets of ordered pairs.

There are, consequently, two equivalence relationships, namely: *bedauern* is equivalent to *regretter* and *bedauern* is equivalent to *plaindre*. In metalexicographical context, only *one* type of equivalence relationship is now formed out of divergence and convergence in each case, that is, out of phenomena both consisting of *several* equivalence relationships (cf., for example, Gouws 1996: 17 and Gouws 2000)! When divergence, for example, is present, then consequently one relation (of a specific type) should exist. This is only the case when one equivalent relation is understood to be a set (cf. Wiegand 2002). When, as is the case with Hausmann, one understands divergence and convergence intensionally, this is conceptually not correct. In the mentioned intensional sense, divergence and convergence should not be interpreted as *one* relationship, even though one would have liked to establish an independent concept of relationships for metalexicography which, in my view, would be totally unsuitable. While the technical analysis in Gouws, which is performed in the context of the concepts *divergence* and *convergence*, is completely acceptable, another line of reception is muddled. Kromann et al. (1991: 2718) writes under the heading of the second version "Partial equivalents": "[...] there is divergence when a lemma, contrasted with the lexical units of the target language, must be divided into several 'sub-meanings' [...]". This means: when, for example, somebody contrasts the German *bedauern* with the French *regretter* and French *plaindre*, then *bedauern* suddenly receives (as if from a higher authority) two "sub-mean-

ings". It should no doubt be clear that one cannot argue like this.

Already in Hausmann, it is not clear (to me) whether the sentence quoted above refers to *langue* or *parole*. This remarkable carelessness when equivalence is dealt with, can be found in a large portion of the metalexicographical literature. Since this carelessness with regard to the consideration of an important distinction does not have an effect on congruence (cf. Figure 9), because here the reference to the *langue* and *parole* levels are analogous, only *congruence* is, in my view, a sufficiently clear concept.

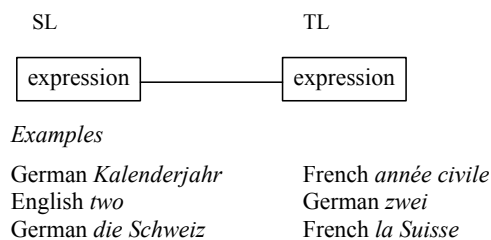


Figure 9: Congruence.

A solution for the above-mentioned conceptual discrepancy may be sought in the direction as indicated below.

Semantic and pragmatic equivalence are, for example, types of equivalence. Following Scholze-Stubenrecht (1995), one can distinguish additional types of equivalence relationships. Convergence and divergence are then, for example, not types of equivalence relationships. Here one is rather dealing with entities on the systemic level which could, for example, be called a *correspondence network*. Koller (2001: 228ff) calls them *correspondence types* (*Entsprechungstypen*). Convergence corresponds with the correspondence type "many-to-one-correspondence"; divergence corresponds to "one-to-many-correspondence". The threefold relationship statement (*x corresponds with y with reference to z with "x" and "y" as variables for noun items of the source and target language and "z" as variable for the correspondence criterion*) belongs to the correspondence relationship. This is given in the relationship of denotation: "x corresponds with y" should be read as "x is the systemic correspondence for y".

If there are *n* systemic correspondences (with $n \geq 2$) in a target language for a lexical item in the source language, so that the correspondence network of divergence is present, the source language item can only be semantically equivalent with the *n* target language items if the *n* target language items are totally synonymous on the lexical-semantic level. If such synonymy is not present in the target language items, then co-text independent semantic equivalence is impossible. On the contrary, every corresponding target language item determines by means of a specific set of features (which can, in comparison to the source language, be considered as a semantic restriction of use) a source language co-text class as condition for equivalence for the *parole*, to which those

usual texts belong in which the target language item, when used in a usual sense, is semantically equivalent to the source language item, when also used in a usual sense. For example, the Swedish *morfar* determines, as condition for equivalence, the class of all co-texts with the German *Großvater* in which *Großvater* refers to a grandfather on the mother's side. If one finds in a dictionary article of a German–Swedish dictionary the entry

(5) Großvater ... (mütterlicherseits) morfar; (väterlicherseits) farfar ...,

then the meaning of the lemma sign *Großvater* is not split up into two "sub-meanings" — as Kromann et al. (1991: 2718) incorrectly believe — but rather, with "mütterlicherseits" and "väterlicherseits", the co-textual conditions for equivalence are presented which have to be fulfilled in order for *Großvater* to be semantically equivalent either to *morfar* or to *farfar*. *Großvater* is consequently not partially semantically equivalent to *morfar* and *farfar* in the sense that a "part" or "sub-meaning" of *Großvater* is partially semantically equivalent with *morfar*, and another "part" or "sub-meaning" with *farfar*. The postulation of such "sub-meanings" makes no sense. It is rather more appropriate to say that *Großvater* can be co-text-specifically equivalent with both *morfar* and *farfar*. It is therefore indeed better motivated and more appropriate to speak of *co-text-specific semantic equivalence* than of *partial semantic equivalence*. The other examples from different language pairs presented in Figure 7 can also be treated in the same way.

The distinction introduced here — even if only very roughly — enables a more appropriate understanding of items differentiating between equivalents of a specific type, such as, for example, "mütterlicherseits" and "väterlicherseits" in (5). With these items specific co-text classes are determined referentially-semanticly.

Here I have to end my reflection for now.

2. Concluding remarks

If my exposition gives rise to a new consideration of the problem of equivalence within the framework of dictionary research, it has fulfilled its purpose. The suggestions made here will be further discussed in Wiegand 2002.

Bibliography

- Atkins, B.T.S. 1996. *Bilingual Dictionaries: Past, Present and Future*. Gellerstam, Martin, Jerker Järborg, Sven-Göran Malmgren, Kerstin Norén, Lena Rogström and Catarina Röjder Papmehl (Eds.). 1996. *Euralex '96 Proceedings. Papers Submitted to the Seventh EURALEX International Congress on Lexicography in Göteborg, Sweden*: 515-546. Göteborg: Göteborg University.
- Baunebjerg Hansen, Gitte. 1990. *Artikelstruktur im zweisprachigen Wörterbuch. Überlegungen zur Darbietung von Übersetzungsäquivalenten im Wörterbuchartikel*. *Lexicographica. Series Maior* 35. Tübingen: Max Niemeyer.

- Duval, Alain.** 1991. L'équivalence dans le dictionnaire bilingue. Hausmann, Franz Josef, Oskar Reichmann, Herbert Ernst Wiegand and Ladislav Zgusta (Eds.). 1991. *Wörterbücher. Ein internationales Handbuch zur Lexikographie / Dictionaries. An International Encyclopedia of Lexicography / Dictionnaires. Encyclopédie internationale de lexicographie*: 2817-2824. Handbücher zur Sprach- und Kommunikationswissenschaft 5.3. Berlin/New York: De Gruyter.
- Fischer, Walther L.** 1973. *Äquivalenz- und Toleranzstrukturen in der Linguistik. Zur Theorie der Synonyma*. Linguistische Reihe 15. Munich: Hueber.
- GIWDS** 1995 = Scholze-Stubenrecht, Werner (Ed.). 1995. *Das große illustrierte Wörterbuch der deutschen Sprache. Das aktuelle Nachschlagewerk der 100 000 Begriffe. Bd. 1: A-L; Bd. 2: M-Z*. Stuttgart/Zürich/Vienna: Verlag Das Beste.
- Gouws, Rufus Hjalmar.** 1996. Bilingual Dictionaries and Communicative Equivalence for a Multilingual Society. *Lexikos* 6: 14-31.
- Gouws, Rufus Hjalmar.** 2000. Strategies in Equivalent Discrimination. 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*: 99-111. *Lexicographica. Series Maior* 103. Tübingen: Max Niemeyer.
- Gouws, Rufus Hjalmar.** 2002. Equivalent Relations, Context and Cotext in Bilingual Dictionaries. *Hermes* 28: 195-209.
- Hartmann, Reinhard Rudolf Karl.** 1994. The Use of Parallel Text Corpora in the Generation of Translation Equivalents for Bilingual Lexicography. 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*: 291-297. Amsterdam: Vrije Universiteit.
- Hausmann, Franz Josef.** 1977. *Einführung in die Benutzung der neufranzösischen Wörterbücher*. Tübingen: Max Niemeyer.
- Hausmann, Franz Josef.** 1988. Grundprobleme des zweisprachigen Wörterbuchs. Hyldgaard-Jensen, Karl and Arne Zettersten (Eds.). 1988. *Symposium on Lexicography III. Proceedings of the Third International Symposium on Lexicography May 14–16, 1986 at the University of Copenhagen*: 137-154. *Lexicographica. Series Maior* 19. Tübingen: Max Niemeyer.
- Hausmann, Franz Josef.** 1995. Von der Unmöglichkeit der kontrastiven Lexikologie. Kromann, Hans-Peder and Anne Lise Kjær (Eds.). 1995. *Von der Allgegenwart der Lexikologie. Kontrastive Lexikologie als Vorstufe zur zweisprachigen Lexikographie. Akten des internationalen Werkstattgesprächs zur kontrastiven Lexikologie, 29.–30.10.1994 in Kopenhagen*: 19-23. *Lexicographica. Series Maior* 66. Tübingen: Max Niemeyer.
- Hausmann, Franz Josef.** 1997. Semiotaxis und Wörterbuch. Konerding, Klaus-Peter and Andrea Lehr (Eds.). 1997. *Linguistische Theorie und lexikographische Praxis. Symposiumsvorträge, Heidelberg 1996*: 171-179. *Lexicographica. Series Maior* 82. Tübingen: Max Niemeyer.
- Jakobson, Roman.** 1981. Linguistische Aspekte der Übersetzung. Wilss, Wolfram (Ed.). 1981. *Übersetzungswissenschaft*: 189-198. Wege der Forschung 535. Darmstadt: Wissenschaftliche Buchgesellschaft. (First published as: On Linguistic Aspects of Translation. Brower, R.A. (Ed.). 1969. *On Translation*: 232-239. New York: Oxford University Press.)
- Jarošová, Alexandra.** 2000. Problems of Semantic Subdivisions in Bilingual Dictionary Entries. *International Journal of Lexicography* 13: 12-28.
- Kade, Otto.** 1968. Zufall und Gesetzmäßigkeit in der Übersetzung. Supplement to *Fremdsprachen* 1.

- Kade, Otto.** 1973. Zur Modellierung von Äquivalenzbeziehungen. Neubert, Albrecht and Otto Kade (Eds.). 1973. *Neue Beiträge zu Grundfragen der Übersetzungswissenschaft*: 157-165. Frankfurt: Athenäum.
- Kade, Otto.** 1975. Möglichkeiten und Grenzen der Anwendung von Valenzmodellen bei der Erforschung und Beschreibung zwischensprachlicher Äquivalenzbeziehungen. *Linguistische Arbeitsberichte* 11: 2-18.
- Karl, Ilse.** 1982. *Linguistische Probleme der zweisprachigen Lexikographie. Eine Nachlese praktischer Wörterbucharbeit*. Linguistische Studien. Reihe A. Arbeitsberichte 96. Berlin.
- Koller, Werner.** 2001⁶. *Einführung in die Übersetzungswissenschaft*. Uni-Taschenbücher 819. Wiesbaden: Quelle und Meyer.
- Kromann, Hans-Peder, Theis Riiber and Poul Rosbach.** 1991. Principles of Bilingual Lexicography. Hausmann, Franz Josef, Oskar Reichmann, Herbert Ernst Wiegand and Ladislav Zgusta (Eds.). 1991. *Wörterbücher. Ein internationales Handbuch zur Lexikographie / Dictionaries. An International Encyclopedia of Lexicography / Dictionnaires. Encyclopédie internationale de lexicographie*: 2711-2728. Handbücher zur Sprach- und Kommunikationswissenschaft 5.3. Berlin/New York: De Gruyter.
- Meyer, Meike and Herbert Ernst Wiegand.** 2000. Gemischt-semiintegrierte Mikrostrukturen für deutsch-spanische Printwörterbücher. Wiegand, Herbert Ernst (Ed.). 2000. *Studien zur zweisprachigen Lexikographie mit Deutsch V*: 87-171. *Germanistische Linguistik* 151-152. Hildesheim/New York: Georg Olms.
- Mylius, Klaus.** 1988. *Wörterbuch Deutsch-Sanskrit*. Leipzig: VEB Verlag Enzyklopädie.
- Neubert, Albrecht and Erika Gröger.** 1991³. *Großes Handwörterbuch Englisch-Deutsch*. Leipzig/Berlin/Munich.
- Petkov, Pavel.** 2001. Zum Problem der Äquivalenzbeziehung und der lexikographischen Lücke in zweisprachigen Wörterbüchern. Iгла, Birgit, Pavel Petkov and Herbert Ernst Wiegand (Eds.). 2001. *Theoretische und praktische Probleme der Lexikographie. 1. Internationales Kolloquium zur Wörterbuchforschung am Institut Germanicum der St. Kliment-Ohridski-Universität Sofia*: 72-81. *Germanistische Linguistik* 161-162. Hildesheim/New York: Georg Olms.
- Rettig, Wolfgang.** 1985. Die zweisprachige Lexicographie Französisch-Deutsch, Deutsch-Französisch. Stand, Probleme, Aufgaben. *Lexicographica* 1: 83-124.
- Scholze-Stubenrecht, Werner.** 1995. Äquivalenzprobleme im zweisprachigen Wörterbuch. Ein Erfahrungsbericht. Wiegand, Herbert Ernst (Ed.). 1995. *Studien zur zweisprachigen Lexikographie mit Deutsch II*: 1-16. *Germanistische Linguistik* 127-128. Hildesheim/New York: Georg Olms.
- Sinclair, John.** 1996. An International Project in Multilingual Lexicography. *International Journal of Lexicography* 9: 179-196.
- Sternemann, Reinhard.** 1983. *Einführung in die konfrontative Linguistik. Von einem Autorenkollektiv unter Leitung von Reinhard Sternemann*. Linguistische Studien. Leipzig: VEB Verlag Enzyklopädie.
- Tognini-Bonelli, Elena.** 1996. Towards Translation Equivalence from a Corpus Linguistics Perspective. *International Journal of Lexicography* 9: 197-217.
- Vietze, Hans-Peter.** 1981. Äquivalenzprobleme bei der Arbeit am deutsch-mongolischen Wörterbuch. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 34: 78-83.
- Werner, Reinhold.** 1999. Das Problem der Äquivalenz im zwei- und mehrsprachigen Fach-

- wörterbuch. Hoffmann, Lothar, Hartwig Kalverkämpfer and Herbert Ernst Wiegand (Eds.). 1999. *Fachsprachen. Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft / Languages for Special Purposes. An International Handbook of Special-Language and Terminology Research: 1853-1884*. Berlin/New York: De Gruyter.
- Wiegand, Herbert Ernst.** 1996. Deutsch–Usbekisches Wörterbuch. Einblicke in die Wörterbucharbeit an der Staatlichen Usbekischen Weltsprachen-Universität in Taschkent. *Lexicographica* 12: 190-254.
- Wiegand, Herbert Ernst.** 1996a. Über usuelle und nichtusuelle Benennungskontexte in Alltag und Wissenschaft. Knobloch, Clemens and Burkhard Schaeder (Eds.). 1996. *Nomination — fachsprachlich und gemeinsprachlich*: 55-103. Opladen: Westdeutscher Verlag.
- Wiegand, Herbert Ernst.** 1998. *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie. 1. Teilbd.* Mit 159 Abbildungen im Text. Berlin/New York: De Gruyter.
- Wiegand, Herbert Ernst.** 2000. Adressierung in der ein- und zweisprachigen Lexikographie. Eine einführende Übersicht über die Forschungs- und Problemlage. *Lexikos* 10: 32-74.
- Wiegand, Herbert Ernst.** 2001. Adressierung in zweisprachigen Printwörterbüchern. Wiegand, Herbert Ernst (Ed.). Forthcoming. *Studien zur zweisprachigen Lexikographie mit Deutsch VIII. Germanistische Linguistik*. Hildesheim/New York: Olms.
- Wiegand, Herbert Ernst.** 2002. *Äquivalenz, Äquivalenzdifferenzierung und Äquivalentpräsentation in zweisprachigen Wörterbüchern*. [Typescript for *Symposium on Lexicography XI* in Copenhagen].
- Wilss, Wolfram.** 1977. *Übersetzungswissenschaft. Probleme und Methoden*. Stuttgart: Klett-Cotta.
- Wotjak, Gerd.** 1982. Äquivalenz, Entsprechungstypen und Techniken der Übersetzung. Jäger, Gert and Albrecht Neubert (Eds.). 1982. *Äquivalenz bei der Translation*: 113-124. *Übersetzungswissenschaftliche Beiträge* 5. Leipzig: VEB Verlag Enzyklopädie.
- Zgusta, Ladislav.** 1971. *Manual of Lexicography*. *Janua Linguarum. Series Maior* 39. The Hague/Paris: Mouton.
- Zgusta, Ladislav.** 1984. Translational Equivalence in the Bilingual Dictionary. Hartmann, R.R.K. (Ed.). 1984. *LEXeter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9–12 September 1983*: 147-154. *Lexicographica. Series Maior* 1. Tübingen: Max Niemeyer.

The African Proverbs Project and After

John S. Mbiti, *Faculty of Theology, University of Bern, Bern, Switzerland*
(john.mbiti@freesurf.ch)

Abstract: The Pew Charitable Trusts financed the African Proverbs Project with Dr Stan Nussbaum as co-ordinator, for three years (1993–1996). He assembled a number of scholars who had collected and studied proverbs. The purpose of the Project was to encourage further work of collecting, studying and publishing proverbs. It organised international and interdisciplinary symposia and embarked on three types of publications.

One kind of publication was the African Proverbs Series, with Prof. John Mbiti as editor. Five volumes were published in 1997, each with proverbs from respectively Ethiopia, Uganda, Lesotho, Burkina Faso and Ghana. Their size ranged from 584 to 1 497 proverbs in African languages, with translations into English or French. Another set of publications with the Rev. Joshua Kudadjie as editor, covered proverbs for preaching and teaching purposes. Three volumes have come from Malawi, Ghana and Liberia. The Project also published an annotated bibliography by Prof. Wolfgang Mieder. The report from one of the symposia appeared in 1997.

The third manner of publication is in the form of a CD-ROM edited by Dr Stan Nussbaum and published in 1996. This is a very comprehensive collection and source of material that includes, inter alia, books from the Project, some 28 000 proverbs, a bibliography with 1 000 items, summaries and extracts from published collections, and a language survey of work on proverbs.

The Project has generated both local and international interest in African paremiography and paremiology. Four academic institutions in South Africa, Kenya, Ghana and the Ivory Coast, with links to others in Africa, Europe and America, are continuing this work on proverbs.

Keywords: AFRICAN PROVERBS PROJECT, AFRICAN PROVERBS SERIES, PROVERB, PAREMIOGRAPHY, PAREMIOLOGY

Opsomming: Die African Proverbs Project en daarna. Die Pew Charitable Trusts het die African Proverbs Project, met dr. Stan Nussbaum as koördineerder, vir drie jaar (1993–1996) gefinansier. Hy het 'n aantal vakkundiges bymekaargebring wat spreekwoorde versamel en bestudeer het. Die doel van die projek was om verdere werk ten opsigte van die versameling, bestudering en publikasie van spreekwoorde aan te moedig. Dit het internasionale en interdisiplinêre simposia georganiseer en drie soorte publikasies onderneem.

Een soort publikasie was die African Proverbs Series, met prof. John Mbiti as redakteur. Vyf bande is in 1997 gepubliseer, elkeen met spreekwoorde van onderskeidelik Ethiopië, Uganda, Lesotho, Burkina Faso en Ghana. Hul omvang het gewissel van 584 tot 1 497 spreekwoorde in Afrikaanse, met vertalings in Engels en Frans. 'n Ander stel publikasies, met eerw. Joshua Kudadjie as redakteur, het spreekwoorde vir preek- en onderwysdoeleindes gedek. Drie dele is afkomstig van Malawi, Ghana en Liberië. Die projek het ook 'n geannoteerde bibliografie deur prof. Wolfgang Mieder gepubliseer. Die verslag van een van die simposia het in 1997 verskyn.

Die derde manier van publikasie is in die vorm van 'n CD-ROM, deur dr. Sam Nussbaum geredigeer en in 1996 gepubliseer. Dit is 'n baie omvattende versameling en bron van materiaal wat

onder andere boeke van die projek, ongeveer 28 000 spreekwoorde, 'n bibliografie met 1 000 items, opsomming en uittreksels van gepubliseerde versamelings en 'n taaloorsig van werk oor spreekwoorde insluit.

Die projek het sowel plaaslike as internasionale belangstelling in die paremiografie en paremiologie van Afrika gewek. Vier akademiese instellings in Suid-Afrika, Kenia, Ghana en die Ivoorkus met skakels met ander in Afrika, Europa en Amerika, sit hierdie werk oor spreekwoorde voort.

Sleutelwoorde: AFRICAN PROVERBS PROJECT, AFRICAN PROVERBS SERIES, SPREEKWOORD, PAREMIOGRAFIE, PAREMIOLOGIE

1. Short Background to the Project

By sheer coincidence, Dr Sivgard von Sicard at the Selly Oak Colleges in Birmingham, England gave a short car ride to Dr Stan Nussbaum from Colorado, United States of America and Prof. John Mbiti from Kenya in the spring of 1992. The latter two had not met before. A casual conversation ensued, in which Dr Nussbaum mentioned that while working in Lesotho as a Mennonite missionary he had been struck by the potency of African proverbs. On his part Prof. Mbiti said that he had a keen interest in such proverbs and had collected about ten thousand of them, both published and unpublished. Would Prof. Mbiti be interested in a project that Dr Nussbaum was planning, to focus on African proverbs? Certainly he would. The two arranged to meet again and discuss the project more thoroughly. This they did in Switzerland a few months later. By then the Pew Charitable Trusts of Philadelphia had formally made a grant for the "African Proverbs Project", with Dr Nussbaum as the co-ordinator. The grant was to run for three years, 1993–1996, without renewal.

2. Setting up the Work

Dr Nussbaum contacted a few scholars who formed the working group for the Project. It comprised: Dan Hoffman (United States of America), Joshua Kudadjie (Ghana), John Mbiti (Kenya), Laurent Nare (Burkina Faso), Stan Nussbaum (United States of America, co-ordinator), John Pobee (Ghana), and Willem Saayman (Republic of South Africa). Apart from Dr Nare who had done a doctoral thesis on Mossi proverbs and the Jewish Bible, none of the committee members were "academic experts" on proverbs as such. But they all shared the necessary interest. The group recruited collectors (contributors) and mapped the direction of work to be covered. The world's leading paremiologist Prof. Wolfgang Mieder of the University of Vermont, United States of America, gave the Project enormous support through a bibliographical publication¹, participating at one of its conferences, personal consultation and ongoing interest.

As the co-ordinator explained it, the main purpose of the Project was "to promote collection, publication and study of African proverbs with particular

attention to their relationship to Christian mission, their role in modern Africa and their significance for a number of academic disciplines." To put the Project into action, the group made plans to cover five areas:

- (1) Two international conferences.
- (2) Publication of five volumes of proverbs from different parts of the continent.
- (3) Publication of three volumes of proverbs for teaching and preaching.
- (4) "Endangered species".
- (5) A bibliography of collections and records of proverbs in articles and books.

2.1 The International Conferences

The first five-day conference met at Ricatla Theological Seminary, Maputo, Mozambique in March 1995. It examined the relationship of proverbs to Christianity. Seventeen theologians and educators attended and presented papers. The second conference in October 1995 brought together thirty-seven participants from different disciplines, including historians, philosophers, cultural researchers, theologians and language scholars. The Missiology Department of the University of South Africa (UNISA), Pretoria hosted the symposium.

The papers presented at the two conferences are included on the CD-ROM that Dr Nussbaum later put together. (More on this, below.) Prof. Willem Saayman edited the papers from the second conference and the University of South Africa (UNISA) Press, Pretoria published them in 1997 under the title *Embracing the Baobab Tree*.

2.2 Publication of Five Volumes

The working group selected five areas where there were scholars that had collected proverbs in local languages. The group felt that the material from those areas would be representative for the whole continent, even if the number published amounted to a small fraction of the total wealth of proverbs in Africa. It chose Ethiopia and Uganda in Eastern Africa, Lesotho in Southern Africa, and Ghana and Burkina Faso in Western Africa. The group chose paremiographers (collectors) who already had enough proverbs or would enlarge their collections to constitute a book. These were Dr George Cotter for the Oromo proverbs in Ethiopia, Prof. Albert Dalfovo for the Lugbara proverbs in Uganda, Dr 'Makali Isabella Mokitimi (the only female) for the Basotho proverbs in Lesotho, Prof. Kofi A. Opoku for the Akan proverbs in Ghana, and both Fr Dr Laurent Nare and Fr François Xavier Damiba for the Mossi proverbs in Burkina Faso. It designated Prof. John Mbiti to be the general editor of the series.

The "writers' group" consisting of the co-ordinator, collectors and general editor, worked out some guidelines for the five volumes. Among other things, the proverbs were to be published in their original African languages, with translations and short explanations in English (for four volumes) and French (for one volume). Each volume would have an introduction about the peoples concerned and their country, the place of proverbs in society, and elaboration on the themes that the proverbs addressed. The general editor's introduction would be included in all the volumes.

The "writers' group" also drew up a long list of themes that emerged from the proverbs. Among them are specifically religious themes like God, the departed (so-called ancestors); values and vices; sociological themes like marriage, the family, behaviour and food; life's journey from pregnancy to death and the hereafter; the human person and activities; environmental settings like plants, animals and birds. The volumes would have indices with alphabetical listings of the proverbs in original languages, topics and key words. Attention was to be given to gender expressions, so that as far as possible, inclusive language would be employed. Each proverb would be numbered for reference purposes. Hand-drawn illustrations were to be included, and each collector was free to find a local artist (illustrator) and to choose the proverbs to be illustrated. The general editor in close liaison with the collectors was to exercise editorial discretion and freedom over the volumes, as would be appropriate and necessary.

The University of South Africa (UNISA) Press, jointly with Asempa Publishers in Accra, Ghana, Daystar and Sefer Presses in Ibadan, Nigeria published the first four volumes in 1997, under the main heading of *African Proverbs Series*. Each has its own title. Volume one with a collection by George Cotter, *Ethiopian Wisdom, Proverbs and Sayings of the Oromo People*, 268 pp., contains a total of 1 497 proverbs. It is in this respect also the largest of the five. Volume two has the collection by Kofi Asare Opoku, *Hearing and Keeping. Akan Proverbs*, 185 pp., with 600 proverbs. Volume three has the collection by Albert Dalfovo, *Lugbara Wisdom*, 185 pp., with 773 proverbs. Volume four is the collection by Makali I. Mokitimi, *The Voice of the People. Proverbs of the Basotho*, 118 pp., with 584 proverbs. The fifth volume with 999 proverbs was published in 1999, being the collection by François Xavier Damiba and Laurent Nare, *Proverbes Mossi. Une Sélection*, 202 pp., in Abidjan, Ivory Coast.²

Thus, the Series itself published a total of 4 453 proverbs in its five volumes. This is obviously a very small portion of an estimated four million proverbs in two thousand African languages. The Project would have made additional publications in the Series, but there were no funds to continue beyond the initial period of three years. However, with this and other publications of the Project, enough interest has been generated for the work to continue, as will be indicated below.

2.3 Publication of Proverbs for Teaching and Preaching

As Dr Nussbaum explains, the purpose of this part of the Project "was much

narrower — to encourage the use of African proverbs as aids to Christian moral instruction. These are intended as textbooks in Bible institutes and seminaries, as well as resources for church leaders. Each book contains an introductory chapter comparing and contrasting the values of African proverbs and the Bible. The book is an explanation of 350 to 500 local proverbs, showing how each could be used in Christian instruction." The following are the three books:

Rev. Joshua Kudadjie, *Ga and Dangme Proverbs for Preaching and Teaching*, Asempa Press, Accra, Ghana 2002. It contains 384 proverbs of which 100 in the Ga and Dangma languages respectively are annotated, 50 in each of the two languages with English translations (but not annotated), and a further 84 are used in the text.

Dr Abba Karnga, *There is Diversity in Wisdom: 150 Bassa Proverbs from Liberia*. A Bassa language edition exists (but publication details are indefinite at time of writing).

Dr David K. Mphande, Nthanthi za Chitonga za Kusambizgiya ndi Kuttaulia (literally, *Tonga Proverbs for Teaching and Preaching*), a Kachere Book by CLAIM (Christian Literature Association in Malawi), Blantyre, Malawi 2001. The book contains 487 proverbs, of which 200 are annotated.

The current plan is to publish all three books in English on the Internet at <http://www.afriprov.org>. English editions are also under consideration by Asempa Press and the Kachere Series.

2.4 "Endangered Species"

One branch of the Project explored or searched for languages in which only a few proverbs, or none, had been collected. Some of these languages were not yet, or only recently became written languages. Six writers were designated to follow up this need. Efforts in this direction brought successful results, as the co-ordinator of the Project later pointed out: "Three lists were made in Burkina Faso and one each in Togo, South Africa and Tanzania. The average is about 400 proverbs per language; French or English translations are included. The project did not pay for printed publication of these; some would require further editing before that would be feasible. However, the existence of ANY list of proverbs can be a starting point from which a researcher may eventually produce a significant work."

This short list here also points to the fact that in many languages, no written collections exist. The Project assembled information on the languages having some published collections, the information about which is in the CD-ROM.

2.5 A Bibliography of Collections of Proverbs in Articles and Books

One major attention of the Project was to search for and assemble paremiographical works on African proverbs. Dr Nussbaum took charge of this area. Prof. Wolfgang Mieder's *African Proverb Scholarship* and the yearly information in *Proverbium: Yearbook of International Proverb Scholarship* (1984–) of which he is editor, gave a boost to this dimension of the Project. Prof. Mieder's book has 279 entries. Dr Nussbaum found more publications, and by the time he put out the CD-ROM on the Project, the bibliography had grown to almost one thousand entries (999). The Project committee was astonished and overjoyed to find this wealth of collections and references. Yet, it was also clear that the list was not exhaustive, as it may have missed other publications especially in Spanish, Portuguese, and local African languages.

3. Outcome and Follow-up of the Project

The African Proverbs Project achieved a number of specific results, some of which took shape during the three-year period of the grant from the Pew Charitable Trusts. We summarise them as follows:

3.1 Printed Publications

We have indicated the five publications in book form under the African Proverbs Series. This small selection invites further work in the area of paremiography.

Two other publications under the same umbrella are the bibliography by Prof. W. Mieder and the conference report edited by Prof. W.A. Saayman — both mentioned above. Mieder's book looks at what has been written, up to the time he collected the information. Saayman's book contains an interdisciplinary examination of proverbs — the methodology in paremiology, the role of proverbs in society and religion. The essays in the book examine individual themes from different disciplines, thus opening up African paremiology for further exploration.

The three books for teaching and preaching purposes also are to serve pedagogical and homiletical ends. At the same time they relate African proverbs to Biblical studies and theological issues that are stirring in contemporary Christianity.

3.2 The CD-ROM

Dr Stan Nussbaum assembled and edited a wealth of information and material relating to proverbs. This was published in the form of a CD-ROM, *The African Proverbs. Collections, Studies, Bibliographies*, by Global Mapping International

(GMI), Colorado Springs, Colorado, 1996. Among other things, the CD-ROM contains some 28 000 proverbs, 17 reprints of collections, the books mentioned above, as well as selections from published collections, and a bibliographical list of 999 items. There are "Key Pages" that are scanned images of the title pages, contents, introductions, bibliographies, etc. from about 200 proverb collections. A list of 33 titles of available collections informs on books that can be purchased. Language maps for each African country show the state of proverb research in each language on a scale of four, from "no known documentation" to "well documented". There are surveys of languages and work done in them on proverbs. The CD-ROM contains a word search for proverbs on particular themes or words. Texts from the CD can be printed or imported to other documents. This is an indispensable tool for the study of proverbs.³

3.3 The Work on Proverbs Continues

The initial grant from the Pew Charitable Trusts for the African Proverbs Project was for three years (1993–1996), without extension or renewal. The Project took shape in different forms as indicated above, and produced results by way of international symposia, publications of books and the CD-ROM. However, work on the proverbs did not stop there. It continues, particularly in East Africa, where two major activities have taken root and grown.

(a) The Project arranged with academic institutions in the Ivory Coast, Ghana, Kenya and South Africa to continue the work. These centres are: The Missiology Department, University of South Africa (UNISA), P.O. Box 392, Pretoria 0001, South Africa; the African Studies Centre, University of Ghana, P.O. Box 66, Legon, Ghana; the Institute Catholique de l'Afrique de l'Ouest (ICAO), 08 B.P. 22, Abidjan 08, Ivory Coast; and The Library, Hekima College, P.O. Box 21215, Nairobi, Kenya. UNISA planned a newsletter of African proverb research. Hekima College has emerged as the leading centre of activity. It has made additional collections of "endangered proverbs" in some of the smaller languages of Kenya, and more are planned. A growing group of about a dozen interested Kenyan scholars is holding occasional meetings to review progress on proverb study. Fr Edward Murphy, the Hekima librarian, is coordinating the effort.

(b) *African Proverbs, Sayings and Stories* at <http://www.afriprov.org> is a website on the Internet, originated, edited and moderated by Fr Joseph Healey, M.M., an American missionary based in Tanzania. An expanding circle of scholars is also involved in this effort. The site publishes one proverb each month, with its explanation and analysis generally linked to Biblical parallels. International writers contribute the proverbs, the largest number being Americans. The website is also a forum

where people can post questions, or make comments connected to proverbs. It has book reviews, a few teaching resources and a 2002 *African Proverbs Calendar*, with a proverb for each day of the year.

4. Conclusion

The African Proverbs Project has sparked interest in individuals and institutions. It has encouraged (or challenged) them to pay closer attention to proverbs, to collect them, use them for pedagogical and preaching purposes, and eventually publish them as collections or in texts. It has stimulated academic scrutiny that may result in the interdisciplinary study of proverbs in various fields such as linguistics, history, philosophy, religion, sociology, literature and others. The two symposia indicated clearly that there are many persons all over Africa, working on proverbs in one way or another. They also brought to the forefront the two interrelated dimensions of collecting and studying proverbs (paremiography and paremiology). As the African soil is very rich in proverbs, these dimensions will occupy scholars for a long time. They go hand in hand, like the Kaguru (Tanzania) proverb describing how marriage gets sealed when both parties know each other: "Magongolo mamanyi gakigoweka sisingo" (Millipedes that know each other, coil their necks around each other).

It seems clear, that what started as a modest project continues to bear fruit much longer than its original life-span. In this way, the Project has acquired a significant and self-generating continuation. It is like the Sesotho proverb that says: "Sejo-senyane ha se fete molomo" (Little food can be shared by many).

Endnotes

1. Mieder, Wolfgang, 1994. *African Proverb Scholarship. An Annotated Bibliography*, Colorado Springs, U.S.A: African Proverbs Project. This work is also reproduced on the Project's CD-ROM. The publication contains summaries of 279 books, dissertations and articles, reproduced from the larger world-wide three volumes: Mieder, Wolfgang, 1982, 1990 and 1993. *International Proverb Scholarship: An Annotated Bibliography*. New York: Garland Publishing.
2. There is no indication of the publisher's name other than *Série 'Proverbes Africains du Projet D'Étude des Proverbes Africains'*. Copies of the book might be available through the two collectors whose addresses (in 2002) are: Évêque Laurent Nare, Diocèse de Koupéla, B.P. 51 Koupéla, Burkina Faso; and Abbé Professeur François Xavier Damiba, Grand Séminaire de Koumi, 01 B.P. 312 Bobo-Dioulasso 01, Burkina Faso.
3. The CD-ROM is distributed by Global Mapping International (website <http://www.gmi.org/store/index.html>) or an expanding network of African distributors (current list available by email from info@gmi.org).

Defining Formats and Corpus-based Examples in the General Ndebele Dictionary, *Isichazamazwi SesiNdebele**

Langa Khumalo, *Ndebele Lexicography Unit, African Languages Research Institute, University of Zimbabwe, Harare, Zimbabwe (langa@arts.uz.ac.zw)*

Abstract: In this article the writer evaluates the defining formats that were used in defining headwords in the first monolingual General Ndebele Dictionary, *Isichazamazwi SesiNdebele* (ISN). The emphasis in the ISN was on the concept of user-friendliness. The article establishes that defining formats in the ISN are a judicious mixture mainly of the defining formats of the Collins Birmingham University International Language Database (COBUILD) and of what has been referred to as *traditional* formats. The first part of this article is an analysis of the decisions taken by the ISN editors in formulating their defining formats. It assesses the COBUILD defining principle vis-à-vis its application in defining headwords in the ISN and the impact of this principle on the user-friendliness of the dictionary. It further discusses other formats, including the decision to retain *traditional* defining formats for defining headwords. One of the traditional defining styles agreed upon was that the editors were to give the hypernym in the case of semantic sets, and then to identify the concept being defined by specifying aspects that distinguish it from others of its type. The second part of the article evaluates the importance and use of the corpus in providing both definitions and examples for the ISN. However, it is further argued that since a corpus has to be "representative" in terms of size in order to be appropriately used as basis for such corpus-based dictionaries, the ISN editors whose corpus was relatively small, could not avoid relying on intuitive knowledge in constructing some examples.

Keywords: COBUILD, CORPUS, DEFINITION, DEFINING FORMATS, TRADITIONAL FORMATS, EXAMPLES, USER-FRIENDLINESS, COTEXT, CONCORDANCE.

Opsomming: *Definieerpatrone en korpusebaseerde voorbeelde in die Algemene Ndebelewoordeboek, Isichazamazwi SeSiNdebele.* In hierdie artikel beoordeel die skrywer die definieerpatrone wat in die eentalige Algemene Ndebelewoordeboek, *Isichazamazwi SesiNdebele* (ISN) gebruik is. Die klem in die ISN was op die konsep gebruikersvriendelikheid. Die artikel toon aan dat definieerpatrone in die ISN 'n oordeelkundige mengsel is hoofsaaklik van die definieerpatrone van die Collins Birmingham University International Language Database

* This article is a slightly adapted version of a paper 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.

(COBUILD) en van waarna verwys word as *tradisionele* patrone. Die eerste deel van hierdie artikel is 'n ontleding van die besluite geneem deur die redakteurs van die ISN by die formulering van hul definieerpatrone. Dit beoordeel die definieerbeginsel van COBUILD ten opsigte van die aanwending daarvan by die definiëring van trefwoorde in die ISN en die uitwerking van hierdie beginsel op die gebruikersvriendelikheid van die woordeboek. Verder bespreek dit ander patrone, inbegrepe die besluit om *tradisionele* definieerpatrone te behou vir die omskrywing van trefwoorde. Een van die tradisionele definieerstyle waaroor ooreengekom is, was dat die redakteurs die hiperniem sou gee in die geval van semantiese stelle, en dan die konsep wat gedefinieer word, te identifiseer deur die spesifisering van aspekte wat dit onderskei van ander van sy soort. Die tweede deel van die artikel beoordeel die belangrikheid en gebruik van die korpus in die verskaffing van sowel definisies as voorbeelde vir die ISN. Daar word egter verder geredeneer dat, aangesien 'n korpus "verteenwoordigend" moet wees in terme van grootte ten einde toepaslik gebruik te word as basis vir sulke korpusgebaseerde woordeboeke, die redakteurs van die ISN wie se korpus relatief klein was, nie kon vermy om op intuitiewe kennis te steun by die maak van sommige voorbeelde nie.

Sleutelwoorde: COBUILD, KORPUS, DEFINISIE, DEFINIEERPATRONE, TRADISIONELE PATRONE, VOORBEELDE, GEBRUIKERSVRIENDELIKHEID, KOTEKS, KONKORDANSIE

1. Introduction

A simple definition of a dictionary is a book or reference work that lists words found in a language in alphabetical order and gives their meanings or definitions, and these definitions may or may not include examples. It is, therefore, common or basic knowledge that a dictionary lists words and their meanings. The importance of a definition in a dictionary cannot, therefore, be overemphasized. It is an intrinsic component of a dictionary. In fact, defining can be considered as one of the most important aspects of the dictionary-making process. It is as a result of this rigorous process that users can interact with the definition and know what a particular word means as explained in it.

This makes defining formats — a format being "the way in which something is arranged and presented" (Sinclair 1990: 218) — very crucial in composing definitions. Defining formats are pertinent in so far as they help the compilers or editors to write clear, specific, systematic and culturally relevant definitions that make dictionaries user-friendly. These are formats that provide the editor with certain guidelines to follow in constructing a definition, hence they contribute to consistency in the treatment of related sets of headwords.

Examples can also be seen as an important part of the definition. This is so because they further help to illuminate the definition. With their aid the editor can write a clear, specific definition, which is culturally and contextually relevant. The article will, therefore, evaluate the importance and use of the corpus in providing both definitions and examples in the *Isichazamazwi SesiNdebele* (ISN). According to Sinclair (1991: 171), "a corpus is a collection of naturally-occurring language texts, chosen to characterize a state or variety of a language". The ISN is largely a corpus-aided dictionary. If instances of usage from

the corpus are studied, it is found that the surrounding words and phrases help considerably in determining the meaning. Each word, therefore, occurs in its own special environment, called its *cotext*, and by studying the cotext it can often be established which meaning is intended. Since concordances provide real text examples, hence have a status which made-up examples do not have, the ISN editors obtained most of their examples from the corpus. This, it will be argued, provided the ISN with "more natural" examples. However, since a corpus has to be "representative" in terms of size in order to be appropriately used as the basis for such corpus-based dictionaries, the ISN editors whose corpus was relatively small, could not avoid relying on intuitive knowledge in constructing some examples.

2. Definitions and Defining Formats in the ISN

As has already been stated in the introduction, a definition is an intrinsic component of any dictionary. A definition is itself "a statement of the meaning of a word" (Waite 1998: 163) which is distinct in outline. It explains the meaning and use of a particular word in a particular speech community. It thus projects the "world-view" of a people. Two facts result from these observations: firstly, that a mother-tongue speaker is usually best equipped to formulate effective definitions of words in a particular language; and, secondly, that it is a rigorous process that follows set principles so as to provide a brief, distinct and conceptually appropriate statement of the meaning of a particular word.

Defining can be considered as one of the most important processes in dictionary making. It is the process through which users may become more aware of the meanings and explanations of words. It is important for users to know the meanings of words so that they can use the words in a *prescribed* manner in both spoken and written media. A dictionary is prescriptive in so far as it distinguishes the formal and colloquial uses of any particular word. A set of defining principles is therefore of paramount importance during the defining stage so that the editors can write clear, concise, consistent, culturally relevant, and specific definitions.

Defining formats are very crucial in that they provide the editor, when defining a particular word, with a scheme to follow. It is a framework from which an editor can construct a definition. Defining formats are also important in that they contribute to consistency in the treatment of related sets of headwords. This view is accentuated by Gellerstam's (1993: 47) observation that "the development of definition formats is advantageous in that once a particular format has been agreed upon, a number of words can be defined with it". The ultimate goal of any defining guideline is to enable an editor to write what is widely referred to as a *user-friendly* definition.

The defining formats designed for the dictionary based on the Collins Birmingham University International Language Database (COBUILD) are well-known. The following are specifications for COBUILD defining formats as out-

lined by Moon (1987) and Hanks (1987):

- (a) the headword should appear in the definition,
- (b) the headword should be shown in a typical context of common usage because "context disambiguates" (Hanks 1987: 125),
- (c) definitions should consist of full sentences,
- (d) explanations should consist of two parts, with the first part showing the word in use and the second part explaining the meaning, and
- (e) the format should suggest a preference rather than a restriction.

For COBUILD, a good definition should be short, user-friendly, culturally relevant, consistent with others of its type and should use superordinate terms (Hanks 1987: 118). When a definition has been made simple and familiar, it helps the dictionary as a whole to be user-friendly. The COBUILD system demonstrates a marked departure from traditional lexicography which, according to Hanks (1987: 117), "heavily relies on the use of parentheses whose function is not based on the standard usage of ordinary written English". COBUILD has done away with the use of parentheses. These are not user-friendly since considerable sophistication is required on the part of the dictionary user to determine what precisely is meant by each set of parentheses.

Evident characteristics of the COBUILD system of defining formats are that:

- (a) definitions should be as natural as possible,
- (b) definitions should project the typical usage of a headword,
- (c) definitions should be easy to understand, and
- (d) a "dictionary as prose" (Hanks 1987: 118) should be created, hence there should be done away with the use of parentheses.

These guidelines are instrumental in organizing, formulating and presenting definitions that are user-friendly. There will be focused on the formats used by the editors of the ISN for defining the noun, verb and the ideophone in Ndebele.

The following are some of the COBUILD defining formats used for explaining the meanings of nouns (cf. Ndlovu 1998: 21):

- (i) **A(n) + Noun Headword is ...**, e.g.
A **pond** is a small area of water that is smaller than a lake, or
An **ostrich** is a large African bird that cannot fly.
- (ii) **Your + Noun Headword is/are ...**, e.g.
Your **father** is your male parent.

- (iii) **If you call someone a(n) + Noun Headword ...**, e.g.
If you call someone a **bitch**, you mean that she behaves in a very unpleasant way.
This format is used to indicate a rude and offensive sense of a word, in this case *bitch*.
- (iv) **Noun Headword is ...**, e.g.
Love is a very strong feeling of affection or liking for someone or something.

The following is a selection of three of the COBUILD defining formats for verbs (cf. Ndlovu 1998: 23):

- (i) **To + Verb Headword + Someone or Something means ...**, e.g.
To **destroy** something means to damage it so much that it is completely ruined or ceases to exist.
- (ii) **If/When you + Verb Headword ...**, e.g.
If you **jump** something such as a fence, you jump over it or across it.
- (iii) **If Someone or Something + Verb Headword ...**, e.g.
If someone or something **activates** something he/she or it spurs it into action.

In the style manual of the ISN, it is clearly stated that the COBUILD formats should be used for styling definitions.

Nouns in the ISN have been defined using the following five defining formats:

- (i) **Noun Headword is ...**, e.g.
indwangu bz 9. *Indwangu* yinyamazana elesiphongo, ifana lenkawu kodwa yona inkulu. FAN imfene. Isib. *Indwangu ziqede umumbu emasimini.*
(**baboon** A *baboon* is an animal that has a forehead, it is like a monkey but is bigger. Example: *Baboons ate and finished all the maize in the maize fields.*)
- (ii) **If you are + Noun Headword ...**, e.g.
iqhawe bz 5. Nxa *uliqhawe* empini uyabe utshengise isibindi lokukhalipha.
(**hero** If you are a *hero* in the army you would have demonstrated bravery and skill.)
- (iii) **You call someone a(n) + Noun Headword ...**, e.g.
isitha bz 7. Uthi umuntu *yisitha* sakho nxa lizondana.
(**enemy** If you call someone your (an) *enemy* you hate each other.)
As in the COBUILD, this format was mainly used to indicate a rude and

offensive sense.

- (iv) **In [specialised field, e.g. medicine, business, law, etc.] + Noun Headword ...**, e.g.
umethironi bz 1a. Kwezokwelapha *umethironi* ngunesi omkhulu.
 (**matron** In a hospital a *matron* is a very senior nurse.)
 This format situates the headword in the context in which it is typically used.
- (v) **Definition + Noun Headword**, e.g.
amachaphaza bz 6. Izibungwana ezimhlophe ezingakabi zinyosi ngama-*chaphaza*.
 (**larva** Small white insects at the stage before they become adult bees are *larvae*.)
 This format does not occur in COBUILD. Its use demonstrates, as has been stated in the introduction, that defining formats in the ISN are a judicious mixture of essentially COBUILD formats and some modifications of traditional formats.

Verbs in the ISN have a typical subject. Giving a typical subject is meant to contextualise the definition and makes it more lucid. The following formats have been used in defining verbs in the ISN:

- (i) **To + Verb Headword + Someone or Something means ...**, e.g.
-tshengisa sz mwa Ukutshengisa umuntu ulutho yikwenza ukuthi alubone.
 (**show** To *show* someone an object is to make them see it.)
 This format was used alternately with the following format which, in the ISN, is a slight modification of the COBUILD format:
- (ii) **To + Verb Headword + is to ...**, e.g.
-labhula sz mwa Ukulabhula yikufuqa ulutho engathi uyaluphosa ngesandla ufuna ukuthi luye khatshana ngoba lukunenga.
 (**shove** To *shove* is to push something as if throwing it contemptuously far away because it irritates you.)
- (iii) **If/When you + Verb Headword ...**, e.g.
-thenga sz mwa Nxa uthenga impahla uyabe uzibhadalela.
 (**buy** When you *buy* property you obtain it by paying money for it.)
- (iv) **If Someone or Something + Verb Headword ...**, e.g.
-sisela sz mwa Nxa umnumzana esisela omunye inkomo uyabe emgcinise ezinye zezinkomo zakhe.
 (**loan** If a rich person *loans* someone cattle he will be asking him to keep some of his cattle.)

The ISN formats are a judicious mixture of COBUILD and traditional defining styles, since the COBUILD dictionary was designed to meet the perceived needs of advanced users of English as second language, while the ISN targeted speakers of Ndebele as first language. Not every aspect of the COBUILD would therefore be relevant to the ISN, hence the need for modification. The ISN editors agreed that consistency in the treatment of items from the same word-class was to be maintained. For the definitions to be user-friendly, the defining style should have to reflect a Ndebele world-view. One of the traditional defining styles agreed upon was that the definers were to give the hypernym or class containing similar animals or plants, and then identify the concept being defined by specifying aspects that distinguish it from others of its type. An example of the treatment of such classes is as follows:

inyamazana bz 9. *Inyamazana* yisidalwa esilenyawo ezine esilesikhumba kanye loboya.
(**animal** An *animal* is a creature that has four legs and has hide and fur.)

Other animals would then be defined in the context of this definition, as is shown by the following definition of *baboon*:

indwangu bz 9. *Indwangu* yinyamazana elesiphongo, ifana lenkawu kodwa yona inkulu. FAN imfene. Isib. *Indwangu ziqede umumbu emasimini.*
(**baboon** A *baboon* is an animal that has a forehead; it is like a monkey but is bigger. Example: *Baboons ate and finished all the maize in the maize fields.*)

One other modification of the COBUILD for defining verbs in the ISN is **Infinitive + Copulative Form ...** which can be exemplified as follows:

-khala sz mwa ukukhala yikulila ikakhulu uphuma izinyembezi ngokuzwa ubuhlungu.
(To *cry* is usually to have tears coming out of your eyes as a result of feeling pain.)

The use of the phrase **Copulative Form ...** does not appear in the COBUILD.

The ISN had to develop definition formats for ideophones since the COBUILD dictionary does not give any definitions for the English ideophones **grr** (sound made by a lion) or **cockadoodle-doo** (sound made by a cock). To create a format for defining ideophones, the ISN had to modify the COBUILD formats for nouns and verbs. The following are formats that were used in the definition of ideophones:

- (i) **This is an ideophone that means ...**, e.g.
vumbu szk. Lesi yisenzukuthi esitsho ukuvela kolutho lungalindelwe.

Isib. *Sithe sisalibele vumbu umfana owayegalule.*

(appear suddenly This is an ideophone that means something appears suddenly and unexpectedly. Example: *When we were relaxed suddenly and unexpectedly there appeared a boy who had drowned.*)

- (ii) **If Something or Someone + Copulative Form + Ideophone Headword ...**, e.g.
phamu szk. Nxa ulutho lugcwele *phamu* luyabe lugcwele lwaze lwaganxa.
(full to the brim If something is *full to the brim* it would have filled its container without leaving space.)
- (iii) **Ukuthi + Ideophone Headword + Verb Form ...**, e.g.
bani szk. Ukuthi *bani* yikuphazima kombani kumbe ukukhanyisa ngokuphangisa.
(spark To *spark* is to appear as a sudden flash of lightning or a short burst of light.)

These formats comply with the specification for defining headwords in contexts of their common usage. They therefore clearly follow the COBUILD practice.

Note that, as Ndebele ideophones usually do not have English equivalents, the above can only be rough translations.

The ISN editors formulated the following as their defining principles, which are essentially those of the COBUILD dictionary:

- (a) definitions should be brief,
- (b) separate entries should be made for each sense,
- (c) the primary sense will determine the sense listed first, and
- (d) definitions should be long enough to explain the concept, but over-particularization should be avoided, e.g. concrete nouns should be defined just far enough to suggest the reality of the object.

The definitions in the ISN are therefore brief, but long enough to capture the main sense.

Words with the same form, but unrelated meanings (homonyms) were entered separately. The following are examples of the treatment of such words:

umuzi bz 3. *Umuzi* yindawo kumbe ikhaya elihlala abantu.

(home A *home* is a place or house in which people live.)

umuzi bz 3. *Umuzi* lichatha elimnyama emzimbeni umuntu azalwa elalo.

(birthmark A *birthmark* is a black spot with which a person is born.)

umuzi bz 3. *Umuzi* butshani obude obukhula ikakhulu emaxhaphozini.

(Umuzi is a tall grass which grows in very wet places.)

Definitions of words with different, but related meanings were given after the particular headword, beginning with the primary definition. The following can serve as an example:

impilo bz 9. 1. *Impilo* yisikhathi esiphilwa ngumuntu kusukela ezalwa aze afe. *UMahlangu waba lempilo ende ngoba waze wabhubha eselekhulu leminyaka.* 2. *Impilo* yindlela abantu abaphila ngayo. *Labana baphila impilo enhle ngoba bafundile.* 3. *Impilo* yikuphila kuhle emzimbeni komuntu engelamkhuhlane omkhathazayo.

(**life** 1. *Life* is the time that a person lives from his birth until his death. Example: *Mahlangu had a long life because he died when he had turned a hundred years old.* 2. *Life* is the way people carry on with their daily chores. Example: *Those people live a good life because they are educated.* 3. *Life* is physical good health without any pain or illness of one's body.)

Concrete nouns were also defined in such a way that the user will understand the structure of the real object referred to by the noun in question. The following is an example:

isihlahla bz. 7. *Isihlahla* yisithombo esihlanyelwayo kumbe esizimilelayo esikhula sibe side, silengatsha lamahlamvu.

(**tree** A *tree* is a tall plant that grows on its own or is grown, it has a firm trunk, branches and leaves.)

It is evident from this that definitions in the ISN are long enough to explain the concept but at the same time avoiding over-particularization.

3. Examples in the ISN

As source of data for linguistic analysis a corpus plays an important role in linguistic research. In lexicography particularly it plays an even greater role. Kennedy (1998: 88) observes that a linguistic corpus, in whatever form, is an important basis for more accurate and reliable descriptions of the structure and use of languages. Sinclair (1991: 171) has defined a corpus as "a collection of naturally-occurring language texts, chosen to characterize a state or variety of a language". It is a body of collected texts of the written or spoken word, which is stored and processed on computer for the purposes of linguistic research (Renouf 1987: 1).

Because of the importance of the corpus in any dictionary-making process, that is, for purposes of headword selection, sense selection and obtaining examples of natural language usage, the ISN team collected a corpus of about a million running words before compiling the dictionary. The Ndebele corpus at the moment comprises about 1,2 million words. The ISN editors benefited from

the Ndebele corpus in the following ways:

- (a) Each word occurs in its own special environment. This is referred to as the *cotext*. By studying the cotext the editors managed to distinguish various senses of each word in these different environments. This made a seemingly difficult task very simple by consulting the corpus for the various instances of the usage of a particular word. The senses of most headwords in the ISN were ascertained through this method.
- (b) Concordances provide real text examples. A text example demonstrates a particular word in a typical sentence in agreement with other naturally co-occurring words. According to Sinclair (1991), concordances have a status which made-up examples do not have. They are typical. The examples for most headwords in the ISN therefore came from the corpus. This provided the dictionary with more "natural" examples.

However, the ISN is not exactly a corpus-based dictionary, but rather a corpus-aided dictionary. Because of the limited size of the corpus at the time of the compilation of the dictionary, the corpus was not considered large enough to be "representative". Since a corpus has to be "representative" in terms of size in order to be appropriately used as the basis for corpus-based dictionaries, the ISN editors whose corpus was then relatively small, could not avoid relying on the intuitive knowledge of the language in constructing some examples. After all, it has been argued that constructed examples are bound to be more captivating and lucid. This observation is, however, open to debate. In the case of the ISN, it provided a balance by including both corpus examples and constructed examples.

Examples in the ISN were viewed as important components of the definitions. The ISN editors resolved that:

- (i) since examples are part of the definition, they be used only when necessary to complement the definition in clarifying the headword, that is, to illustrate meanings or uses,
- (ii) a maximum of one example be given for reasons of saving space, and
- (iii) like definitions, they also be short.

The examples in the ISN were therefore used only as far as they served to illuminate the definitions. Their inclusion was to achieve the goal of producing a user-friendly dictionary.

4. Conclusion

The article has outlined the defining formats for nouns, verbs and ideophones

that were used in the ISN. It is clear from this outline that the formats are essentially those of the COBUILD dictionary although there were very minor modifications especially with regard to verbs. The defining formats for the ideophone were based on the defining styles of the noun and verb. It can be concluded that the ISN mainly followed the COBUILD defining formats as stated in its style manual. While many examples were taken from the corpus to illustrate the meaning of the headwords in instances where it was really necessary, the editors also used their intuitive knowledge to construct some examples. This had the advantage that the example, just like the definition, would be short, direct and lucid. On the other hand, the use of corpus examples provided fresh and natural real text examples. A balance between these two sources can be considered to form a sound basis for examples.

References

- Gellerstam, M.** 1993. The Use of the Corpus. Chimhundu, H. (Ed.). *Report on the African Languages Lexical (ALLEX) Project Planning and Training Workshop*. Harare: University of Zimbabwe.
- Hadebe, S. et al. (Eds.)**. 2001. *Isichazamazwi SesiNdebele*. Harare: College Press.
- Hanks, P.** 1987. Definitions and Explanations. Sinclair, J.M. (Ed.). 1987: 115-136.
- Kennedy, G.** 1998. *An Introduction to Corpus Linguistics*. London/New York: Longman.
- Moon, R.** 1987. The Analysis of Meaning. Sinclair, J.M. (Ed.). 1987: 86-103.
- Ndlovu, F.** 1998. *Considering Factors that are Relevant to Adopting and Adapting COBUILD Defining Formats for the General Ndebele Dictionary*. Unpublished M.A. Thesis. Harare: University of Zimbabwe.
- Renouf, A.** 1987. Corpus Development. Sinclair, J.M. (Ed.). 1987: 1-40.
- Sinclair, J.** 1990. *Collins COBUILD Student's Dictionary*. London: Collins.
- Sinclair, J.** 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sinclair, J.M. (Ed.)**. 1987. *Looking Up: An Account of the COBUILD Project in Lexical Computing and the Development of the Collins COBUILD English Language Dictionary*. London/Glasgow: Collins ELT.
- Waite, M. (Ed.)**. 1998. *The Little Oxford Dictionary*. Oxford: Clarendon Press.

Les proverbes dans *Kongo Proverbs and the Origins of Bantu Wisdom* par Mukumbuta Lisimba

Ludwine Mabika Mbokou, *Département d'afrikaans et de néerlandais, Université de Stellenbosch, Stellenbosch, République d'Afrique du Sud*
(13550721@humarga.sun.ac.za)

Résumé: On emploie tous les jours des proverbes, des maximes, des devinettes et des dictons. Malgré les progrès techniques des dernières décennies en matière d'outils didactiques, le proverbe fait partie du quotidien des sociétés africaines en particulier et de celles du monde en général. Il reste un instrument d'éducation indispensable et vital pour les générations futures, et quiconque désire un conseil ou une orientation dans une certaine situation s'y réfère. Il fait partie intégrante de l'acte de communication.

À la lumière des analyses du Dr Lisimba de certains proverbes de la famille bantu, nous découvrons l'importance que revêt la dynamique du langage offert par les proverbes. Ils peuvent ainsi être employés comme données dans la compilation de dictionnaires. En effet, pour la discipline des sciences du langage qu'est la lexicographie, le traitement des proverbes et idiomes est d'une importance capitale parce que ces derniers sont le véhicule de la culture et des mœurs d'un peuple. Ainsi, pour mener à bien une telle entreprise, le lexicographe devra choisir le traitement à l'intérieur de la liste centrale, sinon à l'extérieur de la liste centrale, l'option la plus profitable pour lui restant une classification thématique qui tient compte non seulement de l'aspect sémantique du proverbe, mais aussi de son aspect formel.

Mots-clés: ANALYSE FORMELLE, ANALYSE SÉMANTIQUE, CLASSIFICATION THÉMATIQUE, DIDACTIQUE, DICTIONNAIRE, GENRE PROVERBIAL, IDIOMES, LEXICOGRAPHE, LEXICOGRAPHIE, LISTE CENTRALE, LITTÉRATURE ORALE, PRÉ-TEXTES, POST-TEXTES, PROVERBES

Abstract: *Proverbs according to Kongo Proverbs and the Origins of Bantu Wisdom by Mukumbuta Lisimba.* Proverbs, maxims, riddles and sayings are used every day. In spite of the technical progress during the last decades as far as didactic tools are concerned, the proverb forms part of the daily life of African societies in particular and those of the world in general. It remains an indispensable and essential educational instrument for future generations, and whenever someone wants advice or direction in a certain situation reference is made to it. It forms an integral part of the act of communication.

In the light of Dr Lisimba's analysis of certain proverbs of the Bantu family, the importance of the linguistic dynamics presented by proverbs will be discovered. They will therefore have to form part of the composition of dictionaries. In fact, for lexicography as discipline of linguistic science, the treatment of proverbs and idioms is of cardinal importance because they are the bearers of the culture and customs of a people. Therefore, to deal satisfactorily with proverbs, the lexicographer will have to choose a treatment either within or outside the main text, the most advantageous being

a thematic classification which allows not only for the semantic, but also for the formal aspects of the proverb.

Keywords: FORMAL ANALYSIS, SEMANTIC ANALYSIS, THEMATIC CLASSIFICATION, DIDACTICS, DICTIONARY, PROVERBIAL GENRE, IDIOMS, LEXICOGRAPHER, LEXICOGRAPHY, MAIN TEXT, ORAL LITERATURE, FRONT MATTER, BACK MATTER, PROVERBS

Résumé de l'ouvrage

Dans le domaine de la littérature orale, une attention particulière a souvent été portée aux proverbes. Il en résulte une abondante littérature dont la plupart des ouvrages sont des répertoires suivis et/ou précédés de brefs commentaires.

Tous les travaux effectués sur les proverbes peuvent être répartis en deux groupes. D'une part il y a ceux dont le contenu se limite à une liste de proverbes suivis de l'explication et/ou d'une traduction de ce dernier. Ce sont pour la plupart des sortes de dictionnaires. Puis il y a les travaux qui se focalisent sur la forme, le sens, l'origine et le contexte d'usage des proverbes. Les analyses effectuées dans ce genre de travaux sont beaucoup plus linguistiques et sont davantage portées vers un public d'initiés. En effet, ces analyses ne se limitent pas seulement à dresser une liste de proverbes pour en démontrer la beauté stylistique et culturelle, mais elles constituent un développement pragmatique et fonctionnel du genre proverbial. C'est dans le même ordre de pensée que plusieurs ouvrages ont déjà été écrits sur la question des proverbes africains en général, et de manière plus particulière sur les proverbes appartenant à la famille de langue Bantu, avec une option spéciale pour la traduction et la classification thématique. La majorité de ces ouvrages ont eu pour support différents proverbes de tradition Bantu.

Dr Lisimba, chercheur et directeur du Centre International de Civilisation Bantu (CICIBA) s'inscrit dans cette lignée, avec une approche quelque peu différente de celle de ces contemporains. De par les analyses entreprises dans son ouvrage *Kongo Proverbs and the Origin of Bantu Wisdom* (Proverbes du Kongo et origine de la sagesse Bantu), le Dr Lisimba apporte certains changements en élargissant la description du proverbe Bantu par une étude comparative et contrastive. Une telle innovation repose sur le développement de la méthode ethnolinguistique, permettant ainsi une rigoureuse analyse formelle et sémantique.

Dans sa préface, Vatomene Kukanda pense que la méthode utilisée par le Dr Lisimba ne montre pas seulement comment se forme le proverbe, mais il retrace aussi les origines naturelles et psychologiques de la métaphore et du symbole dans ce dernier (Lisimba 1999).

Approche proposée par l'auteur

Comme l'indique le titre de l'ouvrage, le champ d'étude couvert par le Dr Li-

simba est le domaine Bantu, plus précisément la zone Kongo. Son œuvre se compose de deux majeures parties. La première est théorique dans le sens où elle fait état de l'aspect fonctionnel et formel du proverbe. La seconde est un répertoire de proverbes de différents pays appartenant à cette zone. C'est dans la première partie de son ouvrage que sont examinées les caractéristiques fonctionnelles de ces derniers. Les deux sections de l'ouvrage constituent non seulement une approche nouvelle en ce qui concerne la description du proverbe dans le domaine Bantu, mais aussi une porte ouverte à toute autre discipline des sciences humaines et des sciences du langage. L'historien, le sociologue et l'anthropologue y trouveraient tout aussi bien leur compte que le sémanticien ou le syntacticien. En effet, chaque proverbe contient un thème dont il apporte une morale, un savoir.

Les analyses que fait le Dr Lisimba tiennent compte à la fois de la structure, de la fonction et du thème sous lequel le proverbe apparaît. Il choisit de faire une description des proverbes en basant son étude sur deux niveaux. Le niveau syntaxique où est compris l'aspect fonctionnel, et le niveau sémantique qui se concentre sur la structure fondamentale du sens du proverbe et sa fonction. La particularité qui en résulte se trouve dans le découpage en caractéristiques fonctionnelles et caractéristiques formelles. Ces dernières sont établies pour renforcer le choix de la classification thématique si souvent utilisée par les auteurs de travaux ayant trait aux proverbes. Il rend cette classification plus explicite dans le sens où il ne répartit plus seulement les proverbes dans les différents thèmes, mais il inclut des critères de classification d'ordre sémantique. Dr Lisimba distingue par ailleurs deux genres de proverbes avec lesquels il faut compter pour une étude explicite du sens des proverbes. La distinction qu'il pose a pour mérite d'amener à mieux cerner la signification des proverbes qui est souvent d'accès difficile.

En s'appuyant sur la définition donnée par l'*Oxford Advanced Learner's Dictionary*. Dr Lisimba présente le genre proverbial avec une identité double comprise entre sa forme et son sens. Il décrit ces deux critères comme étant cruciaux lorsqu'il s'agit de donner une définition générale du proverbe, et fondamentaux quant à l'aspect ethnolinguistique de l'approche.

C'est fort de cette ouverture que nous pensons que l'ouvrage du Dr Lisimba pourrait constituer une approche nouvelle et bénéfique dans le domaine lexicographique, en particulier en ce qui concerne le traitement des idiomes et proverbes dans un dictionnaire.

Notre attention pour cet article a été portée sur la première partie de l'ouvrage, dans laquelle nous avons trouvé des éléments essentiels permettant d'aboutir à des conclusions bénéfiques quant aux problèmes posés par les proverbes dans la compilation des dictionnaires. En effet, le lexicographe, face aux proverbes, se trouve confronté à plusieurs questions auxquelles il est parfois difficile de répondre. Questions qui sont étroitement liées à la définition du proverbe et à sa distinction avec les autres genres de la littérature orale tels que le mythe, la maxime, le conte ou la fable. Nous nous sommes donc intéressés

aux quatre premiers chapitres de cette première partie qui présentent la description du proverbe dans les deux niveaux d'analyses mentionnés en sus. Pour le lexicographe qui a affaire aux proverbes lors de la compilation d'un dictionnaire, une telle approche est favorable et bénéfique. Nous le démontrerons dans les lignes qui suivent en faisant une comparasion entre les principes préconisés par le Dr Lisimba et certaines démarches adoptées au cours de la compilation de dictionnaires.

Presentation des chapitres

L'auteur commence par une définition générale du genre proverbial. Cette définition implique la fonction et la forme du proverbe, ainsi que l'environnement dans lequel le proverbe est conçu, les destinataires et les utilisateurs de la communauté dont il est issu.

Le premier chapitre est un résumé du contenu de tous les chapitres qui suivront. Il sert d'introduction, et les différents points qui seront examinés tout au long des chapitres suivants y sont évoqués et clairement énoncés. Les chapitres deux et trois exposeront les analyses relatives à la classification thématique et la structure formelle du proverbe tandis que dans le chapitre quatre, l'aspect sémantique est développé.

Le proverbe

Dans les anciennes sociétés africaines à tradition orale, les proverbes constituaient un moyen d'instruction pour véhiculer toute la connaissance socio-culturelle de la communauté. C'est de l'expérience de la vie humaine que les proverbes tirent leur source. Aujourd'hui encore, ils sont façonnés et modelés par la communauté linguistique qui les produit. Ils naissent de la société, en sont le produit, et sont par la même occasion utilisés par cette dernière de sorte qu'au fur et à mesure que les époques changent, les proverbes restent les mêmes (tout au moins dans leur fond).

Dans ces sociétés, "savoir parler" se rapportait à faire un usage habile des proverbes, et ce dans quelques circonstances que ce soit. Aujourd'hui encore, ils s'utilisent comme moyen d'instruction dans le sens où ils s'emploient lorsqu'il s'agit de véhiculer un savoir, communiquer une morale, ou prodiguer un conseil. Les proverbes traduisent la culture d'un peuple, d'une communauté linguistique quelle qu'elle soit. Ils représentent un moyen d'instruction et véhicule un savoir, particulièrement pour les communautés d'Afrique. Ils renferment les coutumes et mœurs d'un pays aussi bien qu'ils constituent un héritage pour les générations présentes et futures.

De plus, lorsqu'il s'agit de réflexion sur les principes de vie, sur des situations très difficiles à appréhender, ou encore lorsqu'il s'agit de choix à faire, on est souvent à la recherche de sages paroles qui mettraient de la lumière dans

nos réflexions. Les proverbes par le savoir qu'ils véhiculent font jaillir la lumière quant à de telles situations. C'est dans ce sens que Akporobaro et Amovon (1994: 1) disent:

Of the varied literary forms, the proverb is the form which has proved itself to be of great continuing relevance to modern man. It has been, and remain a most powerful and effective instrument for the transmission of culture, social morality manners and ideas of a people from one generation to another.¹

Exemple 1:

- *Jette ton pain à la surface des eaux, avec le temps tu le retrouveras*; se dit pour signifier à une personne qu'il est toujours utile de faire du bien autour de soi (proverbe de Salomon);
- *Quand les singes et les écureuils s'amuse sur les branches, les brindilles tombent sur le dos des éléphants*; se dit pour signifier que ce sont toujours les plus grands qui reçoivent les coups lorsqu'il y a des bévues (proverbe gabonais);
- *Le léopard inspire toujours la crainte dans la forêt même lorsqu'il a perdu ses dents*; se dit pour signifier qu'un chef reste un chef quelles que soient les circonstances (proverbe bakongo);
- *Celui qui monte sur un arbre doit savoir distinguer une branche sèche d'une branche verte*; se dit pour avertir à la prudence, ou pour inciter à la réflexion avant un projet (proverbe yipunu);
- *L'abeille est honorée parce qu'elle travaille non pour elle seule, mais pour tous*; se dit pour marquer les bénéfices de la générosité et du partage (proverbe chinois).

Dr Lisimba, dans son ouvrage, définit le proverbes comme étant essentiellement un état destiné à être appliqué à une situation sociale, dans un contexte présent. Maputa Mboukou, pour sa part, le définit par rapport à sa fonction dans la littérature orale: Parce que les proverbes traduisent les réalités de l'activité humaine par rapport à des situations actuelles et ponctuelles, il les décrit comme une situation qui véhicule une morale, mais à cette différence que cette situation n'est ni épique, ni une légende, mais un vécu.

Un proverbe ne peut donc être compris et/ou utilisé que dans un contexte bien précis. Les utilisateurs tout comme les destinataires de ce dernier doivent en connaître le code. Les auteurs qui travaillent sur ce type de littérature doivent savoir comment en présenter le code afin de permettre aux lecteurs de le déchiffrer. À cette fin, ils doivent tenir compte du contexte d'utilisation dans lequel le proverbe est utilisé, de sa forme, et du genre auquel il appartient. En effet, on distingue plusieurs forme de proverbes qui vont des plus complexes aux simples proverbes phrases. Les proverbes phrases sont les plus facile à appréhender. Ils se retrouvent le plus souvent sous forme de phrases éliptiques et leur structure profonde diffère très peu de leur structure superficielle (Lisimba 1999: 77-88). Certains de ces proverbes sont utilisés pour exprimer des états de faits et sont des assertions soient positives, soient négatives.

Exemple 2:

- *Kinzu kiabila, nkuni* (proverbe bakongo)
La marmite boue à cause du feu de bois
- *Kisadi ni kidie* (proverbe bakongo)
La personne qui travaille beaucoup sera récompensée
- *Lela mboma, mboma akuminyi* (proverbe lunda)
Élevez un python, et il vous avalera un jour
- *Buzoza bwa yoka, nsongo* (proverbe bakongo)
Trop d'ignorance rend malade
- *Bole bantu* (proverbe bakongo)
Il faut deux personnes pour faire un peuple

D'autres sont utilisés pour exprimer des cautions et des réprimandes. Dans ces derniers, la forme éliptique tend à disparaître. Les sociétés africaines utilisent ce type de proverbes pour parler des thèmes liés à l'éducation de l'enfant et la conduite d'une femme ou d'une jeune fille.

Exemple 3:

- *A crab does not beget a bird* (proverbe nigérian)
Une crabe ne donne pas naissance à un oiseau
- *Kgomo go gatana tsa saka le le lengue* (proverbe tswana)
Seules les vaches d'une même ferme se marchent dessus
- *Musiru u wara na miri* (proverbe yipunu)
La forêt se perpétue par ses arbres
- *Mwana gè dji, mu bè odjougala* (proverbe yipunu)
L'enfant qui refuse de manger doit être conduit dans un endroit désert
- *Mwana wa munyako ngwako* (proverbe tumbuka)
L'enfant de ton voisin est aussi le tien

Ces deux types de proverbes appartiennent à la première distinction faite par le Dr Lisimba, à savoir, les proverbes dont le thème est explicite à l'intérieur de la structure de surface. Ces derniers sont le plus souvent utilisés dans les dictionnaires comme exemples et illustrations. Ils ne nécessitent pas un traitement trop particulier car leur portée sémantique sert à renforcer le sens de notion abstraite comme la patience, la joie, l'amour.

Quant aux proverbes les plus complexes, ils sont ceux qui expriment des conseils et des directives. Leur structure superficielle est très différente de la structure profonde. Ils peuvent être constitués d'une ou de plusieurs phrases.

Exemple 4:

- *Ntsila gu saka dji gwè gnangu, go didjulu mabaga* (proverbe yipunu)
En suivant un sentier en plein jour, on en reconnaît pas les bifurcations

- *Bóka gnama gu kè la biburu, pa gu ma bóka mutu, go la biburu* (proverbe yipunu)
Tuez un animal et vous en verrez les propriétaires, tuez un être humain vous ne verrez personne
- *Mutu na mutu bunono mu mumbali* (proverbe liwandji)
La force d'un homme se trouve en relation avec un autre
- *Ongo omori are bongin'igomi nyimanga* (proverbe nkomi)
Une main ne peut pas cueillir dix noix de coco
- *Bari bwolo ngo ha dza nyi* (proverbe téké)
Deux hommes ne peuvent pas être mangés par un léopard

Dans les pages introductives d'un ouvrage sur les proverbes nigériens intitulé *Nigerian Proverbs: Meaning and Relevance Today*, les auteurs posent comme objectif principal la présentation du proverbe comme instrument actif de la société nigérienne. Le succès de ces derniers s'explique par la pensée qui fait du genre proverbial un fait démodé, dépassé. Pourtant, tous les jours, il se dit des proverbes, des dictons, des devinettes, des maximes (Maloux 1980: I). Malgré le développement des temps modernes, les proverbes ont traversés les âges et les époques. Et même la dépravation que subissent les mœurs dans ces dernières décennies apparaît comme un tremplin pour la promotion et l'utilisation des proverbes dans la vie communautaire, en général, et chez les jeunes en particulier. Les proverbes dans ce sens sont comme des gardes-fous.

Dr Kuzwayo, professeur de littérature orale sud-africaine, parle des proverbes comme une solution à la montée de violence et de délinquance juvénile à laquelle fait face son pays. Pour elle, ils sont les garants de la culture et de l'héritage des peuples. En tant que tel, ils protègent les valeurs morales et se chargent de les retransmettre aux générations futures.

As I grappled with the frightening violence and I racked my mind for remedies, I had to conclude that any effective remedy would have to combine a variety of solutions. And the language of proverbs struck me as one of the instruments, which could help.²

Dans la société africaine comme dans toutes autres sociétés, l'aspect géographique et environnemental influence les caractères utilisés dans les proverbes. Les personnifications d'êtres appartenant soit au règne animal, soit au règne végétal sont très prisées. Elles traduisent l'activité humaine de la communauté linguistique dans et pour laquelle les proverbes sont conçus tout aussi bien que l'environnement, le climat, le relief, etc. On retrouvera ainsi des personnifications d'animaux de la savane, de la forêt, du désert, etc. Même l'aspect religieux sera pris en compte. Par exemple au Nigéria et au Cameroun, les proverbes utilisés par les populations du nord auront une influence provenant de l'islam parce qu'elle domine sur le plan culturel dans ces régions. Ainsi, le contenu d'un proverbe dépendra non seulement de la communauté linguistique qui l'utilise, mais aussi des mythes, des croyances et coutumes, et de l'environne-

ment géographique de cette dernière. Pour bien comprendre et/ou expliquer un proverbe, il faut donc tenir compte de tous les aspects sus-mentionnés. En effet, la portée sémantique contenue à l'intérieur du proverbe ne saurait être appréciée dans sa totalité. Car n'oublions pas que c'est la vie de toute la société et les comportements humains qui sont représentés dans les proverbes.³

Un autre point à ajouter est que tout proverbe, ou du moins certains d'entre eux, porte un double manteau. Il peut véhiculer une assertion à la fois positive et négative par rapport au fait que les peuples diffèrent dans leur coutume les uns des autres. Par conséquent, la compréhension des différents concepts traités dans les proverbes peut être différente d'une communauté à une autre, mais la fonction reste la même. Ils véhiculent tous une morale, un conseil, une sagesse, un savoir. C'est dans ce sens que Akporobaro et Emovon (1994: 10) disent:

The proverbs are crystallized forms of human experiences which are extremely diversified. Since there is no single solution to any one human problem, proverbs are not offered as unique inviolable solution. (...) they present very clearly possible ready-made answers and advices to the multiplicity of human situations.⁴

Exemple 5

- *Dress up the little piece of wood, it will become pretty* (proverbe nigérian)
Habillé, un vulgaire morceau de bois paraît beau.

Pour les nigériens, ce proverbe est employé pour montrer le ridicule, ou pour tourner en dérision. Dans la culture européenne par contre, il invite au courage, à l'optimisme. Tout peut être beau, il suffit de le vouloir. Ce proverbe se lira donc selon deux points de vue. D'une part, le point de vue négatif du nigérian qui l'utilisera pour montrer le grotesque d'une jeune fille laide qui prétend à la coquetterie, d'autre part, celui d'un européen pour qui ce sera le moyen de démontrer un tout autre point de vue: Un clochard peut paraître repoussant dans ses loques, mais vêtu de manière décente, il devient un prince charmant.

La double charge que peut porter le proverbe selon les cultures se retrouve aussi dans la symbolique de l'imagerie. Un tel animal peut représenter la force dans une société alors qu'il représente tout le contraire dans une autre.

Les proverbes dans les dictionnaires

Dans le domaine de la lexicographie pratique, le traitement des proverbes, s'il n'est pas bien organisé, entraîne une confusion chez le lecteur, beaucoup plus qu'il ne contribue à l'aider. Certains lexicographes, pour éviter ce genre de problèmes, préfèrent ne pas du tout les placer comme données (les proverbes sont des données de nature complexe). D'autres par contre les utilisent soit comme exemple, soit comme sens polysémique d'un lemme, voir même comme syno-

nyme. La confusion qui peut se créer dans l'esprit du lecteur vient du fait que les proverbes sont traités dans les dictionnaires sans indications aucunes.

Tout proverbe se rapporte à un thème, un sens bien donné. Le recours à une classification thématique paraît alors plus judicieux pour le traitement des proverbes. Cette classification se fait au préalable lors du décorticage des informations par le lexicographe. Elle constitue ainsi un début de réponse à la question de savoir quelle place donner aux proverbes et idiomes dans un dictionnaire. Partant du genre de classification proposée par le Dr Lisimba, nous avons essayé de proposer une adaptation qui permettrait d'améliorer les deux approches couramment utilisées en lexicographie pratique. Le lexicographe a le choix entre:

- traiter les proverbes à l'intérieur de la liste centrale, ou
- traiter les proverbes dans la structure textuelle de l'ouvrage.

Les dictionnaires ont longtemps été utilisés comme instruments dans l'acquisition et l'usage correct du langage. L'un de leurs attributs est de donner une somme d'informations aux usagers pour leur permettre d'approfondir leurs connaissances. Ce sont des ouvrages de référence. Ils ont toujours joui d'une certaine autorité car ils permettent *l'usage des mots, des constructions et des sens* d'une langue *en les intégrant dans la communauté*. Ils leur donnent ainsi le droit d'exister ou les condamnent en rejetant leur usage (Dubois et Dubois 1971: 51). Ils représentent par là même, la norme d'une langue. Cependant, ils demeurent moins rigides que les grammaires traditionnelles. Cette réalité du dictionnaire dans une communauté linguistique fait ressortir son caractère pédagogique. En effet, le système éducatif d'une société a pour objectif d'amener tout individu à se fondre dans une certaine culture, garant de la société dans laquelle il est appelé à vivre. Le dictionnaire transmet alors une série de règles avec lesquelles il doit compter pour être reconnu comme membre de la communauté linguistique à laquelle il appartient.

De plus, le langage que véhicule le dictionnaire est par essence didactique. Il contribue donc à l'instruction de ceux qui l'utilisent.

Le dictionnaire est un outils fiable quant à la standardisation des langues. Par son aspect normatif, il en impose aux usagers parce qu'il offre, ou encore propose les règles d'usage des différents items lexicaux par leur traitement macro- et microstructurel. Conséquemment, le traitement et l'utilisation des proverbes dans les dictionnaires ne servent pas uniquement à préserver les proverbes, ou à les utiliser comme exemples, mais ils contribuent aussi à l'aspect pédagogique inhérent au fait de dictionnaire. Le lexicographe prévoit ainsi pour les générations présentes et futures un moyen d'instruction qui, dans l'Afrique moderne, tend à disparaître. Le dictionnaire permet alors de mettre à la disposition de tout locuteur, natif ou non-natif, un savoir universel. Car, bon nombres de proverbes partagent les mêmes thèmes, l'activité humaine étant la même de part et d'autre du globe terrestre. La portée sémantique et fonctionnelle du proverbe reste donc inchangée d'un continent à un autre, tandis que l'aspect formel varie. Par ailleurs, dans le dictionnaire, les informations qui sont

traitées peuvent être d'ordre géographiques, historiques, voire même dialectal, la langue étant considérée comme la vision de la communauté linguistique qui la parle.

Les proverbes à l'intérieur de la liste centrale

Le traitement des proverbes comme exemples n'est pas une approche nouvelle, bien au contraire, c'est l'une des plus utilisées par les lexicographes. Elle revêt un caractère pratique en ce qu'elle satisfait au souci d'économie du lexicographe, c'est-à-dire, mettre le maximum d'informations à la disposition de l'utilisateur, sans pour autant "gonfler" le contenu du dictionnaire. Elle constitue aussi un moyen de garage en ce sens qu'elle peut éviter au lexicographe de fournir trop d'explications quant à la portée sémantique du proverbe traité. Ce qui, dans certains cas, peut porter à confusion. Par exemple, dans les prétextes du *Micro Robert* (dictionnaire d'apprentissage du français), il est spécifié que les proverbes sont donnés en exemple et traités avec des étiquettes parce qu'il est nécessaire de connaître certaines tournures. Ils sont ainsi différenciés des exemples usuels que les auteurs de ce dictionnaire nomment *exemples libres*.

Il est aisé pour certains proverbes d'en retrouver le sens. Pour d'autres par contre, il faut toute une somme d'informations pour arriver à en déchiffrer le code. Dans le traitement de ce genre de proverbe, le sens se trouve souvent confondu avec le sens premier du mot clé, et semble parfois sortir de nul part. Le locuteur ne sait pas si il a affaire à un sens polysémique, à un synonyme, à une locution, etc.

Maloux (1980: X) écrit que "tous les jours il se dit des proverbes". Pour lui, "ils jaillissent de la verve populaire". C'est cette même "verve" qui est le dénominateur commun entre les usagers et le dictionnaire. Ces derniers entendent et parlent le langage que le lexicographe est chargé de rendre dans le dictionnaire, parfois en rejetant, ou en approuvant l'usage. Et ce, de la manière la plus accessible qu'il soit. Le dictionnaire a aussi pour but de développer le langage de ceux qui le consultent. En utilisant la distinction proposée dans l'ouvrage du Dr Lisimba, le lexicographe peut ainsi sélectionner les proverbes les plus faciles d'accès, et souvent les plus fréquents, et les utiliser pour illustrer les définitions des notions abstraites telles l'amour, la patience, la joie, etc. Ce procédé est souvent utilisé dans les dictionnaires scolaires.

Les proverbes, comme mentionné en sus, prennent leur source dans l'observation du règne animal et végétal, et dans l'environnement immédiat de la communauté auxquels ils appartiennent. Dans le proverbe "petit à petit l'oiseau fait son nid", comme illustration de la définition du lemme "nid", il est à noter que le sens du proverbe n'a rien à voir avec la définition même du mot, mais plutôt avec le comportement de l'animal face à la confection de son habitat (ce dernier use de beaucoup de patience et prend le temps de chercher les brindilles qui conviennent à un tel ouvrage). Par contre, comme illustration ou exemple du lemme "patience", ou celui de "persévérance", il serait plus facile à l'utilisateur non seulement de comprendre son sens, mais aussi de savoir l'utiliser dans

un contexte approprié. D'où l'intérêt d'un traitement thématique tel que celui proposé par le Dr Lisimba.

Nous tenons à signaler qu'il est aussi possible d'avoir recours à une procédé autre que les exemples pour le traitement des proverbes à l'intérieur de la liste centrale. Il consiste à ouvrir une zone dans l'article de façon à traiter tous les proverbes découlant de l'item lexical retenu comme lemme dans la structure du dictionnaire. Cette approche est très prisée dans la compilation de dictionnaires élargis parce qu'elle est beaucoup plus appropriée. En effet elle permet de donner un nombre considérable d'informations sur un traitement "complet" de la langue étudiée. Ces informations qui doivent nécessairement être fournies par ce type de dictionnaire parce qu'il a pour but de traiter le vocabulaire de la langue étudiée dans sa quasi-totalité, tout en permettant d'en comprendre le fonctionnement. Ils sont plus descriptifs que la plupart des autres types de dictionnaires.

Bien que traités sous la bannière d'un lemme, les proverbes dans cette procédure le sont par rapport à leur sens. Nous n'avons donc pas choisi de présenter un examen complet de cette approche parce qu'elle ne saurait être adaptée aux analyses qui ont retenu notre attention. En outre présentent-elles très souvent un examen complet des proverbes dans les dictionnaires élargis.

Les proverbes dans la liste centrale peuvent encore être utilisés comme substitut dans la définition d'un lemme. Surtout lorsqu'il s'agit d'un terme abstrait. Mais dans ce cas comme dans celui du traitement séparé, il convient à un type bien défini de dictionnaire. Il est préférable de l'utiliser dans le cadre des dictionnaires bilingues et des dictionnaires de débutants.

Les proverbes à l'extérieur de la liste centrale

Pour cette approche, le procédé est le même que pour le traitement séparé à l'intérieur la liste centrale. La différence se trouve au niveau de la zone. Cette dernière se trouve généralement en fin de dictionnaire dans les post-textes, et non plus à l'intérieur de la liste centrale. Le plus souvent, cette zone est le lieu où le lexicographe choisit de placer les éléments de culture générale (cartes, noms de villes/pays, noms propres, citations, etc.) qu'il juge nécessaire à la connaissance du lecteur. Il les traite de manière séparée pour ne pas surcharger le contenu de la liste centrale et/ou en souligner l'importance. Sur cette lancée, les proverbes se trouvent dressés en répertoire et toujours par ordre alphabétique.

Exemple 6

À cœur vaillant rien d'impossible, avec du courage, on vient à bout de tout.

À la Sainte-Luce, les jours croissent du saut d'une puce, les jours commencent à croître un peu à la Sainte-Luce (autref. 13 décembre, auj. 23 décembre).

À l'impossible nul n'est tenu, on ne peut exiger de quelqu'un ce qu'il lui est impossible de faire.

Bien mal acquis ne profite jamais, on ne peut jouir en paix du bien obtenu par des voies illégitimes.

Bon chien chasse de race, on hérite généralement les qualités de sa famille.

Bonne renommée vaut mieux que ceinture dorée, mieux vaut jouir de l'estime publique que d'être riche.

Il n'y a que la vérité qui blesse, les reproches vraiment pénibles sont ceux que l'on a mérités.

Il n'y a que le premier pas qui coûte, le plus difficile en toute chose est de commencer.

Loin des yeux, loin du cœur, l'absence détruit ou affaiblit les affections.

Les loups ne se mangent pas entre eux, les méchants ne cherchent pas à se nuire.

Mauvaise herbe pousse toujours, se dit pour expliquer la croissance rapide d'un enfant de mauvais caractère.

Qui aime bien châtie bien, un amour véritable est celui qui ne craint pas d'user d'une sage sévérité.

Rira bien qui rira le dernier, qui se moque d'autrui risque d'être raillé à son tour si les circonstances changent.

Si jeunesse savait, si vieillesse pouvait, les jeunes manquent d'expérience, les vieillards de force.

Tel père, tel fils, le plus souvent, le fils tient de son père.

Le temps, c'est de l'argent, traduction de l'adage anglais *Time is money*, le temps bien employé est un profit.

(extrait tiré du *Dictionnaire de Français Contemporain*)

Ce genre de procédure est beaucoup plus astucieux que de traiter les proverbes comme exemples, surtout lorsqu'il s'agit de dictionnaires pédagogiques. Elle permet en effet un traitement complet des proverbes, en donnant le maximum d'informations à tous les niveaux (informations d'ordre sémantiques, culturels, formels, etc.). Elle joue dans un certain sens le rôle de la mini-grammaire dans un dictionnaire. Cependant, dans un souci d'économie et pour un rendement meilleur de la production, il serait préférable pour le lexicographe, de laisser de côté l'ordre traditionnel alphabétique, et de procéder par classification thématique. Car, plusieurs proverbes peuvent partager le même thème; dans ce cas le contexte d'usage est l'un des facteurs qui permet de saisir la distinction sémantique de chacun. Puis, il y a des proverbes qui se retrouvent sous la bannière de plusieurs thèmes à la fois. Traiter les proverbes dans les post-textes d'un dictionnaire offre donc un plus large champ d'action. La présentation qui y est faite se révèle plus explicite, et s'avère plus claire et plus utile pour l'utilisateur.

De plus, cette façon de traiter les proverbes nécessite plus d'effort de la part du lecteur, et parfois même du lexicographe parce que le fait de partager le même "mot-clé" n'a généralement rien à voir avec la portée sémantique du proverbe (comme avec l'exemple de "nid" dans "petit à petit l'oiseau fait son nid"). Le lecteur doit alors s'engager dans un tri pénible pour retrouver l'information dont il a besoin (rappelons au passage que les usagers des dictionnaires ne lisent pas les directions d'usage proposées par les auteurs dans les pré-textes). L'exemple 7 qui suit en est une illustration. Les auteurs de ce dictionnaire bilingue de proverbes ne présentent aux usagers qu'une somme de proverbes liés au lemme principal. Par exemple, aucun contexte d'usage n'est proposé, et

la distinction entre le sens du proverbe et la définition du lemme, s'il y en a une, est inexistante. Les auteurs comptent avec l'intuition des usagers, ce qui est parfois trompeur dans le domaine de la lexicographie pratique.

Exemple 7

FATHER:

To be gathered to one's fathers: To die: *Tot jou vaders vergader wees.*

LIKE father, like SON; like mother, like daughter: Resembling the parents: *So vaartjie, so moertjie.*

FAVOUR:

To COUNT in one's favour: To be regarded as a merit: *Ten goede kom; tot voordeel strek; as verdienste toereken.*

To FIND favour with: To meet with approval: *Genade vind; instemming vind.*

IN favour of: On behalf of; in support of; to the advantage of: *Ten gunste van.*

To be OUT of favour: To have lost the goodwill of the others: *Uitgebak wees; in onguns wees.*

To WIN (general) favour: To become popular: *Gewild word; die goedkeuring wegdra; opgang maak.*

To favour SOMEONE: To treat someone with partiality: *Iemand begunstig; vir iemand party trek.*

FEATHER:

A feather in one's CAP: Something one may be proud of: *'n Veer in jou hoed.*

That is a feather in his CAP: That is something he can be proud of: *Dis 'n pluimpie vir hom.*

To have one's feathers CROPPED: To be humiliated: *Verneder word; sy stertvere kwyf wees.*

FINE feathers make fine birds: People are judged by their outward appearance: *Die klere maak die man; die vere maak die voël.*

To be in HIGH feather: To be in good spirits: *In 'n goeie bui wees; in jou noppies voel (wees).*

You could have KNOCKED me down with a feather: I was most surprised: *Jy kon my omgeblaas het; ek was skoon verbaas.*

As LIGHT as a feather: Having little weight: *So lig soos 'n veer.*

To show the WHITE feather: To be a coward: *Lafhartig wees; papbroekig wees.*

As you feather your NEST, so you must lie on it: You must accept what you have been responsible for: *Wat jy saai, sal jy maai.*

(extrait tiré d'*English Proverbs and Expressions with Afrikaans equivalents*)

Perspectives et conclusion

Un bon dictionnaire est un dictionnaire qui présente ses données de manière à ce que le l'utilisateur en comprenne le sens et l'utilisation. L'idéal serait pour le lexicographe que les usagers se réfèrent aux pré-textes pour savoir comment se

diriger dans la voie qui les conduiraient aux informations dont ils ont besoin. Bien de problèmes seraient alors vite résolus. Malheureusement, ce n'est pas toujours le cas. Il arrive même très rarement que les usagers passent par les pré-textes avant de consulter la liste centrale. Ils ne se préoccupent même pas de leur utilité. Cette situation est encore plus dramatique lorsqu'il s'agit de lecteur tels que les élèves, les étudiants. La présentations des informations à l'intérieur du dictionnaire revêt alors un caractère capital dans le travail du lexicographe.

Nous avons exposé dans nos propos l'esquisse d'une approche profitable des proverbes quant à leur traitement dans la lexicographie pratique. Les différents procédés examinés nous ont amené à proposer la classification thématique proposée par le Dr Lisimba comme un moyen plus tenable. Voici quelques avantages qui justifieraient notre choix:

1. Un plus large champ d'action. Par conséquent, le lexicographe est libre de donner des analyses plus détaillées sur la question sans devoir se limiter à de simples traductions ou définitions comme le font la plupart des auteurs sur les proverbes. Il a ainsi la possibilité d'envisager l'aspect culturel du proverbe en même temps que les aspects sémantiques et formels.
2. Résoudre certains problèmes posés par le choix de l'ordre alphabétique. Par exemple, quel est le mot sous lequel il faut placer le proverbe, ou encore sur quel mot s'appuyer s'il faut placer le proverbe au début comme dans les dictionnaires spécialisés de proverbes. Réduire la présentation des données. Les proverbes sont groupés sous différents thèmes bien précis, ce qui a l'avantage d'en retracer plus facilement le sens contrairement au procédé qui consiste à les mettre sous un "mot-clé".
3. Enfin, se servir de la classification proposée par le dr Lisimba permettrait de restaurer la fonction du proverbe, c'est-à-dire savoir *bien parler*, et par là même, présenter le discours poétique de la littérature orale, discours souvent méconnu du public.

En conclusion, il nous semble que le but poursuivit dans la composition de dictionnaires en général est non seulement d'informer les usagers sur le langage de leur communauté, mais aussi de leur donner des armes pour en faire de bon citoyens, sachant s'exprimer et respectant les valeurs de la société. Bien plus qu'un ouvrage didactique, le dictionnaire, de part l'autorité qui lui est conférée, est un mode d'emploi pour l'intégration des individus dans une société. Dans les sociétés traditionnelles africaines, les proverbes étaient utilisés pour transmettre le savoir, la connaissance, la sagesse. Bien intégrer ces derniers dans l'usage du dictionnaire restituerait certaines valeurs morales cor-

rompues depuis l'avènement de la modernisation, ce qui est un fléau parmi les jeunes de ces dernières décennies.

Notes de références

1. Parmi la variété de genres littéraires appartenant à la littérature orale, le proverbe est l'une des formes des plus pertinentes pour les temps modernes. Il est et demeure un puissant facteur de transmission de la culture, des valeurs sociales et morales, et des idées d'un peuple d'une génération à une autre.
2. Horrifiée par la montée terrifiante de la violence, et cherchant un remède, je conclus que tout remède effectif devrait combiner diverses solutions. Le langage des proverbes m'apparaît alors comme l'un des instruments qui pourrait aider.
3. En Afrique du Sud, les peuples sont généralement pasteur ou fermier. La vache est source de richesse et est symbole de pouvoir et d'autorité. On retrouvera donc une forte concentration de proverbes l'ayant comme objet principal, tandis que dans les pays d'Afrique centrale comme le Gabon, elle serait quasi inexistante. L'autorité et le pouvoir seront symbolisés par les animaux de la forêt, tels que le lion ou le léopard.
4. Les proverbes sont cristallisés par les diverses expériences humaines. Parce qu'il n'existe pas d'unique solution aux problèmes des êtres humains, les proverbes n'offrent ni d'uniques, ni de stables solutions. (...) Ils présentent par contre des réponses et des conseils précis et prêts à être utilisés.

Bibliographie

Dictionnaires

- Dubois, J. (Éd.). 1966. *Dictionnaire du Français Contemporain*. Paris: Larousse.
- Maloux, M. 1980. *Dictionnaire des Proverbes, Sentences et Maximes*. Montréal: Larousse.
- Rey, A (Éd.). 1988. *Le Micro Robert de Poche: dictionnaire d'apprentissage de la langue française*. Montréal: Icorobet.

Autres lectures

- Akporobaro, F.P.O. et J.A. Emovon. 1994. *Nigerian Proverbs: Meaning and Relevance Today*. Lagos: Nigeria Magazine.
- Dubois J. et C. Dubois. 1971. *Introduction à la lexicographie: Le dictionnaire*. Paris: Larousse.
- Kuzwayo, E.K. 1998. *African Wisdom*. Cape Town: Kwela.
- Lisimba, M. 1999. *Kongo Proverbs and the Origins of Bantu Wisdom*. Pésence Africaine.
- Prinsloo, Anton F. 1997. *Afrikaanse Spreekwoorde en Uitdrukings*. Pretoria: J.L.van Schaik.
- Raponda-Walker, A. 1993. *3 000 proverbes du Gabon*. Paris: Saint-Paul.
- Sabbagha, N.G. et M.S.B. Kritzinger. 1980. *English Proverbs and Expressions with Afrikaans Equivalents*. Pretoria: J.L. van Schaik.

The Systematic Development of Wiegand's Metalexigraphy as Demonstrated in *Kleine Schriften*

Maria Smit, *Department of Music, University of Stellenbosch, Stellenbosch, Republic of South Africa (msmit@sun.ac.za)*

Abstract: This review article deals with the contents of the publication *Kleine Schriften*, a selection of articles written by Herbert Ernst Wiegand and compiled by Matthias Kammerer and Werner Wolski. It purports to illustrate the topics with which Wiegand engaged himself over the years, and his gradual development over a span of thirty years. Wiegand has written on many topics, but in this article mainly two aspects are discussed. Firstly, Wiegand's spelling out of an "actional-semantic approach" to lexicography is explained. He investigates the works of many authors, sometimes drawing on them, and sometimes refuting their arguments. Secondly, Wiegand's construction of a detailed text theory for lexicographical texts is discussed. Wiegand is one of the most important theoretical lexicographers of our time, and his suggestions towards metalexigraphical terminology and distinctions form an essential part of present-day lexicographical practices and planning across the world.

Keywords: ACTIONAL-SEMANTIC APPROACH, DICTIONARY RESEARCH, METALEXICOGRAPHY, SEMANTICS, TEXT THEORY, UNUSUAL TEXTS, USUAL TEXTS

Opsomming: Die sistematiese ontwikkeling van Wiegand se metaleksikografie soos geïllustreer in *Kleine Schriften*. Hierdie resensieartikel handel oor die inhoud van die publikasie *Kleine Schriften*, 'n seleksie van artikels geskryf deur Herbert Ernst Wiegand, en versamel deur Matthias Kammerer en Werner Wolski. Dit poog om die onderwerpe waarmee Wiegand homself deur die jare besig gehou het, aan te toon, en sy geleidelike ontwikkeling oor 'n verloop van dertig jaar. Wiegand het oor baie onderwerpe geskryf, maar in hierdie artikel word veral twee aspekte bespreek. Eerstens word Wiegand se uiteensetting van 'n "handelingssemantiese benadering" tot leksikografie verduidelik. Hy bestudeer die werk van baie skrywers, soms deur van hulle sienings te gebruik, en soms deur hulle te weerlê. Tweedens word Wiegand se opbou van 'n gedetailleerde teksteorie vir leksikografiese tekste bespreek. Wiegand is een van die belangrikste teoretiese leksikograwe van ons tyd, en sy voorstelle vir metaleksikografiese terminologie en onderskeidings vorm 'n noodsaaklike deel van hedendaagse leksikografiese praktyke en beplanning regoor die wêreld.

Sleutelwoorde: GEWONE GEBRUIK, HANDELINGSEMANTIESE BENADERING, METALEKSIKOGRAFIE, ONGEWONE GEBRUIK, SEMANTIEK, TEKSTEORIE, WOORDEBOEKNAVORSING

1. Introduction

It is an almost impossible task to publish a two-volume collection of Herbert Ernst Wiegand's articles written through the years, and at the same time do justice to the wide range of topics he has dealt with. Wiegand's oeuvre currently comprises more than 390 titles. In *Kleine Schriften* Kammerer and Wolski collected 37 of Wiegand's articles written over a span of 30 years. They used distinct criteria in determining the selection of articles:

(a) Articles that were too long had to be omitted (for example, Wiegand 1977 and 1985) even though they may be some of his better known publications;¹

(b) No articles from the series of comprehensive *Handbücher zur Sprach- und Kommunikationswissenschaft* were included (for example Wiegand 1984a, 1989a, 1989b, 1989c, 1989d, 1989e and 1990), which actually form the axis of Wiegand's metalexigraphical theory;

(c) No articles that Wiegand wrote together with other authors were included. Round about 80 articles were published in this way through the years, for example, Henne and Wiegand 1969, Ripfel and Wiegand 1988, Hausmann and Wiegand 1989, and Konerding and Wiegand 1994, and those mentioned on p. XXV of *Kleine Schriften* in connection with bilingual dictionary projects;

(d) Series such as *Lexicographica Series Maior* and *Studien zur neuhochdeutschen Lexikographie* (cf. p. XXIII) had been omitted;

(e) And, of course, it was not possible to include the contents of the comprehensive first volume of Wiegand's *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie* (1998).²

Even though Kammerer and Wolski published the articles in chronological order, they explain that recurring themes in Wiegand's work through the years can be traced in the two volumes. Their aim is not so much to honour a specific person and his individual accomplishments, although it is clear to anybody working in the field of dictionary research that Herbert Ernst Wiegand made an invaluable contribution in this regard. It is also not a general orientation to Wiegand's works. They rather want to indicate and unite in one publication different directions pointed out by Wiegand in course of time which reflect trends of thought in linguistics, especially with regard to sub-disciplines such as pragmatics and text linguistics. They strive to disclose Wiegand's connections with research traditions in linguistics by selecting articles that demon-

strate the developments which took place in dictionary research and Wiegand's positioning of himself within the wider field of linguistics (Kammerer and Wolski 2000: X-XII). For example, during the 1970s he developed arguments with regard to the use of semantic features and theories about the so-called "lexicographic definition", in the process making use of many references to the work of other linguists. Kammerer and Wolski (2000: XXX) also remind the reader that Wiegand's most recent works link up with newer conceptions in the philosophy of linguistics, the cognitive sciences and cognitive psychology (frames, scripts, scenarios, prototypes, stereotypes, etc).³

Through the years Wiegand has specifically excelled at dealing in a critical way with various diverging and controversial theories, and usually demonstrated his acquired insights with carefully chosen examples, sometimes even revising his own former conceptions in reaction to other scholars' work (p. XII).⁴ One of Wiegand's favourite approaches is to construct his arguments by looking critically at the work of other scholars, using citations from their work, in many cases to refute or correct their arguments.⁵ He then usually engages in complicated reasoning, finding solutions and making suggestions on which he can build his comprehensive theory of lexicography. This *modus operandi* is also perceptible in *Kleine Schriften*.

According to their introduction, the various themes Kammerer and Wolski distinguish in *Kleine Schriften* are Wiegand's investigations into (i) semantic theories including "actional-semantic" theory,⁶ (ii) practical lexicology, (iii) text linguistics, (iv) metalexigraphy, (v) dictionary typology, (vi) special-field lexicography, (vii) bilingual lexicography, (viii) dictionary research, and (ix) computer lexicography. They also included several isolated articles, such as "Dialekt und Standardsprache im Dialektwörterbuch und im standardsprachlichen Wörterbuch" (1986) and "Der frühe Wörterbuchstil Jacob Grimms" (1986), which they claim do not belong under any of the above-mentioned topics.

In order to unlock the contents of *Kleine Schriften* more comprehensibly, this article mainly focuses on two themes with which most of the selected articles can be associated. These themes are (i) Wiegand's views on meaning, or his "actional-semantic approach", and (ii) his text theory for lexicographical texts. The main aim is to demonstrate Wiegand's argumentation and gradual development of these two basic themes through the years, and how his entire lexicographical theory is based on these two important principles.

To place the articles of *Kleine Schriften* within the greater framework of Wiegand's theory, two illustrations are given. Figure 1 presents an overview of Wiegand's metalexigraphical theory. Figure 2 features a list of the topics treated in *Kleine Schriften* (an adapted exposition of Kammerer and Wolski's above-mentioned classification) and the particular components of Wiegand's theory to which the topics belong. Since some articles deal with more than one topic, Figure 2 therefore only gives a very broad classification.

Figure 1

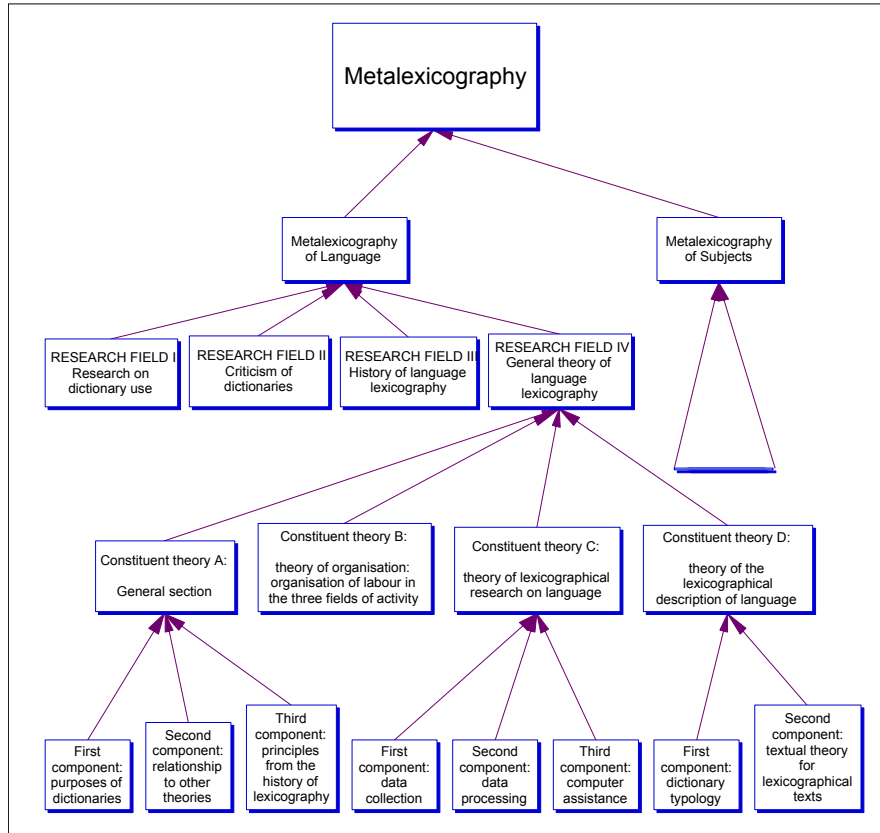


Figure 2

TOPICS INCLUDED IN <i>KLEINE SCHRIFTEN</i>	ARTICLE NUMBER(S)	COMPONENT WITHIN WIEGAND'S THEORY
Semantic Theory	1, 2, 3, 37	Research Field IV: Constituent Theory A: Second component
Practical Lexicology	4, 8, 9, 18	Research Field IV: Constituent Theory A: Second component
Actional-semantic Theory	6, 14, 20, 22, 26, 30, 31, 37	Research Field IV: Constituent Theory A: Second component
Text Theory	5, 6, 7, 10, 15, 23, 24, 25, 27, 28, 29, 32, 34, 35	Research Field IV: Constituent Theory D: Second component

Dictionary Typology	11, 12, 13, 16, 17, 18, 19, 21, 25, 26, 27	Research Field IV: Constituent Theory D: First component
Special-field Lexicography	5, 7, 22, 26, 30	Research Field IV: Constituent Theory D: First component
Bilingual Lexicography	29, 32	Research Field IV: Constituent Theory D: First component
Computer Lexicography	15, 36	Research Field IV: Constituent Theory C: Third component
Learners' Lexicography	27	Research Field IV: Constituent Theory D: First component
Research on Dictionary Use	4, 14, 20	Research Field I
Metalexigraphy	10, 15, 25, 28, 33	Metalexigraphy of Languages
Criticism of Dictionaries	11, 12, 13, 19, 25	Research Field II

Of the 37 articles in *Kleine Schriften*, only the following are available in English: (i) numbers 15 and 24 (originally published in English); and (ii) numbers 3, 9 and 26 (included in Wiegand 1999).

2. Wiegand's "actional-semantic" approach

Even though Kammerer and Wolski (2000: XIII) assert that the first two articles in Volume 1 ("Synchronische Onomasiologie und Semasiologie. Kombinierte Methoden zur Strukturierung der Lexik" (1970) (pp. 1-98) and "Einige Grundbegriffe der lexikalischen Semantik" (1972) (pp. 99-124)) mainly have only historical value from a scientific point of view, it is interesting to note that the very last article in Volume 2 ("Mit Wittgenstein über die Wortbedeutung nachdenken. Gebrauch? Regel im Kopf? Ein Etwas im Kopf?" (1999) (pp. 1507-1552)), deals with exactly the same issues. It is clear that the compilers connect Wiegand's most early contributions with his most recent ones, and rightly so.

In the first two articles of Volume 1 of *Kleine Schriften*, Wiegand uses the "structuralist terminology" popular during those days, but already also starts forming new, more appropriate terminology specific to the study of lexicographical works. Terms he uses, are, for example, "langue" and "parole",⁷ "Sem", "Noem", "Plerem", "Signem", "Monem", "onomasiology", "semasiology"; references are made to discussions about triangular and trapeze models to demonstrate relationships between "Inhaltsform" ("content form") and "Ausdrucksform" ("expression form"); etc. These two articles were written during the same period as Henne and Wiegand (1969), in which they scrutinize the

then existing literature dealing with terminology on meaning, for example, those terms used by De Saussure, Baldinger, Heger, Katz and Fodor, and Chomsky. The main idea at that time was to get a clear picture of what the concepts "meaning" and "notion" ("Begriff") entail, and to what these terms, generally used in semantics, actually refer. Wiegand strives towards a very exact, determinable description of meanings ("Berechenbarkeit von Bedeutungen" (Kammerer and Wolski 2000: XIV)), and in the process he often, but not exclusively, uses the prevalent tools of that period, such as semantic features. By the mid-1970s, he also gradually starts using other methods in his investigations, such as communication models, which later became more popular approaches in linguistics (Kammerer and Wolski 2000: XIV). But, even in his latest articles, Wiegand still uses many of the earlier structuralist concepts and insights (cf., for example, Wiegand 2002). Nowadays Wiegand's terms are widely used in dictionary research.

In the last article of Volume 2, "Mit Wittgenstein über die Wortbedeutung nachdenken. Gebrauch? Regel des Gebrauchs? Ein Etwas im Kopf?" (1999), Wiegand also distinguishes between a "realistic" and "non-realistic" approach to semantics. Different scholars, such as Austin, Ayer, Carnap, and others are surveyed. According to him the "realistic" approach considers languages as abstract sign systems. Linguistic signs are viewed as semiotic phenomena, which convey meaning, and to which certain entities such as ideas, notions, concepts, abstract objects, classes, presentations and propositions, are assigned. Realistically conceptualised meanings can then be analysed or decomposed into smaller parts, such as semantic features and semes. The use speakers make of linguistic signs when communicating, does not play a big role in realistic approaches to meaning (p. 1511).

Wiegand explains that he does not disapprove of realistic approaches to semantics (p. 1511, note 3), but that he rather prefers a so-called "non-realistic" approach in which languages are viewed as "arsenals of tools for action" (p. 1512). According to this view, language signs are primarily tools by means of which linguistic actions can be performed. They convey meanings because they are employed according to collectively accepted rules of usage, and fulfil certain purposes while being used.

In "Synonymie und ihre Bedeutung in der einsprachige Lexikographie" (1976) (pp. 125-173), Wiegand illustrates his position on actional-semantics by means of the by now famous "Kajak" ("kayak") example (pp. 130-134), where he demonstrates that, in everyday dialogues, speakers of a language (and, by implication, dictionary users) normally do not distinguish between a particular word and the extra-linguistic entity to which that word refers. In this example, two boys, Matthias and Bobby, have a conversation in which Matthias speaks about the "kayak" he received as a birthday present. Bobby reacts by asking what a kayak is, and Matthias then answers by giving some information on what a kayak actually is, describing what it looks like.⁸

Wiegand argues that a dictionary user will not be able to use a word in a dictionary without some extra-linguistic knowledge (or encyclopedic knowl-

edge) of the entity to which the word refers. That is why lexical paraphrases in monolingual dictionaries should not be based solely on the analysis of semantic features. Lexical paraphrases should also inform dictionary users of the "extension" (p. 133, note 22), or encyclopedic aspects, of a word to enable them to use the particular word appropriately and in the right context(s), according to the usual semantic rules.

Another example Wiegand uses to illustrate his concept of meaning, is found in the fourth article of Volume 1, "Einige grundlegende semantisch-pragmatische Aspekte von Wörterbucheinträgen" (1977) (pp. 174-236). Here, he tries to explain what people speak about in everyday conversation when referring to the meaning of linguistic expressions, and how they speak about this. Normally, he contends (p. 201), people do not distinguish between language and world. However, they need reference rules in order to use words correctly or "in a usual way". Reference rules are "intra-individual" rules (p. 671), by means of which people can communicate with each other.

Wiegand often uses the concepts "usual texts" and "unusual texts" (for example, p. 204). To illustrate what he means by these expressions, he uses another very famous example, namely the word "Rappe". Wiegand (p. 204) subscribes to Grice's approach, and explains that if a speaker uses a text in an "unusual" way, he/she says something but may mean something different. In his example, the father of the boy uses the word "Rappe", which normally (under "usual" circumstances) refers to a black horse. The father, in a game, calls his son "Rappe", because the son indicated that he is imitating a horse. This, of course, does not mean that the boy is a horse. In this example, the father breaks the normal reference rules and does not use the word "Rappe" in the "usual" way. Dictionary compilers could not keep this type of "unusual" texts in mind when they formulate lexical paraphrases.

Wiegand's argumentations in connection with "usual texts" and "usual contexts of naming" are also developed in "Pragmatische Informationen in neu-hochdeutschen Wörterbüchern" (1981) (pp. 335-432), included as the eighth article in Volume 1. Here Wiegand starts to distinguish between "knowledge 1" (i.e. competence with regard to the semantic rules of reference) and "knowledge 2" (i.e. competence with regard to pragmatic rules in order to use expressions appropriately), which gives a useful tool for future explorations and could be used instead of referring to "denotative" and "connotative" features.⁹

In the article "Was eigentlich ist Fachlexikographie? Mit Hinweisen zum Verhältnis von sprachlichem und enzyklopädischem Wissen" (1988) (pp. 458-511), Wiegand not only structures special-field lexicography, but also develops important theoretical assumptions on action theory and text theory, and lexicographical concepts such as "genuine purposes of dictionaries". Although the main focus of the article is to classify special-field dictionaries into distinct types ("fachliches Sprachwörterbuch" (special-field language dictionary), "fachliches Sachwörterbuch" (special-field encyclopedia), and "fachliches Allbuch" (a combination of the first two)), Wiegand sheds further light on the issue that was investigated in the article "Pragmatische Informationen ..." (1981), namely

types of knowledge. "Knowledge 1" now becomes "gegenstandskonstitutives Bedeutungswissen" (p. 871).

A later article concentrating on special-field lexicography included in Volume 2 is "Zur Unterscheidung von semantischen und enzyklopädischen Daten in Fachwörterbüchern" (1994) (pp. 1106-1127), in which the earlier distinctions between the different types of special-field dictionaries are investigated further. In the field of special-field lexicography, Wiegand distinguishes different types of knowledge, namely (i) special-field encyclopedic factual knowledge ("fach-enzyklopädisches Sachwissen"), (ii) special-field object-constituting meaning knowledge ("fach-enzyklopädisches gegenstandskonstitutives Bedeutungswissen"), and (iii) non-encyclopedic meaning knowledge ("nicht-enzyklopädisches Bedeutungswissen").¹⁰ It is, however, also useful to apply these distinctions in general lexicography, especially the notion of "object-constituting meaning knowledge".

Wiegand (p. 870) illustrates this notion when he states that, from the point of view of an actional-semantic approach, a lexical paraphrase can be seen as a "lexicographical rule formulation". In a lexical paraphrase, one formulates the reference rules and predicate rules for the use of a non-special-field lemma sign, so that the correct use of that lemma sign will be compatible with the usual contexts of naming. The language user will, therefore, have knowledge that something is something specific, and not something else. This type of knowledge Wiegand calls "object-constituting knowledge", because it brings the specific object to mind when the expression is uttered.

In the article "Über usuelle und nichtusuelle Benennungskontexte in Alltag und Wissenschaft" (1996) (pp. 1278-1310), included in Volume 2, Wiegand once again discusses the difference between everyday dialogues and scientific communication. Everyday objects are named, and there exists a socially collective accepted agreement among language users about the names of objects.¹¹ This agreement enables people to know what other people mean when they use certain words "in a usual way". When people perform a referential action ("Referenzhandlung"), it means that they use a referential expression ("Referenzausdruck"), in other words, they refer to something. In Wiegand's terms, the action of the type TO REFER TO SOMETHING ("AUF ETWAS REFERIEREN") is a linguistic action, performed by one speaker who utters a referential expression with the purpose of at least one other speaker identifying the object of reference (p. 1291). Reference actions are, therefore, genuine actions of dialogue ("genuine dialogische Handlungen") which have to be oriented towards the knowledge of at least one other speaker if they are to be successful. For example, under normal circumstances, a person A would not perform a reference action which he/she knows in advance would not be understood by person B.

In this article, Wiegand once again uses some examples. He claims, for instance, that using the name "Pferd" ("horse") to refer to a horse, is governed by rules of correct, "usual" usage, and when people know these collective rules of naming and know how to use this name, it follows that at least the "object-

constituting meaning knowledge" (which forms part of the encyclopedic knowledge)¹² is conveyed (pp. 1297-1298). If a speaker A is then successful in using an utterance, it means that the other communicator, speaker B, could derive the meaning of the particular utterance. In order to be able to do this, speaker B has to have a special "linguistic action knowledge" ("Sprachhandlungswissen"), which is possible because the "object-constitutive meaning knowledge" of the utterance was present and the usual reference rules for the particular utterance had previously been acquired (p. 1301).¹³

It could, of course, happen that the meaning knowledge ("Bedeutungswissen") of two speakers differs, so that an utterance could have a different actual meaning for the two communicating speakers, even when it is used according to the "usual" rules. In every communication process where language is used between two speakers, the individual meaning knowledge of one speaker is aligned with the exteriorized meaning knowledge of the other, present in the rule-bound usage of both. As long as this process of comparison is free of disturbances, speakers are not concerned about word meanings. But when communication conflicts arise, word meanings become important, because it is usually impossible to solve such conflicts without solving the problem of word meanings (p. 1534).

In the third article of Volume 1, "Synonymie und ihre Bedeutung in der einsprachige Lexikographie", first published in 1976 (pp. 125-173), the implications of Wiegand's semantic approach for lexicography begin to become clear. Wiegand considers the formulation of dictionary entries as a "linguistic action", but is opposed to the idea that a dictionary entry contains a so-called "definition" in the same way as this term is used in scientific language. Wiegand refutes the idea of using the term "definition" in general lexicography. He states that dictionaries are the result of lexicographers' "writing actions" about the codified language and also about objects (p. 201). He rather prefers the expression "lexical paraphrase" to denote a dictionary entry containing the meaning description, but even then, he does not merely rely on the analysis of semantic features (p. 131, note 18) for the formulation of this lexical paraphrase. Later, in his text theory, Wiegand prefers the expression "Bedeutungsangabe" ("meaning item") for this section of the dictionary article.¹⁴

Some of Wiegand's earlier articles, a number of which are included in Volume 1, also deal with special-field lexicography. These are, for example, "Kommunikationskonflikte und Fachsprachengebrauch" (1979) (pp. 237-263), and "Bemerkungen zur Bestimmung metakommunikativer Sprechakte" (1979) (pp. 264-291), which respectively encompass investigations into types of communication conflicts in special-field languages and general communication conflicts and interruptions. In the latter article, Wiegand takes the findings of discourse analysis and speech act theory of that time into account by discussing the viewpoints of, amongst others, Meyer-Hermanns, Searle, Wunderlich, Bühler, Grice, Bales and Betten, in order to make certain important distinctions with regard to speech acts and successful reciprocal communication between humans. He bases his discussion on about 150 everyday dialogues which he recorded. In

this article, Wiegand once again discusses the "Kajak" example as representative of many other everyday dialogues, and draws some important conclusions. Some of these are: (i) human speech acts or communication by means of language can result in communication conflicts, because people may have certain knowledge gaps ("Wissenslücken"), (ii) humans "acquire" certain ways of dealing with such communication conflicts by developing "interactional-reflexive" actions ("interaktionsreflexive Handlungen"), (iii) humans want to be successful when communicating with other people, and (iv) humans know when to use these "interactional-reflexive" actions when there is danger of communication not being successful.

When these are applied to lexicography, as in the situation of consulting a dictionary, users may have a search question of the type "WAS BEDEUTET X?" ("WHAT DOES X MEAN?") (with X as variable for lexical items). In usual naming contexts, a dictionary would then ideally fill the individual's knowledge gap, keeping the everyday lexicon stable in the process (p. 1303).

To demonstrate the types of information that he deems necessary for dictionary users, Wiegand (pp. 212-219) discusses the example of "Weberei" ("weaving") in the early article "Einige grundlegende semantisch-pragmatische Aspekte von Wörterbucheinträgen" (1977) (pp. 174-236). The word "Weberei" is used in a general as well as a specialized sense, and Wiegand suggests that one could include more information than mere semantic features. To indicate the "special-field" use of "Weberei", one could even give a short historical introduction to the art of weaving, together with bibliographical references. An interesting idea is Wiegand's suggestion about using a so-called "lexicographical narration". This means that one systematically relates how a loom is built (for example, an expert telling non-experts), so that the user can get an idea of what the components are, and what they look like. Even though this process is specifically useful for learners' dictionaries, it illustrates Wiegand's emphasis on the fact that users need extra-linguistic knowledge in addition to knowledge about semantic features.

In lexicography, one has additionally to take the purpose of the particular dictionary into account to determine the type of knowledge that should be conveyed to the reader. Wiegand claims that different purposes of description need different methods of description (p. 1541). That which is considered the "description of meaning", is mostly nothing more than a description of the rule-bound usage, or an example of this. This is particularly true when the addressees are individuals of whom it is accepted that they do not know the meaning of the lexical items. These "descriptions" can then enable them to obtain knowledge about a word's meaning and to know how to use the word correctly.

A very complete exposition of Wiegand's views on action theory can be found in the article "Zur handlungstheoretischen Grundlegung der Wörterbuchforschung" (1987) (pp. 704-748). In this article, Wiegand, approaching dictionary use as a set of user actions, explains and develops this area of research by drawing up an "actional-theoretical" framework (p. 709). In the process, he formulates numerous definitions and distinctions, at the same time giving a

critical survey of the texts on analytical action theory available at the time.

One of the best-known examples used by Wiegand is the one of "Frosch" ("frog") (pp. 709-712).¹⁵ Wiegand indicates that he deliberately chooses examples which are not specifically related to dictionary research. The example illustrates how, in an everyday dialogue, two persons speak about what a certain person Hans has done: *Hans hat den Frosch gequält* ("Hans has tortured the frog"). This activity by Hans is described by means of a linguistically uttered interpretation of his activity. That which Hans has done, can be considered as a case which can be referred to by means of the usual rules of usage, utilising the expression *einen Frosch quälen* ("to torture a frog") (p. 710). All such activities, to which one can refer in usual texts by using the expression *einen Frosch quälen*, belong, as actions, to the action type EINEN FROSCH QUÄLEN ("TORTURE A FROG").

In the same way, one can also get user actions relating to dictionary use, which can belong to an action type such as EIN WÖRTERBUCH BENUTZEN ("USING A DICTIONARY"). For every action, there is an internal action context (p. 716) such as the particular reason why it is performed, and the purpose. For example, if one wants to find out the reason why a person has used a dictionary, one can formulate a why-question such as WHY HAS HE USED THE DICTIONARY?¹⁶ In the same way, one can also interpret the consequences of the user action, for example, by formulating a question such as WITH WHICH CONSEQUENCES HAS HE USED THE DICTIONARY? (p. 717). Wiegand concludes that user actions belonging to the action type USING A DICTIONARY have many features (p. 718). These are: (i) subject of the action (WHO?), (ii) result of the action (which determines the action type) (WHAT?), (iii) modality of the action (HOW?), (iv) external context of the action (including circumstances under which the action is performed (UNDER WHICH CIRCUMSTANCES?), time of the action (WHEN?), duration of the action (HOW LONG?), and place of the action (WHERE?)), (v) internal context of the action (including the purposes of the action (which are attained when the intended results are obtained) (FOR WHAT?), reasons for the action (WHY?), and incentive for the action (OUT OF WHICH CAUSE?)), and (vi) consequences of the action (WITH WHICH CONSEQUENCES?). Each of these features is included in Wiegand's so-called actional-theoretical structure of dictionary use (pp. 719-721).

In *Kleine Schriften*, one gets a good overview of Wiegand's gradual development of his actional-semantic approach to lexicography, even though many of his articles which give more detail on this topic, are not included.

3. Text theory

In early articles such as "Synonymie und ihre Bedeutung in der einsprachigen Lexikographie" (1976) (pp. 125-173), Wiegand already states that there is a lack of an "empirically-based sociology about dictionary users". He gradually develops such an empirically based theory, which later becomes one of the

principles of his metalexigraphy.

The fourth article in Volume 1 has as subtitle "Ein Beitrag zur praktischen Lexikologie" ("A Contribution to Practical Lexicology"), as does the article "Pragmatische Informationen in neuhochdeutschen Wörterbüchern ..." (1981) (pp. 335-432). Wiegand's aim was to devise, by means of investigating plausible ways of linguistic explanation and description, more precise suggestions for designing lexicographical texts. Apparently there has always been a huge gap between research in lexicology and lexicographical practice. This also has led to questions about the relationship between theory and practice: whether lexicology and lexicography both are theories.¹⁷

One of Wiegand's approaches is to analyse different types of dictionaries, in order to determine which types of information they contain, and which different methods these dictionaries use in presenting data. These analyses were used to draw up, amongst others, his (i) dictionary typology, (ii) his theory of dictionary use, and (iii) his theory of lexicographical texts. Examples of articles in *Kleine Schriften* which contain such analyses, are (i) 4, 8, 9, and 26 (content of dictionary articles), (ii) 10, 23, 24, 25, 27, 28, 29, 32, 34 and 35 (structure of dictionary articles), and (iii) 11, 13, 17, 19, 21, and 22 (dictionary characteristics and typology). It is impossible to give attention to all of these issues within the scope of this article. Only some short comments are therefore made on Wiegand's framework for research on dictionary use, and then more detail is given about his approach to text theory.¹⁸

3.1 Wiegand's framework for dictionary use

The article "Fragen zur Grammatik in Wörtbuchbenutzungsprotokollen. Ein Beitrag zur empirischen Erforschung der Benutzung einsprachiger Wörterbücher" (1985) (pp. 560-618) is a precursor to "Zur handlungstheoretischen Grundlegung der Wörterbuchforschung" (1987), in which Wiegand outlines a framework for research on dictionary use. The latter article explains many terms such as *experienced users*, *potential users*, *ways of dictionary use*, *user questions*, etc., and different types of dictionary use are systematically classified.¹⁹ These articles also connect with the much later article "Über primäre, von Substantiven 'regierte' Präpositionen in Präpositionalattributkonstruktionen" (1996) (pp. 1311-1338), included in Volume 2, in which Wiegand constructs a framework of types of user questions and many terms that should be used in this regard (cf., for example, pp. 724 and 746).

3.2 Wiegand's text theory for lexicographical texts

Wiegand formulates certain hypotheses with regard to the lexicographical layout of different types of dictionaries by analyzing many different dictionaries and dictionary articles. These deliberations enable him to draw up a very detailed framework for the structure of dictionary articles, with the aim of

making future dictionaries more user-friendly.²⁰ He uses the "mathematical structure concept" of Bourbaki as well as language from set theory to marshal his findings (p. 968).²¹ According to Wiegand, his theory of lexicographical texts has three parts: (i) the production of texts, (ii) the structure of texts, and (iii) the reception of texts (p. 968).²²

In the article "Was ist eigentlich ein Lemma? Ein Beitrag zur Theorie der lexikographischen Sprachbeschreibung" (1983) (pp. 458-511), Wiegand introduces numerous new terms which refer to lexicographical practice. This well-known article is an important contribution to his text theory for lexicographical texts. Wiegand strongly indicates his preferences for certain expressions, for example, "dictionary article" instead of "dictionary entry". He defends his viewpoints by citing numerous examples from dictionaries, and by giving citations from the work of other authors. Other terms discussed are, for example, "lemma", "lemma collection", "guiding element", "guiding form", "micro- and macrostructure", "partial lemma", "nesting lemma" and "niching lemma". He introduces no less than 52 metalexicographical terms in this article.

In his article "Metalexicography. A Data Bank for Contemporary German" (1986) (pp. 619-634), Wiegand writes about computer lexicography. But he also deals with important aspects regarding text theory for lexicographical texts. He lists text type segments (cf., for example, p. 623) which can be used in dictionaries, and makes suggestions on how one could proceed in compiling an electronic database, based on different text types. This article in English is a good directive to illustrate Wiegand's line of thought.

The approach towards text linguistics in the article "Zur handlungstheoretischen Grundlegung der Wörterbuchforschung" corresponds with that in "Wörterbuchartikel als Text" (1988) (pp. 877-950). These two very important articles laid the foundation for argumentations used in many of Wiegand's following publications, for example, his treatment of special-field lexicography, his text theory for the structural organization of dictionaries, and his *magnum opus*, *Wörterbuchforschung* (1998).

Not only is the article "Wörterbuchartikel als Text" (1988) connected with Wiegand's earlier attempt towards structuring a general theory of lexicography,²³ but he also states here that new and fruitful insights for dictionary research can be obtained if text theory categories are used (p. 950). Dictionaries are carriers of text types: the "lemma" could, for example, be compared to the text type "title" (p. 932). Different types of dictionaries show different article structures and different degrees of standardisation. Dictionary articles consist of partial texts which he calls "data types" (p. 907).²⁴ For the sake of scientific inquiry, it is possible to present such data types by means of tree structures, as he often does (cf., for example, pp. 916-917).

Lexicographical partial texts should be seen as "functional parts in a greater context" (p. 950). One should therefore not consider the individual items in lexicographical texts as isolated linguistic units, but dictionary articles (the most important type of partial texts) should be investigated with regard to

their integration within the text as a whole. Wiegand states that a text theory for lexicographical texts also offers a tool for the development of new article structures, and the examination of lexicographical instruction books and sample articles. Worthwhile insights that can enhance the computerisation of existing dictionaries, or help with the planning of a lexicographical system can also be obtained.

The article "Printed Dictionaries and their Parts as Texts" (1991) (pp. 951-1062) clearly takes these issues further. By means of graphic illustrations and different notational conventions, Wiegand gives a very comprehensive overview of the growing interest in the textual properties of lexicographical texts in Europe.²⁵ Being in English, this article is of course most useful to English-speaking readers. Wiegand lists aspects of his text theory for lexicographical texts: (i) criteria for the textuality of lexicographical texts and representation form of textuality, (ii) text condensation, propositional density and expansion of texts, (iii) kinds and degrees of standardisation of lexicographical texts, and (iv) order structures of lexicographical texts, namely: textual book structure, textual word list structure; inner and outer access structures including rapid access structures; hierarchical and precedential article structures; hierarchical and precedential microstructures and item structures; kinds of microstructures, such as simple, expanded, composed, rudimentary, listing, integrated, partially integrated, non-integrated and the possible combinations; partial structures of microstructures, item structures and article structures; and the microstructure programme of a dictionary and the grammar for establishing microstructures.²⁶ Other structures studied are scope structures, cross-reference structures, addressing structures, cohesion structures, theme-rheme structures and coherence structures.

In "Über die Mediostrukturen bei gedruckten Wörterbüchern" (1996) (pp. 1163-1192), Wiegand sketches the developments within the field of systematic dictionary research since the article "Was ist eigentlich ein Lemma?" (1983). He proceeds from the assumption that dictionaries are text type carriers of which the texts, text parts and text segments have a specific relationship to each other. Clear distinctions are made between ordering structures, such as macrostructures, access structures, text constituent structures, addressing structures, cohesion structures, etc., leading to intricate illustrations of structures.

The term "mediostructure" is a newer addition to Wiegand's arsenal of lexicographical concepts. The "dictionary-internal mediostructure" links the knowledge elements which are represented by means of texts in the various sections of the dictionary when it is not possible to observe them all at the same time, as, for example, in cross-references (pp. 1164, 1168). In this article, Wiegand distinguishes between different types of mediostructures, such as (i) dictionary-internal mediostructures (cross-referring within the same dictionary), (ii) dictionary-linking mediostructures (cross-references linking lexicographical data in one dictionary by means of references to data in another dictionary), (iii) source-related mediostructures (cross-referring to external sources), and

(iv) literature-related mediostructures (cross-referring to literature). Wiegand systematically discusses various types of "cross-reference mediating items" ("Verweisvermittelnde Angaben") and methods of cross-referring (cf., for example, p. 1191 for a list of classes of cross-references, and the accompanying terminology). He concludes that it is better nowadays to use a computer system to set up cross-references, because formerly, when cross-references were written manually, there were many mistakes and inconsistencies. This could lead to the qualitative improvement of dictionaries and to enriched mediostructural programmes (p. 1192).

The article "Das Konzept der semiintegrierten Mikrostrukturen. Ein Beitrag zur Theorie zweisprachiger Printwörterbücher" (1996) (pp. 1193-1277) is very complicated. To understand it well, one has to be very familiar with Wiegand's metalexigraphy. Even though the article uses examples from bilingual lexicography, it is an important contribution to the text theory of lexicographical texts in general. Wiegand argues that it is not necessarily true that one would need at least four translation dictionaries for each language pair, as some translation scholars believe.²⁷ If one could compile a dictionary in such a way that it contains all the necessary data, ordered in a clearly observable and easily retrievable way, one dictionary could serve all the different purposes of active and passive language use. This means that it may be fruitful to use semi-integrated article structures.

Wiegand explains what these are by means of discussing several well-known types of microstructures, such as integrated, non-integrated and partially integrated microstructures (pp. 1197-1203). For example, a "simple hierarchical microstructure is integrated when (i) each item of a particular monolingual dictionary not belonging to the comment on form lies in the lexical-semantic area of a specific meaning, and when (ii) each of these items belong to the same semantic sub-comment as the one by means of which the meaning is conveyed, so that each of the items is in the direct scope of this 'item giving the meaning'" (p. 1198).²⁸ This means that, in simple cases of integrated microstructures, the semantic sub-comments are part of the comment on semantics.

On the other hand, a simple microstructure of a monolingual dictionary is called "non-integrated when (i) all the 'items giving the meaning' which are addressed to the 'lemma sign form item', are present in the first semantic sub-comment of the semantic comment, and when (ii) all co-text items (such as example items, collocation items and all other types of phraseological items) as well as all items which are addressed to these, are allotted according to a system (which is explained in the metatext of the dictionary) to different semantic sub-comments which follow each other and which are called 'semantic sub-comments on the co-text'" (p. 1220). This means that there are different sets of items after the "item giving the meaning", following each other, each with different types of data.

Other microstructures Wiegand discusses are in monolingual dictionaries, partially-integrated microstructures (with pre-integrate and with post-inte-

grate) and over-annexed and under-annexed microstructures, and in bilingual dictionaries, mixed-integrated microstructures and non-integrated microstructures.

Wiegand claims that there are no monolingual dictionaries in German containing semi-integrated microstructures. This type of structure, however, would mean that integrated and non-integrated microstructures are combined, having the advantages of both types of microstructures, and avoiding their disadvantages (p. 1232). Firstly, there are advantages regarding the internal access time: it takes considerably less time to find the lexical text data. Wiegand points out that one can prove this empirically (pp. 1232-1233). This would enhance the user-friendliness of monolingual dictionaries, especially general comprehensive dictionaries and learners' dictionaries (cf. also p. 1240).

Wiegand compiles his own semi-integrated article as illustration (cf. p. 1234): he uses different "text blocks" which display "text block beginning signals" ("Textblockanfangsignale") such as numbers printed in bold, and "text block ending signals" ("Textblockendsignale") such as a full-stop. This means that, being bordered at the beginning and the end from the other article text constituents in the particular dictionary article, they can be detected more easily. The article text is, therefore, demarcated by means of typographical micro-architectural indicators ("typographische Mikroarchitekturanzeiger") (p. 1233). Wiegand notes that this type of article is "internally bi-accessible" ("intern biakzessiv") (p. 1240), which means that there is more than one search path: one leads to the first part of the article which contains semi-specific data, and the other leads to the items containing the co-text classes. In computer lexicographical terms, one would therefore be able to compile "multiple window articles" ("Mehrfensterartikel"), in which each window would contain a distinct search area which exhibits the specific text data to convey answers on search questions specifically belonging to the window classes (p. 1241).

In the development of the theory of dictionary research, Wiegand has coined numerous important terms for theoretical concepts. Over the years, however, he has sometimes changed certain details in view of newer insights and in reaction to criticism by other authors. In one such article, "Altes und Neues zur Makrostruktur alphabetischer Printwörterbücher" (1998) (pp. 1428-1453), he more precisely defines terms he introduced in earlier publications, such as "grouping", "article nest", and "straight-alphabetical macrostructures" ("glattalphabetische Makrostrukturen"). He points out that in earlier publications, he worked with Carnap's structural approach, but that since 1989, he started to prefer the structural approach of the Bourbaki mathematician group (p. 1429). The latter approach enables the development of a more precise, concrete structure in which the carriers of the guiding element ("Leitelementträger"), for example the lemmata, also form part of the macrostructure of dictionaries. In the former approach, this was not possible.

Other articles dealing with dictionary typology also exhibit early traces of Wiegand's investigation of text condensation. One example included in Vol-

ume 1 is "Zur Geschichte des deutschen Wörterbuchs von Hermann Paul" (1983) (pp. 512-527). However, especially in the article "Lexikographische Textverdichtung. Entwurf zu einer vollständigen Konzeption" (1998) (pp. 1454-1489), Wiegand deals extensively with the concept of text condensation in dictionaries.²⁹ He claims that it is possible precisely to determine the degree of textual condensation in dictionary articles by using mathematical calculations (p. 1455). Because of this possibility, the extent of user-friendliness of dictionaries can be determined by devising formulas based on these calculations. This will, furthermore, enable the planning of dictionaries, and ensure that the compilation of dictionary articles could be a learnable skill. Wiegand distinguishes between (i) internal text condensation (which concerns all lexicographical partial texts with a carrier of the guiding element (e.g. lemmata) as well as register entries of which the keyword is the carrier of the guiding element) and (ii) external text condensation (which concerns all lexicographical partial texts with an outer access structure, i.e. the central word list especially) (p. 1456).

To illustrate Wiegand's line of thought, the following article of *Flöte* ("flute") can serve as example:

Flöte, die; -, -n *rohrförmiges Blasinstrument (aus Holz): die F., auf der F. blasen; er spielt (die) F.*

Certain "full texts" can be deduced from this article, some of which are:

- Dictionary article on *Flöte*.
- The correct spelling of *Flöte* is |Flöte|.
- The form of the noun singular is *Flöte*.
- *Flöte* has the accent on the first syllable.
- The accent syllable of *Flöte* is long.
- The correct article for *Flöte* is *die*.
- The gender of *Flöte* is feminine.
- *Flöte* is a noun.
- The form of the possessive singular is *Flöte*.
- The form of the noun plural is *Flöte*.
- *Flöte* is not pragmatically marked.
- *Flöte* means something like *rohrförmiges Blasinstrument* or *rohrförmiges Blasinstrument aus Holz* ("pipe-shaped wind instrument" or "pipe-shaped wind instrument made of wood")
- Examples for the use of *Flöte* are: *die Flöte blasen, auf der Flöte blasen, er spielt die Flöte, er spielt Flöte*.

Most dictionaries standardize the way in which they use text condensation. This has to be explained in the metatext of the dictionary (e.g. the user's guide or the list of abbreviations) (p. 1463). Users have to be acquainted with the conventions in order to be able to extract information from the presented data in

the condensed texts. The full texts are of course omitted to save space, and certain symbols are used instead.

To calculate the degree of condensation, Wiegand suggests that the number of full text segments is calculated. For example, in the case of **Flöte**, the full text part, "The correct spelling of *Flöte* is |Flöte|.", has ten text segments (including the full-stop). According to Wiegand, the entire full text of this dictionary article has 77 segments in German. One therefore divides 77 by 10, which means that the proportional condensation of the dictionary article at hand is 7,7. The greater the value of this quotient, the less the proportional condensation of the full text or full text parts.

This section dealt with Wiegand's gradual development of a text theory for lexicographical texts. Even though it is not possible to deal with all the intricate distinctions Wiegand has made, it was attempted to show how Wiegand went through different stages of reasoning in order to devise his text theory.

4. Concluding remarks

In this article, the content of the publication *Kleine Schriften* was discussed. An attempt was made to illustrate Wiegand's lines of thought through the years, which shaped his theory of dictionary research. Firstly, Wiegand used structuralist terminology, but gradually started coining his own. He approaches semantics from the point of view of action theory, and this affects his view of which types of information different dictionary types should contain. Secondly, Wiegand develops a comprehensive theory for lexicographical texts, based on his belief that dictionaries should be useful. His impressive investigations into dictionaries and dictionary articles led him thoroughly to understand the different textual structures in dictionaries and dictionary articles. Because of various subtle distinctions, his theory enables lexicographers to compile better dictionaries in future. Wiegand's classifications and characterisation of text types within the text theory for lexicographical texts are now widely used in lexicographical practices and in the planning of new dictionaries.

Notes

1. English translations of Wiegand (1977 and 1985) are, however, included in *Semantics and Lexicography. Selected Studies (1976–1996)*, edited by Immken and Wolski (1999). Cf. also Louw (2000) for an article on this English publication.
2. Cf. Smit (2001) for a review article of this publication.
3. Cf., for example, Konerding and Wiegand (1994), where the issue of frames is investigated to determine their usefulness for lexicography.
4. Cf., for example, p. 1528, note 15 and p. 1531, note 16 in Volume 2, where Wiegand admits that he changed his viewpoints after having read Keller.
5. Cf. also Henne and Wiegand (1969) and *Wörterbuchforschung* (Wiegand 1998) for examples of

this approach.

6. It is very difficult to translate the German word "handlungssemantisch". Immken and Wolski (1999: 4) use the expression "actional-semantic approach". This follows the expression used in Wiegand (1992: 236). For lack of a better translation, the expression "actional semantics" will therefore be used in this article.
7. Although Wiegand uses the terms "langue" and "parole" in his early publications as well as in the more recent ones (cf., for example the last article in Volume 2 of *Kleine Schriften*, and also in "Zur Äquivalenz in der zweisprachigen Lexikographie. Kritik und Vorschläge" (2002)), Henne and Wiegand (1969: 137) use the expressions differently from De Saussure by making even finer distinctions.
8. For an English version of Wiegand's thoughts in this regard, cf. also Wiegand (1992) or Wiegand (1999: 175-289).
9. Cf. also p. 1287, where Wiegand states that he has abandoned the use of the predicate "denotative" after Lyons had made finer distinctions.
10. In Immken and Wolski's translation included in Wiegand (1999: 297ff), the expression "domain-specific dictionaries" is used as translation for "Fachwörterbücher". In the present article, the expression "special-field" is preferred to refer to "subject-field dictionaries".
11. Wiegand (p. 1535) calls these agreements "silent conventions" amongst a community of persons.
12. Cf. also p. 871 for a discussion of this type of encyclopedic knowledge.
13. Wiegand (p. 1533) notes that this knowledge about the rules of "usual" usage need not be complete.
14. Cf. especially Wiegand (1989e: 539-552) for a detailed argumentation on why he finds the expressions "lexicographical definition" and "lexicographical explication" problematic. Wiegand (1992: 235-243) is an excellent English version dealing with the same issues (also published in Wiegand 1999: 253-260).
15. Wiegand also expanded greatly on the so-called "Frosch" example in *Wörterbuchforschung* (1998).
16. Here Wiegand's examples refer only to males.
17. Cf. Wiegand (1998: 13-256) for a detailed discussion on this issue.
18. Cf. Wiegand (1984b) for an English version of his exposition of the general theory of lexicography.
19. In *Wörterbuchforschung* (1998), Wiegand expands greatly on this publication. Cf. also Ripfel and Wiegand (1988).
20. Cf. p. 1455, note 1, for references to Wiegand's publications which dealt with lexicographical text theory.
21. Cf. Bourbaki, N. 1958. *Éléments de mathématique*. Paris: Hermann. (English translation: Bourbaki, N. 1968. *Elements of Mathematics: Theory of Sets*. Paris: Hermann/Reading, MA: Addison-Wesley.)
22. Cf. also Wiegand (1992) or Wiegand (1999: 203-282) for an English version by Immken and Wolski of Wiegand's text theory.
23. Cf. also Wiegand (1983 and 1989b) as well as article 10 "Deutsch–Usbekisches Wörterbuch" in *Kleine Schriften* for similar discussions.
24. Cf., for example, p. 907, where Wiegand illustrates some data types of the dictionary article "Bestand" from the *Great German–Chinese Dictionary*. It is important to note that Wiegand distinguishes between "data" and "information". The former are the items in a dictionary article,

- and these data types convey the information that a user needs. Cf. Wiegand (1998: 160-171) for a discussion on this distinction.
25. Cf. also Wiegand (1991) for a detailed exposition of article texts.
 26. Cf., for example, p. 974, pp. 1058-1062 and 1405-1409 for lists of symbols and descriptions that Wiegand uses in his text theory to present different items in dictionary articles.
 27. Cf. p. 1241 for a list of authors on translation theory who dealt with the "active-passive" issue in translation and its implications for bilingual lexicography.
 28. Cf. also Wiegand (1989d) for a detailed discussion of microstructures.
 29. "Text condensation" is the expression used here, even though Wiegand distinguishes between "Textverdichtung", "Textkondensierung" and "Textkomprimierung" in German (p. 1488). However, he states clearly that he is not interested in forming terminology in English. He is only concerned with the formation of terminology for German lexicography. He feels that the formation of English terminology should be done by native speakers of English (p. 1454, note*).

Bibliography

- Hausmann, F.J. and H.E. Wiegand.** 1989. Component Parts and Structures of General Monolingual Dictionaries: A Survey. Hausmann, F.J. et al. (Eds.). 1989-1991: 328-360.
- Hausmann, F.J., O. Reichmann, H.E. Wiegand and L. Zgusta (Eds.).** 1989-1991. *Wörterbücher: Ein internationales Handbuch zur Lexikographie / Dictionaries: An International Encyclopedia of Lexicography / Dictionnaires: Encyclopédie internationale de lexicographie*. Berlin/New York: De Gruyter.
- Henne, H. and H.E. Wiegand.** 1969. Geometrische Modelle und das Problem der Bedeutung. *Zeitschrift für Dialektologie und Linguistik* 36: 129-173.
- Kammerer, M. and W. Wolski (Eds.).** 2000. Zur Einführung. Wiegand, H.E. 2000: IX-XXX.
- Konerding, H.-P. and H.E. Wiegand.** 1994. Framebasierte Wörterbuchartikel. Zur Systematisierung der lexikographischen Präsentation des Bedeutungswissens zu Substantiven. *Lexicographica* 10: 100-170.
- Louw, P.** 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.
- Ripfel, M. and H.E. Wiegand.** 1988. Wörterbuchbenutzungsforschung. Ein kritischer Bericht. Wiegand, H.E. (Ed.). 1986. *Studien zur neuhochdeutschen Lexikographie VI, 2. Teilband. Mit einem Namen- und Sachregister zu den Bänden I-VI sowie einer Bibliographie zur Wörterbuchforschung*. Germanistische Linguistik 87-90: 490-520. Hildesheim/New York/Zürich: Georg Olms.
- Smit, M.** 2001. Herbert Wiegand with a "Metalexigraphical Panga" in the Jungle: An Unlocking of Wörterbuchforschung. *Lexikos* 11: 296-310.
- Wiegand, H.E.** 1977. Nachdenken über Wörterbücher: Aktuelle Probleme. Drosdowki, G., H. Henne, H.E. Wiegand (Eds.). *Nachdenken über Wörterbücher*: 51-102. Mannheim: Bibliographisches Institut.
- Wiegand, H.E.** 1983. Überlegungen zu einer Theorie der lexikographischen Sprachbeschreibung. Hyldgaard-Jensen, K. and A. Zettersten (Eds.). 1983. *Symposium zur Lexikographie / Symposium on Lexicography. Proceedings of the Symposium on Lexicography, September 1-2 1982, at the Uni-*

- versity of Copenhagen. Germanistische Linguistik* 5–6: 35-72. Hildesheim/New York: Georg Olms.
- 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. Berlin: De Gruyter.
- Wiegand, H.E.** 1984b. 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. Tübingen: Max Niemeyer.
- Wiegand, H.E.** 1985. Eine neue Auffassung der sog. lexikographischen Definition. Hyldgaard-Jensen, K. and A. Zettersten (Eds.). *Symposium on Lexicography II. Proceedings from the Second International Symposium on Lexicography May 16–17 1984 at the University of Copenhagen*: 15-100. Tübingen: Max Niemeyer.
- Wiegand, H.E.** 1989a. Der gegenwärtige Status der Lexikographie und ihr Verhältnis zu anderen Disziplinen. Hausmann, F.J. et al. (Eds.). 1989–1991: 246-280. Berlin: De Gruyter.
- Wiegand, H.E.** 1989b. Aspekte der Makrostruktur im allgemeinen einsprachigen Wörterbuch: Alphabetische Anordnungsformen und ihre Probleme. Hausmann, F.J. et al. (Eds.). 1989–1991: 371-409.
- Wiegand, H.E.** 1989c. Der Begriff der Mikrostruktur: Geschichte, Probleme, Perspektiven. Hausmann, F.J. et al. (Eds.). 1989–1991: 409-462.
- Wiegand, H.E.** 1989d. Arten von Mikrostrukturen im allgemeinen einsprachigen Wörterbuch. Hausmann, F.J. et al. (Eds.). 1989–1991: 462-501.
- Wiegand, H.E.** 1989e. Die lexikographische Definition im allgemeinen einsprachigen Wörterbuch. Hausmann, F.J. et al. (Eds.). 1989–1991: 530-588.
- Wiegand, H.E.** 1990. Die deutsche Lexikographie der Gegenwart. Hausmann, F.J. et al. (Eds.). 1989–1991: 2100-2246.
- Wiegand, H.E.** 1991. Über die Strukturen der Artikeltexte im *Frühneuhochdeutschen Wörterbuch*. Zugleich ein Versuch zur Weiterentwicklung einer Theorie lexikographischer Texte. Goebel, U. and O. Reichmann (Eds.). 1991. *Historical Lexicography of the German Language. Volume 2*: 341-673. Lewiston, NY: Edwin Mellen.
- Wiegand, H.E.** 1992. Elements of a Theory towards a So-called Lexicographic Definition. *Lexicographica* 8: 175-289.
- Wiegand, H.E.** 1998. *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie. 1. Teilband*. Mit 159 Abbildungen im Text. Berlin/New York: De Gruyter.
- Wiegand, H.E.** 1999. *Semantics and Lexicography. Selected Studies (1976–1996)*. Edited by A. Immken and W. Wolski. Tübingen: Max Niemeyer.
- Wiegand, H.E.** 2000. *Kleine Schriften: Eine Auswahl aus den Jahren 1970 bis 1999 in zwei Bänden. Band 1 1970–1988. Band 2 1988–1999*. Edited by M. Kammerer and W. Wolski. Berlin/New York: De Gruyter.
- Wiegand, H.E.** 2002. Zur Aequivalenz in der zweisprachigen Lexikographie. Kritik und Vorschläge. Wiegand, H.E. (Ed.). 2002. *Studien zur zweisprachigen Lexikographie mit Deutsch VIII. Germanistische Linguistik* 166: 93-110. Hildesheim/New York: George Olms. (English translation: Wiegand, H.E. 2002. Equivalence in Bilingual Lexicography: Criticism and Suggestions. *Lexikos* 12: 239-255.)

Igor Burkhanov. *Linguistic Foundations of Ideography: Semantic Analysis and Ideographic Dictionaries.* 1999, 388pp. ISBN 83 87288 98 5. Rzeszów: Wydawnictwo Wyższej Szkoły Pedagogicznej. Price: 16 PLN.

In his book, Igor Burkhanov assumes that cognitive semantics, as developed mainly in the United States of America from about 1980, is a linguistic innovation which lends itself to a thorough renovation of pedagogical dictionaries. In order to show this convincingly, he discusses lexicography as an applied discipline of linguistics with the study of meaning at its centre. The various types of ideographical (onomasiological) dictionaries extant are presented and the development of semantics is unravelled from the field theory of the early 20th century to the present. After a general sketch of the learner's ideographical dictionary, the possible contribution of cognitive semantics is shown in a number of sample entries. The book is evaluated as being highly informative and equally highly honest, because the many shortcomings and deficiencies of pedagogical metalexicography are pointed out. A careful weighing of its statements gives occasion for some critical ideas about the relationship between linguistics and philology, the terminology and typology of ideographical dictionaries, and the feasibility of the proposed new dictionaries for didactic purposes.

In the language-related sciences, abstract innovations in theory often stimulate concrete innovations in related fields of practice. New ideas in syntax, for example, may lead to new descriptive grammars and eventually to new teaching methods. New ideas in semantics may lead to new ways of meaning determination (lexicology) and eventually to new dictionaries. Such projections from theory to practice are more likely to occur the more radical the abstract innovations are, i.e. the more they introduce a shift of paradigm. Igor Burkhanov assumes that *cognitive semantics* is an innovation of this sort. He regards its principles of the constitution of word meanings as seminal for lexicology and lexicography because they deviate from the older clear-cut and binary demarcations of structural semantics. In particular, the concept of the prototype is understood by him to have the potential for streamlining a rather neglected old (and therefore in its revival now new) dictionary type as such and the methods of its entry structure. He calls it by the name *ideography/ideographic*.

It is to the credit of the author that he does not jump hastily from cognitive semantics to lexicographical conclusions. Rather, he reflects painstakingly on the general presuppositions of lexicography as an academic discipline and the role of the linguistic description of the lexicon in it (chapter 1), then moves on to discussing the scope of ideography (chapter 2) plus the types of relevant dictionaries (chapter 3), and finally unravels the developments in linguistic semantics and their significance for the dictionaries in question (chapter 4). The general foundations thus being laid, he demarcates the design of a learner's ideographical dictionary of the general vocabulary (chapter 5), i.e. the type of dictionary he has in mind for some innovative proposals, and finally applies

everything said and discussed so far to 'problems and perspectives' as well as to 'results and their representation' (chapter 6). The reader is, thus, guided from an abstract positioning of lexicography as an applied linguistic discipline to a representative sample of dictionary entries in a linguistically new vein. All the reflections that demand to be present in a lexicographer's mind when at work are laid out critically, and this is done in a clear and intelligible sequence of scientific thought.

Generally speaking, applied disciplines (for example in mathematics, theology, linguistics, etc.) are subject to the idea that the theory behind them can provide an adequate solution to problems which originate in practice, but all adherents of this idea find it difficult to admit that these problems of practice are of a genuine nature which is different from that which a theory can meet. It would therefore be advisable to approach these difficult 'applications' at least in an interdisciplinary way. Several disciplines in co-perspective will certainly be successful where one must fail. This does not preclude the fact that, occasionally, there is one discipline which is more centrally involved in the approach than others. Such is, for instance, the case with linguistics and the solution to lexicographical problems. Therefore, it is legitimate to discuss the linguistic foundations of lexicography (and inside lexicography of ideography) as the author does. Notwithstanding this fact, lexicography should not be regarded as theoretical linguistics in application but as lexicography in its own right. There are many disciplines of this sort which, though looking for support elsewhere, have their own constitution. (Think of medicine, mathematics, and others.) Therefore, I fully agree with the author, when he informs his readers about these queries over demarcation (chapter 1), that lexicography is a practice-oriented (i.e. applied) discipline 'of not only a linguistic but also a historical and philological nature' (p. 22). Philology in its diachronic perspective is indeed rather interdisciplinary and quite different from today's methods of system and form driven linguistics. I only regret that the book as a whole does not make wider use of this insight, that 'the ways of life, traditions, common beliefs, institutions and collective activities' (p. 23), which are so important for understanding dictionaries, are not mentioned further. (Of course, I realise that this would have changed the character and the volume of the book to a larger degree than the author was obviously prepared to do.) But lexicographers must themselves become aware of the fact, and inform their readers accordingly, that they are dealing with a highly culture-sensitive section of linguistic activities (e.g. Green 1996, Hüllen 1999, Haß-Zumkehr 2001). If this is so, they must maintain more clearly than is usually done that lexicography is not only the discipline of compiling dictionaries (process-oriented) and of analyzing them (product-oriented). It is also the discipline dealing with the linguistic needs of people (in a given culture and at a certain time) and of the use they make of dictionaries in order to answer them. This pertains to the dictionaries of their own language as well as to those of foreign ones.

Furthermore, I agree with the author that, as a consequence of this, a typology of dictionaries, as far as meaning determination is concerned, must

not be taken solely from the methods of linguistics. An imposing overview of contemporary semantics and lexicology, in their own right and in relation to pragmatics and stylistics, is given. Again the cognitive approach earns the praise of the lexicographer, because of its culturally relevant information, including the important difference between common sense and expert knowledge (p. 48). Generally speaking, functional rather than structural divisions are explained as being pertinent. In spite of the wealth of semantic concepts, information on the actual use of lexemes is not sufficiently provided. But this is what dictionaries need. So the results of this *tour d'horizon* are actually rather disappointing (e.g. p. 68). Only the expectation that the new developments will help, emerges. This is why the question now becomes urgent for the reader of what the information is which a dictionary has to provide for its users and how this information can be catered for by linguistics.

This answer is (at least partly) given in the discussion of the typological features of an ideographical dictionary (chapter 2). In my view, it is regrettable in general that there are so many terms for naming those dictionaries whose entries are not arranged alphabetically. Although some reasons are given for the differences between dictionaries called *ideological*, *ideographical*, *analogical*, *semantic*, *conceptual*, *thematic*, *topical*, or *onomasiological*, and for names like *thesaurus* and *The Wordtree*, the impression of an embarrassing terminological abundance (of course not of the author's making) prevails. I myself prefer the term *onomasiological* as a synonym for *topical*, because it was used and satisfactorily discussed in opposition to *semasiological* in a long drawn-out debate in Germany on the *wissenschaftliches Wörterbuch* towards the end of the 19th century and later (Baldinger 1960). For philosophical reasons (Hüllen 1999), I describe the onomasiological dictionary before (roughly) 1700 as being of the *speculative*, the one after this date as being of the *mental* type. The watershed is the work of John Locke with its new concept of the meaning of words as signs for simple and complex ideas (Hüllen 2000). If we leave dictionaries of languages with other than alphabetical letter systems aside (because they require special consideration anyway) and furthermore neglect a possible phonological entry arrangement in a dictionary (because to my knowledge such dictionaries do not exist) and also the reverse dictionary (because of its limited, purely academic, purpose), all the names used and mentioned above mean the same: A dictionary whose entries are not arranged alphabetically from left to right and from top to bottom but according to meaning. One can call this the dichotomy between a formal and non-formal arrangement in the macrostructure.

This purely superficial property is in fact indicative of two possible ways of processing language as they already appear in language learning. Children either point to X (e.g. a toy) asking 'What's that?' (onomasiological), or they hear a word /toy/ and ask 'What does it mean?' (semasiological). The difference between the two types of dictionaries therefore pertains to mental processing and the authors' intention of meeting it adequately. This may be more important than the fact of non-/alphabetization. The two methods of processing are perfectly common to every language user. They have their ultimate

motivation, of course, in the two sides of the linguistic sign (p. 90). They constitute the difference between, for example, *extracting* meanings *from* signs by reading and listening, and *expressing* meanings *in* signs by writing and speaking. The two ways of organizing dictionaries are meant to serve them in daily practice.

Igor Burkhanov rightly observes that the compromises between the extremes of the dichotomy are perhaps the, for metalexigraphy, more important phenomena than the 'pure' cases (provided they exist at all, see below). Analogical dictionaries, although of very limited importance outside the French-speaking world, are a good case in point. They make it clear that the difference between the macrostructure (the arrangement of lemmata or keywords) and the microstructure (the arrangement of explaining elements complementing the keywords) is of importance for deciding between the two principles. In present-day dictionaries, there is no compromise possible between them in the macrostructure. Our dictionaries are either semasiological or onomasiological, either alphabetical or ideographical (etc.), but cannot be both. A mixed procedure, as is occasionally to be found in the past (e.g. Adrianus Junius' *Nomenclator*, 1567) is no longer acceptable. But for the microstructure of entries a compromise between the two principles is frequent or even the common case, at least in explanatory dictionaries. We should distinguish between two cases:

(a) The explanation of word meanings by definition, paraphrase, quotation, etc. usually indicates overarching domains of which the lexeme in question is a member, as does the definition of a species by referring to the genus. Pragmatic tags like *terminology*, *medical language*, *poetry*, *nautical term*, etc. refer the reader to a general order which is helpful for finding the meaning of an individual word. In such cases, the mental processing goes both ways, *top-down* as well as *bottom-up*, which means it is onomasiologically as well as semasiologically oriented. We recognize an individual item by its features, but we also locate its place relative to higher ranking or coordinated ones.

(b) Synonym dictionaries with an alphabetical ordering of key words complement them with various lexemes which have partly overlapping and slightly differing meanings (according to the usual definition of a synonym). As the boundaries of networks of synonyms vary between small groups and large word fields, the entry articles of these dictionaries can also vary enormously in length. They can consist of just a few lexemes complementing the first one, but also of lengthy enumerations of lexemes which list all the pertinent words of a field (as analogical dictionaries do). These entry articles, whether small or large, function as onomasiological (or ideographical) self-contained units on a small scale. Synonym dictionaries of this sort are in fact a sequence of content-oriented (ideographical, onomasiological) word-lists arranged in an alphabetical succession of their key-words. The user must engage in both methods of processing when consulting them.

The difference between these types of compromises is that (a) is inevitable for semanticization, but (b) is an artificial and deliberately planned product. In present-day lexicography, synonym dictionaries have been developed into highly sophisticated blends as can, for example, be seen in *Longmans Language Activator* of 1993 where a simple alphabetical dictionary and a complex synonym dictionary have been mixed in such a way that the latter functions as the explanatory part of the former. To sum up: The alternative between alphabetical (formal) and non-alphabetical (onomasiological, ideographical) dictionaries works only in their macrostructure. In their microstructure they are of a mixed nature because the techniques of semanticizing work in both ways. The only dictionary which is purely onomasiological in the macro- and also in the microstructure is the cumulative synonymy of the Roget-type *Thesaurus*. It does not semanticize at all but leaves this to the reader. (To my knowledge a synonym dictionary whose microstructure is alphabetical does not exist outside the usual natural and scientific taxonomies where, for example, the names of trees, etc. are given alphabetically.)

For synonym dictionaries the development of modern cognitive semantics is indeed of the highest importance. Its concepts come closer to the clustering of synonyms than any others do, because they were originally developed as units of the mental lexicon. They are psychologically rather than logically orientated. This brings our argument back to Igor Burkhanov's book.

The author gives a broad picture of the recent development of semantics (chapter 4) which the readers will welcome highly because their attention has been directed to it from the beginning of the book. (For chapter 3 see below.) He outlines field-theory in its structural, comparative, and cognitive elements. It has its counterpart in Eastern European linguistics in the concept of the (thematic, functional) word-group. It favours the onomasiological approach to lexis and, together with communicative grammar, has influenced language teaching enormously. It was followed by componential analysis where meanings appear as the configuration of features and this means as the combination of, possibly universally valid, atomic elements. With its clear-cut boundaries and methodical binarism it was more appropriate to logical thought than to common language use. This kind of language analysis took the expert's language as its model rather than the common man's. The fact that everyday performance and expert performance do not coincide itself triggered the ideas of cognitive linguistics with its concepts of prototypes (in word meanings and also in categorization) and of frames, scripts, scenarios, etc. instead of feature configurations and fields. A realistic classification was to replace the former idealistic one. Consequently, the idealistic classifications of world-knowledge in onomasiological dictionaries must now give way to new ordering principles which follow the order of folk knowledge. 'Frame', the most common one, is a complex of socially relevant scenes, situations, states, etc. in common life, where the multiple relations are explained with the help of a presupposed meaningful whole (as Fillmore, for example, pointed out by explaining words

for selling and buying). 'Script', in many respects its counterpart, is a series of events where, again, each action or step becomes meaningful as part of a higher unit of behaviour (as was explained by the example of going to a restaurant). 'Schema' is a more abstract unit of structural knowledge (for example, of the chemical elements). A script is to a cluster of actions what a frame is to a cluster of concepts. A frame is to common knowledge what a schema is to the more scientific type. Here again, a regrettable abundance of scientific terminology (correctly reported by the author) is to be found, because, besides using the terms mentioned, linguists also speak of 'idealized cognitive models', 'image schemas', 'domains', 'scenarios', 'interactive networks', etc. Moreover, categorization is now understood to be prototypical. This allows for fuzzy borders of meanings, membership of classes on the basis of family resemblances, and graded categorial status (insofar as, for example, a robin is much more a prototypical bird than a penguin, or an agentive noun is much more a prototypical subject of a sentence than *it* (in *It is raining*)). All these concepts have in common that they come from experience rather than from thought. It is obvious that such new constructs have the potential for giving dictionaries a new shape. They impose a new order on the presentation of words, provided this order is not determined by the empty formalism of the alphabet, and they demand new guidelines for meaning definition. The condition, however, on which these innovations can be accepted in lexicography is that they meet the needs of dictionary users who, after all, do not consult their books in order to look at linguistic systematization but to improve their linguistic competence.

Igor Burkhanov discusses three types of meaning-oriented dictionaries (chapter 3). The first, the thesaurus, has a classification according to the standards of accepted philosophy and the sciences. The ideological bias that inevitably goes with this system is obvious. But it is interesting to reflect on the idea that this bias is less strong on the lower levels of abstraction, where the entries of a dictionary appear, than on the higher ones, because even the ideological deduction of closed systems leaves the compilers some freedom for the choice and the arrangement of lexemes. The second, the thematic dictionary, is an ideographical (onomasiological) work curtailed to the needs of foreign language teaching. It is a dictionary type in which the practical purposes dominate scientific planning. The third, finally, is the systematic dictionary. (In the course of this chapter, the text alternates between calling this dictionary type 'systemic' and 'systematic'. Vis-à-vis the fact that there is a linguistic school which calls itself 'systemic' in opposition to 'structuralist' (e.g. M.A.K. Halliday), it should be made clear which adjective is meant — the rather general 'systematic' or the more specific 'systemic'). It takes as its principles of classification linguistic features between lexemes of one word-class, as they serve, for example, to subcategorize verbs and adjectives. They are regularly abstract semantic notions such as, for example, 'inherent property', 'dynamic property', 'relational property', and 'class membership' as features of adjectives (p. 135). It is, however, difficult to imagine a practical use for such dictionaries (of adjectives, of verbs, etc.) beyond the interest of linguists.

The characterization of the learner's ideographical dictionary of the general vocabulary (chapter 5) is the actual focus of the whole book. It starts with the sentences: 'The main purpose of this chapter is not an attempt to provide a comprehensive survey of the state of the art in pedagogical ideography with respect to foreign language teaching. Instead, it aims to raise questions, formulate some objectives, and illustrate the way in which this field of lexicography should be developing' (p. 199). After the long journey through the meanders of the semantics and lexicography of the last fifty years or so, this is a rather modest aim — and the author is to be praised for it, because he avoids the impression that ideal dictionaries for learners fall from the tree of linguistics like a ripe fruit.

The types of dictionaries discussed here are coursebooks, usage guides, minimal dictionaries, etc. Of course, they demand an overall classification and a format of meaning definition which serves the needs and the capabilities of their users — difficult as their assessment may be. For example, the method of *Collins Cobuild Dictionary* (1987) of giving meanings not by definitions (in metalanguage) but by quotations (in object language) has the advantage of showing the learners language-in-use and is, thus, presumably, what they need. Here the dictionary partly takes on the function of the coursebook. The main postulate of Igor Burkhanov, however, is that the new pedagogical dictionary, which he calls the *Concept-Word Dictionary*, has to interface grammar and meaning description, the conceptual domains, as, for example, described by Ronald Langacker, serving as the coordinates of a content-oriented dictionary. This means, among other demands, inserting learner-oriented explanations of those categories which structure lexis into the dictionaries. An example, the presentation of the category *time*, is given (p. 215). It is highly interesting because of two points, namely (a) the assumed possibility of explaining abstract concepts discursively so that they can be grasped by foreign language learners (who after all are not all of them academics), and (b) the assumed power of such reflections for boosting the learning process. Neither assumption has been tested so far, although it is certainly worth doing so. To explain word meanings by explaining the domains in which a lexeme is situated demands a high degree of metalanguage which will be difficult to control in its pedagogical effects. Ultimately, the proposed method means that cognitive linguistics in itself is the best method of foreign language teaching.

Many further deliberations on learning dictionaries deal with what has been part and parcel of so-called communicative language teaching for some decades now. It is valuable to find these criteria in a concise overview and to be reminded that they are rather a bundle with loose ends than a system of thought.

Finally, Igor Burkhanov discusses three procedures in the preparation of a dictionary, namely (a) the observation of daily communication (primary material) and of other dictionaries (secondary material), (b) the intuition of the native speaker, and (c) experiment in the sense of Labov and Eleanor Rosch. These are the ways that lead to the entry articles whose new shapes are given

in several examples (chapter 6). These are lexemes denoting, for example, 'landscape features' (like *mountain, island, district*) or physical objects, domains like 'similarity', 'smoking', or 'artefacts' (the latter in a crosslinguistic comparison). The defining categories which separate them from neighbouring lexemes (like altitude, shape, relative position, etc. in the case of *mountain*) are to be found and tested by the three procedures mentioned. This leads to highly interesting analyses by the author. It is important that he also has a clear idea of their pedagogical presentation: 'In addition to cross-references intended to account for the general conceptual affinities [...], a learner-oriented ideographic dictionary should provide adequate explications of the organizing concepts [...]. Those descriptive definitions of the conceptual domains in question can be presented after the headings and before the lemmata assigned to the semantic domains in question' (p. 303). Various entries are offered as models for the envisaged *Concept-Word Dictionary*, either as a frame or as a script. They are fascinating to read. In fact, the dictionary as a word-list with attached explanations is here replaced by a meta-lexicographical explanation of how to arrive at these explanations by observation, intuition, and experiment. The idea is intriguing, the examples are convincing. The question is, however, open of whether this 'explaining how to explain' is really helpful for foreign language learning.

Igor Burkhanov's book makes a very interesting and stimulating read. It gives plenty of information and also food for critical thought. Not the least profit for the reader versed in the Western tradition of language pedagogy is to be gained from the author's frequent references to publications on the topic in Central and Eastern Europe which otherwise (alas) would remain unknown.

Endnotes

- Baldinger, K.** 1960. Alphabetisch oder begrifflich gegliedertes Wörterbuch? *Zeitschrift für Romanische Philologie* 76: 521-36.
- Green, J.** 1997. *Chasing the Sun. Dictionary-Makers and the Dictionaries They Made*. London: Pimlico.
- Haß-Zumkehr, U.** 2001. *Deutsche Wörterbücher. Brennpunkt von Sprach- und Kulturgeschichte*. Berlin/New York: Mouton de Gruyter.
- Hüllen, W.** 1999. *English Dictionaries 800–1700: The Topical Tradition*. Oxford: Clarendon.
- Hüllen, W.** 2000. *John Locke, Semanticist*. Dalton-Puffer, C., and N. Ritt (Eds.). *Words: Structure, Meaning, Function*: 141-52. Berlin/New York: Mouton de Gruyter.

Werner Hüllen
University of Essen
Herchenbachstrasse 1
40470 Düsseldorf
Germany
(werner.huellen@uni-essen.de)

Julie Coleman and Christian J. Kay (Eds.). *Lexicology, Semantics and Lexicography: Selected Papers from the Fourth G.L. Brook Symposium, Manchester, August 1998.* 2000, xiv + 249 pp. Current Issues in Linguistic Theory Volume 194. ISBN 90 272 3701 8 (Eur.), 1 55619 972 4 (US). Amsterdam/Philadelphia: John Benjamins. Price US\$ 75,00 (Hb).

Introduction

Volume 194 of the series *Current Issues in Linguistic Theory*, entitled *Lexicology, Semantics and Lexicography*, comprises a selection of eleven academic contributions that were originally presented at the Tenth International Conference of English Historical Linguistics. The disciplinary focus of this volume of texts is the diachronics of English vocabulary, and the representation thereof in dictionaries. From a theoretical point of view prototype theory assumes a central position, and the methodological approach is mainly corpus-analytical. The above-mentioned foci are highlighted as follows by the editors (p. vii):

Two points stand out particularly. The first is the impact of prototype theory and cognitive approaches generally in lexical studies. The second is the very positive effects of the remarkable range of electronic resources now available to historical linguists, notably corpora, dictionaries, bibliographies and thesauruses. These are important both quantitatively, in the amount of data they make available, and qualitatively, in the versatility of their searches.

These points of emphasis do not seem to be coincidental as they converge with the theoretical and methodological fundamentals of the Cognitive Linguistics enterprise, more specifically with those of usage-based theories and models of language (Barlow and Kemmer 2000; Langacker 1987, 1988, 2000). Three characteristic assumptions of usage-based models that coincide with the abovementioned foci are: the relation between language usage and diachronic change; the importance of usage data; and the central role that categorization plays in our understanding of language structure. On the grounds of these shared focal points, one could view *Lexicology, Semantics and Lexicography* not only as a welcome addition to the growing body of work on usage-based models of language change (e.g. Croft 2000; Geeraerts 1997; Kellermann and Morrissey 1992; Verhagen 1998), but also as a demonstration of the value of usage-based models of the lexicon for lexicographical work.

It is regrettable that the term "usage-based" does not feature in the title of the book, in the introduction, or in the subject index (pp. 241-249). The absence of an overt reference may be ascribed to the fact that, although the term was coined by Langacker as early as 1987 in his book, *Foundations of Cognitive Grammar Volume I*, it has only recently started to gain currency (Kemmer and Barlow 2000: vii). This assumption is substantiated by an overview of the programmes for the 1999 and 2001 International Conferences on Cognitive Linguistics. While none of the presentation titles for the 1999 conference contained the term

"usage-based" (despite the fact that some of them were inherently usage-based), this label appears overtly in 6 of the 2001 titles.

The title of the book under review is also slightly problematic from the point of view of readers' expectations. The syntactic co-ordination of *lexicology*, *semantics* and *lexicography* suggests that an equal (or close to equal) amount of emphasis and space has been devoted to each of the three disciplines, which would amount to a thematic spread more or less comparable to that of Jerzy Tomaszczyk and Barbara Lewandowska-Tomaszczyk's *Meaning and Lexicography* (1990). Yet, with the exception of the articles by Kay and Coleman, the majority of contributions in the book are in essence diachronic descriptions of lexical meaning, based on lexicographical data. But then again, there are mitigating circumstances: titles of publications are merely labels that serve as cues to assist the prospective reader in forecasting content. Moreover, the prefatory notes by the authors delimit the scope of the book fairly accurately.

In this review, we shall evaluate the contributions in *Lexicology, Semantics and Lexicography* in terms of the overtly expressed focus (p. vii), with particular emphasis on the ways in which they contribute to the affirmation and/or elaboration of usage-based models of lexical change.

Evaluation of the respective contributions

Andreas Fischer. *Lexical Gaps, Cognition and Linguistic Change.*

This contribution gives a well-expounded, theory-based overview of the structure and origins of lexical gaps. It could be a useful source of reference for students of lexicology, semantics and lexicography. Lexicographical notions, particularly those of zero equivalence and surrogate equivalence (although these metalexicographical concepts have not been explicated in the article), could be explained cogently against the background of Fischer's exposition.

The claim that "[t]he paper breaks new ground when it attempts to adduce cognitive factors to explain lexical gaps" (p. 15) may be somewhat of an over-appraisal, given that the author does not proceed beyond a fairly superficial description of the roles that psychological salience, perceptive salience and prototypicality may play in constituting lexical gaps. Moreover, the distinction made between psychological and perceptive salience is not quite clear, and the author gives no real evidence to clarify this distinction.

However, Fischer's explanation of lexical gaps in terms of more general psychological capacities makes a lot of sense, and could form the basis for further usage-based research on the topic in question.

Gabriella Rundblad and David B. Kronenfeld. *Folk-Etymology: Haphazard Perversion or Shrewd Analogy?*

Rundblad and Kronenfeld offer a comprehensive and multi-faceted account of

the phenomenon of folk-etymology by departing from a simple, yet motivated, typology. Unfortunately cognitive aspects of this type of language change have not been addressed adequately. Rundblad and Kronenfeld argue that "[j]ust as analogies do, folk-etymologies reveal how speakers regard linguistic and cultural matters" (p. 21), yet they do not show that analogical change is essentially a cognitive process (cf. Winters 1997); neither do they indicate the implications of such processes for the genesis of folk-etymologies.

Another point of criticism is that, although the authors combine a data-driven approach with systematic sampling techniques, the representativity of the data across the spectrum of language usage may be questionable as their database was rather small (100 folk-etymologies taken from Palmer's *Dictionary of Folk-etymology*). Moreover, examples are restricted to lexicalized etymologies.

From a positive angle it could be said that this article is accessible to a fairly wide readership, and may serve as a foundation for further research on the topic of folk-etymology.

Päivi Koivisto-Alanko. *Mechanisms of Semantic Change in Nouns of Cognition: a General Model?*

This contribution calls to mind the early structuralist work of Jost Trier, who studied the Middle High German terms of knowledge *kunst*, *wisheit* and *list*. However, the author explicitly states that his starting point is very different from that of Trier. His point of departure is not the semantic field of particular words constituting a semantic field, but the concept delimiting a specific semantic field, namely COGNITION. A second difference between his approach and that of Trier is that he does not focus on semantic fields per se but on the directionality of semantic change, with specific reference to the English word *wit*. Two large corpora as well as the OED served as databases.

Apart from the intrinsic value of his study, Koivisto-Alanko's corpus-based findings authenticate previous research on the directionality of semantic change resulting from metonymic transfer (cf. Carstens 1992).

On the negative side, however, it is a pity that the author did not take into account recent cognitive insights on semantic change, such as those presented in Kellerman and Morrissey (1992).

Christian J. Kay. *Historical Semantics and Historical Lexicography: will the twain ever meet?*

Kay summarizes some of the possibilities opened up by cognitive semantics, including usage-based approaches, for description and problem-solving in the domain of historical lexicography. In line with current thinking in mainstream cognitive linguistics he stresses the advantages of an approach that discards rigid boundaries in favour of prototypicality, fuzziness and interdisciplinarity. In our opinion, this article is one of the most important contributions in the vol-

ume under review, in that it sets the scene for further explorations in usage-based models of lexicology and lexicography. We agree with Kay (p. 65):

This approach will not solve all our problems, and we are still far from achieving a comprehensive theory of lexical semantics, but it offers enlightening ways for looking at some of them.

Julie Coleman. *Strange Linguists: The Cant and Slang Dictionary Tradition*

Coleman's contribution is one of the few that do not invoke a usage-based approach to explain specific lexical phenomena or to support particular lexicographical practices. Besides the entertainment provided by her examples and quotes, Coleman captures interesting assumptions and "linguistic theories" underpinning dictionaries of cant and slang published between the sixteenth and nineteenth centuries. This is accomplished through in-depth analyses of the prefatory matter of the dictionaries in question. She cogently argues that although the prefaces may be criticized for their subjectivity, they constitute entry points to the otherwise opaque editorial policies underlying these lexicographical products. Coleman thereby highlights the fact that the value of qualitative research should not be underestimated in today's corpus-dominated world of lexicography.

Her research should also be valued from a sociolinguistic point of view as it sheds light on social factors that contributed towards the development of cant and slang between the sixteenth and nineteenth centuries.

Maurizio Gotti. *Lexical Choices in an Early Galilean Translation*

In our opinion Gotti's most important contribution is not his critical reflection on Salusbury's translation of Galileo's scientific works, but the implications of his views on translation as part of corpus planning for developing languages, and for the development of scientific knowledge through translation. The following quotations from Gotti's article (p. 87) are offered in support of this view:

This work [Salusbury's translations — authors], which can be considered fundamental for the growth of scientific thought in this period and which has greatly influenced the evolution of several disciplines, not only in England but all over Europe, introduced many new terms, which were soon included in the specialized terminologies of the various European languages.

and

A great contribution to this lexical growth was provided by translations of foreign texts, particularly those pertaining to the various scientific branches.

Of particular interest to Gotti is Salusbury's innovative translation of metaphor and his use of certain grammatical constructions in the target language (Eng-

lish) to render a satisfying translation of the LSP of the source language (Italian) without producing stylistic clashes or inconsistencies in the target language (p.92):

[O]n the whole the form of Salusbury's translation follows the standard rules of the language and — although it succeeds in maintaining the characteristics of Galileo's prose — it reads in a very natural English style.

A question that comes to mind is whether this methodology could be viable with regard to developing LSP's for the African languages in branches of science and technology that have hitherto remained outside the scope of serious corpus development. In other words, Gotti's reflections on Salusbury's work triggers the question of whether a decontextualised methodology of finding (and coining) translation equivalents for English scientific terms should not be discarded in favour of a methodology that departs from the translation of authentic texts, a methodology that has the creation of new terms as a natural by-product.

The methodology that has just been described is particularly promising in that it does not entail a simple process of borrowing from the source language, but a process of enrichment through "the processes of borrowing and calquing or [...] the adoption of existing lexical elements endowed with new meanings, thus making use of the semantic redefinition processes of specialization, semantic innovation, metonymy and conversion" (p. 97) according to contextually motivated procedures.

In South Africa, the translation of examination papers for Physical Science into Sepedi by P.F. Ntake and P.T. Pare (Ntake and Pare 2001) could be regarded as a modern-day parallel to Salusbury's work. Their translations, and the resulting English–Sepedi glossaries, may indeed have an influence on the development of Sepedi similar to that which Salusbury's translations of the scientific works of Galileo had on the development of an English LSP for Science and Mathematics.

Carole P. Biggam. *Grund to Hrof: Aspects of the Old English Semantics of Building and Architecture*

In a fascinating contribution, Biggam illustrates how much scientific inquiry has benefited from the rapprochement between semantics and pragmatics (or language and culture) since the cognitive revolution. She demonstrates that knowledge of the world does not only contribute towards shaping our understanding of language and linguistic structure, but that language is also a gateway to understanding culture and its physical manifestations. Biggam combines the "forces of archaeological and semantic research" by consulting sources such as Taylor's three volume *Anglo-Saxon Architecture* as well as *A Thesaurus of Old English* (which lists more than 630 words connected with building, building

materials, types of building, parts of buildings, and rooms) to reconstruct the world of Anglo-Saxon buildings.

Her contribution is a clear demonstration of the benefits of transdisciplinary usage-based research. The strategy to "build" a stone church and a timber house using Old English vocabulary in the course of the article, offers a truly usage-based strategy to improve our understanding of cross-linguistic influences on language development and change.

Heli Tissari. *Five Hundred Years of Love: a Prototype-Semantic Analysis*

Similar to the contribution by Koivisto-Alanko, this article combines semantic fields with prototype theory, supporting claims with both usage-based and dictionary-based data. It describes the semantic microfield constituted by the English lexeme LOVE by making a detailed corpus-based diachronic comparison between Early Modern and Present-Day English. The well-known distinction between 'family love', 'friendship', 'sexual love', 'religious love' and 'love of things' forms the point of departure. On the basis of semasiological evidence from English dictionaries, these distinctions are then taken as the five major clusters or senses which are situated in different conceptual domains.

Although the author concedes that there are strong family resemblances between these 'loves', she indicates (on the basis of corpus statistics) that 'sexual love' is dominant in both periods, becoming even more frequent in Present-Day English. Tissari's numerical analyses indicate that the relative frequencies of the five 'loves' changed between the two periods studied: 'family love' and 'friendship' became less frequent, while the relative frequencies of 'love of things' and 'religious love' remained largely the same.

Similar to other contributions in the volume under review, this article is accessible to a fairly wide readership, including students of linguistics.

The research on which this contribution reports could be furthered by combining Tissari's views with the insights of other cognitive linguists on the genesis of metaphorical expressions and conceptual metaphors for love (e.g. Botha 1998, Gibbs 1994, Lakoff 1987, etc.).

Louise Sylvester. *The Vocabulary of Consent in Middle English*

Sylvester combines a semasiological and an onomasiological approach by meticulously examining the definitions, ordering of senses, and use of citations regarding the term *consent* in major dictionaries of English, while also looking at synonyms and near synonyms of the term around the time of its adoption into Middle English.

Her primary focus is the definition language used in major dictionaries of English. A secondary focus is the lexicalization of the notion of CONSENT in Old English, involving lexical-field development.

On the basis of her data, Sylvester concludes that although the concept of

CONSENT was lexicalized in Old English, the resulting lexical items may largely be regarded as rarely used, and hardly central to the Old English corpus. Unfortunately, no usage-based data (apart from dictionary data) is offered to substantiate her claims.

Claire Cowie. *The Discourse Motivations for Neologising: Action Nominalization in the History of English*

Cowie's article is certainly one of the highlights of the book, in that it is clearly usage-based, and invokes quantitative as well as qualitative methods of research. Along with scholars such as Baayen and Renouf (1996) she argues that word-formation is conceptually driven, rather than morphologically motivated. Cowie further argues that neologising is "fundamentally a social and cultural practice", and contends that her contribution merely "begin[s] to explore the extralinguistic motivations for neologising" (p. 180). This is done through a context-sensitive, corpus-based study, followed by discourse analysis.

The extralinguistic factor targeted in her study is register. She focuses on the English deverbal nominalizing suffix *-(t)ion*. What makes her approach so unique is the fact that she adopts a position between that of "passive correspondence" (between words and the concepts, p. 182) and a "constructivist" approach, in which language construes human experience rather than simply reflecting it. Her focus is the "pragmatic or functional role that a new derivation plays in structuring discourse, and the use of word-formation to achieve a certain stylistic effect" (p. 182). She challenges the clear-cut distinction between register markers and common or core linguistic features. A common linguistic feature such as nominalization may, for instance, lose its functionality and become associated with a marker of a particular style or register. A corpus-based analysis reveals that medicine and science consistently produce more neologisms in *-(t)ion* than the other registers, such as letters, legal prose, news, sermons and fiction.

Through in-depth discourse analysis Cowie also proves that nominalization in scientific discourse fulfills different functions, both on a lexical pole (providing a designation for a 'nameworthy' segment of extralinguistic reality) and as a marker of style, e.g. by effecting a learned style and creating a distancing effect.

Cowie's contribution emphasizes the importance of bridging the gap between a structuralist (modernist) approach aimed at describing existing structures and processes of change, and a critical (post-modernist) approach aimed at describing the socially constitutive role of such processes. Cowie's analyses could indeed be viewed as a fulfillment of the ideals of Critical Discourse Analysis and Critical Linguistics, as worded by Sandig and Selting (1997: 153):

A further area of research [...] is concerned with the ways in which different world views, predominantly with respect to phenomena such as power and status, manifest themselves implicitly and subconsciously in the style of texts.

R.W. McConchie. *The Vernacularization of the Negative Prefix dis- in Early Modern English*

This article explores the assimilation of the negative prefix *dis-* in Early Modern English, a prefix that has changed from being restricted to usage in borrowings from Latin and French, to a fully productive prefix by the end of the sixteenth century.

An impressive number of usage-based sources were invoked to study the assimilation process, including electronic corpora, dictionaries, and concordances, thereby creating the expectation that a satisfactory answer would be found to the question of whether there is a connection between the rarity of a word and the sociolinguistic significance of its occurrence. The paper then specifically reports on the significance of the statistical rise in the appearance of words starting with the prefix *dis-* in the sixteenth century, and the sociolinguistic implications of this phenomenon.

Although the methodology used in the article is sound, the results of the research (and the conclusions) are, however, somewhat disappointing. After wading through pages of discussions, data tables and graphs the reader is rewarded with "conclusions" such as (p. 223):

[T]hat the rise of *dis-* takes place in the sixteenth century and forms an s-curve is *probable* within the limits of data available [authors' emphasis];

and

[T]he association with literary humanism is *still possible but* must be demonstrated in detail [authors' emphasis].

These rather inconclusive conclusions are followed by the concession (or confession!) that "the result is inevitably unsatisfactory, but may be sufficient to reveal the need for more complete corpora, and more extensive studies at the same level of detail", which is then topped by the following summary of the sociolinguistic implications of the study (p. 223):

The sociolinguistic implications seem to be that idiosyncrasy may be so marked in certain registers and levels of rarity as to distort present data.

Conclusion

The series *Current Issues in Linguistic Theory* is overtly aimed at publicizing current research findings in the field of linguistics. Its target market is the scholarly fraternity, yet the books published in this series are also widely consulted

by students of linguistics. Volume 194, *Lexicology, Semantics and Lexicography*, comfortably fits into this frame. It introduces various topical issues in the field of lexical semantics, particularly focusing on the diachronics of the lexicon within a corpus-based (usage-based) framework. In those instances where we have criticized individual authors for not invoking the most recent theories and terms, our criticism should be partially offset against issues such as comprehensibility, clarity and accessibility.

In terms of methodology and theoretical underpinnings, *Lexicology, Semantics and Lexicography* clearly stands in the sign of the corpus-analytical, usage-based era. While *Meaning and Lexicography*, published just over a decade ago (Tomaszczyk and Lewandowska-Tomaszczyk 1990), contains only one contribution (out of nineteen) invoking corpus evidence (cf. Hanks 1990: 31-42), the work under review contains only one contribution that is *not* overtly corpus-based (and, as indicated above, the overt aim of this "exception" was to demonstrate the relevance of qualitative research).

This evidence indicates that in the course of a decade linguistic theory has moved a galaxy away from "the ideal speaker-hearer" and has embraced "the real user", as instantiated by a representative sample (in terms of both frequency and spreading) of authentic usage. *Lexicology, Semantics and Lexicography* demonstrates that usage-based approaches are most valuable in terms of their explanatory value with regard to lexical phenomena. Coupled with "the meticulous scholarship characteristic of philology" (p. vi), historical lexicology and historical lexicography has a lot to benefit from the heuristic tools of a usage-based approach.

A particular contribution, namely that of Claire Cowie, moves even further away from traditional approaches in that she provides a glimpse of the increasing rapprochement between quantitative and qualitative research, and between modernist and postmodernist approaches in the humanities and social sciences, a view that stresses the dialectical relationship between the descriptive and socially constitutive uses of language. Also in the contribution by Biggam, traces of postmodernism can be found (in the format of constructionism), and it will not be surprising if critical and constructionist approaches feature even more overtly in future lexicological studies.

References

- Baayen, R.H. and A. Renouf.** 1996. Chronicling The Times: Productive Lexical Innovations in an English Newspaper. *Language* 72(1): 69-96.
- Barlow, M. and S. Kemmer (Eds.).** 2000. *Usage-based Models of Language*. Stanford: CSLI Publications.
- Biebuyck, B., R. Dirven and J. Ries (Eds.).** 1998. *Faith and Fiction. Interdisciplinary Studies on the Interplay between Metaphor and Religion*. Frankfurt am Main: Peter Lang.
- Botha, W.J.** 1998. The Love Frame in the Bible. A Cognitive Linguistic Analysis. Biebuyck, B., R. Dirven and J. Ries (Eds.). 1998: 55-82.

- Carstens, A.** 1992. Metonimie, polisemie en die leksikografie. *SA Journal of Linguistics* 10(3): 109-113.
- Croft, W.** 2000. *Explaining Language Change: An Evolutionary Approach*. Harlow: Longman.
- Geeraerts, D.** 1997. *Diachronic Prototype Semantics: A Contribution to Historical Lexicology*. Oxford: Clarendon Press.
- Gibbs, R.W. Jr.** 1994. *The Poetics of Mind: Figurative Thought, Language and Understanding*. Cambridge: Cambridge University Press.
- Hanks, P.** 1990. Evidence and Intuition in Lexicography. Tomaszczyk, J. and B. Lewandowska-Tomaszczyk (Eds.). 1990: 31-42.
- Kellermann, G. and M.D. Morrissey (Eds.)**. 1992. *Diachrony within Synchrony: Language History and Cognition*. Frankfurt am Main: Peter Lang.
- Kemmer, S. and M. Barlow.** 2000. Introduction: A Usage-based Conception of Language. Barlow, M. and S. Kemmer (Eds.). 2000: vii-xxviii.
- Lakoff, G.** 1987. *Women, Fire, and Dangerous Things*. Chicago: University of Chicago Press.
- Langacker, R.W.** 1987. *Foundations of Cognitive Grammar Volume I: Theoretical Prerequisites*. Stanford: Stanford University Press.
- Langacker, R.W.** 1988. A Usage-based Model. Rudzka-Ostyn, B. (Ed.). 1988: 127-161.
- Langacker, R.W.** 2000. A Dynamic Usage-based Model. Barlow, M. and S. Kemmer (Eds.). 2000: 1-63.
- Ntake, P.F. and P.T. Pare.** 2001. *Questions and Answers Matric Physical Science / Dipotšišo Le Dikarabo Maramatlou Thutamahlale*. Pretoria: Pan South African Language Board.
- Rudzka-Ostyn, B. (Ed.)**. 1988. *Topics in Cognitive Linguistics*. Amsterdam: John Benjamins.
- Sandig, B. and Selting, M.** 1997. Discourse styles. Van Dijk, T. (Ed.). 1997: 138-156.
- Tieken-Boon van Ostade, I., M. van der Wal and A. van Leuvensteijn (Eds.)**. 1998. *'Do' in English, Dutch and German*. Amsterdam: Stichting Neerlandistiek VU.
- Tomaszczyk, J. and B. Lewandowska-Tomaszczyk (Eds.)**. 1990. *Meaning and Lexicography*. Amsterdam: Benjamins.
- Van Dijk, T. (Ed.)**. 1997. *Discourse as Structure and Process. Discourse Studies: A Multidisciplinary Introduction Volume I*. London: Sage Publications.
- Verhagen, A.** 1998. Changes in the Use of Dutch 'doen' and the Nature of Semantic Knowledge. Tieken-Boon van Ostade, I., M. van der Wal, and A. van Leuvensteijn (Eds.). 1998: 103-119.
- Winters, M.E.** 1997. Kuryłowicz, Analogical Change, and Cognitive Grammar. *Cognitive Linguistics* 8(4): 359-386.

Adelia Carstens
School for Languages
University of Pretoria
Pretoria
Republic of South Africa
(acarsten@postino.up.ac.za)

Gerhard B. van Huyssteen
School for Languages
Potchefstroom University for C.H.E.
Potchefstroom
Republic of South Africa
(afngbvh@puknet.puk.ac.za)

G.J. van Wyk (Tegn. Red.), A.E. Cloete, A. Jordaan, H.C. Liebenberg en H.J. Lubbe. *Etimologiewoordeboek van Afrikaans (eerste fase)*. 2002. 271 pp. Stellenbosch: WAT-Trust.

Daar is in die onlangse paar dekades bloedweinig oor die Afrikaanse etimologie gepubliseer. Die "jongste" samevatting is Boshoff en Nienaber 1967; vergelyk ook — naas 'n verskeidenheid voortreflike studies deur Scholtz, waaronder Scholtz 1972 —, Smith 1962 ('n versameling koerantartikels uit die laat negentiendertigerjare), Combrink 1969 en Franken 1953. Die mees resente bydrae tot die Afrikaanse etimologie kom uit 'n heel onverwagte oord, die Engelse leksikografie, met name Silva 1996, 'n bron wat ruim deur *Etimologiewoordeboek van Afrikaans* benut word.

Etimologiewoordeboek van Afrikaans (EWA) baat in verskeie opsigte by 'n verbintenis met die Nederlandse taalgebied:

- ten eerste deur die Nederlandse etimologie soos dit neerslag vind in verskeie woordeboeke wat gereeld bygewerk word, deur omvangryke publikasies soos Van der Sijts 2001 en deur die uitgebreide historiese gegewens in die pas voltooide WNT;
- ten tweede deur ruim befondsing waarsonder die huidige projek glad nie uit die wegspringblokke sou kon kom nie; en
- ten derde deur die samewerking met die Instituut voor Nederlandse Lexicologie wat in die persoon van prof. A.M.F.J. Moerdijk met die opleiding van die redaksionele span behulpsaam was.

Boshoff en Nienaber se *Afrikaanse etimologieë* van 1967 wou 'n naslaanwerk wees: dit bevat 'n baie groot aantal lemmas wat so bondig moontlik aan die hand van die literatuur bespreek word. Die boek lees moeilik, selfs as naslaanwerk, weens die struikgewas van geleentheidsafkortings en bronverwysings. EWA slaan 'n ander koers in: dit wil in die heel eerste plek 'n boeiende, leesbare teks wees.

Dit gebeur wel dat verklarende woordeboeke ter hand geneem word om deurgeblaai en gelees te word, maar so iets is tog uiters ongewoon. Verklarende woordeboeke is immers in die eerste plek werke wat gou ten opsigte van 'n bepaalde probleemkwessie nageslaan en nie vir stigting of afleiding gelees word nie. Met etimologiese woordeboeke is dit heeltemal anders gesteld: hulle is die enigste lid van die woordeboekfamilie wat nie net nageslaan word nie, maar ook, soekend of aaneenlopend, gelees word. Op hierdie behoefte speel EWA in deur 'n beperkte versameling goedgekose lemmas so uitvoerig en onderhoudend moontlik sonder taalkundige omhaal te bespreek.

Veral van die langer artikels bevat verwysings na bronne en selfs ook na polemieke. Soms hinder dit. Die omslagtige bespreking van die uitroep *alla* met die talryke bronverwysings bevestig 'n enkele feit: dat die woord 'n ontlening

aan Kaapse Maleis is. Die kritiek op die etimologie van *klops* in Silva 1996 lewer geen wesenlike bydrae tot hierdie artikel nie. Die uitvoerige bespreking en bronneapparaat by *paai* en *rondawel* sal 'n leser afskrik.

Nietemin, EWA bied vir die grootste deel sy ryk skat van inligting op uiters bevatlike en onderhoudende wyse aan. Ek noem hier 'n los versameling van die artikels waaraan ekself groot vreugde gehad het: *amper*, *aspoester*, *biltong*, *bokmakierie*, *fiskaal*, *flennie*, *galberou*, *geelwortel*, *gentoe*, *gorrel*, *halfmens*, *hart-beeshuisie* ('n probleemgeval wat besonder keurig bespreek word), *jammerlappie*, *janbruin*, *kankerbos*, *kapater*, *maaifoedie*, *manel*, *nadroe*, *naklip*, *nerina*, *oes*, *omgeëllie*, *oorlams*, *pad*, *pampoen*, *patrys*, *piering*, *rabbedoe*, *renostervoël*, *salpeter*, *sardonies* en *seidissel*.

Wat die makrostruktuur betref, meld die Inleiding dat hierdie uitgawe, die eerste fase van die projek, sowat 3 400 lemmas bevat en dat die finale uitgawe uit meer as twee maal soveel lemmas sal bestaan. Selfs tien duisend lemmas sal geensins 'n volledige versameling "etimologiseerbare" Afrikaanse lemmas verteenwoordig nie, en dit is goed dat EWA geensins makrostrukturele volledigheid op die agenda het nie. Nogtans, hoewel daar in hierdie stadium niks oor ontbrekende lemmas gesê kan word nie, vertrou ek dat bv. *hoekom*, *kierie*, *koedoe*, *skaduwee* en *stukkend* wel in die finale uitgawe opgeneem sal word.

Die etimologiese inligting in 'n aantal artikels is so skraal dat hierdie lemmas gerus geskrap kan word, bv.: *bek-en-kloueer*, *boetie*, *boetie-boetie*, *bossiedokter*, *bossiekop*, *dek*, *dekgras*, *dik*, *doek*, *dol*, *eetservies*, *eikeboom*, *erf*, *erfdeel*, *fanfare*, *fantasties*, *fraai*, *geiser*, *geit*, *geselligheid*, *gewas*, *gier*, *gifappel* (en ander samestellings met gif-), *glasogie*, ¹*glip*, *glyjakkals*, *gneis*, *goddelik*, ¹*gom*, *gordyn*, *gorilla*, *grasdak*, ¹*grief*, ³*grou*, ⁴*grou*, *gru*, *gruis*, *handlanger*, *hardepeer*, *hartomleiding*, *hekel*, *helpeper* en *huilboom*.

Die mikrostruktuur is duidelik en word oorsigtelik in die voorwerk uiteengesit. Die mikrostrukturele plan word konsekwent deur die vier redaksionele medewerkers gevolg. 'n Onderdeel van hierdie mikrostruktuur is die (sinkroniese) leksikale definisie, met die eventuele onderskeiding van die poliseme waarna in die etimologiese bespreking verwys word. Onder *mannetjie* word daar bv. drie poliseme onderskei, waarvan twee volgens die etimologiese bespreking pre-Afrikaans is en een 'n vernuwing in Afrikaans is. Hierdie sisteem werk oor die algemeen goed, maar soms is die sinkroniese apparaat veels te uitgebreid. Onder *galjoen* se vier poliseme is net een etimologies relevant: sou dit nie toegankliker wees om net hierdie enkele poliseem te vermeld en te bespreek nie? Dieselfde wanbalans tussen sinkronie en diakronie kenmerk lemmas soos *aanbied*, *basaar*, *garnaal*, *geilsiekte*, *hofschoen* en *jakkalsbessie*.

Die etimologiese bespreking van komplekse bevat sinkroniese inligting in die vorm van die verklaring van die betrokke eenhede. Onder *dikbekkie* word bv. vermeld: "Samestellende afleiding met *-ie* van *dik* en *bek*, so genoem omdat die vis *dik*, vlesige lippe het." In hierdie geval en ook wat betref bv. *gompou*, *kisklere*, *kleinhuisie* en *renostervoël* werk die prosedee goed, aangesien die verkla-

ring inligting bevat waarvan die leser waarskynlik nie bewus is nie, maar dan weer is die verklaring so voor die hand liggend dat die etimologiese bydrae daarvan minimaal is, vgl. bv. *bossiedokter*, *grasdak* en *huilboom*.

In die beskrywing van die leksikale ontwikkeling van Afrikaans moet weldeeglik van die Nederlandse en Germaanse agtergrond kennis geneem word. Oor die ouer Germaanse agtergrond bied EWA weinig inligting. Een moontlike rede is die veronderstelling dat hierdie soort gegewens in Nederlandse etimologiese woordeboeke voorhande is. Maar dit is 'n groot vraag of die Afrikaanse gebruiker nie baie ontnem word nie. Hoe toeganklik is die Nederlandse bronne immers hier?

Die historiese verband van Afrikaans met Nederlands vertoon verskillende fasette. In die eerste plek het Afrikaans uit gewestelike Nederlands (in die besonder die Hollandse dialek) ontwikkel; dus nie uit formele of Standaardnederlands nie. Die dialektiese herkoms van Afrikaans word mooi belig in die artikel oor die bywoord *nou* (bl. 101): "Uit gewestelike Ndl. *nou* (al Mnl.). Veral in N.Nederland, o.a. Holland, is *nou* die gewone vorm, terwyl *nu* die formele vorm is." In die tweede plek het formele Nederlands wel, sedert die negentiende eeu, invloed op Afrikaans uitgeoefen, wat gelei het tot 'n groot aantal redelik resente Nederlandse kultuurontleninge in Afrikaans.

Die formulering "uit Ndl.", wat by herhaling gebruik word, is verwarrend. Onder *brandsiekte* word bv. gestel: "Uit Ndl. *brandziekte* ...", waar *brandziekte* die huidige Standaardnederlandse vorm is. Afrikaans het nie uit Standaardnederlands ontwikkel nie, en daar was nooit 'n vorm met *z-*, soos *ziekte*, in Afrikaans se verlede nie. Die Nederlandse variëteit waaruit Afrikaans ontwikkel het, het nes hedendaagse Afrikaans 'n *s-* gehad, dus *siekte* en *brandsiekte*. Vir sover *brandziekte* in ouer Afrikaanse dokumente gebruik word, is dit 'n skryfvorm in ooreenstemming met die norme van die formele Nederlandse skryftaal. Die verwarring kan voorkom word deur die uitdrukking "uit Ndl." te vermy, dus deur bv. te sê: "Ndl. *brandziekte* (1869) is tans 'n siekte onder varke."

Van *dankie* word gesê dat dit 'n sametrekking is van Ndl. (*ik*) *dank jy/you*. In der waarheid stam Afrikaans *dankie* uit die gewestelike Hollandse vorm *dankie*, waarin *-ie* 'n vorm is van die voornaamwoord vir die tweede persoon wat bv. in Standaardnederlands *je* is. Hoe dit ook al sy, *u* kom hier glad nie in die spel nie.

Ook in gevalle soos die volgende is die gelykstelling tussen die Afrikaanse en die Nederlandse vorme misleidend: *eiergeel* – *eigeel*; *erf* – *erven*; *flerrie* – *flerie*; *flou* – *flauw*; *goël* – *goochelen*; *hakskeen* – *hakzeen*; *pag* – *pacht*; *peul* – *puilen*; *senuwee* – *zenuw*. Vir die grootste deel kan die woord *uit* geskrap word en uitdrukkinge soos "uit Ndl. *eigeel*, uit Ndl. *flerie*" vervang word deur die blote vermelding: "Ndl. *eigeel*, Ndl. *flerie*". Soms is meer ingrypende herformulering nodig. Ek volstaan met 'n paar voorbeelde.

genugtig. Uit *genugte* (Ndl. 1635), 'n wisselvorm van *geneugte* 'genieting, genoeg', afgelei van *genoeg*. In Afr. is die uitgang *-ig* aangeheg.

naastenby. *Naaste* is die oortreffende trap van *na* soos in *naby*. *Naastenby* het ontwikkel uit 'n verbinding met die voorsetsel *te(n)*, soos in Ndl. *te(n)naaste(n)bij* (1645).

perdeby. 'n Gewestelike vorm in Ndl., vgl. bv. *peerdsbie* (1596). Die Standaardnederlandse vorm is *horzel*.

skinkbord. Standaardnederlands *dienblad*. Afr. het die samestelling uit gewestelike Ndl.; vgl. S.Ndl. *schenkbord* (1714).

Ook van later Afrikaanse kultuurontleninge aan Standaardnederlands word gesê dat hulle "uit Ndl." kom, bv. *broodwinner*, *eikeboom*, *fanfare*, *fiets*, *gru*, *grusaam* en *ontgogel*. Maar hulle staan nie op een lyn met die Nederlandse basisvorme wat in die sewentiende eeu reeds in Afrikaans gevestig is nie. Hierdie kultuurontleninge behoort as sodanig aangedui te word.

Te min word in EWA oor die fonologiese ontwikkeling van leksikale eenhede gesê. Die lemmas *aardvark*, *aars*, *erd* en *ertappel* gee nie genoegsaam blyke van die *ê/aa*-wisseling in die geskiedenis van Afrikaans nie. *Aardvark* behoort net as kruisverwysing opgeneem te word, as die geskrewe teenhanger van *erdvark*. By *aars* ontbreek die kruisverwysing na *ners-* en *blikners*. Die *-sie* van *bagasie* en *gasie* kom uit pre-Afrikaans *bagagie*, *gagie* (wat uiteindelik uit *bagage*, *gage* ontwikkel het). Afrikaans *-wee* in *skaduwee* (nie opgeneem nie), *senuwee* en *weduwee* kom uit 'n ou verboë vorm van hierdie naamwoorde. Met die wisseling tussen *i* en *e* moet rekening gehou word o.a. in *grenadella*, *fletters* en *skinkbord*; met metatesis in bv. *ferweel* en *kaperjol*. *Ekspres* se variant *aspris* (met die verlagings van die sjwa in die eerste sillabe) word nie behandel nie.

Leksikale eenhede kleiner as die woord behoort opgeneem te word, soos *-a* (waarna in *boeta* verwys word), *-erig* (vgl. *nukkerig*) en *dees-/dus-* (*deeskant*, *duskant*).

Die historiese verband (met Nederlands) sal duideliker wees indien gevalle soos die volgende eksplisiet as Afrikaanse vernuwings gekenmerk word: *agteros*, *agterryer*, *binnemuurs* (na aanleiding van Engels *intramural*), *bevraagteken* (na aanleiding van Engels *query*), *bloedpens*, *dassie*, *domastrant*, *elsbekkie*, *elsie*, *foeitog*, *geilsiek(te)*, *hardekool*, *hardepad*, *harpuisbos*, *hierjy*, *lawaaewater*, *nieshout*, *nooiensboom*, *padkos*, *pan* 'komvormige holte in die grond', *partykeer*, *piet-my-vrou*, *plakkie*, *pofadder*, *sekretarisvoël*, *skaapsteker*, *skurwejantjie*.

Die teks is skoon en leesbaar. Die verbinding van die werkwoord *ontleen* met die voorsetsel *uit* (bl. 131, 161) is hinderlik. Daar is weliswaar 'n tendens in die spreektaal om *aan* hier deur *uit* te vervang, maar in 'n teks soos hierdie, met so 'n hoë frekwensie van die betrokke uitdrukking, is *aan* verkieslik. Die uitdrukking "... het in Afr. self ontwikkel" word deurlopend gebruik (vgl. bv. *aster*, *fletter*, *gesout* en *hamerkop*). Verkieslik is: "... is 'n eie ontwikkeling in Afr./'n vernuwing in Afr./'n Afr. vernuwing." "Weeshoofde" behoort uitgeskakel te word, o.a. *1abba*, *allemagtig*, *bewe*, *Dinsdag*, *eendag*, *erfporsie*, *gala* en *gerieflik*.

Hier volg spesifieke kommentaar op 'n aantal artikels.

- afjak*. Die oorgang van *g* tot *k* is problematies.
- agtelosig*. Vermeld die (volksetimologiese) variant *agterlosig*. *Agte-* kom uit *agt* (Ndl. *acht*) 'aandag'.
- akkedis*. Dit is goed moontlik dat *hage* in Ndl. *hagedis* volksetimologies is. Die Afr. vorm kom in elk geval nie uit *hagedis* nie.
- apiesdoring*. Wydlopige bespreking. Die onderskeid tussen nagapies en blou-apies is nie ter sake nie.
- arrie*. 'n Probleemgeval, wat sonder die omslagtige bespreking as sodanig vermeld of weggelaat moet word.
- blikners*. Oorspronklik 'n verbinding van 'n sg. stofadjektief *blikken* en die naamwoord *ers*; verwys na *aars*.
- botterbroodjie*. Die artikel kan ingrypend verkort word deur 'n kruisverwysing na *skon* in te bou.
- dierbaar*. Die verwysing na Vlaams is ondeursigtig sonder vermelding van die (Vlaamse) ontronding van *duur* tot *dier*.
- Dinsdag*. Die variant *Dingsdag* werp nuwe lig op die etimologie van hierdie woord. Die pedantiese herhaling by elke dag van die week: "Nde (bv. vyfde, sesde) dag van die week" met die verwysing na die hernommring van dae van die week kan vervang word deur: "dag van die week", punt.
- ditsem*. Nie aan Engels ontleen nie; uit ouer Afrikaans *dit is em*, wat die *em* 'hom' uit gewestelike Nederlands het.
- domkrag*. Die verband met *duim* word nie vermeld nie.
- flenters*. Die adjektief het ontwikkel uit 'n setselkonstruksie *aan flenters* en loop parallel met *stukkend* (nie opgeneem nie) uit *aan stukkend* uit *aan stukken*.
- fles*. Verwys na die konkurrensie met *bottel*.
- geilsiekte*. Reeds in Trichardt.
- ghoen*. Die herkomsforme is te uiteenlopend om daaruit 'n sinvolle etimologie af te lei.
- grein*. Die poliseem 'draad van hout' is waarskynlik 'n Engelse ontlening.
- haasbek*. Die bespreking is onduidelik.
- kapok*. Maleise ontlening.
- kardoes*. Die uitvoerige bespreking in Scholtz 1972: 130-1 moet ingewerk word.
- karkoer*. Die etima is wyd uiteenlopend en die etimologie dus spekulatief.
- klip*. Belangrike gegewens in Scholtz 1972: 134.
- kokkedoor*. Verwysing na Nederlandse vorme is nie ter sake nie.
- naboom*. Rig die bespreking meer pertinent op die Koi-vorm *!na*.
- naweek*. Ingegee deur Engels *weekend*.

nè. Kom in die Rynse dialekte van Duits voor.

pamperlang. 'n Probleemgeval wat sonder werklike uitsluitel bespreek word.

partytjie. Ingegee deur Engels *party*. Standaardnederlands het *feest(je)*.

ru. Die Nederlandse vorm het ontwikkel uit 'n vorm met *uu*. Die *ui* het eers later ontstaan.

skemerkelkie. Ingegee deur Engels *cocktails*.

skielik. Die bespreking is omslagtig. Die Afr. variant *skierlik* laat blyk dat dit hier gaan om die verbinding van *skier* 'vinnig' en die agtervoegsel *-lik*.

Ontleningsgeskiedenis is 'n belangrike aspek van die etimologie wat in EWA breed uitgemeet word, met betrekking tot onder andere Maleis (*amok*, *baadjie*, *baie* ...), Koi (*arrie*, *dagga*, *nadroejakkals* ('n voortreflike artikel) ...), Duits (*gaip*), Nguni (*dagha*, *Fanagalo*), en Sotho (*marog*). Lesers behoort die groot aantal leenvertalings aan Engels, waaronder *bandopnemer* en *dubbeldekker*, interessant te vind. Wat Koi betref, moet ons roei met die rieme wat ons het, om van San te swyg — vgl. die kleitrappery met die etimologie van *karkoer* en *ghantang*.

EWA maak 'n boeiende stuk Afrikaanse kultuur toeganklik. Ek vertrou dat hierdie interessante woordeboek die belangstellende mark vind wat hy verdien.

Bronne

- Boshoff, S.P.E. en G.S. Nienaber.** 1967. *Afrikaanse etimologieë*. Pretoria: Suid-Afrikaanse Akademie vir Wetenskap en Kuns.
- Combrink, Johan.** 1969. *Waar kom daardie woord vandaan?* Johannesburg: SAUK.
- Franken, J.L.M.** 1953. *Tentatiewe etimologieë*. Annale van die Universiteit van Stellenbosch 27B(1).
- Scholtz, J. du P.** 1972. *Afrikaans-Hollands in die agtiende eeu*. Kaapstad: Nasou.
- Silva, P. (hoofred.).** 1996. *A Dictionary of South African English on Historical Principles*. Oxford: Oxford University Press.
- Smith, J.J.** 1962. *Op ons taalakker*. Pretoria: Van Schaik.
- Van der Sijs, N.** 2001. *Chronologisch woordenboek. De ouderdom en herkomst van onze woorden en betekenissen*. Amsterdam/Antwerpen: L.J. Veen.

Fritz Ponelis
Departement Afrikaans en Nederlands
Universiteit Stellenbosch
Stellenbosch
Republiek van Suid-Afrika
(fap@sun.ac.za)

Publikasieaankondigings / Publication Announcements

- Monika Elisabeth Bründl. *Lexikalische Dynamik: Kognitiv-linguistische Untersuchungen am englischen Computerwortschatz*. 2000, XII + 293 pp. ISBN 3 484 30443 X. Tübingen: Max Niemeyer. Price: €72.00.
- Igor Burkhanov. *Linguistic Foundations of Ideography: Semantic Analysis and Ideographic Dictionaries*. 1999, 388pp. ISBN 83 87288 98 5. Rzeszów: Wydawnictwo Wyższej Szkoły Pedagogicznej. URL: <http://www.univ.rzeszow.pl>. E-mail: wydaw@univ.rzeszow.pl. Price: 16 PLN. (Review in this issue.)
- Julie Coleman and Christian J. Kay (Eds.). *Lexicology, Semantics and Lexicography: Selected Papers from the Fourth G.L. Brook Symposium, Manchester, August 1998*. 2000, xiv + 249 pp. Current Issues in Linguistic Theory Volume 194. ISBN 90 272 3701 8 (Eur.), 1 55619 972 4 (US). Amsterdam/Philadelphia: John Benjamins. Price: US\$ 75,00 (Hb.). (Review in this issue.)
- Mukumbuta Lisimba. *Kongo Proverbs and the Origins of Bantu Wisdom*. 1999, 251 pp. Gabon: CICIBA (International Centre for Bantu Civilisations). (Review article in this issue.)
- Ulrike Rothe. *Das einsprachige Wörterbuch in seinem soziokulturellen Kontext: Gesellschaftliche und sprachwissenschaftliche Aspekte in der Lexikographie des Englischen und des Französischen*. 2001, vii + 282 pp. ISBN 3 484 39108 1. Tübingen: Max Niemeyer. Price: €68.00.
- G.J. van Wyk (Tegn. Red.), A.E. Cloete, A. Jordaan, H.C. Liebenberg en H.J. Lubbe. *Etimologiewoordeboek van Afrikaans (eerste fase)*. 2002, 271 pp. Stellenbosch: WAT-Trust. (Resensie in hierdie uitgawe.)
- 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. (Review article in this issue.)

VOORSKRIFTE AAN SKRYWERS

(Tree asseblief met die Buro van die WAT in verbinding (wat@wat.sun.ac.za) vir 'n uitvoeriger weergawe van hierdie instruksies of besoek ons webblad: <http://www.sun.ac.za/wat>)

A. REDAKSIONELE BELEID

1. Aard en inhoud van artikels

Artikels kan handel oor die suiwer leksikografie of oor implikasies wat aanverwante terreine, bv. linguistiek, 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 evalueerend 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 feitlike 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 evalueerende 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-tekst, as volledig geformateerde Microsoft Word (DOS of Windows) lêers of as WordPerfect (DOS of Windows) lêers op rekenaarskyf (360 KB tot 1.44 MB) voorgelê word. 'n Rekenaardrukstuk van die artikel moet die skyf vergesel. Elke artikel moet voorsien wees van 'n Engelse opsomming van tussen 150 en 250 woorde, sowel as tussen 10 en 30 Engelse sleutelwoorde.

2. Grafika

Een stel duidelike oorspronklike illustrasies, tabelle, grafieke, diagramme, of kwaliteitsafdrukke daarvan, moet voorgelê word. Die plasing van grafika binne die teks moet duidelik aangedui word.

3. Bibliografiese gegewens en verwysings binne die teks

Kyk na onlangse nommers van *Lexikos* vir meer inligting.

INSTRUCTIONS TO AUTHORS

(For a more detailed version of these instructions, please contact the Bureau of the WAT
(wat@wat.sun.ac.za) or refer to our website: <http://www.sun.ac.za/wat>)

A. EDITORIAL POLICY

1. Type and content of articles

Articles may deal with pure lexicography or with the implications that related fields such as linguistics, general linguistics, computer science and management have for lexicography.

Contributions may be classified in any one of the following categories:

- (1) **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 Woordeboek van die Afrikaanse Taal. Authors are free however to use their material elsewhere provided that *Lexikos* (AFRILEX Series) is acknowledged as the original publication source.

5. Originality

Only original contributions will be considered for publication. Authors bear full responsibility for the originality and factual content of their contributions.

6. Free offprints and copies

Authors will receive five free offprints of each of their 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 (wat@wat.sun.ac.za) für eine ausführlichere Wiedergabe dieser Hinweise oder besuchen Sie unsere Webseite: <http://www.sun.ac.za/wat>)

A. REDAKTIONELLE ZIELSETZUNGEN

1. Art und Inhalt der Artikel

Es können Artikel aufgenommen werden, die sich mit Themen der Lexikographie befassen oder mit Zusammenhängen, die zwischen der Lexikographie und benachbarten Fachgebieten wie z.B. Linguistik, allgemeiner Sprachwissenschaft, Lexikologie, Computerwissenschaft und Management bestehen.

Die Beiträge sollten einer der folgenden Kategorien entsprechen:

- (1) **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.

INSTRUCTIONS AUX AUTEURS

(Pour une version plus détaillée de ces instructions, contacter le Bureau du WAT (wat@sun.ac.za)
ou consulter notre website: <http://www.sun.ac.za/wat>)

A. POLITIQUE ÉDITORIALE

1. Caractéristiques et contenu des articles

Les articles seront consacrés à la lexicographie pure, ou aux rapports entre la lexicographie et les disciplines voisines telles que la linguistique, la linguistique générale, l'informatique et le management.

Les contributions pourront appartenir à l'une des catégories suivantes:

- (1) **Articles de recherche:** Recherches scientifiques originales, avec leurs résultats;
- (2) **Articles de réflexion:** Présentations originales, synoptiques, interprétatives, comparatives, évaluatives et critiques des résultats de recherches en cours;
- (3) **Articles bilans:** Articles de recherche présentés sous forme de bilan critique de travaux scientifiques déjà publiés.

Les contributions appartenant aux catégories (1) à (3) seront soumises de manière anonyme à des experts spécialistes indépendants afin d'en assurer la qualité scientifique au niveau international.

- (4) **Recensions:** Analyses et évaluations critiques de travaux de recherche et de productions scientifiques, telles que livres ou logiciels;
- (5) **Projets:** Présentations de projets lexicographiques;
- (6) **'Lexiconotes':** Textes contenant des informations pratiques, ou des suggestions, des problèmes, des questions, des commentaires et des solutions concernant des activités lexicographiques;
- (7) **'Lexicovaria':** Articles, annonces, communiqués de presse émanant de centres de lexicographie et qui revêtent un intérêt particulier pour les lexicographes;
- (8) **Rapports:** Rapports sur des colloques et ateliers.

Les contributions dans les catégories (4) à (8) devront répondre aux exigences de qualité des publications scientifiques et seront évaluées dans cette optique.

2. Critères et procédures d'évaluation

La revue *Lexikos* est reconnue et subventionnée par le Ministère de l'Éducation du gouvernement Sud-Africain comme revue devant générer des revenus.

Les articles seront évalués selon les critères suivants: langue et style, concision et clarté, formulation de la problématique, raisonnement et conclusion, référence aux travaux les plus importants et les plus récents, contribution substantielle à la discipline.

3. Langue des contributions

Afrikaans, allemand, anglais, français ou néerlandais.

4. Copyright

Le Bureau du WAT ou l'*African Association for Lexicography* (AFRILEX) décline toute responsabilité en cas de réclamations motivées par l'utilisation d'autres sources par les auteurs.

Les droits d'auteurs des documents publiés dans *Lexikos* appartiennent au Board of Control du *Woordeboek van die Afrikaanse Taal* (WAT). Cependant, les auteurs sont libres d'utiliser leurs textes dans d'autres publications, à condition d'indiquer *Lexikos* (collection AFRILEX) comme source de la publication originale.

5. Originalité

Seules les contributions originales seront acceptées pour la publication. Les auteurs conservent l'entière responsabilité de l'originalité et du contenu factuel de leur texte.

6. Tirés-à-part et exemplaires gratuits

Les auteurs recevront gratuitement cinq (5) tirés-à-part de chaque article de recherche, de réflexion ou de bilan, ainsi qu'un exemplaire gratuit de la publication contenant cet article.

Les auteurs des recensions et des publications dans les catégories 'Projets', 'Lexiconotes', 'Lexicovaria' et Rapports recevront cinq (5) tirés-à-part de leur texte. Dans ces quatre dernières catégories, le responsable de la revue pourra néanmoins décider, en fonction de la nature et des dimensions des publications, d'accorder un exemplaire gratuit de la revue à leurs auteurs.

7. Adresse de la revue

Les auteurs intéressés sont invités à soumettre leurs propositions à:

L'éditeur: LEXIKOS
Bureau du WAT
Boîte postale 245
7599 STELLENBOSCH
République d'Afrique du Sud

B. PRÉSENTATION DES MANUSCRITS

Les manuscrits se conformeront aux exigences suivantes:

1. Longueur et format

Les contributions ne devraient pas excéder 20 pages dactylographiées, de format A4, avec double espacement et marges suffisantes (environ 2,5 cm), si possible en format électronique (disquette 360KB à 1,44MB), sous forme de fichier ASCII, complètement formaté sous Microsoft Word ou sous WordPerfect (DOS ou Windows). La disquette sera accompagnée d'un tirage papier. Chaque article sera pourvu d'un résumé en anglais de 150 à 200 mots, et de 10 à 30 mots-clés.

2. Tableaux et graphiques

Les dessins, tableaux, graphiques et diagrammes seront envoyés, soit sous leur forme originale soit sous forme d'une copie de bonne qualité. Leur place dans le texte devra être clairement indiquée.

3. Bibliographie et références dans le texte

Voir les exemplaires récents de *Lexikos*.