Lexikos 13

Lexikos 13

Redakteur

Editor

J.C.M.D. du Plessis

Resensieredakteur

Review Editor

T. Harteveld

Adviesraad / Advisory Board

Redaksiekomitee / Editorial Committee

Inhoud / Contents

Projekte / Projects

Lexikonotas / Lexiconotes

Voorwoord

Met sowel AFRILEX as die NLE's goed gevestig in Suid-Afrika en vanweë groter kontak met die res van Afrika, ontvang *Lexikos* 'n konstante stroom artikels wat oor leksikografiese en metaleksikografiese aspekte en probleme handel. In elke uitgawe verskyn daar artikels nie net van vorige medewerkers nie, maar ook van nuwes. Ek wil hulle almal bedank vir hul ondersteuning omdat *Lexikos* nie sonder artikels kan bestaan en groei nie. My waardering gaan aan hulle vir hul bereidheid om artikels te herbewerk volgens die opmerkings en voorstelle van die keurders. Vanweë die baie artikels wat *Lexikos* ontvang, moet baie medewerkers ook 'n aansienlike tyd wag voordat hul artikels gepubliseer kan word. Hulle moet bedank word vir hul begrip.

Aan die keurders moet 'n woord van dank gaan vir hulle bereidheid om hul kennis en kundigheid te deel by die beoordeling van die artikels, dikwels op kort kennisgewing. Hulle sluit, behalwe lede van die Redaksiekomitee, ook verskeie ander in wat bereid was om artikels te keur wat binne hul gebied van kennis en kundigheid val.

Daar was 'n aantal veranderinge in die Redaksiekomitee. Ek wil dr. M.M.R. Dube bedank vir haar bydrae tot *Lexikos*. Dit is aangenaam om proff. I.M. Kosch en L.J. Louwrens op die Redaksiekomitee te verwelkom. Hulle het reeds artikels vir *Lexikos* beoordeel, en ek hoop hul meer permanente verbintenis met *Lexikos* sal net so aangenaam in die toekoms wees soos dit in die verlede was. Ek is ook bly om prof. B.M. Mini na 'n afwesigheid van drie jaar, terug te verwelkom op die Redaksiekomitee.

Me. Tanja Harteveld moet bedank word vir haar hulp by die oplossing van rekenaar- en ander tegniese probleme wat die setwerk betref het. Me. Riette Ruthven was verantwoordelik nie net vir die professionele en bekwame setwerk van die dikwels moeilike artikels nie, maar ook vir die hantering van die uitgebreide korrespondensie. Ek wil haar bedank vir haar geesdrif en bedrewenheid by die finalisering van hierdie nommer van *Lexikos*.

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

Foreword

With both AFRILEX and the NLUs well established in South Africa and because of greater contact with the rest of Africa, *Lexikos* receives a constant stream of articles dealing with lexicographical and metalexicographical aspects and problems. In every issue there appear articles not only from previous contributors, but also from new ones. I want to thank them all for their support, because without articles *Lexikos* cannot exist and extend. My appreciation also goes to them for their willingness to rework articles according to comments and suggestions from the referees. Because of the many articles *Lexikos* receives, many contributors must also wait a considerable time for their articles to be published. They must be thanked for their understanding.

To the referees must go a word of thanks for their willingness to lend their knowledge and expertise in judging the articles, often on short notice. They include, besides members of the Editorial Board, also several others who were willing to judge articles falling within their sphere of knowledge and expertise.

There have been some changes to the Editorial Board. I want to thank Dr M.M.R. Dube for her contribution to *Lexikos*. It is a pleasure to welcome Proff. I.M. Kosch and L.J. Louwrens on the Editorial Board. They have already refereed articles for *Lexikos*, and I hope their more permanent association with *Lexikos* will be as pleasant in future as it has been in the past. I am also happy to welcome Prof. B.M. Mini back on the Editorial Board after an absence of three years.

Ms Tanja Harteveld must be thanked for her support in solving computational and other technical problems as far as the typesetting was concerned. Ms Riette Ruthven was responsible not only for the professional and accomplished typesetting of the often difficult articles, but also for handling the extensive correspondence. I want to thank her for her enthusiasm and efficiency in finalizing this issue of *Lexikos*.

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

'n Woord van AFRILEX

Een van die doelstellings van AFRILEX is om leksikografie op 'n breë front in Suidelike Afrika te bevorder. AFRILEX stel hom ten doel om leksikograwe, terminoloë en ander taalpraktisyns op te lei in verskeie aspekte rakende leksikografiese beginsels en praktyk. Dit streef ook daarna om jaarliks internasionale konferensies en plaaslike seminare te organiseer oor onderwerpe wat op 'n bepaalde tydstip tersaaklik is. Die Agste Internasionale AFRILEX-konferensie wat van 7 tot 9 Julie 2003 in Windhoek, Namibië gehou is, het baie aktuele onderwerpe behandel en was 'n groot sukses. As mondstuk van AFRILEX bevorder *Lexikos* die leksikografie in sy breedste betekenis.

As lede van AFRILEX is ons almal by die samestelling van woordeboeke betrokke, alhoewel in verskillende soorte woordeboekwerk of -projekte, in adviserende hoedanighede, of bloot weens die liefde vir of verslaafdheid aan leksikografie:

— sommige is lede van die leksikografiespanne van die elf Nasionale Leksikografie-eenhede (NLE's);
— sommige is besig met die samestelling van terminologielyste of vakwoordeboeke vir die vaktale;
— sommige is lede van adviesliggame (bv. lede van die Pan-Suid-Afrikaanse Taalraad (PanSAT), lede van PanSAT-subkomitees vir verskeie fokusgebiede, lede van die Beheerrade van die NLE's, lede van die Nasionale Taalliggame (NTL'e), en lede van die Provinsiale Taalkomitees (PTK's));
— sommige is lede van taalkantore, bv. die Nasionale Taaldiens (NTD), die Terminologiekoördineringsafdeling (TKA), ens.;
— sommige is dosente verbonde aan tersiêre instellings wat kursusse in metaleksikografie en terminologie aanbied;
— sommige is studente in leksikografie en terminologie; en
— sommige is taalpraktisyns soos vertalers, redigeerders, tolke, taalonderwysers en joernaliste — dié mense wat nie hul spesifieke werk sonder die hulp van woordeboeke kan doen nie.

Die punt wat ek graag wil stel, is dat nie een van ons onpartydig teenoor leksikografie staan nie en ons is almal geïnteresseerd in die artikels wat in hierdie dertiende uitgawe van *Lexikos* opgeneem is om meer te leer van leksikografie, woordeboeke en hoe om woordeboeke saam te stel en ten beste te benut.

Die inhoud van hierdie uitgawe van die vaktydskrif is baie aktueel, veral vir ons NLE's. Ons weet almal dat 'n leksikograaf en terminograaf in die spesifieke behoeftes van die woordeboekgebruiker moet voorsien. Woordeboeksamestellers moet weet wie hul teikengebruikers is en moet hulle op hoogte bring van die gebruikers se behoeftes. Dit moet beklemtoon word dat woordeboeke

gebruikersvriendelik moet wees en dat gebruikers maklike toegang tot en verkryging van woordeboekinligting moet hê. Dit is slegs enkele van die brandende vraagstukke wat onder andere deur die NLE's ervaar word. U sal sekerlik artikels in *Lexikos* vind wat hierdie sake aanspreek, want 'n blik op die inhoudsopgawe toon groot belofte: bekwame navorsers behandel verskeie baie tersaaklike onderwerpe.

Ter afsluiting wil ek graag beklemtoon dat ons 'n verantwoordelikheid het teenoor die verskeie spraakgemeenskappe wat ons dien. 'n Woordeboekkultuur behoort ontwikkel te word en spraakgemeenskappe behoort bewus te wees van AFRILEX, en van PanSAT en sy strukture soos die elf Nasionale Leksikografie-eenhede (NLE's), die dertien Nasionale Taalliggame (NTL'e), die nege Provinsiale Taalkomitees (PTK's), en van die Terminologiekoördineringsafdeling (TKA) van die Nasionale Taaldiens (NTD), Departement van Kuns en Kultuur (DKK), asook van hul verskillende doelstellings. Die NLE's behoort ook by skole en tersiêre instansies bekendgestel te word. Woordeboekgebruiksvermoëns kan ontwikkel en aangeleer word as opvoeders soos onderwysers en dosente en diegene wat opgelei word, soos leerders en studente, bewus is van die werk wat by die NLE's gedoen word. Indien al die taalstrukture aan die algemene publiek bekend is en al die werk wat deur hierdie liggame gedoen word op die beoogde virtuele netwerk vir Mensliketaaltegnologie beskikbaar kan wees, sal hierdie streek in 'n posisie wees om ons tale te bewaar en hulle as funksionele tale in alle lewensfere te ontwikkel. Eers dan sal die verskillende taalgemeenskappe in staat wees om hul kommunikasievermoëns te ontwikkel.

Ek wil graag van die geleentheid gebruik maak om die Buro van die *Woordeboek van die Afrikaanse Taal* (WAT) te bedank vir die publisering van *Lexikos* en 'n spesiale woord van dank aan dr. J.C.M.D. du Plessis vir sy uitstekende werk as redakteur van die vaktydskrif.

Dr. Mariëtta Alberts
President: AFRILEX
PanSAT
Privaat sak X08
ARCADIA 0007
Tel: +27 (0)12 341 9638   Faks: +27 (0)12 341 5938
E-pos: marietta@pansalb.org.za
AFRILEX-tuisblad:
<http://www.up.ac.za/academic/libarts/afrilang/homelex.html>

A Few Words from AFRILEX

One of the aims of AFRILEX is to promote lexicography on a wide front in Southern Africa. AFRILEX aims to train lexicographers, terminologists and other language practitioners in various aspects relating to lexicographical principles and practice. Its objective also is to organise annual international conferences and local seminars on topics relevant at a specific time. The Eighth International AFRILEX Conference held in Windhoek, Namibia, from 7 to 9 July 2003, as with previous AFRILEX conferences and seminars, dealt with very topical issues and proved to be a great success. *Lexikos*, as mouthpiece of AFRILEX, also serves to promote lexicography in its broadest sense.

As members of AFRILEX we are all involved in dictionary compilation although in different kinds of dictionary work or projects, in advisory capacities, or for the mere love of or addiction to lexicography:

— some are members of the lexicography teams for the eleven National Lexicography Units (NLUs);
— some are busy compiling terminology lists or technical dictionaries for Languages for Special Purposes (LSP);
— some are members of advisory bodies (e.g. members of the Pan South African Language Board (PanSALB), members of PanSALB Subcommittees for various focus areas, members of the Boards of Directors of the NLUs, members of National Language Bodies (NLBs), and members of Provincial Language Committees (PLCs)) ;
— some are members of language offices, e.g. the National Language Service (NLS), the Terminology Coordination Section (TCS), etc.;
— some are lecturers at tertiary institutions who teach courses in metalexicography and terminology;
— some are students of lexicography and terminology; and
— some are language practitioners such as translators, editors, interpreters, language teachers and journalists — those people who cannot do their specific jobs without the assistance of dictionaries.

The point I would like to make, is that none of us are impartial towards lexicography and we are all interested in the articles included in this thirteenth issue of *Lexikos* to learn more about lexicography, dictionaries and how to compile dictionaries and use them to our best advantage.

The contents of this issue of the journal are very relevant, especially to our NLUs. We all know that a lexicographer or terminographer have to address the specific needs of the dictionary user. Dictionary compilers should know who their target users are and should familiarise themselves with the users' needs. It should be stressed that dictionaries should be user-friendly and that users

should be able easily to access and retrieve dictionary information. These are only a few of the burning issues experienced by *inter alia* the NLUs. You will certainly find articles in *Lexikos* that will address these issues because a glance at the table of contents shows great promise: competent researchers address various very relevant topics.

In conclusion I would like to stress that we have a responsibility towards the various speech communities we serve. A dictionary culture needs to be developed and speech communities should be aware of AFRILEX, of PanSALB and its structures such as the eleven National Lexicography Units (NLUs), the thirteen National Language Bodies (NLBs), the nine Provincial Language Committees (PLCs), and of the Terminology Coordination Section (TCS) of the National Language Service (NLS), Department of Arts and Culture (DAC), as well as their different objectives. The NLUs should also be promoted in schools and tertiary institutions. Dictionary-using skills can be developed and acquired if educators such as teachers and lecturers and trainees such as learners and students are aware of the work being done at the NLUs. If all the language structures are known to the general public and all the work of these bodies can be available on the envisaged Human Language Technology virtual network, this region would be in a position to preserve our languages and develop them as functional languages in all spheres of life. Only then will the various language communities be in a position to develop their communication skills.

I would like to take this opportunity to thank the Bureau of the *Woordeboek van die Afrikaanse Taal* (WAT) for the publication of *Lexikos* and a special word of thanks to Dr J.C.M.D. du Plessis for his excellent work as editor of the journal.

Dr Mariëtta Alberts
President: AFRILEX
PanSALB
Private Bag X08
ARCADIA 0007
Tel: +27 (0)12 341 9638   Fax: +27 (0)12 341 5938
E-mail: marietta@pansalb.org.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 AFRI-LEX-reeks uitgegee. "AFRILEX" is 'n akroniem vir "leksikografie in en vir Afri-ka". 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 leksiko-grafiese 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 be-werkstellig en te bevorder;

(4) om die interdissiplinêre aard van die leksikografie, wat ook terreine soos die taalkunde, algemene taalwetenskap, leksikologie, rekenaarweten-skap, 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* (AFRI-LEX) 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 han-del, sal vir opname in die AFRILEX-reeks kwalifiseer.

(4) Die wetenskaplike standaard van die bydraes sal gewaarborg word deur hulle aan 'n komitee van vakspesialiste van hoë akademiese aansien voor te lê vir anonieme keuring.

Lexikos sal jaarliks verskyn, terwyl verdienstelike monografiese studies spora-dies 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 AFRI-LEX-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 Lexiko-graphen und Linguisten;

(3)   Kontakt herstellen und fördern zwischen südafrikanischen und aus-ländischen lexikographischen Projekten;

(4)   die Aufmerksamkeit lenken auf die interdisziplinäre wissenschaftliche Praxis der Lexikographie, die Beziehung aufweist zur Linguistik, allge-meinen 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 publiées sous forme de publications séparées dans la collection AFRILEX.

Online Dictionaries on the Internet: An Overview for the African Languages*

Gilles-Maurice de Schryver, *Department of African Languages and Cultures, Ghent University, Ghent, Belgium and Department of African Languages, University of Pretoria, Pretoria, Republic of South Africa (gillesmaurice.deschryver@UGent.be)*

Abstract: The main purpose of this research article is rather bold, in that an attempt is made at a comprehensive overview of all currently available African-language Internet dictionaries. Quite surprisingly, a substantial number of such dictionaries is already available, for a large number of languages, with a relatively large number of users. The key characteristics of these dictionaries and various cross-language distributions are expounded on. In a second section the first South African online dictionary interface is introduced. Although compiled by just a small number of scholars, this dictionary contains a world's first in that lexicographic customisation is implemented on various levels in real time on the Internet.

Keywords: LEXICOGRAPHY, TERMINOLOGY, DICTIONARIES, INTERNET, ONLINE, LOOK‑UP MODE, BROWSE MODE, AFRICAN LANGUAGES, SESOTHO SA LEBOA, SIMULTANEOUS FEEDBACK, FUZZY SF, CUSTOMISATION

Senaganwa: Dipukuntšu tša online tše di lego mo Inthaneteng: Ponokakaretšo ya maleme a Afrika. Morero wo mogolo wa taodišwana ye ya nyakišišo ke wo o tiilego ka ge teko e dirilwe ka tebelelo ya kakaretšo ye e tletšego go dipukuntšu ka moka tša Inthanete tšeo di šetšego di le gona mo malemeng a Afrika. Sa go makatša ke gore go šetše go na le palo ye ntši ya dipukuntšu tše bjalo mo malemeng a mantši gape di na le badiriši ba bantši. Go hlalošwa dipharologantšho tše bohlokwa tša dipukuntšu tše le ka moo diphatlalatšo di dirwago ka gona gare ga maleme a mantši a go fapana. Mo karolong ya bobedi go tsebišwa pukuntšu ya online ye e lego ya pele gape e lego ya makgonthe ya Afrika Borwa. Le ge e le gore pukuntšu ye e hlamilwe ke dirutegi di se kae, e šetše e tšea sefoka lefaseng ka bophara. Se ke ka lebaka la gore pukuntšu ye e dirilwe ka tsela yeo e lego gore dilo di ka beakanywa gore di itšweletše ka botšona gomme tša lokela batho ka moka bao ba e dirišago mo Inthaneteng ka yona nako yeo.

Mantšu a bohlokwa: TLHAMO YA DIPUKUNTŠU, TLHAMO YA MAREO, DIPUKUNTŠU, INTHANETE, ONLINE, MOKGWA WA GO NYAKA, MOKGWA WA GO LEKOLA, MALEME A AFRIKA, SESOTHO SA LEBOA, SIMULTANEOUS FEEDBACK, FUZZY SF, GO BEAKANYA DILO GORE DI BE KA MOKGWA WO O LEGO GORE O TLA GO LOKELA

African-language lexicography goes electronic

The face of dictionaries is changing. Rapidly. Whereas paper dictionaries still ruled the market until a decade ago, the large-scale commercial production of electronic dictionaries has boomed since the mid-1990s. Today, dictionaries on CD-ROM typically come in the back pocket of their hardcopy counterparts, while the number of dictionaries on the Internet already runs into tens of thousands. Lexicographers working on the South African languages should not merely watch the unrolling of these events from the sidelines; they *must* and *can* take an active part in writing the future. The wildest futuristic dreams revolve around multimedia Internet dictionaries, for which space restrictions disappear, and for which the output can be tailored to suit each unique user. The aim of this article is therefore twofold. On the one hand, the results are presented of a detailed study of current African-language Internet dictionaries. On the other, the first truly South African online dictionary interface is introduced and contrasted with what is already available on the Internet.

In South Africa, the *Woordeboek van die Afrikaanse Taal* (WAT), a multi-volume overall-descriptive historical dictionary with a paper past of three quarters of a century, recently went electronic. The entire text of the first eleven volumes was computerised and made available on CD-ROM in mid-2003. It is expected that subscription-based intranet and Internet versions will follow suit. The focus in this article will however be on the African languages, and more particularly on those languages that belong to the same language family as the official African languages of South Africa. Outside South Africa, these languages are known as '*Bantu* languages'. Given that this term is stigmatised in South Africa, this language family will be referred to as '*African* languages' below.

When embarking on this research project, colleagues wondered whether just two, a good twenty, or perhaps as many as two hundred African-language Internet dictionaries were already available. Probably the most surprising result of this study is that the actual number is closest to the last estimate. There are indeed nearly two hundred of them, for nearly one hundred and twenty *different* African languages. Considering that there are roughly five to six hundred languages that belong to this family, this means that one fifth are represented. The downside of this extremely positive and, admittedly, surprising outcome of this study is that the sizes of the current African-language Internet dictionaries are generally small, and the contents not often of a high quality.

Guthrie's classification and beyond

Over half a century ago, Malcolm Guthrie 'identified' and 'classified' the languages that are the focus of this article (Guthrie 1948). He used two main and two subsidiary linguistic criteria to identify the languages, and came to the conclusion that this language family covers that part of the African continent

starting from an imaginary line north of the current Democratic Republic of the Congo (DRC) all the way down to the southern tip of the African continent. Roughly speaking, only the languages spoken in the Cape region and north of it (Afrikaans and the Khoesaan languages) do not belong to this family. He then set out to classify all the languages within this region, a classification mainly based on geographical contiguities, and much less on linguistic features. The result consisted of 16 'zones' covering nearly 80 'groups'. The zones start in the northwest (A), go to the northeast (B, C, D and E), then south (F and G), again from west to east (H, K, L, M, N and P), and once more from west to east (R, S and T). These zones are made up of groups (A10, A20, ...; ...; M10, M20, ...; ...), with each group bringing together so-called related languages (A11, A12, ..., A21, A22, ...; ...; M11, M12, ..., M21, M22, ...; ...). Over the years he extensively revised zones A, B and C (Guthrie 1953), and also — all of a sudden, but apparently in response to criticism (Cope 1971: 218) — collapsed the Southern African zones S and T into a single zone S. Guthrie's 'final' classification can be found in the third volume, pages 11 to 15, of his Magnum Opus (Guthrie 1967, 1971, 1970, 1970a).

In Tervuren, Belgium, which soon became the mecca of Central-African language studies, a new zone was introduced around the region of the Great Lakes, zone J, consisting of Guthrie's groups E10, E20 and E30, as well as of sections of D40, D50 and D60. The numbering was simply transferred to J10 up to J60 respectively (Bastin 1978). In order to distinguish between neighbouring languages/dialects, extra letters are sometimes added (e.g. L31a for Cilubà spoken by the Balubà, L31b for Cilubà spoken by the Beena Luluwà, etc.). Since Guthrie, some languages have become extinct, while previously undocumented ones have been documented. Languages not originally in Guthrie's list mostly start with the linguistic group to which the extra language seems to be most affiliated, say E40, to which a third digit is added, e.g. E402 for Ikizu. At least, the latter is done by most scholars, such as for instance Lowe and Schadeberg (1996) or Maho (2003).

Nonetheless, in both the Guthrie and Tervuren checklists, the same code sometimes covers different languages. Furthermore, not everyone uses the Tervuren zone J. The result of this state of affairs is that there is considerable confusion as to which language has which code, and *vice versa*. Moreover, many languages often have numerous alternate spellings and/or are simply referred to by means of different names. The existence and status of dialects further complicate the issue. The exact number and location of languages is therefore still not known half a century after Guthrie's pioneering work, yet one generally accepts that there are at least five hundred and less than six hundred. Given all this confusion, it is obviously not truly possible to quantify any claims regarding this family of languages. For one, there is not even a fixed upper limit.

Apart from Guthrie's final classification, and Tervuren's latest checklist (Bastin, Coupez and Mann 1999), there is also a third classification that is often

consulted, viz. the one found in *Ethnologue* (Grimes and Grimes 2000). A highly useful comparison of the three classifications was compiled by Maho (2002). In the discussion below, however, certain decisions had to be made in order to provide for a scientific framework. These decisions were as follows: (1) *Ethnologue* was used as arbiter on language names, (2) the codes for the languages were mainly taken from the Tervuren checklist, (3) wherever Guthrie's data seemed more precise, his 'language name + language code' pair was kept, and (4) where applicable, the current official language names overruled the above.

Internet dictionaries for the South African languages

Now that the term 'African languages' has been delimited for the purposes of this article, one can turn to the concept 'Internet dictionaries'. Such reference works form part of the larger family of human-oriented electronic dictionaries and, within a three-step access dictionary typology, can be characterised as reference works for which 'users worldwide use laptops/desktops to access a dictionary stored on an online server' (De Schryver 2003: 151). Reformulated, these are thus *online* dictionaries for which the data are stored in databases, no matter where these databases are located, and which can be consulted from a *search screen* by anyone from anywhere through the *Internet*. Intranet dictionaries, another type of online electronic dictionary, will thus not be considered. For convenience, however, the terms 'online dictionary' and 'Internet dictionary' are used interchangeably in this article. A comprehensive overview of the features of the various electronic dictionaries, as well as a detailed discussion of their advantages over paper dictionaries, can be found in De Schryver (2003). Suffice it to say here that an electronic dictionary is much more than 'a dictionary in electronic form'. At the very least, the data are stored in a database, to which various (search) indexes are added, with a multitude of links to multimedia, as well as, increasingly, Natural Language Processing (NLP) extensions.

Rather surprisingly, these various aspects *already* exist for some of the African languages spoken in South Africa, albeit not yet all together in one integrated Internet dictionary package. An online dictionary for Tshivenda (S21), for example, is available from *CBOLD*. It contains 8 900 lemma signs, all of them searchable from a search screen, yet only with textual output. Sound files were added to various basic travellers' phrases for Sesotho (S33), among others, at *TravLang*, while full multimedia (i.e. text, audio and computer graphics) can be found at *eLanguage* for isiZulu (S42). Lastly, an example of an online NLP aspect that has been developed for a South African language is the machine translation (MT) system running between isiXhosa (S41) and English at *Xhosa on the Web!* (O'Kennon 1996–2003).

As argued by Varantola (2002: 35) and De Schryver (2003: 167, 169-172), multimedia corpora will increasingly become part and parcel of future electronic dictionaries. This NLP aspect does not yet exist for South African languages, but across the border Internet-searchable *text* corpora are already

available for ChiShona (S11-S12-S14) and SiNdebele (or Zimbabwean Ndebele (S44)). These online corpora of respectively 2.2 million and 0.7 million running words were originally assembled with dictionary compilation in mind, and have now been made available to the wider linguistic community (Ridings 2002).

Although most of the online dictionaries for South African languages have been online for quite some years now, it is somewhat disturbing to note that relatively few people know about their existence. Apart from the fact that the full Internet potential is not used within a single integrated package in any one of them, one of the reasons for their shadowy presence could be that none of these existing online dictionaries was made *in* South Africa, *by* South Africans, *for* South Africans. All these aspects are niches that can be filled by prospective lexicographers, besides the fact that such lexicographers can of course also improve on current size, quality and functionality.

A systematic overview of online African-language dictionaries

In this section, a systematic overview will be presented of currently available Internet dictionaries for the African languages. One immediately notices an uneasy balance between the concepts 'currently available' and 'Internet' here. Indeed, the Internet being an organic medium, its contents literally change every single second. One must therefore put a timestamp on the study, with all claims referring to that time frame. The timestamp is 'April 2003', as this is the period during which the Internet was trawled (with the help of search engines such as *Google*) to trace all available African-language Internet dictionaries. What follows is a summary and a discussion of the main findings, with all claims thus 'valid' for April 2003.

Before the results themselves are presented, it is important to recall that 'Internet dictionaries' in this study are *only* those online dictionaries that can be accessed from a search screen. This thus means that one must be able to type in words or sections of words, potentially including wildcards, followed by a mouse click or 'enter', upon which one or more articles are presented ensuing a page-reload. Based on this premise, the following two types of dictionaries that can be found *en masse* on the Internet have not been included in this study: (1) dictionaries in pdf (Portable Document Format), word processor, or any other downloadable text format — such as for instance Odden's (2002) *Kikerewe–English Dictionary* (J24) in pdf; and (2) dictionaries which are simply plain online HTML (HyperText Markup Language), or HTML-like, files — such as *Ikuska Libros*'s (1997–2003) *Diccionario Lingala–Español–Lingala* (C36d) in HTML, or dictionaries such as those from the *TravLang* series mentioned above, which have no search facilities and can only be 'browsed'.

The following three types, on the other hand, were considered for this study: (1) *online* dictionaries, i.e. dictionaries stored in databases over the Internet; (2) *pop-up* dictionaries, i.e. dictionaries with which, once one has down-

loaded a small piece of software, one can move the mouse over words online, upon which the relevant articles pop up in dedicated screens; and (3) *PC* dictionaries, i.e. dictionaries for which a piece of software *cum* one or more lexica are downloaded from the Internet, to be used as offline PC dictionaries. Note that the lexica in (2) can also be downloaded to the hard drive of a PC, at which point they become, in addition, functional as *offline* pop-up PC dictionaries.

Following the investigation, an impressive number of 182 African-language Internet dictionaries were found, 165 of the 'online' type, 8 of the 'pop-up' type, and 9 of the 'PC' type. All major characteristics of these 182 dictionaries have been tabulated in the Appendix, and as such this appendix — which is sorted by the *names* of the languages — should be considered the basis of the analysis that is to follow. These 182 dictionaries cover 117 *different* languages, as well as Common Bantu (CB) and Proto Bantu (PB). PB is the hypothetical language to which all current languages within this family can be traced back, while CB are the c. 2 800 series of comparative forms that were used by Guthrie to reconstruct PB. The distribution of the number of Internet dictionaries per language is as follows: Swahili (G42): 20 x, Chagga (E62): 14 x, Lingala: 5 x, Ganda (J15) and isiZulu: 4 x each, Meru (E61): 3 x, 18 other languages + PB: 2 x each, and 93 other languages + CB: 1 x each. As for many other real-world phenomena, one notices a Zipfean distribution, i.e. the number of Internet dictionaries is extremely high for just a small number of languages, while the frequency for the great majority is very low. That there are relatively many dictionaries for languages such as Swahili, Lingala and isiZulu is understandable; these are the languages that also receive much academic (and other) attention. That a language such as Chagga scores high, however, is out of proportion.

Indeed, there is some serious skewing in the geographical dispersion as a result of one single source that contains over a hundred African-language Internet dictionaries. In the early 1970s Derek Nurse and Gérard Philippson surveyed the languages of Tanzania and neighbouring countries — their study is known as the *Tanzania Language Survey* (TLS, Nurse and Philippson 1975) — and this resulted in 124 parallel c. 1 000-word wordlists. For some of the languages, however, different dialects were recorded — in the case of Chagga, 14, in the case of Meru, 3, etc. In all, there are lexica for 97 *different* languages, as well as one for PB and one for English. Given this, it is thus clear that there is a significant bias towards the languages of Tanzania and East Africa. The fact moreover that Swahili is mainly spoken in Tanzania, pushes the distribution even more into that region of the African continent.

Despite the bias, and despite the small size of the TLS lexica, they are as a whole an interesting application of the *hub-and-spoke model* (Martin 1996: 209, 214). Indeed, with English/Swahili as hub, all the other 122 lexica are linked to it as spokes, and as a result an online dictionary for each and every language pair, triple, quadruple, etc. can now also be 'created', passing through the hub. The number of permutations, and thus the potential number of different *multilingual* dictionaries one can generate in this way, is virtually unlimited. The

basic hub-and-spoke framework is actually becoming ever more popular online for dictionaries involving the languages used in the European Community (EC). In one set of applications, viz. *Ergane* and *Majstro*, Esperanto was chosen as hub with, besides mostly EC languages, Swahili, isiZulu and Setswana (S31) as spokes. From a sound metalexicographic point of view, there are many good reasons to have reservations when it comes to the hub-and-spoke model. Yet choosing an artificial language as hub, thus one where the level of polysemy is virtually non-existent, definitely goes some way to avoid a number of the theoretical problems.

While learners might find it most useful that English was included as one of the parallel lexica of TLS, comparative linguists surely appreciate the fact that Guthrie's PB reconstructions were also added, so that reflexes across the various languages can be directly compared. From the time when Guthrie worked on PB, reconstructions have mainly been drawn up in Tervuren, with Meeussen's (1980, based on a manuscript from 1969) *BLR* and Coupez, Bastin and Mumba's (1998) *BLR 2* the two major releases so far. BLR 2, with 9 800 reconstructed forms, is the backbone of the ambitious *CBOLD* project, originally located in Berkeley, now transferred to Lyon. This research team collected a manifold of dictionaries, mostly as downloadable text files only however, and containing many errors resulting from the use of optical character recognition (OCR) on poor-quality scans. As pointed out above, such dictionaries have not been considered in this study. A total of 22 other dictionaries, as well as BLR 2, *can* be queried online though. Reconstructions to PB for these dictionaries, with BLR index numbers and Guthrie codes, is still ongoing. Note that, at the time of writing, a web site dedicated to BLR 3 is in the making (Bastin et al. 2003).

The CBOLD web site also houses the TLS data, which effectively makes this single site the 'major collection', at least quantity-wise, with 146 online dictionaries for 111 different languages and 2 for PB. In April 2003, the largest Internet dictionary for this language family, however, was located at Yale University, where *The Kamusi Project* contained 58 038 Swahili and 58 041 English 'articles' (Kamusi 1994–2001). These values were arrived at by simply counting the number of entries, and do not reflect the true sizes since a new entry is used for each new synonym, for each new sense, etc. If the number of truly unique lemma signs is summed, regardless of part of speech (POS), then the Swahili to English side turns out to contain 18 411 items, and the English to Swahili side 26 970 items. This dictionary is a prototypical example of bottom-up lexicography (Carr 1997: 214), which means that it is being compiled by Netizens. The contents should thus be consulted with caution.

The second-largest online African-language dictionary, for Lozi (S34), contains 24 000 items. Then follow dictionaries for ChiShona with 15 000 items, for Nyankore (J13) with 12 500 items, etc. At the other end of the spectrum, some of the online dictionaries contain as few as 100 items (for Ganda), 186 items (for Setswana), 300 items (for Lingala), etc. The average number of items in the 182 online African-language dictionaries is 1 978.

It has already been pointed out that the 182 dictionaries cover 117 differ-ent languages. Many of these languages are spoken across country borders, such as Chewa (N31b) which is spoken in both Malawi and Botswana, or Fipa (M13) in Tanzania and Malawi, Luyia (J32) in Kenya and Uganda, Yaka (H31) in the DRC and Angola, etc. If one studies the distribution of the number of languages that have online dictionaries *per country*, the data shown in Table 1 are arrived at.

Table 1: Distribution of the number of African languages with Internet dic-tionaries per country

#C	Country	#Ls with Ds	%
1	Angola	1	0.66
2	Botswana	3	1.99
3	Burundi	2	1.32
4	Congo	2	1.32
5	DRC	6	3.97
6	Gabon	2	1.32
7	Kenya	14	9.27
8	Malawi	7	4.64
9	Mozambique	5	3.31
10	Namibia	2	1.32
11	Rwanda	2	1.32
12	South Africa	4	2.65
13	Tanzania	81	53.64
14	Uganda	10	6.62
15	Zambia	5	3.31
16	Zimbabwe	5	3.31
		151	100.00

From Table 1 it is clear that the greatest allocation is once more to be found in Tanzania, with as many as 81 languages covered. Neighbouring countries such as Kenya with online dictionaries for 14 languages, and Uganda for 10 lan-guages, also score high. In Southern Africa, countries like Zambia, Zimbabwe, Malawi and Mozambique, each cover more languages than South Africa, where there are but 4 languages with Internet dictionaries.

Based on the data found in *Ethnologue*, the 117 languages are spoken by over 100 million people. The dispersion once more moves between extremes. At one extreme, some languages covered are nearly extinct (Geviya (B30)), or are spoken by only a few (Zalamo (G33)), up to a few thousand people (Mpo-ngwe (B11a), Kahe (E64), etc.). At the other extreme, some languages are spo-ken as primary language by over 5 million (Swahili, Sukuma (F21) and Gikuyu (E51)), over 6 million (Rundi (J62) and isiXhosa), over 7 million (ChiShona and Rwanda (J61)), up to over 9 million (isiZulu) people. Very roughly speaking, the average number of primary speakers per language for which there is at least one Internet dictionary is 1 million.

If one looks at dictionary typology, one notices that all but one of the 182 dictionaries is bilingual or multilingual. The only monolingual dictionary is the *Duramazwi ReChiShona* 'General Shona Dictionary' (Chimhundu 1999). Ironically, however, the interface of this monolingual dictionary is entirely in English. A full breakdown of the gloss and/or hub languages is shown in Table 2.

Table 2: Breakdown of the gloss and/or hub languages for all African-language Internet dictionaries

Gloss/Hub language	#	%
English/Swahili	123	67.58
English	25	13.74
French	17	9.34
Esperanto	7	3.85
English/French	4	2.20
German	3	1.65
English/etc. (9 in all)	1	0.55
Italian	1	0.55
— (*monolingual*)	1	0.55
	182	100.00

As one could have expected, roughly nine out of ten dictionaries use English, and only one out of ten use French as the gloss/hub language. Unexpectedly, however, is the relatively large number of dictionaries that involve Esperanto.

None of the 182 dictionaries is stored on a computer in Africa. Even the electronic version of the *Duramazwi ReChiShona* was developed by The Norwegian Documentation Project, and is stored on a server in Oslo. Moreover, very few Africans were involved in the computerisation and creation of these online dictionaries. If one studies the various providers, one notices a clear bias towards academic institutions, which are responsible for eight out of every ten dictionaries. Dotcoms provide one out of seven dictionaries, and less than five percent are personal efforts. The exact distribution has been calculated in Table 3.

Table 3: Providers of African-language Internet dictionaries

Provider	#	%
Academic	149	81.87
Dotcom	24	13.19
Personal	9	4.95
	182	100.00

In general, the soundest contents can be found for the Internet dictionaries compiled by academics, while the most versatile and appealing interfaces are those brought together by dotcoms. The average compilation year is 1981, with the distribution per decade as listed in Table 4.

Table 4: Number of African-language Internet dictionaries compiled per decade

Decade	#Ds	%
1930s	1	0.55
1940s	—	—
1950s	3	1.65
1960s	2	1.10
1970s	127	69.78
1980s	—	—
1990s	18	9.89
2000s	30	16.48
s.d.	1	0.55
	182	100.00

The number of users of the current online dictionaries is much higher than anticipated. For Swahili, for example, *The Kamusi Project* has received over 1.1 million visitors since mid-1995, the *Freedict* dictionary handles 700 visitors per day, while the *Kamusi Kiswahili–Kiesperanto* (Vessella 2001) is accessed at least 1 000 times per month. The online pop-up dictionaries for African languages available from *Babylon* have an average number of 1 400 users each. Lastly, *Xhosa on the Web!* (O'Kennon 1996–2003) has welcomed nearly 30 000 visitors so far.

The first South African online dictionary interface

From the overview presented above, at least two conclusions can be drawn. On the one hand, African-language lexicographers will have to admit that quite a substantial body of Internet dictionaries is *already* available. On the other hand, and this primarily from a South African perspective, one cannot deny the fact that the South African languages should and could be better represented as far as Internet dictionaries are concerned. *TshwaneDJe*, a Human Language Technology (HLT) development team, based in Pretoria and consisting of David Joffe, Gilles-Maurice de Schryver, D.J. Prinsloo and Salmina Nong, therefore decided to bring together all the material for the first South African Internet dictionary.

The choice fell on Sesotho sa Leboa (S32) as the first language for which to compile a dictionary, given that no online dictionaries were found for this language during the course of the Internet study summarised above. The gained expertise would then be applied to the compilation of other African-language Internet dictionaries. The starting point was Prinsloo and De Schryver's (2000) *SeDiPro 1.0*, a Sesotho sa Leboa to English dictionary available to the team in Microsoft Word format. Joffe wrote a parser to transfer the data to *TshwaneLex*, a novel and professional South African software application for dictionary compilation (Joffe, De Schryver and Prinsloo 2003, 2003a). TshwaneLex was

designed in such a way that it can be used to produce hardcopy, CD-ROM as well as online dictionaries. On 22 April 2003, the first version of an *Online Sesotho sa Leboa–English Dictionary* was uploaded (De Schryver and Joffe 2003). Two months later, on 20 June 2003, the online dictionary was officially launched at the University of Pretoria.

Between the first upload and the launch, several adaptations were made and numerous extra features were added to the online dictionary. As such this dictionary is a direct implementation of the concept known as *Simultaneous Feedback* (De Schryver and Prinsloo 2000, 2000a), a methodology whereby especially *indirect feedback* is near-instantly 'fed back' into the compilation process of a dictionary. The lexicographic contents are currently being updated by Nong.

During the first two months, users primarily learned about the new online dictionary through word of mouth. On the eve of the launch, 366 different users had searched for 3 341 items, or on average 9.12 searches per person. This was equivalent to more than 50 searches by more than 7 different users per day. The first media release appeared two weeks later, on 4 July 2003 (cf. e.g. *Mail and Guardian Online* 2003). At the end of that day, the number of searches had already reached 5 779 by 802 different users, or an average of 78.09 searches by 12.15 persons per day. The great majority of these searches had been made from hosts *in* South Africa. This clearly exceeded even the wildest expectations at TshwaneDJe.

From a metalexicographic perspective, this online dictionary deserves some extra discussion. Firstly, it is the first African-language Internet dictionary that can be accessed in all languages covered by the dictionary. In this case, this means that all interface pages are available in both Sesotho sa Leboa and English. Primary speakers of Sesotho sa Leboa can thus for the first time consult a dictionary in their own language.

Secondly, although actually only the direction Sesotho sa Leboa to English exists, an English search index (which also includes support for multi-word units) has been added which makes it possible to search the dictionary as if the reverse side were also available. The layout of the output is also a first, as it shows how the senses in one language are spread all over the lexicon in another, and how these then again spread out, etc. With 24 921 items on the Sesotho sa Leboa side and 28 198 in the English index, this online dictionary becomes the largest African-language Internet dictionary.

Thirdly, besides a general-language dictionary, this is also the first online dictionary that includes a dedicated terminology list for an African language. The terminology list that has currently been added is one for linguistics, containing over 300 terms, and more terminology lists are planned.

Fourthly, when consulting the terminology list, users can choose between look-up and browse mode. This is thus an original implementation of Atkins's (1996) innovative view of future electronic dictionaries. According to her, "the user is in search of a specific piece of information" in look-up mode, while "a more relaxed reading takes place" in browse mode (1996: 529). In look-up mode

users are furthermore re-routed from (potentially) incorrectly to correctly spelled items for words involving the letters s/š, e/ê and o/ô.

Lastly, and also most importantly, the terminology list contains a world's first for an online dictionary, namely the customisation of the output of part-of-speech (POS) tags, usage labels and cross-references depending on the language chosen. As such, this is the first step towards one concept of the dictionary of the future, viz. *Fuzzy SF* (De Schryver and Prinsloo 2001). In Fuzzy SF, or Fuzzy Simultaneous Feedback, "log-file based Artificial Intelligence components enable the implicit retrieval of personalised user feedback with which the package customises each user's own and unique dictionary" (De Schryver 2003: 189).

Conclusion

In this article a near-exhaustive overview was presented of the current state-of-the-art of African-language Internet dictionaries. The concepts 'African languages' and 'Internet dictionaries' were first defined for the purposes of this article. All currently available African-language Internet dictionaries were then reviewed, listed and compared to one another. Various statistics were calculated and distributions shown, from which one may conclude that there is a geographic bias towards the languages of East Africa, especially Tanzania. Among the most successful implementations one must count the hub-and-spoke model as used for the presentation of the data from the *Tanzania Language Survey*, now part of the *CBOLD* web site.

A surprising number of 182 dictionaries were uncovered, for 117 different languages. The South African share was shown to be small. Although an estimated 100 million people speak the languages covered, just one of the dictionaries is a monolingual one. None of the dictionaries is stored in Africa, and few Africans contributed to the computational creation of these dictionaries. Most dictionaries are the output of academic institutions, are relatively recent, and have a higher-than-expected number of users. The most popular dictionaries are those for Swahili, for which there are as many as 20.

In order to turn the relatively inactive online lexicographic tide for the languages spoken in South Africa, it was indicated how the HLT development team *TshwaneDJe* decided to produce the first truly South African online dictionary interface. The language embarked upon is Sesotho sa Leboa. Compilation is undertaken within the frameworks of *Simultaneous Feedback* (SF) and *Fuzzy SF*, and it was shown how, in less than three months, the number of searches and users had already reached unexpected heights. The dictionary is currently the largest online African-language Internet dictionary. Among the novelties of the online Sesotho sa Leboa dictionary, the dual dictionary interface language (including the first in an African language), a layout inherently departing from an African language, the first searchable African-language Internet terminology list, the optional look-up and browse modes, as well as

the first steps towards user customisation, were highlighted. As such, South African lexicography is already writing the future.

References

Atkins, B.T. Sue. 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 I–II, Papers Submitted to the Seventh EURALEX International Congress on Lexicography in Göteborg, Sweden*: 515-546. Gothenburg: Department of Swedish, Göteborg University.

Babylon. 1997–2003. Online Dictionary Software by Babylon [online]. Available: <http://www.babylon.com/>.

Bastin, Yvonne. 1978. Les langues bantoues. Barreteau, Daniel (Ed.). 1978. *Inventaire des études linguistiques sur les pays d'Afrique noire d'expression française et sur Madagascar*: 123-185. Paris: Conseil international de la langue française.

Bastin, Yvonne, André Coupez and Michael Mann. 1999. *Continuity and Divergence in the Bantu Languages: Perspectives from a Lexicostatistic Study.* Human Sciences Annals of the Royal Museum for Central Africa, Series in -8°, 162. Tervuren: RMCA.

Bastin, Yvonne, André Coupez, Evariste Mumba and Thilo C. Schadeberg. 2003. *Reconstructions lexicales bantoues 3 / Bantu Lexical Reconstructions 3* [online]. Available: <http://linguistics.africamuseum.be/BLR3.html>.

Carr, Michael. 1997. Internet Dictionaries and Lexicography. *International Journal of Lexicography* 10(3): 209-230.

CBOLD. 1997–2003. Comparative Bantu OnLine Dictionary [online]. Available: <http://www.cbold.ddl.ish-lyon.cnrs.fr/>.

Chimhundu, Herbert (Ed.). 1999. *Duramazwi ReChiShona* (General Shona Dictionary) [online]. Available: <http://www.dokpro.uio.no/allex/gsd.html>.

Cope, Anthony T. 1971. A Consolidated Classification of the Bantu Languages. *African Studies* 30(3-4): 213-236.

Coupez, André, Yvonne Bastin and Evariste Mumba. 1998. *Reconstructions lexicales bantoues 2 / Bantu Lexical Reconstructions 2* [online]. Available: <http://www.cbold.ddl.ish-lyon.cnrs.fr/>.

De Schryver, Gilles-Maurice. 2003. Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography* 16(2): 143-199.

De Schryver, Gilles-Maurice and David Joffe. 2003. *Online Dictionary Interface Sesotho sa Leboa (Northern Sotho)–English* [online]. Available: <http://africanlanguages.com/sdp/>.

De Schryver, Gilles-Maurice and D.J. Prinsloo. 2000. Dictionary-Making Process with 'Simultaneous Feedback' from the Target Users to the Compilers. Heid, Ulrich, Stefan Evert, Egbert Lehmann and Christian Rohrer (Eds.). 2000. *Proceedings of the Ninth Euralex International Congress, EURALEX 2000, Stuttgart, Germany, August 8th–12th, 2000*: 197-209. Stuttgart: Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.

De Schryver, Gilles-Maurice and D.J. Prinsloo. 2000a. The Concept of 'Simultaneous Feedback': Towards a New Methodology for Compiling Dictionaries. *Lexikos* 10: 1-31.

De Schryver, Gilles-Maurice and D.J. Prinsloo. 2001. Fuzzy SF: Towards the Ultimate Customised

Dictionary. *Studies in Lexicography* 11(1): 97-111. [Also published in Lee, Sangsup (Ed.). 2001. *ASIALEX 2001 PROCEEDINGS, Asian Bilingualism and the Dictionary*: 141-146. Seoul: Center for Linguistic Informatics Development, Yonsei University.]

eLanguage. 2002–2003. Riverdeep Interactive Learning Limited [online]. Available: <http://www.elanguage.com/eng/default.asp>.

Ergane. 2003. Ergane Free Multi-lingual Dictionary [online]. Available: <http://www.travlang.com/Ergane/>.

Freedict. 2003. Free Online Swahili–English Dictionary [online]. Available: <http://www.freedict.com/onldict/swa.html>.

Freelang. 1997–2003. Free Dictionaries to Download and Free Human Translation [online]. Available: <http://www.freelang.net/>.

Glossword. 2002–2003. Glossword Development [online]. Available: <http://dev.glossword.ru/?il=en>.

Google. 2003. Google: Internet Search Engine [online]. Available: <http://www.google.com>.

Grimes, Barbara F. and Joseph E. Grimes (Eds.). 2000. *Ethnologue. Volume 1: Languages of the World. Volume 2: Maps and Indexes* (14th Edition) [online]. Dallas: SIL International. Available: <http://www.ethnologue.com/>.

Grouselle, Pascal. 2002–2003. *Initiation au lingala, langue bantoue* [online]. Available: <http://perso.wanadoo.fr/pascal.grouselle>.

Guthrie, Malcolm. 1948. *The Classification of the Bantu Languages*. London: Oxford University Press.

Guthrie, Malcolm. 1953. *The Bantu Languages of Western Equatorial Africa*. London: Oxford University Press.

Guthrie, Malcolm. 1967. *Comparative Bantu 1: An Introduction to the Comparative Linguistics and Prehistory of the Bantu Languages*. Farnborough: Gregg Press.

Guthrie, Malcolm. 1970. *Comparative Bantu 3: An Introduction to the Comparative Linguistics and Prehistory of the Bantu Languages*. Farnborough: Gregg International Publishers.

Guthrie, Malcolm. 1970a. *Comparative Bantu 4: An Introduction to the Comparative Linguistics and Prehistory of the Bantu Languages*. Farnborough: Gregg International Publishers.

Guthrie, Malcolm. 1971. *Comparative Bantu 2: An Introduction to the Comparative Linguistics and Prehistory of the Bantu Languages*. Farnborough: Gregg International Publishers.

Ikuska Libros. 1997–2003. Diccionario Lingala–Español–Lingala [online]. Available: <http://www.ikuska.com/Africa/Lenguas/lingala/diccionario.htm>.

Joffe, David, Gilles-Maurice de Schryver and D.J. Prinsloo. 2003. Introducing TshwaneLex — A New Computer Program for the Compilation of Dictionaries. De Schryver, Gilles-Maurice (Ed.). 2003. *TAMA 2003 South Africa: CONFERENCE PROCEEDINGS*: 97-104. Pretoria: (SF)² Press.

Joffe, David, Gilles-Maurice de Schryver and D.J. Prinsloo. 2003a. Tree Structure, I/O Architecture, Database Loading and Cross-referencing in the Dictionary Application "TshwaneLex". *Southern African Linguistics and Applied Language Studies* 21(3).

Kamusi. 1994–2001. The Kamusi Project. Internet Living Swahili Dictionary [online]. Available: <http://www.yale.edu/swahili/>.

Lowe, John B. and Thilo C. Schadeberg. 1996. *Bantu MapMaker version 3.1* [online]. Available: <http://www.cbold.ddl.ish-lyon.cnrs.fr/Maps/BMM1.html>.

Maho, Jouni F. 2002. *Bantu Line-up: Comparative Overview of Three Bantu Classifications* [online].

Gothenburg: Department of Oriental and African Languages, Göteborg University. Available: <http://www.african.gu.se/maho/jfmpub.html>.

Maho, Jouni F. 2003. *The Languages of Tanzania: A Web Links Collection* [online]. Available: <http://www.african.gu.se/tanzania/weblinks.html>.

Mail and Guardian Online. 2003. Online Sotho–English Dictionary Launched [online]. Friday, July 4, 2003; South African Press Association. Available: <http://www.mg.co.za/Content/l3.asp?ao=16666>.

Majstro. 2003. A Multilingual Translation Dictionary [online]. Available: <http://www.majstro.com/>.

Maniacky, Jacky. 2003. *Bantu Web Site, Site Web Bantu* [online]. Available: <http://www.bantu.ovh.org>.

Martin, Willy. 1996. Lexicographical Resources in a Multilingual Environment: An Orientation. *Lexikos* 6: 199-214.

Meeussen, Achille E. 1980 [1969]. *Bantu Lexical Reconstructions.* Archives d'Anthropologie 27. Tervuren: RMCA.

Nurse, Derek and Gérard Philippson. 1975. The North-East Bantu Languages of Tanzania and Kenya: A Classification. *KiSwahili* 45(2): 1-28.

Odden, David. 2002. *Kikerewe–English Dictionary* [online]. Available: <http://www.ling.ohio-state.edu/~odden/>.

O'Kennon, Martha. 1996–2003. *Xhosa on the Web! Xhosa to English and English to Xhosa Translator* [online]. Available: <http://mokennon.albion.edu/>.

Prinsloo, D.J. and Gilles-Maurice de Schryver (Eds.). 2000. *SeDiPro 1.0, First Parallel Dictionary Sepêdi–English.* Pretoria: University of Pretoria.

Ridings, Daniel. 2002. *Shona and Ndebele Corpus Resources from the African Languages Lexicon Project (ALLEX)* [online]. Available: <http://www.ling.gu.se/~ridings/africanlang.html>.

TLS. 1999 [1970s]. Tanzania Language Survey [online]. WWW version 1.0. Available: <http://www.cbold.ddl.ish-lyon.cnrs.fr/Docs/TLS.html>.

TravLang. 1995–2003. Foreign Languages for Travellers [online]. Available: <http://www.travlang.com/languages/>.

Varantola, Krista. 2002. Use and Usability of Dictionaries: Common Sense and Context Sensibility? Corréard, Marie-Hélène (Ed.). 2002. *Lexicography and Natural Language Processing: A Festschrift in Honour of B.T.S. Atkins*: 30-44. EURALEX. Information: <http://www.ims.uni-stuttgart.de/euralex/>.

Vessella, Nino. 2001. *Kamusi Kiswahili–Kiesperanto / Vortaro Esperanto–Suahila* [online]. Available: <http://kamusi.vessella.net/>.

WAT. 1926– . Woordeboek van die Afrikaanse Taal. Information: <http://www.sun.ac.za/wat/>.

Appendix: African-language Internet dictionaries – An overview (April 2003)

#D	#L	Language/Dialect	Guthrie code	Country(ies)	Speakers	Type	Lemma signs	Gloss/Hub language	Compiler(s)	Year	Provider
1	1	Bena	G63	Tanzania	568 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
2	2	Bende	F12	Tanzania	20 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
3	3	Bondei	G24	Tanzania	80 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
4						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
5	4	Bukusu	J31c	Kenya	565 000	Online	5900	English	KWL	1998	CBOLD
6						Online	900	English	Deblois	s.d.	CBOLD
7	5	Bungu (Wungu)	F25	Tanzania	36 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
8						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
9			E62	Tanzania	400 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
10						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
11						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
12						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
13		Chagga (Keni, Kibosho, Kimochi, Kiseri, Lema, Machame, Mamba, Mkuu, Siha, Vunjo)				Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
14						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
15						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
16	6					Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
17						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
18						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
19						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
20						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
21						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
22						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
23	7	Chewa	N31b	Malawi, Botswana	3 958 000	Online	6200	English	Scott & Hetherwick	1957	CBOLD
24	8	Chiga (Rukiga)	J14	Uganda	1 391 442	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
25	9	ChiShona	S11-S12-S14	Zimbabwe, Zambia	7 000 000	Online	15000	— *(monolingual)*	Chimhundu	1999	Norwegian Doc. Project
26	10	Chuka	E53	Kenya	70 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
27	11	Doe	G301	Tanzania	24 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
28	12	Embu	E52	Kenya	429 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
29	13	Fipa	M13	Tanzania, Malawi	200 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
30						Online	6300	English	Snoxall	1967	CBOLD
31	14	Ganda (Luganda)	J15	Uganda, Tanzania	3 025 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
32						PC	100	French	Ramel	2001	Freelang
33						PC	100	(French →)	Ramel	2001	Freelang
34	15	Geviya	B30	Gabon	*near extinct*	Online	1450	French	Van der Veen	1994	CBOLD

No	#	Language	Code	Country	Population	Type	Entries	Languages	Author	Year	Source
35	16	Gikuyu (Kikuyu)	E51	Kenya	5 347 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
36	17	Gusii	E42	Kenya	1 582 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
37	18	Gweno	E65	Tanzania	*unknown*	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
38	19	Gwere	J17	Uganda	275 608	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
39	20	Ha (Kiha)	J66	Tanzania	800 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
40	21	Hangaza	J65	Tanzania	150 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
41	22	Haya	J22	Tanzania	1 200 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
42	23	Hehe	G62	Tanzania	750 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
43	24	Ikizu	E402	Tanzania	28 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
44	25	Ikoma (Nata)	E45	Tanzania	15 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
45	26	isiXhosa	S41	South Africa	6 876 000	Online	3000	English	O'Kennon	1996	Personal
46				South Africa		Online	3000	(English →)	O'Kennon	1996	Personal
47	27	isiZulu	S42	South Africa	9 142 000	Online	2000	English	Anon.	2002	eLanguage
48						Online	1001	Esperanto	Van Wilgen	2003	Majstro
49						PC	835	Esperanto	Van Wilgen	2002	Ergane
50						Pop-up	805	Italian	Toscano	2001	Babylon
51	28	Jita	J25	Tanzania	217 000	Online	2200	English	Downing	1999	CBOLD
52						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
53	29	Kahe	E64	Tanzania	2 700	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
54	30	Kalanga	S16	Botswana, Zimbabwe	321 000	Online	3000	English	Mathangwane	1994	CBOLD
55	31	Kamba (Kamba Kitu/Mach)	E55	Kenya	2 448 302	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
56						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
57	32	Kami	G36	Tanzania	10-20 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
58	33	Kara (Kilegi)	J252	Tanzania	86 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
59	34	Kerebe (Kerewe)	J24	Tanzania	100 000	Online	1500	English	Odden	1994	CBOLD
60						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
61	35	Kimbu	F24	Tanzania	78 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
62	36	Kinga	G65	Tanzania	65 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
63	37	Kinyaturu (Nyaturucha, Nyaturuwii)	F32	Tanzania	556 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
64						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
65	38	Kisi	G67	Tanzania	13 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
66	39	Koyo	C24	Congo	*unknown*	Online	1700	French	Gazania & Hyman	1996	CBOLD
67	40	Kuria (Kuria Mago/Tari)	E43	Tanzania, Kenya	348 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
68						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
69	41	Kutu	G37	Tanzania	45 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
70	42	Kwaya (Mkwaya)	J251	Tanzania	102 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
71	43	Lambya	N26	Tanzania, Malawi	81 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
72	44	Langi	F33	Tanzania	310 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
73	45	Lingala	C36d	DRC	309 100	Online	539	French	Grouselle	2002	Personal

#	No.	Language	Code	Countries	Population	Format	Entries	Languages	Author	Year	Source
74						Online	693	(French →)	Grouselle	2002	Personal
75						PC	508	French	Grouselle	2002	Freelang
76						PC	656	(French →)	Grouselle	2002	Freelang
77						Online	300	English/French	Maniacky	2002	Personal
78	46	Logooli (Maragoli)	E41	Kenya	197 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
79	47	Lozi	S34	Zambia, Namibia, Zimbabwe	557 000	Online	24000	English	Jalla	1937	CBOLD
80	48	Luyia	J32	Kenya, Uganda	3 593 461	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
81	49	Makhua (Emakhua)	P31	Mozambique, Tanzania	2 500 000	Online	7200	English	Kisseberth	1996	CBOLD
82	50	Makonde	P23	Tanzania, Mozambique	1 260 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
83	51	Malila	M24	Tanzania	52 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
84	52	Mambwe–Lungu	M15-M14	Zambia, Tanzania	359 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
85		(Mambwe, Rungu)				Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
86	53	Manda	N11	Tanzania	18 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
87	54	Masaba (Lumasaaba)	J31	Uganda	751 253	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
88	55	Matengo	N13	Tanzania	150 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
89	56	Matumbi	P13	Tanzania	72 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
90	57	Maviha (Mabia)	P25	Tanzania, Mozambique	*cf. Makonde*	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
91	58	Mbunga	P15	Tanzania	29 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
92	59	Meru (Meruimenti,	E61	Kenya	1 305 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
93		Merutig)				Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
94						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
95	60	Mpongwe	B11a	Gabon	1-4 000	Online	2600	French	Mouduiama	1994	CBOLD
96	61	Mpoto	N14	Tanzania, Malawi	80 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
97	62	Munukutuba	Kongo (H16) based Creole	Congo	1 156 800	Online	300	English/French	Maniacky	2002	Personal
98	63	Mwanga (Namwanga)	M22	Zambia, Tanzania	256 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
99	64	Mwera	P22	Tanzania	400 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
100						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
101	65	Nandi (Nande)	J42	DRC	903 000	Online	2100	French	Kavutirwaki	1978	CBOLD
102	66	Ndali (Ndari)	M301	Tanzania	150 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
103						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
104	67	Ndamba	G52	Tanzania	55 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
105	68	Ndengereko (Ndengeleko)	P11	Tanzania	110 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
106	69	Ngh'wele (Kwere)	G32	Tanzania	98 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
107	70	Ngindo	P14	Tanzania	220 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
108	71	Ngoni	N12	Tanzania, Malawi, Mozambique	205 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
109	72	Ngulu (Nguungulu)	G34	Tanzania	132 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS

#	#	Language	Code	Country	Speakers	Medium	Entries	Languages	Author	Year	Source
110	73	Ngurimi (Ngoreme)	E401	Tanzania	32 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
111	74	Nyakyusa	M31	Tanzania, Malawi	1 050 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
112	75	Nyambo (Runyambo)	J21	Tanzania	7 000	Online	1500	English	Rugemalira	1993	CBOLD
113						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
114	76	Nyamwezi	F22	Tanzania	926 000	Online	2000	English	Maganga & Schadeberg	1992	CBOLD
115						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
116	77	Nyankore (Runya-nkore, Kiga-Nkore)	J13	Uganda	1 643 193	Online	12500	English	Taylor	1959	CBOLD
117						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
118	78	Nyiha	M23	Tanzania, Zambia	626 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
119	79	Nyiramba (Nilamba)	F31	Tanzania	440 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
120	80	Nyoro (Runyoro)	J11	Uganda	495 443	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
121	81	Pangwa	G64	Tanzania	177 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
122	82	Phende (Pende)	K52	DRC	420 000	Online	8200	French	Gusimana	1972	CBOLD
123	83	Pimbwe	M11	Tanzania	29 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
124	84	Pogolo (Pogoro)	G51	Tanzania	185 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
125	85	Rufiji	P12	Tanzania	200 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
126	86	Ruguru (Luguru)	G35	Tanzania	520 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
127	87	Rundi	J62	Burundi, Rwanda, Tanzania, Uganda	6 000 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
128	88	Rungwa (Lungwa)	M12	Tanzania	18 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
129						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
130	89	Rwanda (Munyarwanda)	J61	Rwanda, Burundi, DRC, Tanzania, Uganda	7 362 800	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
131	90	Saamia (Samialugwe)	J34	Uganda	50 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
132	91	Safwa	M25	Tanzania	158 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
133	92	Sangu	G61	Tanzania	75 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
134	93	Setswana	S31	South Africa, Botswana, Namibia	4 000 000	PC	186	Esperanto	Van Wilgen	1998	Ergane
135	94	Shambala (Sambaa)	G23	Tanzania	550 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
136	95	Shi	J53	DRC	654 000	Online	2500	French	Polak-Bynon	1978	CBOLD
137	96	Shubi	J64	Tanzania	153 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
138	97	SiNdebele (Zim-babwean Ndebele)	S44	Zimbabwe	1 502 000	Online	5000	English	Pelling	1971	CBOLD
139	98	Sizaki (Shashi Siz)	E404	Tanzania	82 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
140	99	Soga (Lusoga)	J16	Uganda	1 370 845	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
141	100	Suba	E403	Kenya, Tanzania	159 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
142	101	Sukuma (Ntuzu)	F21	Tanzania	5 000 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
143						Online	1000	Swahili/English	Nurse & Philippson	1975	TLS
144	102	Sumbwa	F23	Tanzania	191 000	Online	1000	Swahili/English	Nurse & Philippson	1975	TLS

No.	ID	Language	Code	Region	Speakers	Type	Lemmas / Language	Author	Year	Source
145			G42	Tanzania, Kenya	5 000 000	Online	18411 English	Netizens	2001	Kamusi
146						Online	26970 (English →)	Netizens	2001	Kamusi
147						Online	8545 Esperanto	Vessella	2001	Personal
148						Online	6547 (Esperanto →)	Vessella	2001	Personal
149						Online	5000 English/etc. (9)	Anon.	2002	eLanguage
150						Pop-up	2503 German	Alim	2001	Babylon
151						Pop-up	684 French	Omankoy	2001	Babylon
152						Pop-up	1213 (French →)	Omankoy	2001	Babylon
153						Pop-up	1092 German	Madete	2002	Babylon
154	103	Swahili				Pop-up	1045 (German →)	Madete	2002	Babylon
155						Online	1000 Swahili/English	Nurse & Philippson	1975	TLS
156						Pop-up	644 English	Van der Meijden	2000	Babylon
157						Pop-up	949 (English →)	Van der Meijden	2000	Babylon
158						Online	750 English	Anon.	2003	Freedict
159						Online	750 (English →)	Anon.	2003	Freedict
160						PC	570 French	Anon.	1999	Freelang
161						PC	749 (French →)	Anon.	1999	Freelang
162						PC	736 Esperanto	Van Wilgen	2002	Ergane
163							665 Esperanto	Van Wilgen	2003	Majstro
164						Online	645 English	Shilnikov	2002	Glossword
165	104	Taita (Dawida)	E74a	Kenya	203 389	Online	1000 Swahili/English	Nurse & Philippson	1975	TLS
166	105	Temi (Sonjo)	E46	Tanzania	20 000	Online	1000 Swahili/English	Nurse & Philippson	1975	TLS
167	106	Tharaka	E54	Kenya	112 000	Online	1000 Swahili/English	Nurse & Philippson	1975	TLS
168	107	Tooro (Rutooro)	J12	Uganda	488 024	Online	1000 Swahili/English	Nurse & Philippson	1975	TLS
169	108	Tshivenda	S21	South Africa, Zimbabwe	750 000	Online	8900 English	Murphy	1997	CBOLD
170	109	Vinza (Kivinza)	J67	Tanzania	10 000	Online	1000 Swahili/English	Nurse & Philippson	1975	TLS
171	110	Wanda	M21	Tanzania	24 000	Online	1000 Swahili/English	Nurse & Philippson	1975	TLS
172	111	Wanji	G66	Tanzania	60 000	Online	1000 Swahili/English	Nurse & Philippson	1975	TLS
173	112	Yaka	H31	DRC, Angola	150-200 000 000	Online	3900 French	Ruttenberg	1969	CBOLD
174	113	Yao	P21	Malawi, Mozambique, Tanzania	1 597 000	Online	7400 English	Sanderson	1954	CBOLD
175						Online	1000 Swahili/English	Nurse & Philippson	1975	TLS
176	114	Zalamo (Zaramo)	G33	Tanzania	a few	Online	1000 Swahili/English	Nurse & Philippson	1975	TLS
177	115	Zanaki	E44	Tanzania	62 000	Online	1000 Swahili/English	Nurse & Philippson	1975	TLS
178	116	Zigula (Zigua)	G31	Tanzania	355 000	Online	1000 Swahili/English	Nurse & Philippson	1975	TLS
179	117	Zinza	J23	Tanzania	138 000	Online	1000 Swahili/English	Nurse & Philippson	1975	TLS
180	118	* Common Bantu	* CB	N.A.	N.A.	Online	2783 English/French	Maniacky	2002	Personal
181	119	* Proto Bantu	* PB	N.A.	N.A.	Online	9800 English/French	Tervuren Bantu Group	1998	CBOLD
182						Online	1000 Swahili/English	Nurse & Philippson	1975	TLS

The Lemmatisation of Adverbs in Northern Sotho[*]

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

Abstract: To date Northern Sotho metalexicographers have focused their attention on lemmatisation problems in respect of the so-called main or primary part of speech categories, viz. nouns and verbs. See, for example, Prinsloo and De Schryver (1999) and Prinsloo and Gouws (1996). No attention has been given to the lemmatisation of *adverbs*. The latter are regarded by Ziervogel and Mokgokong (1975: 114, Introduction) as a "secondary part of speech". The treatment of adverbs in Northern Sotho dictionaries is marred by inconsistencies such as omissions from the macrostructure, insufficient and inconsistent labelling, inferior treatment in the microstructure, under-utilization of the mediostructure and outer texts, and reflects a lack of a strategy of selection of items for lemmatisation. Linguistic descriptions of adverbs in currently available grammars vary substantially and therefore confuse learners of the language and inexperienced lexicographers[1]. The aim of this article is to offer solutions to the lemmatisation problems regarding adverbs in Northern Sotho and to propose guiding entries for paper and electronic dictionaries which could serve as models for future dictionaries. The treatment of adverbs in Northern Sotho dictionaries will also be critically evaluated, especially in terms of frequency of use and target users' needs.

Keywords: LEXICOGRAPHY, LEMMATISATION, ADVERBS, INFORMATION RETRIEVAL, ELECTRONIC DICTIONARY, MACROSTRUCTURE, MICROSTRUCTURE, CROSS-REFERENCING, MEDIOSTRUCTURE, DICTIONARY, AFRICAN LANGUAGES, BACK MATTER, NORTHERN SOTHO

Opsomming: Die lemmatisering van bywoorde in Noord-Sotho. Tot dusver het Noord-Sotho metaleksikograwe hulle aandag bepaal by lemmatiseringsprobleme ten opsigte van die sogenaamde primêre woordkategorieë, naamlik naamwoorde en werkwoorde. Vergelyk byvoorbeeld, Prinsloo en De Schryver (1999) en Prinsloo en Gouws (1996). Geen aandag is gegee aan die lemmatisering van *bywoorde* nie. Laasgenoemde word deur Ziervogel en Mokgokong (1975: 72, Inleiding) as 'n "sekondêre rededeel" beskou. Die bewerking van bywoorde in Noord-Sotho woordeboeke word bederf deur inkonsekwenthede soos weglatings uit die makrostruktuur, onvoldoende en inkonsekwente etikettering, minderwaardige bewerking in die mikrostruktuur, onderbenutting van die mediostruktuur en buitetekste, en vertoon 'n gebrek aan 'n strategie vir seleksie van items vir lemmatisering. Taalkundige beskrywings van bywoorde in tans beskikbare grammatikas verskil grootliks en verwar dus aanleerders van die taal en onervare leksikograwe.[2] Die doel van hierdie artikel is om oplossings aan die hand te doen vir die lemmatiseringsprobleme rakende bywoorde in Noord-Sotho en gidsinskrywings voor te stel vir papier- en elektroniese woordeboeke

---

wat as modelle vir toekomstige woordeboeke kan dien. Die bewerking van bywoorde in Noord-Sotho woordeboeke sal ook krities geëvalueer word, veral ten opsigte van gebruiksfrekwensie en teikengebruikers se behoeftes.

Sleutelwoorde: LEKSIKOGRAFIE, LEMMATISERING, BYWOORDE, INLIGTINGSONT-SLUITING, ELEKTRONIESE WOORDEBOEK, MAKROSTRUKTUUR, MIKROSTRUKTUUR, KRUISVERWYSING, MEDIOSTRUKTUUR, WOORDEBOEK, AFRIKATALE, AGTERWERK, NOORD-SOTHO

Introduction

According to Prinsloo and Gouws (1996: 103), the lexicographer is the mediator between theoretical linguistics and the everyday language user. In practical terms, this often means that the African-language lexicographer has to take great pains in lemmatising grammatically complex systems in a user-friendly way on the level of the target user. Typical examples are the lemmatisation of nouns, verbs, reflexives, adjectives and especially copulatives (cf. Prinsloo 2002). A dictionary should not primarily reflect the attitude of the lexicographer; it should rather be aimed at specific needs of a well-defined target user. It will be illustrated in terms of adverbs that lexicographers should strive to lemmatise adverbs in Northern Sotho in such a way that the whole spectrum of occurrences of adverbs is covered with maximum utilization of all lexicographic mechanisms at their disposal. The user-perspective, and especially the need for modern dictionaries to be user-friendly, has been prominent in lexicographic studies of the past decade (cf. Gouws and Prinsloo (1998), Hartmann and James (1998), Prinsloo and De Schryver (1999), Gouws (2000), etc.) and will be regarded as a given in this article. The South African situation moreover often demands dictionaries to be accessible to a wider user group than originally envisaged by the compiler. Lexicographers should therefore strive towards maximum poli-functionality of their dictionaries. Special attention should be given to the encoding needs of learners, in this case to the need to find enough information in dictionaries in order to actively use adverbs in speech and writing.

The aim of this article is to offer solutions to the lemmatisation problems regarding adverbs in Northern Sotho. It will also be attempted to show how macrostructural and microstructural strategies as well as the mediostructure can be maximally utilized in order to reach this objective. The different kinds of adverbs distinguished for Northern Sotho appear thousands of times in the Pretoria Sepedi Corpus. These enormous overall counts clearly indicate not only that they should be included as lemmas but also that an exhaustive treatment is required and/or justified especially for the encoding needs of inexperienced target users. Prerequisites will be to obtain an overall picture of the adverbial system and to find appropriate lemmatisation strategies for the different types of adverbs in Northern Sotho. The question is therefore what the lexicographer has to know about the adverb in Northern Sotho in order to em-

bark on successful lexicographic treatment of adverbs and how to lemmatise them in a user-friendly way. It cannot be expected from him/her, however, to solve deeply-rooted theoretical differences between linguists on the approaches to the description of adverbs.

It will also be emphasized that in order to lemmatise adverbs successfully, the lexicographer should not hesitate to go beyond 'word boundaries'[3] in the selection of lemmas. Lexical elements smaller than words, such as affixes, and lexical elements larger than words, such as adverbial phrases, should be considered for lemmatisation. Gouws (1989: 84) correctly emphasizes that the traditional focus on the word as representative of the lexicon should be shifted to lemmas representing the lexical items of the particular language.

Although general definitions of adverbs vary, they all formulate the core function of adverbs as describing or modifying a clause or action in terms of especially time, place and manner.

> An **adverb** is a word such as 'slowly', 'now', 'very', 'politically' or 'fortunately' which adds information about the action, event, or situation mentioned in a clause. (Sinclair 1995: 27)

> … a word used for describing a verb, an adjective, another adverb, or a whole sentence. Adverbs in English often consist of an adjective with '-ly' added, for example 'quickly', 'mainly', and 'cheerfully'. (Rundell 2002: 20)

> … to describe how, where, when or how often something happens … (Procter 1995: 20, textbox)

> Adverbs are words which qualify or describe verbs, adjectives and other adverbs in some or other way. (Van Wyk et al. 1992: 118)

> … adverbs describe the nature of the action in terms of *time*, *place* and *manner*. (Louwrens 1991: 26)

It could be argued that learners and prospective, inexperienced compilers find the description and treatment of adverbs in currently available dictionaries and grammars of Northern Sotho unsatisfying and even confusing.

Firstly, a wide range of terminology is used to refer to the different kinds of adverbs, viz. basic adverbs, genuine adverbs, common adverbs, secondary derivations, derived adverbs, adverbs that developed from other categories, adopted adverbs, descriptive adjuncts and pseudo-adverbs. With particular reference to adverbial phrases, the terms particles, prepositions and prefixes are used to describe the same kind of lexical elements, depending on the theoretical framework favoured by the author in question. On the one hand different terms such as *basic adverbs* and *genuine adverbs* are used to refer to the same type of adverb while on the other, a single term, for example *derived adverbs*, is used to refer to different types of adverbs by different compilers. The learner can also easily mistakenly assume adverbs derived from other categories, and adverbs developed from other categories to be the same type of adverbs. The latter, however, refers to adopted adverbs. Louwrens (1991) regards *ka*, *le*, *go*,

etc. which introduce adverbial groups, as particles, but Poulos and Louwrens (1994) call them prefixes.

Secondly, Louwrens (1991: 26) says "it is preferable not to regard particle groups as adverbs …" but in Poulos and Louwrens (1994) these groups are indeed regarded as adverbs (see main and subcategories 1 to 6 in Table 2).

The potential confusion for the learner and the prospective lexicographer can also be illustrated by means of *kudu* 'mainly'. Lombard (1985: 166) says it is a basic adverb not related to any other word category. Poulos and Louwrens (1994: 341) agree and add that it is not derived from any other word category and that it has an inherent adverbial meaning. Ziervogel and Mokgokong (1975: 114, Introduction) refer to it as a noun which is a common adverb, and in the central text indicate the part of speech of *kudu* as adverb. Kriel and Van Wyk (1989) label it as a noun of class 9 and offer no treatment of its adverbial characteristics in the entire article of the lemma *kudu*. Van Wyk et al. (1992) and Lombard (1985) recognize 3 basic types of adverbs. Louwrens (1991) distinguishes the categories time, place and manner. Lombard (1985: 168) does not make provision for adverbs of place and says that the so-called adverbs of place are not adverbs. Van Wyk et al. (1992) only say adverbs qualify "in some or other way". Louwrens (1991: 26), in contrast to Lombard (1985) and Van Wyk et al. (1992), does not categorise adverbs in terms of basic, derived and adopted.[4] Poulos and Louwrens (1994) describe adverbs in terms of their derivations and distinguish not less than 9 main categories and up to 17 subcategories. Ziervogel and Mokgokong (1975), in contrast to the other linguists, disregard the category "basic adverb". In fact they describe the nature of adverbs in a rather clumsy way. A dead reference in respect of the final category *ga-* adds to the user's predicament since vital information required to complete the paradigm cannot be retrieved at this point.

> Other parts of speech are used as adverbs, or adverbs may be formed by affixing prefixes or suffixes to other parts of speech. Nouns are often used unchanged as adverbs. … secondary derivations with the secondary formatives ka-, le-, ga-, go- may also be regarded as adverbs. … Adverbs, usually those of quality, are derived from adjective and relative stems by means of ga-. (Ziervogel and Mokgokong 1975: 114-115, Introduction)

Such inconsistencies, whether justified or not, have a negative effect on the learner's and/or user's information retrieval efforts. The issue here is not the validity of their views — criticism on linguistic grounds lies beyond the scope of this article. Furthermore, one should also accept that the adverb can be described from more than one angle and that progressive linguists have the academic right to change their minds. The concern lies with the learner who tries to master the nature and use of adverbs in Northern Sotho and with the lexicographer in his/her role as mediator who finds it difficult to obtain a comprehensive overview of the adverb in order to treat it satisfactorily on the macrostructural and microstructural levels in dictionaries.

Thirdly, a single glance at the treatment of adverbs in Northern Sotho dictionaries reveals far too many inconsistencies and errors. Kriel (1983) includes the lemma *ga(n)nyane* which means that the lemma could either be *ganyane* or *gannyane*. This lemma is placed in the wrong alphabetical position for either *ganyane* or *gannyane*. There is also another treated lemma *gannyane*, again in an incorrect alphabetical position. He gives *ga n'.nyane* as comment on form of *ga(n)nyane* but *ga nya.ne* as comment on form for *gannyane*. Kriel (1950) is inconsistent in respect of circumflexes and POS indication regarding adverbs. As an example of the latter, he labels *gatee* 'once' and *gararo* 'three times' as adverbs but not *gabedi* 'twice'. Ziervogel and Mokgokong (1975) lemmatise the question particles *afa*, *na* and *naa* but indicate the POS of *afa* as adverb. Incorrect alphabetical sorting of lemmas is a common problem in Kriel and Van Wyk (1989), e.g. for *gakale*, compare De Schryver and Lepota (2001, Note 6). Missing punctuation, for example a question mark at *gakakang*, and typing errors such as *by.* instead of *byw.* at *gakalo* are unfortunate. In the latter case the user can interpret the incorrectly spelt label as a translation equivalent, he/she may incorrectly conclude that *gakalo* means *by* 'at' instead of 'so many'.

Form and meaning of adverbs in Northern Sotho

A prerequisite to successful lemmatisation strategies for and treatment of adverbs, is a thorough understanding of the nature of adverbs in Northern Sotho. Poulos and Louwrens (1994: 328) say:

> The analysis of the adverb can be approached in different ways. One could, for example, classify adverbs according to whether they express the concepts of time, place, manner, etc. Or one could describe them in terms of their derivation, that is, in terms of the prefixes and/or suffixes that are used.

Louwrens (1991: 26) says "adverbs describe the nature of the action in terms of *time*, *place* and *manner*" and gives the following examples.

— *Adverbs of time:* Pula e nele *maabane* It rained yesterday

— *Adverbs of place:* Ba dutše *moriting* They are sitting in the shade

— *Adverbs of manner:* Masogana a ja *kudu* The young men eat a lot

Van Wyk et al. (1992: 118) distinguish three types of adverbs namely basic adverbs, derived adverbs and adverbs that have been adopted from other word categories.

— *Basic adverbs* refer to words that are not derived from other words or stems and solely function as adverbs. Examples include *ruri* 'really', *kudu* 'much, a lot', *bjale* 'now', *bjalo* 'like that', *kae?* 'where?', and *neng?* 'when?'.

— *Derived adverbs* are derivations by means of the prefix *ga-*, from nouns and adjectives, e.g. *gabotse* 'well', *gatee* 'once', *gabohloko* 'painful', *gašoro* 'cruelly', etc.

—   *Adopted adverbs* are words, such as nouns, which are overwhelmingly or even exclusively used as adverbs such as *maabane* 'yesterday', *bošego* 'at night', *godimo* 'above', *Tshwane* 'Pretoria', etc.

In addition to their discussion in terms of basic adverbs, derived adverbs and adopted adverbs, Van Wyk et al. (1992: 121) also mention that particle groups, such as *ka mehla* 'always', *le gatee* 'not at all', "often function as adverbs".

Poulos and Louwrens (1994: 328) describe adverbs "according to the way in which they are formed". They distinguish 9 categories, viz. adverbs formed by using the prefixes *ka-* (instrumental) (7 subcategories), *le-* (associative) (2 subcategories), *go-* (locative), *ga-* (locative), *mo-* (locative), *kua-* (locative), the suffix *-ng* (locative), the prefix *ga-* (adverbial) and word categories which may function as adverbs (without the addition of any prefixes or suffixes) (8 subcategories).

The nature of the description (time, place and manner), the 3 basic types of adverbs (basic, derived, adopted and particle groups) and the way in which they are formed will now be interlinked in two ways in Tables 1 and 2. Table 1 interlinks the categories of time, place and manner with the three basic types of adverbs that occur in Northern Sotho, and with the way in which they are formed. Table 2 is based upon the way in which adverbs are formed, thus reflecting the viewpoint of Poulos and Louwrens (1994), and interlinked with the three basic types of adverbs as well as with the categories time, place and manner.

In this way the viewpoints of all the above-mentioned authors as well as most of their examples are catered for.[5] The purpose of the compilation of Tables 1 and 2 is threefold. Firstly, either or both of these tables can assist the lexicographer in obtaining a comprehensive overview of the adverb in Northern Sotho. Secondly, these tables can be used in the back matter of a paper dictionary, or, thirdly, in pop-up information boxes in an electronic dictionary. It is for the lexicographer to decide whether he/she prefers to base the back matter entry (entries) and pop-up box(es) on say, Table 1 or Table 2 or both, or whether to use these tables as they are or to adapt them to the level of the target user of the dictionary.

Table 1: Time, place and manner linked to basic types of adverbs and the way in which they are formed (P&L = Poulos and Louwrens 1994)

Time	Place	Manner
bošego 'at night' ADOPTED P&L-9(v)	*ka* **toropong** 'in town' GROUP P&L-1(vi)	**kudu** 'very much' BASIC P&L-9(vii)
neng? 'when?' BASIC P&L-9(vii)	*Go* **Madika** 'to Madika' GROUP P&L-3	**gagolo** 'mainly' DERIVED P&L-8
lehono 'today' ADOPTED P&L- 9(v)	*Ga* **Madika** 'to/at Madika's place' GROUP P&L-4	**ruri** 'really' BASIC P&L-9(vii)

ka **Labobedi** 'on Tuesday' GROUP P&L-1(iv)	*mo* **tafoleng** 'on the table' GROUP P&L-5	*ka* **sefatanaga/lerato** 'with or by means of a car/love' GROUP P&L-1(i), (ii) and (vii)
ka **letšatši** 'per day' GROUP P&L-1(iv)	*kua* **Amerika** 'over there (far away) in America' GROUP P&L-6	*ka ga* **molato** wo 'about this problem' GROUP P&L-1(iii)
maabane 'yesterday' ADOPTED P&L-9(v)	**toropo***ng* 'in/at the town' DERIVED P&L-7	**ka fao** 'because' GROUP P&L-1(v)
nkgapela 'shortly' ADOPTED P&L- 9(v)	**godimo** 'above' ADOPTED P&L-9(iii);	*le* **Tate** *'together with* father, *to* father' GROUP P&L-2(i) and (ii)

Thus, for example, *kudu* in Table 1 is a basic adverb of manner belonging to the subcategory (vii) "basic, non-derived adverbs with an inherent adverbial meaning" within the main category 9 "word categories which may function as adverbs without the addition of any prefixes or suffixes" of Poulos and Louwrens (1994).

Table 2: The way in which adverbs are formed, linked to the basic types of adverbs and the categories time, place and manner (P&L = Poulos and Louwrens 1994)

P&L	The way in which adverbs are formed		
1	**Adverbs formed by using the prefix ka- (instrumental)**	(i) 'by means of' **ka sefatanaga** 'by car' GROUP: MANNER	(ii) 'with' **ka thipa** 'with a knife' GROUP: MANNER
		(iii) 'about' **ka ga molato** wo 'about this problem' GROUP: MANNER	(iv) Time **ka Labobedi** 'on Tuesday' GROUP: TIME
		(v) 'because of, on account of' **ka fao** 'because' GROUP: MANNER	(vi) Place **ka toropong** 'in town' GROUP: PLACE
		(vii) Miscellaneous **ka lerato** 'with love' GROUP: MANNER	
2	**Adverbs formed by using the prefix le- (associative)**	(i) 'together with' **le Tate** 'together with father' GROUP: MANNER	(ii) translating the English preposition 'to' **le Tate** 'to father' GROUP: MANNER
3	**Adverbs formed by using the prefix go- (locative)**	**Go Madika** 'to Madika' GROUP: PLACE	
4	**Adverbs formed by using the prefix ga- (locative)**	**Ga Madika** 'to/at Madika's place' GROUP: PLACE	
5	**Adverbs formed by using the prefix mo- (locative)**	**mo tafoleng** 'on the table' GROUP: PLACE	

6	**Adverbs formed by using the prefix kua-**	**kua Amerika** 'over there (far away) in America' GROUP: PLACE	
7	**Adverbs formed by using the suffix -ng (locative)**	**toropong** 'in/at town' DERIVED: PLACE	
8	**Adverbs formed by using the prefix ga- (adverbial)**	**gagolo** 'mainly' DERIVED: MANNER	
9	**Word categories which may function as adverbs — (without the addition of any prefixes or suffixes)**	(i) Place names **Tshwane** 'Pretoria' ADOPTED: PLACE	(ii) Other nouns indicative of place **mošate** 'to/at/from the chief's place' ADOPTED: PLACE
		(iii) Nouns of classes 16-18 **godimo** 'above' ADOPTED: PLACE	(iv) Demonstratives of classes 16-18 **mo** 'here' ADOPTED:PLACE
		(v) Time **Bošego** 'at night' ADOPTED: TIME	(vi) Certain possessive forms **la mathomo** 'for the first time' GROUP: TIME
		(vii) Inherent adverbial meaning **kudu** 'very much' BASIC: MANNER	(viii) Either conjunctions or adverbs **fela** adverb 'merely' BASIC: MANNER

Given the presentations of the different linguists of adverbs in Northern Sotho, as well as Tables 1 and 2, it is for the lexicographer to decide on the best angle of approach for lemmatisation of these adverbs. He/she can decide to approach the lemmatisation of adverbs departing from the way in which they are formed, or from the basic types of adverbs or even in terms of their function. Whatever the preferred angle might be, sound decisions regarding lemmatisation, treatment in the microstucture, utilization of the mediostructure, and treatment in the user's guide and back matter have to be taken. In this article, lemmatisation will be attempted on the basic types of adverbs.

Lemmatising basic adverbs

Basic adverbs, in terms of Van Wyk et al. (1992), Louwrens (1991), Ziervogel and Mokgokong (1975), Lombard (1985), as well as Poulos and Louwrens' (1994) Categories 9(vii) and 9(viii), are less problematic. Only a limited number of basic adverbs exist in Northern Sotho and since they are all frequently used, they should all be lemmatised. Consider all the basic adverbs listed by Van Wyk et al. (1992: 118), Louwrens (1991: 26), Ziervogel and Mokgokong (1975: 114), Poulos and Louwrens (1994) and Lombard (1985: 166) and their respective frequency counts in the 6.1 million-word Pretoria Sepedi Corpus, henceforth simply referred to as "the corpus".

Table 3: Overall frequencies of *basic adverbs* in the corpus

Word	Freq.	Word	Freq.	Word	Freq.	Word	Freq.	Word	Freq.
kudu	4,996	ntshe	2,566	kudukudu	285	fela	17,337	bjale	9,397
ruri	3,582	bjalo	15,468	ruriruri	206	neng?	1,794	kae?	5,222

In addition to information on frequency, corpus lines and information on collocates obtained from the corpus can be very useful to the lexicographer. Compare for example information on the most frequent collocates of *ruri* in Table 4.

Table 4: Collocates of *ruri* 'really'

	L3	L2	L1		R1	R2	R3
sa			357	**ruri**			
e		376		**ruri**			
le		139	330	**ruri**			

From this extract from the collocates table for *ruri* it is clear that the possessive concord, classes 7/8, *sa* occurs very frequently one position to the left (L1) of *ruri*, thus 357 occurrences of *sa ruri*. Likewise, the copulative stem *-le*, and in fact the entire copulative verb *e le*, are indicated as frequent collocates of *ruri* in the positions L1 and L2 respectively. *Sa ruri* and *e le ruri* are therefore prime candidates for inclusion in the microstructural treatment of *ruri*. Let there furthermore be no doubt that the corpus is a most valuable source for, among others, sense distinction, typical examples, collocations, decisions on inclusion in or omission from the dictionary. Compare De Schryver and Prinsloo (2000, 2000a and 2000b) for an exhaustive overview of corpus compilation and corpus utilization on macro- and microstructural levels.


Lemmatising derived adverbs

Since the number of derived adverbs is unlimited or open ended, it is not possible to lemmatise all forms separately. From a lexicographic angle, a number of issues are at stake here. Firstly, there is a need for selection in the case of such paradigms. Secondly, the lexicographer has to take decisions in terms of paradigm completion. Thirdly, the lexicographer has to consider certain affixes and particles/prepositions for inclusion in the macrostructure, i.e. as lemmas in their own right. The following analysis is a typical example of how such instances should be approached.

Consider firstly the open-ended paradigm *gatee* 'once', *gabedi* 'twice', … *galesome* 'ten times', … *galekgolo* 'hundred times', … *gadiketekete* 'thousands of times' … . The lexicographer in his/her role as mediator has to take certain decisions, e.g. in respect of inclusion into or omission from the dictionary, after having studied corpus data and available dictionaries.

Table 5: Overall frequencies of the numeral paradigm *gatee, gabedi, … gadimi-lione* in the corpus

		Freq.	NEnSeD (1950)	GrNSoW (1975)	Pukuntšu (1983)	Pukuntšu (1989)	Sediba (1992)	SeDiPro (2000)
gatee	once	799	yes	yes	yes	yes	yes	yes
gabedi	twice	438	yes	yes	yes	yes	no	yes
gararo	three times	224	yes	yes	yes	yes	yes	yes
gane	four times	7	yes	no	yes	yes	no	yes
gahlano	five times	7	yes	yes	yes	yes	no	yes
gaselela	six times	3	no	no	yes	yes	no	yes
gašupa	seven times	7 (40)[6]	yes	no	yes	yes	no	yes
gaseswai	eight times	4	no	no	yes	yes	no	yes
gasenyane	nine times	1	no	no	no	no	no	yes
galesome	ten times	13	no	yes	no	no	no	yes
gamasome	tens of times	8	no	no	no	no	no	no
gamasomesome	multiple/several tens of times	5	no	no	no	no	no	no
galekgolo	a hundred times	7	no	no	yes	yes	no	yes
gamakgolo	hundreds of times	1	no	no	no	no	no	no
gamakgolokgolo	multiple/several hundreds of times	15	no	no	no	no	no	no
gasekete	a thousand times	4	no	no	no	no	no	no
gadiketekete	multiple/several thousands of times	1	no	no	no	no	no	no
gadimilione	millions of times	0	no	no	no	no	no	no

Figure 1: Overall frequencies of the numeral paradigm *gatee, gabedi, … gadimi-lione* in the corpus
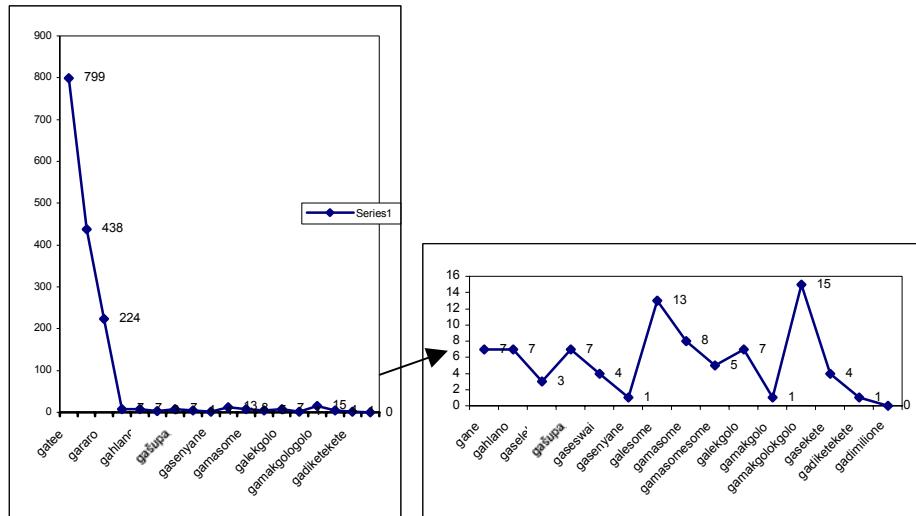
The frequency counts reveal a rather interesting pattern. From Table 5 and Figure 1 it is clear that, from a frequency angle, *once*, *twice* and *three times* are much

more frequently used than *four times* up to *nine times* with relative frequency for "rounded off" numerals such as *ten times* and *hundreds of times*. Treatment in existing dictionaries indicates that the compilers did fairly well on intuition but did miss out on frequently used items such as especially Sediba (Lombard et al. 1992) for *twice*, *four times* and *five times*, and NEnSeD (Kriel 1950), Pukuntšu (Kriel 1983), Pukuntšu (Kriel and Van Wyk 1989) as well as Sediba for *ten times*.

As far as the principle completing-a-paradigm is concerned, two strategies are suggested. Firstly, the lexicographer could complete the 1-to-10 paradigm by also entering *four times* up to *ten times* as separate lemmas, although in terms of frequency counts, this cannot wholly be justified. Secondly, the rest of the open-ended paradigm could be addressed by lemmatising the outstanding "beacons" such as *ten times*, *hundreds of times*, *thousands of times*, etc. Guidance in respect of the paradigm as a whole could be done by appropriate cross-referencing to the back matter. The back matter section would then explain the normal (rather complicated) numerical system of Northern Sotho for expressing numbers from say 1 to 10 and 11 up to 10 000 000 and/or contain references to grammar books where this system is described. Thirdly, the prefix *ga-* (used to derive these adverbs) should be entered as a separate lemma, cf. (1). Compare Gouws (1989: 84) for the importance of lemmatising elements bigger than words and also elements smaller than words.

(1) **ga-** *adv prefix* **gatee** 'once' < **tee** 'one' *Ba mmethile gatee fela They hit him only once*. **gabotse** 'well' < **botse** 'lovely' *Sepela gabotse! Go well!* **gantši** < **-ntši; gammogo < -mmogo; gagolo < -golo; gabedi < -bedi** ► BM 2.8 [7]

This suggested entry not only caters for the numerical paradigm but also covers the other most frequent typical adverbs formed by means of this derivation strategy, cf. Poulos and Louwrens' Category 8, either as a treated sublemma in the case of *gabotse* or as untreated sublemmas such as *gantši*, *gammogo* and *gagolo*. This brings us to another open-ended paradigm, namely *all adverbs* derived by means of the adverbial prefix *ga-* (of which the numerals just discussed, are only a subsection). Once again the lexicographer has to find a strategy for inclusion or omission. Consider the most frequently used adverbs in this broader category.

Table 6: The most frequently used adverbs derived by means of the prefix *ga-*

		NEnSeD (1950)	Pukuntšu (1983)	Pukuntšu (1989)	Sediba (1992)	SeDiPro (2000)	
gabotse	4,905	yes	yes	yes	yes	yes	nicely, well, carefully
gantši	1,265	yes	yes	yes	yes	yes	often, frequently
gammogo	1,229	yes	yes	yes	yes	yes	together (with), simultaneously
gagolo	804	yes	yes	yes	no	yes	greatly; especially; mostly
gatee	799	yes	yes	yes	yes	yes	once
gannyane	466	no	yes	yes	yes	yes	little, sparingly
gabedi	438	yes	yes	yes	no	yes	twice
gabotsebotse	315	no	yes	no	no	yes	very well, clearly

gararo	224	yes	yes	yes	yes	yes	three times, thrice
gabonolo	235	no	yes	yes	no	yes	easily, with ease
gampe	112	yes	yes	yes	no	yes	badly
gabohloko	71	yes	yes	yes	no	yes	painfully
gabotsana	67	no	yes	no	no	no	somewhat beautiful/nice
gagologolo	51	no	yes	no	yes	yes	especially, chiefly

From this table it is clear that NEnSeD (Kriel 1950) missed out on very frequently used adverbs such as *gannyane*, *gabotsebotse*, *gabonolo*, Sediba (Lombard et al. 1992) on *gagolo*, etc. The high frequency counts for *gabotsebotse* and *gabotsana* furthermore urge the lexicographer to venture beyond the boundaries of the "basic word", viz. those consisting of a prefix and a stem, and also to consider forms with *reduplicated* stems and *diminutive* forms for lemmatising and not merely the basic forms ga + stem.

The next step is to look into *all other* derived adverbs especially Poulos and Louwrens' Categories 1-7 and 9(vi).

Here, each of the particles *ka*, *le*, *go*, *ga*, *mo* and *kua* as well as the suffix *-ng* should be lemmatised with *elaborate* attention in the microstructure of each article to its function as initiator of adverbial groups. For example, one should attempt to cover all of the P&L categories 1(i) to 1(vii) in the treatment of the lemma *ka*. The lexicographer should, preferable in the user's guide, take a clear stand on the use of the terms *preposition* versus *prefix* versus *particle*, and should not burden the user with grammatical labels such as *prep./pref./part.*, cf. (2).

(2)   **ka** part. [intr. adv. phrases], **o sepela ka sefatanaga** she goes by car; **ka Labobedi** on Tuesday; **ka toropong** in town. ► BM 1.1-1.3

As in the case of the prefix *ga-* above, the lexicographer should not hesitate to lemmatise the locative suffix *-ng* as an article in its own right.

Poulos and Louwrens' Categories 1(iii), 1(v) and 9(vi) also require special attention. Here the lexicographer should be prepared to lemmatise multiword lemmas such as *ka ga*, *la mathomo*, 'beginning' *la bobedi* 'for the second time' and even, not mentioned by Poulos and Louwrens, *ka mo*, *ka kua*, etc. Furthermore, in the case of *la bobedi* for example, appropriate cross-references should be given to *Labobedi* 'Tuesday' and *bobedi* 'second'. Consider their frequencies in the corpus:

Table 7: Frequencies of multiword adverbs that are candidates for lemmatisation

Lemma candidate	Freq.		Lemma candidate	Freq.		Lemma candidate	Freq.
ka ga	3,709		la bobedi	274		ka kua	851
la mathomo	487		ka mo	4,795			

(3)   **la bobedi** *adv*. for the second time, secondly; ► **Labobedi, bobedi**

(4)   **la mathomo** *adv*. for the first time, firstly; ► **mathomo**

It should be reiterated that the lexicographer should also and always use the corpus as an invaluable aid to the lexicographic treatment of all types of adverbs of which the study of concordance lines like those in Table 8 generated for the adverb *galesome* 'ten times', is a typical example.

Table 8: Concordance lines for *galesome* 'ten times'

Tataweno yena o mphorile, **a fetoša moputso wa ka**	**galesome**	fela, Modimo o mo thibetše go ntira bošul
apa le gorogile maabane. **Malome o bethile Lesibana**	**galesome**	. Naa mošomo wa mantšu ao a ngwadilwe
ditaba tšeo ke tša Chabalala. **O ile a ingwaya hlogo**	**galesome**	, a hloka karabo, fela a tlelwa ke kgopolo y
kilwe maabane. BoMakotlo ba tlile. **Ke tlo go lebalela**	**galesome**	fela. Modimo o tlo go lebalela gamasome a

Such concordance lines are the ideal point of departure for microstructural treatment (cf. De Schryver and Prinsloo (2000b) for a detailed discussion).

Lemmatising adopted adverbs

In the case of *adopted* adverbs, the lexicographer is once again confronted by limited or even open-ended *paradigms* but also with difficult decisions regarding the functions as adverbs versus nouns, especially in terms of part-of-speech indication. Firstly a number of paradigms, this time mostly on a semantic level, have to be dealt with such as *lehono : maabane : maloba*, 'today : yesterday : the day before yesterday', *fase : godimo : morago* 'below : above : behind', *leboa : borwa : bohlabela : bodikela* 'north : south : east : west', etc. Frequency of use and the obligation to complete such semantic paradigms should be the norm.

Lemmatising nouns that are often or even exclusively used as adverbs, *twice*, once with POS-label *adverb* and again with POS-label *noun*, will be totally redundant. In the microstructural treatment, lexicographers often opt for indicating the POS in such cases as *noun* with no reference to a possible adverbial function. Neglecting the POS *adverb* in this way can however only be tolerated up to a point where the labelling of adverbs as nouns becomes artificial and questionable, especially in those cases where nouns are exclusively used as adverbs. The question here is whether the part of speech of nouns that are exclusively used as adverbs should be indicated as *noun*, *adverb* or *both*. What should definitely be avoided is a situation where the same adverb is labelled differently in different dictionaries, or even in different editions of the same dictionary, or where clearly "related" adverbs (i.e. belonging to the same paradigm), are labelled differently in the same dictionary. Consider the treatment of the three words listed by Lombard (1985: 167) as *adverbs that developed from class 6 nouns*, i.e. *maabane*, *maloba* and *mantšiboa*, as a case in point.

(5)   (a) *Pukuntšu* (1989)
              maabane, byw. … gister; *ka maabane*, in die aand …
       (b) *New English–Sesotho Dictionary* (1950)
              maabane, adv., yesterday …

(6)   (a)   *Pukuntšu* (1989)
              malôba, snw. kl 6, … eergister, die ander dag …
        (b)   *New English–Sesotho Dictionary* (1950)
              malôba, adv., the day before yesterday …

(7)   (a)   *Pukuntšu* (1989)
              mantšiboa, snw./byw. … aand, in die aand, teen sononder, saans.
        (b)   *New English–Sesotho Dictionary* (1950)
              mantšiboa, n., evening, in the evening, …
        (c)   *New English–Northern Sotho Dictionary* (1967)
              mantšiboa, adv., n., evening, in the evening, …

All these dictionaries offer a single entry for each of these words. In (5) both dictionaries label *maabane* as an adverb, in (6)(a) *maloba* is labelled as a noun with no separate entry or reference whatsoever to *adverb* but in (6)(b) as an adverb. In (7)(a) a single entry is given for *mantšiboa* but with *dual* labelling of its function. In (7)(b), in contrast to (5)(b) and (6)(b), the lemma is now labelled as a *noun* but in a later edition of the same dictionary, i.e. (7)(c), also labelled as an adverb.

Different options can be considered here. The lexicographer could simply ignore the overwhelming or even exclusive function of such nouns as adverbs and consistently label them as nouns (coupled with an explanation in the user's guide and/or back matter of the dictionary) as in (6)(a) and (7)(b). Alternatively, the lexicographer could decide to label the POS in cases where nouns are exclusively used as adverbs, as in (5) and (6)(b) or even in addition to the label *noun*, as in (7)(c). A third possibility, which would represent a sound application of the metalanguage could be to order the POS-labels according to the dominant function, i.e. *n./adv.* if the nominal function is more frequent or *adv./n.* if the word is more frequently used as an adverb. This has to be clearly explained in the front matter of the dictionary. The dominant function can be determined on the basis of frequency counts in the corpus.

Electronic dictionaries

Generally speaking, many more options are available to the lexicographer in electronic dictionaries and fewer restrictions exist in terms of access, available space, mediostructure, etc. See Prinsloo (2001) and De Schryver (2003) for detailed discussions of electronic dictionaries. For example, pop-up screens alone can instantly provide the user with a wealth of information on various aspects of adverbs. This could for instance be done as shown in (8) by simply momentarily resting the cursor on the label *adverb*. Note that all this information, brought together in an instant, also narrows the gap between dictionary and grammar, which is generally believed to be "unbridgeable" (cf. Geeraerts 2000: 77).

> ***Basic adverbs*** refer to words such as *ruri* 'really', *kudu* 'much, a lot', *bjale* 'now', *bjalo* 'like that', *kae?* 'where?', and *neng?* 'when?'.
> ***Derived adverbs*** are derivations by means of the prefix *ga-*, e.g. *gabotse* 'well', *gatee* 'once', *gabohloko* 'painful', *gašoro* 'cruelly', etc.
> ***Adopted adverbs*** are words, such as nouns, which are overwhelmingly or even exclusively used as adverbs such as *maabane* 'yesterday', *bošego* 'at night', *godimo* 'above', *Tshwane* 'Pretoria', etc.

(8) **ga-** *adv prefix* **gatee** 'once' **< tee** 'one'

Time	Place	Manner
bošego 'at night' ADOPTED	**ka toropong** 'in town' GROUP	**kudu** 'very much' BASIC
neng? 'when?' BASIC	**Go Madika** 'to Madika' GROUP	**gagolo** 'mainly' DERIVED
lehono 'today' ADOPTED	**Ga Madika** 'to/at Madika's place' GROUP	**ruri** 'really' BASIC
ka Labobedi 'on Tuesday' GROUP	**mo tafoleng** 'on the table' GROUP	**ka sefatanaga/lerato** 'with, by means of a car/love' GROUP
ka letšatši 'per day' GROUP	**kua Amerika** 'over there (far away) in America' GROUP	**ka ga molato wo** 'about this problem' GROUP
maabane 'yesterday' ADOPTED	**toropong** 'in/at town' DERIVED	**ka fao** 'because' GROUP
nkgapela 'shortly' ADOPTED	**godimo** 'above' ADOPTED	**le Tate** 'together with father, to father' GROUP

Conclusion

Compiling user-friendly dictionaries of a high lexicographic standard for African languages poses a great challenge to prospective lexicographers. They often are the mediators between complicated grammatical structures and the decoding and encoding needs of their target users. Adverbs should not be lemmatised haphazardly as they cross the compiler's way. They should be carefully researched and lemmatised in a structured way. Lexicographers should be aware of the fact that different subcategories of the same phenomenon might require different lexicographic treatments as in the case of basic adverbs versus derived adverbs versus adopted adverbs. This is even true for subcategories within a given category such as the various approaches required for different categories of adverbs derived by *ga-*. 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 the needs of both encoding and decoding users are met and the medio-structure should be maximally utilized. The ultimate aim should be to ensure an unimpeded information retrieval process in respect of

— easy access to the lemma,

— successful information retrieval in the microstructure,

— added value obtained in following up on cross-references,

— useful guidance from the user's guide in the front matter,

— a comprehensive overview of adverbs in the back matter, and

— appropriate references to external sources such as grammar books.

Notes

1. An estimated 80% of freelance lexicographers and lexicographers employed by the National Lexicography Units in South Africa have little or very limited lexicographic experience.
2. Na beraming het 80% van alle vryskut leksikograwe en leksikograwe in diens van die Nasionale Leksikografie eenhede min of beperkte ervaring van leksikografie.
3. "Word boundaries" should here be interpreted as for orthographic words.
4. The term "adopted adverb" is used in terms of Van Wyk et al. (1992) in this article and should not be interpreted in the more general sense of *adopted* 'borrowed from another language'.
5. Note that the status of ideophones as adverbs is not recognized in these classifications and requires further research. Compare also Poulos and Louwrens (1994: 351).
6. If Christian religious data in the corpus are taken into account.
7. The symbol ► is a reference marker referring the user to the reference address which in this case is Section 2.8 in the back matter (BM).

References

De Schryver, G.-M. 2003. Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography* 16(2): 143-199.

De Schryver, G.-M. and B. Lepota. 2001. The Lexicographic Treatment of Days in Sepedi, or When Mother-Tongue Intuition Fails. *Lexikos* 11: 1-37.

De Schryver, G.-M. and D.J. Prinsloo. 2000. The Compilation of Electronic Corpora, with Special Reference to the African Languages. *Southern African Linguistics and Applied Language Studies* 18(1-4): 89-106.

De Schryver, G.-M. and D.J. Prinsloo. 2000a. Electronic Corpora as a Basis for the Compilation of African-language Dictionaries, Part 1: The Macrostructure. *South African Journal of African Languages* 20(4): 290-309.

De Schryver, G.-M. and D.J. Prinsloo. 2000b. Electronic Corpora as a Basis for the Compilation of African-language Dictionaries, Part 2: The Microstructure. *South African Journal of African Languages* 20(4): 310-330.

Geeraerts, D. 2000. Adding Electronic Value: The Electronic Version of the *Grote Van Dale*. Heid, Ulrich et al. (Eds.). *Proceedings of the Ninth EURALEX International Congress. EURALEX 2000. Stuttgart, Germany, August 8th–12th, 2000*: 75-84. Stuttgart: Stuttgart University.

Gouws, R.H. 1989. *Leksikografie*. Pretoria: Academica.

Gouws, R.H. 2000. Toward the Formulation of a Metalexicographic Founded Model for National Lexicography Units in South Africa. Wiegand, H.E. (Ed.). 2000. *Wörterbücher in der Diskussion IV. Vorträge aus dem Heidelberger Lexikographischen Kolloquium*: 109-133. Tübingen: Max Nie-meyer.

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/New York: Routledge.

Kriel, T.J. 1950. *The New English–Sesotho Dictionary*. Johannesburg: APB.

Kriel, T.J. 1967. *The New English–Northern Sotho Dictionary*. Johannesburg: Educum.

Kriel, T.J. 1983. *Pukuntšu Woordeboek*. Pretoria: J.L. van Schaik.

Kriel, T.J. and E.B. van Wyk. 1989. *Pukuntšu Woordeboek, Noord-Sotho–Afrikaans, Afrikaans–Noord-Sotho*. Pretoria: J.L. van Schaik.

Lombard, D.P. 1985. *Introduction to the Grammar of Northern Sotho*. Pretoria: J.L. van Schaik.

Lombard, D.P., R. Barnard and G.M.M. Grobler. 1992. *Sediba, Practical List of Words and Expressions in Northern Sotho, Northern Sotho–Afrikaans–English, English–Northern Sotho*. Pretoria: Via Afrika.

Louwrens, L.J. 1991. *Aspects of Northern Sotho Grammar*. Pretoria: Via Afrika.

Prinsloo, D.J. 2001. The Compilation of Electronic Dictionaries for the African Languages. *Lexikos* 11: 139-159.

Prinsloo, D.J. 2002. The Lemmatization of Copulatives in Northern Sotho. *Lexikos* 12: 21-43.

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 G.-M. de Schryver (Eds.). 2000. *SeDiPro 1.0, First Parallel Dictionary Sepêdi–English*. Pretoria: University of Pretoria.

Prinsloo, D.J. and Rufus 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.

Poulos, G. and L.J. Louwrens. 1994. *A Linguistic Analysis of Northern Sotho*. Pretoria: Via Afrika.

Procter, Paul. (Ed.). 1995. *Cambridge International Dictionary of English*. Cambridge: Cambridge University Press.

Rundell, M. (Ed.). 2002. *Macmillan English Dictionary for Advanced Learners*. Oxford: Macmillan.

Sinclair, J.M. (Ed.). 1995. *Collins COBUILD English Dictionary*. London: HarperCollins.

Van Wyk, E.B., P.S. Groenewald, D.J. Prinsloo, J.H.M Kock and E. Taljard. 1992. *Northern Sotho for First-Years*. Pretoria: J.L. van Schaik.

Ziervogel, D. and P.C. Mokgokong. 1975. *Groot Noord-Sotho Woordeboek*. Pretoria: J.L. van Schaik.

On the Lexicographical Description of Equivalent Open Class Expressions[*]

Herbert Ernst Wiegand, *Germanistisches Seminar, University of Heidelberg, Heidelberg, Germany (herbert.ernst.wiegand@gs.uni-heidelberg.de)*

Abstract: The contribution proceeds from the assumption that linguistic equivalence is already determined, and therefore only deals with questions of the lexicographical description of equivalent open class expressions in completely condensed dictionary articles in bilingual printed dictionaries, with the purpose of establishing a notion of lexicographical equivalence. Firstly, the most important differences between non-condensed, partially condensed and completely condensed dictionary articles are explained. In completely condensed dictionary articles, which are not natural language texts, addressing relationships are present. These relationships have a vital role in the lexicographical description of equivalent open class expressions. They are, therefore, explained in more detail. A further assumption for the development of a notion of lexicographical equivalence is the notion of semantic-pragmatic equivalence, which is therefore introduced briefly. Next, there is a stepwise development, by means of examples, of what one can understand by the notion of lexicographical equivalence. The developed notion of lexicographical equivalence can on the one hand form the point of departure for the lexicographical textualisation of semantic-pragmatic equivalence, and on the other hand it takes into consideration that the linguistic equivalence relationship has to be inferred first by the user-in-action by means of complex results of lexicographical textualisation. The developed notion of lexicographical equivalence presents a firm foundation for the optimisation of the presentation of lexicographical equivalence.

Keywords: ADDRESSING, SEMANTIC-PRAGMATIC EQUIVALENCE, LEXICOGRAPHICAL EQUIVALENCE, LEXICOGRAPHICAL TEXT CONDENSATION, CONDENSED DICTIONARY ARTICLES

Abstrakt: Zur lexikographischen Beschreibung nennlexikalischer äquivalenter Wortschatzeinheiten. Der Beitrag setzt voraus, dass die sprachliche Äquivalenz bereits ermittelt ist und behandelt entsprechend nur Fragen der lexikographischen Beschreibung äquivalenter nennlexikalischer Wortschatzeinheiten in vollständig kondensierten Wörterbuchartikeln zweisprachiger Printwörterbücher mit dem Ziel, einen Begriff von lexikographischer Äquivalenz zu etablieren. Zuerst werden die wichtigsten Unterschiede zwischen nichtkondensierten, partiell und vollständig kondensierten Wörterbuchartikeln erläutert. In vollständig kondensierten Wörterbuchartikeln, die keine natürlichsprachlichen Texte sind, treten Adressierungsbeziehungen auf. Diese spielen für die lexikographische Beschreibung äquivalenter Wortschatzeinheiten eine

---

zentrale Rolle. Sie werden daher genauer erklärt. Eine weitere Voraussetzung für die Entwicklung eines Begriffs der lexikographischen Äquivalenz ist der Begriff der semantisch-pragmatischen Äquivalenz, der deshalb kurz eingeführt wird. Danach wird schrittweise an Beispielen entwickelt, was unter lexikographischer Äquivalenz zu verstehen ist. Der entwickelte Begriff der lexikographischen Äquivalenz kann einerseits den Ausgangspunkt bilden für die lexikographische Vertextung der semantisch-pragmatischen Äquivalenz, und andererseits berücksichtigt er, dass die sprachlichen Äquivalenzbeziehungen vom Benutzer-in-actu anhand der komplexen Ergebnisse der lexikographischen Vertextung erst erschlossen werden müssen. Der entwickelte Begriff der lexikographischen Äquivalenz bietet eine stabile Grundlage für die Optimierung der lexikographischen Äquivalentpräsentation.

Schlagworte: ADRESSIERUNG, SEMANTISCH-PRAGMATISCHE ÄQUIVALENZ, LEXIKOGRAPHISCHE ÄQUIVALENZ, LEXIKOGRAPHISCHE TEXTVERDICHTUNG, KONDENSIERTER WÖRTERBUCHARTIKEL

1.   Some assumptions from lexicographical theory

Firstly, some assumptions from lexicographical theory, which have been developed in more recent works (cf. Wiegand 2002, 2002a, 2002b, 2002c and 2003) and which are necessary in order to understand the following expositions, are briefly explained or at least referred to. This article does not deal with genuine linguistic aspects of equivalence. On the contrary, it is taken for granted that the equivalence of the open class expressions is already determined linguistically, so that — according to the title of the article — specific questions of the lexicographical description of equivalent open class expressions (and here the topic is being narrowed) in completely condensed dictionary articles in bilingual printed dictionaries are dealt with. Therefore, some explanations will be given of the type of completely condensed dictionary articles and of the textual features of articles of this article type, preferred in modern bilingual lexicography.

Fig. 1-1 shows two non-condensed dictionary articles (wa); wa_1 and wa_2 are only of interest in view of the present context because they contain lexicographical texts in natural language which exhibit completely well-marked sentence and textual syntactic structures in accordance with the syntax of a specific natural language, so that they can as such be read continuously, in the same way as other natural language texts for which only linguistic competence and not any specific competence for dictionary use is needed.

wa_1: **Plutokratie**
Zu den beliebten Diffamierungen demokratischer Staatsformen, insbesondere der angelsächsischen Demokratien durch die NS-Propaganda, zählte der Begriff *Plutokratie* als »Herrschaft der Reichen«.

wa_2: *Action*
gehört zu den Lieblingsvokabeln der Jungendszene und ist zugleich Inbegriff eines erfüllten Daseins. Interessant ist es in erster Linie da, wo voll Action ist oder wo man Action machen kann – da ist eben auch Leben. Das Wort steht für eine Betriebsamkeit, die vor allem durch Spontaneität und Kreativität gekennzeichnet ist.

Fig. 1-1:   Non-condensed dictionary articles wa_1 and wa_2 from Schlosser 2000 and Müller-Thurau 1983

Articles wa$_1$ and wa$_2$ do not exhibit any items, microstructures or addressing structures. These articles contain item texts. Item texts, which do not always have to be complete dictionary articles, contain at least one complete sentence (cf. Wiegand 2000d and 2003); one should note here that item texts are lexico-graphical texts containing language-reflexive descriptions on a meta-level of the subject of the particular dictionary domain (in the sense of Wiegand 1998: 302) (in the second way of using the language treated in the dictionary, in the sense of Wiegand 1983: 416). That means that vouched example items which mention a sentence or more than one sentence, as in "*Die Glocke ruft […] ent-rückt*" in dictionary article wa$_3$ in Fig. 1-2 are not item texts.

wa$_3$:   **Blutstuhl**
Stuhl für einen Hinzurichtenden
Die Glocke ruft, das Stäbchen bricht. / Wie
sie mich binden und packen! / Zum
Blutstuhl bin ich schon entrückt.
Faust I / Kerker / Vers 4590 ff.
HA 3, 144

Fig. 1-2:   Completely condensed dictionary article wa$_3$ from Müller 1999

Completely condensed dictionary articles do not exhibit item texts. The bilin-gual articles wa$_4$ and wa$_5$ in Fig. 1-3 are also completely condensed dictionary articles. They can be clearly distinguished from wa$_3$, because they do not ex-hibit a micro-architecture and accordingly also no architectonically phased-out microstructures. Therefore wa$_4$ and wa$_5$ are compressed to a greater extent than wa$_3$, which considerably complicates the perception of the text form.[1]

wa$_4$:   **meat** [mi:t] *s* Fleisch *n* ⟨cold ~ kalter Braten; assorted cold ~s Aufschnitt *m*; chilled / frozen ~ Gefrierfleisch *n*; pre-served ~ Fleischkonserve *f* ⟩ | (Frucht-) Fleisch *n* | *arch dial* Speise *f* ⟨~ and drink Speise und Trank; after ~ nach dem Essen⟩ | *übertr* Genuß *m*, Vergnügen *n* ⟨this is your ~ das wird Ihnen gefallen; to be ~ and drink to s.o. jmdm. größtes Vergnügen bereiten; one man's ~ is another man's poison des einen Tod ist des anderen Brot⟩ | *übertr* innerer Gehalt, Substanz *f* ⟨a book full of ~ ein gehaltvol-les Buch⟩ ◇ ~ **and potatoes 1.** *s Sl* entscheidende Grund-lage, das, worauf es ankommt; **2.** *adj* entscheidend ⟨≈ in-formation⟩; ˡ~ **ball** *s* Fleischklößchen *n*, Frikadelle *f* ; ˡ~ ˌ**chopper** *s* Hackmesser *n*, Fleischwolf *m*; ˡ~**ed** *adj* flei-schig ⟨well- ≈ reich an Fleisch; nahrhaft⟩; ˡ~ **fly** *s* Schmeiß-fliege *f* ; ˡ~**less** *adj* fleischlos ⟨≈ days fleischfreie Tage *m/ pl*⟩; ˡ~ **pie** *s* Fleischpastete *f* ; ˡ~ ˌ**plat·ter** *s* Bratenplatte *f*, -teller *m* | kalter Braten, Bratenplatte *f* ; ˡ~ ˈ**tea** *s selten* frü-hes Abendessen (kalte Platten und Tee); ˡ~**y** *adj* fleischig | *übertr* gehaltvoll, markig, kräftig

wa$_5$:   **stumpf** [ʃtʊmpf] *adj* **1.** (*nicht scharf*) desa-filado; (*nicht spitz*) romo **2.** (*glanzlos*) opa-co **3.** (MATH: *Winkel*) obtuso **4.** (*teilnahms-los*) apático

Fig. 1-3:   Completely condensed dictionary articles wa$_4$ and wa$_5$ from Neubert and Gröger 1991 and from DE. Dt.–Span./Es.–Al. 1999.

Partially condensed dictionary articles exhibit, apart from items, at least one

item text and thus microstructures which are heterogeneous with regard to their elements (in the sense of Wiegand 2003). An example is wa_6 in Fig. 1-4. The item text appears between the asterisk and the double asterisk.

wa_6:   **Parteidisziplin,** die / Schon vor 1933 / : Die für jedes Mitglied stets verbindliche Pflicht, alle Parteibeschlüsse zu befolgen. ∗ Die P. ist eine der wesentlichen Normen für alle Mitglieder und → Kandidaten der Partei. Die Einhaltung der P. wird von der → Parteikontrollkommission überwacht, Verletzungen der P. werden mit → Parteistrafen belegt. ∗∗ Kein Plural.

Fig. 1-4:   Partially condensed dictionary article wa_6 from Kinne and Strube-Edelmann 1980

After the exemplified explanations by means of wa_1 to wa_6, it is possible to determine the typological location of completely condensed dictionary articles in a typological system which has in the meantime been drawn up as shown in Fig. 1-5 (cf. Wiegand 2002d: 510 and 2003).
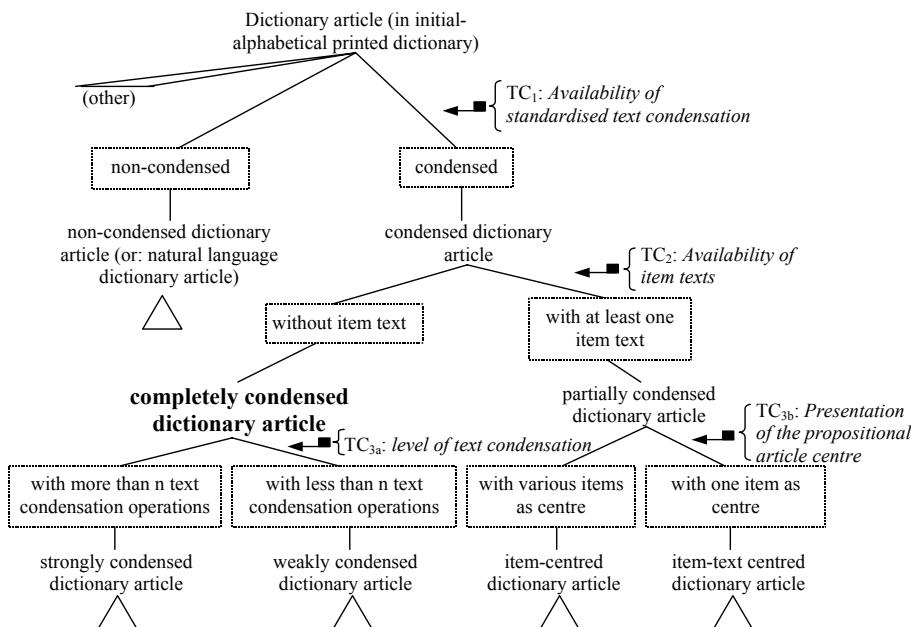
Fig. 1-5:   Excerpt from a typological system for dictionary articles in initial-alphabetical printed dictionaries; *Abbreviations and notation conventions*: TC = typology criterion; "◄▪" means *the application of TC leads to the subdivision*

As opposed to non-condensed dictionary articles, completely condensed dictionary articles such as wa₃, wa₄ and wa₅ are not natural language texts, because they do not exhibit natural language syntax. The syntactical relationships and therefore the cohesion-forming phenomena are omitted in the process of inner text condensation (in the sense of Wiegand 1998: 13ff and 2002: 126ff). They are replaced by the article-internal addressing relationships which contain items indicating their reference addresses. The non-natural article syntax of a completely condensed dictionary article is not only recognizable on the basis of linguistic competence; competence in the use of dictionaries is also necessary. When designing dictionary articles, one should always proceed from the assumption that a user may not be in a position to relate the text constituents of a dictionary article to each other on the basis of semantic features of the lexicographically treated linguistic expressions. It should rather be obvious to a competent user (in the sense of Wiegand 1998: 506) to which reference addresses an item is addressed by means of the features of the article form. If that is not the case, the textual prerequisites ensuring that a user-in-action can infer lexicographical information by means of the article texts and thereby systematically obtain lacking knowledge, are not provided. For it is necessary for successful information retrieval in the process of user actions that the user-in-action, who does not know the dictionary subject sufficiently, can link the items on the basis of features of the article form to their reference addresses.

The article-internal addressing also plays a vital role in the lexicographical description of equivalent open class expressions. Therefore, some more introductory explanations will be given by means of wa₇ in Fig. 1-6, which especially should help one not to confuse the addressing relationship with other relationships or, as recently happened in Engelberg and Lemnitzer (2001: 137ff; compare with Wiegand 2002), misunderstand it.

wa₇: **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.** (ˡ~ded, ˡ~ded) *vt* (jmdn.) ins Bett legen, zu Bett bringen | betten (*auch übertr*) | *Tech* betten, einlegen, einmörteln, festlegen (**in** in); ~ **down** (Pferd) in der 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

Fig. 1-6:   Completely condensed dictionary article wa₇ from Neubert and Gröger 1991

The following explanations are illustrated by means of wa₇ in Fig 1-7. In wa₇, the target language item giving the word equivalent (ÄA.W) "Bett" is addressed to the item giving the form of the lemma sign (LZGA) "**bed**"[2]; it is a case of non-adjacent left addressing that is three times expanded. The textual distance from an item to its reference address has the value 3, because three elementary items are placed between the reference address "**bed**" and the addressed item "Bett", namely: the item giving the pronunciation (AusA) "bed", the item giving the word class distinction (WUntA) "**1.**" and the item giving the word class (WAA) "*s*" (cf. Wiegand 2002: 144ff). The addressing relationship (cf. (1) in Fig. 1-7) should not be confused with the item relationship (cf. (4) in Fig. 1-7). For the item giving the word equivalent "Bett" is not by chance an item for another item, namely for the item giving the form of the lemma sign "**bed**", it is rather an item to the lemma sign *bed*, which is mentioned with "**bed**". The item giving the form of the lemma sign belongs to the items giving the form(s), with which something else is thus given, namely that at least one linguistic form is mentioned (cf. the relationship of mentioning (3) in Fig. 1-7). The converse addressing relationship is called *reference addressing relationship* (cf. (2) in Fig. 1-7); "**bed**" is the non-adjacent left-situated reference address for "Bett". The converse item relationship is the treatment relationship (cf. (5) in Fig. 1-7); the lemma sign *bed* is treated lexicographically by means of the item giving the word equivalent "Bett", so that the 2-tuple (*bed*, Bett) forms a bilingual lexicographical treatment unit.
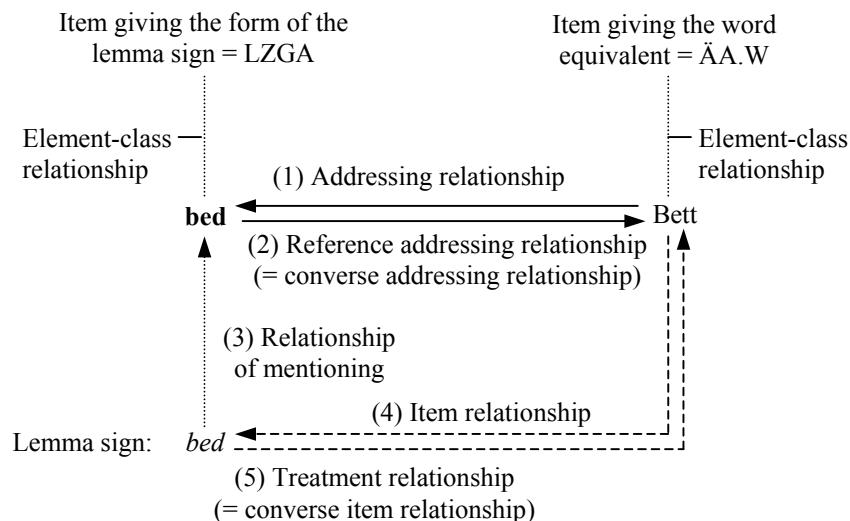

Fig. 1-7:   Illustration of textual relationships in dictionary articles according to wa₇; *Notation conventions*: "y ⋯⋯⋯ Y" means *y is an element of Y* (= y∈Y); "y ⟶ x" means *y is the reference address for x*; "x ⟶ y" means *x is addressed to y*; "u ⋯⋯▶ y" means *u is mentioned by y*; "x ------▶ u" means *x is an item for u*; "u ------▶ x" means *u is treated lexicographically by means of x*

The following abbreviations should now be introduced:

— RT = two-place relation term
— R_{Ad} = addressing relation
— R_{BeAd} = reference addressing relation
— R_{Erw} = relation of mentioning
— R_{An} = item relation
— R_{Bea} = treatment relation

Then the following propositions, ordered into five groups, are valid:

(1) $RT_1(x,y)$ = x is addressed to y; $R_{Ad}(wa_7) = \{(x,y) \mid RT_1(x,y)\}_{wa_7}$;
"Bett" is addressed to "**bed**"; (Bett, **bed**) $\in R_{Ad}(wa_7)$.

(2) $RT_2(y,x)$ = y is the reference address for x; $R_{BeAd}(wa_7)$ =
$\{(y,x) \mid RT_2(y,x)\}_{wa_7}$; "**bed**" is the reference address for
"Bett"; (**bed**, Bett) $\in R_{BeAd}(wa_7)$.

(3) $RT_3(u,y)$ = u is mentioned with the item giving the form;
$R_{Erw}(wa_7) = \{(u,y) \mid RT_3(u,y)\}_{wa_7}$; "*bed*" is mentioned with the item
giving the form "**bed**"; (*bed*, **bed**) $\in R_{Erw}(wa_7)$.

(4) $RT_4(x,u)$ = x is an item for u; $R_{An}(wa_7) = \{(x,u) \mid RT_4(x,u)\}_{wa_7}$;
"Bett" is an item for "*bed*"; (Bett, *bed*) $\in R_{An}(wa_7)$.

(5) $RT_5(u,x)$ = u is treated by means of the item x; $R_{Bea}(wa_7)$ =
$\{(u,x) \mid RT_5(u,x)\}$; "*bed*" is treated by means of the item
"Bett"; (*bed*, Bett) $\in R_{Bea}(wa_7)$.

Finally, the lexicographical-theoretical assumptions include the notion of semantic-pragmatic equivalence for open class expressions which has been developed in recent works (cf. Wiegand 2002a, 2002b and 2002c). The following brief remarks should be given about this notion: In interlingual equivalence relations, one finds lexical-semantic units; these are those abstract units which Alan Cruse called *lexical units* (cf. Cruse 1987: 76f). They consist of an abstract lexical form, according to which all concrete forms of an accompanying inflection paradigm can be realised, and also of an accompanying meaning (in the sense of one "sememe"). Accordingly, a lexeme is a family of lexical-semantic units (cf. Wiegand 2002b: Fig. 2-3). When lexical-semantic units are referred to below, italic capital letters are used (which can be numbered with subscripts). If a form realised according to a lexical-semantic unit of a language A (e.g. English *bed* is a realisation of BED_1) is valid as a semantic-pragmatic equivalent to a form realised according to a lexical-semantic unit of a language B (e.g. German *Bett* realises to $BETT_1$), they should be in two four-place equivalence relations to which one belongs to the type of lexical-semantic (ls) equivalence relation and the other to the type of lexical-pragmatic (lp) equivalence relation, so that the first-mentioned relation is determined by the relation term

$RT_{ls}(x,y,z,k)$ = "x is a lexical-semantic equivalent for y with reference to z relating to the co-text class k", and the latter relation is determined by the relation term $RT_{lp}(x,y,u,k)$ = "x is a lexical-pragmatic equivalent to y with reference to u relating to the co-text class k". The variables "x" and "y" can be inserted with lexical-semantic units; "z" in $RT_{ls}(x,y,z,k)$ and "u" in $RT_{lp}(x,y,u,k)$ are provided for the equivalence criteria: For "z" one can insert markers for reference objects and for "u" one can insert pragmatic labels. The variable "k" is provided for the co-textual equivalence stipulation. With wa_7 as a lexicographical reference text, one can now make the following statements (claiming truth-value):

(i)   *BED₁* is lexically-semantically equivalent to *BETT₁* with reference to "Bett" relating to standard-language habitual usage of all accompanying forms.

(ii)   *BED₁* is lexically-pragmatically equivalent to *BETT₁* with reference to the pragmatic zero marking relating to standard-language habitual usage of all accompanying forms.[3]

If (i) and (ii) are valid, then *BED₁* (and therefore *bed* and *beds*) and *BETT₁* (and therefore *Bett*, *Bett(e)s*, *Betten*) are semantically-pragmatically equivalent, so that a case of word-word equivalence (short: word equivalence) and therefore a case of lexical-internal (or lexicon-specific) equivalence is effected. These explanations should be sufficient for now (cf. further in Wiegand 2002b).

2.   On the notion of lexicographical equivalence

In the following section, I want to give an impression of what a concept of lexicographical equivalence that represents one of the assumptions for an appropriate description of equivalent open class expressions in bilingual printed dictionaries should look like. Such a concept should be understood in such a way that it can, on the one hand, provide lexicographers with a starting point for lexicographical textualisation of semantic-pragmatic equivalence in completely condensed dictionary articles, and, on the other hand, take into account that the linguistic equivalence relationship should be inferred by the user-in-action by means of the absolutely complex lexicographical textualisation outcomes in condensed texts.

     The reason for constructing a concept of the described kind lies in the following facts: Dictionary research not only deals with linguistic equivalents, but also deals with the fact that linguistic equivalence must be, among others things, presented in a completely condensed dictionary article and commented on by means of further items. Equivalents should always be mentioned by means of items giving the equivalent; items giving the equivalent are therefore — just like other items — microstructural text constituents — text segments with a specific item form, at least one specific genuine item purpose and one specific item position; because equivalents which are mentioned should also be

commented on, the items giving the equivalent, function as article-internal reference addresses for other equivalence-relevant items and are themselves article-internally addressed to other equivalence-relevant items, with the result that the particular semantic-pragmatic equivalence (which is conveyed lexicologically and textualised lexicographically) has to be inferred as a specific element of knowledge by the user-in-action first by means of textual facts of a condensed dictionary article in the process of user actions (which are not only reading actions based on linguistic competence).

Next we illustrate graphically the following exerpt (e) from wa_7:

e_1: **bed** […] Bett […]

In e_1, a semantic-pragmatic word-word equivalence (short: word equivalence) is textualised lexicographically. The competent user who wants to understand this textualisation completely, does not only have to know that "Bett" is addressed to "**bed**"; he/she also, according to the accompanying metatext in Neubert and Gröger 1991, has to know that a zero item should be taken into account, since there are no items in e_1 in the positions provided for pragmatic labels. This means: *bed* and *Bett* are pragmatically non-marked (in all the dimensions of marking which the dictionary takes into account). Therefore, one should take into account two blank items ($_iAB_j$). A corresponding representation, in which the positional neighbouring variables "i" and "j" are filled, so that both the blank items are distinguished, then has the following form:

e'_1: **bed** […] $[_{[bed]}AB_1.]$ […] Bett $[_nAB_{\langle\sim and…\rangle}]$ […]

The item giving the pragmatic zero marking (A-pragNM) that is named with "$[_{[bed]}AB_1.]$" is source language-internally addressed to "**bed**"; the item giving the pragmatic zero marking that is named with "$[_nAB_{\langle\sim and…\rangle}]$" is target language-internally addressed to the item giving the equivalent, "Bett". In both cases, left addressing occurs. However, in the domain of the two-place relation of the type of lexicographical equivalence relations which belong to e_1 — it is called $R_{lexÄ}(e_1)$ — appears a set {(**bed**, $[_{[bed]}AB_1.]$)}, of which the item pair (**bed**, $[_{[bed]}AB_1.]$) is an element[4]; as can easily be seen, this pair is an element of the reference addressing relation $R_{BeAd}(wa_7)$ (cf. (2) above, so that (**bed**, $[_{[bed]}AB_1.]$) $\in R_{BeAd}$ is valid. In the range of $R_{lexÄ}(e_1)$ also appears a set with an element of $R_{BeAd}(wa_7)$, namely the element: (Bett, $[_nAB_{\langle\sim and…\rangle}]$); accordingly, (Bett, $[_nAB_{\langle\sim and…\rangle}]$) $\in R_{BeAd}$ is valid. The lexicological fact that BED_1 and $BETT_1$ are semantically-pragmatically equivalent and therefore that BED_1 and $BETT_1$ stand in two four-place relations, namely in a type of lexical-semantic equivalence and a type of lexical-pragmatic equivalence, is thus textualised lexicographically in such a way that a two-place relation of the type of lexicographical equivalence relation is given. In Fig. 2-1, this fact which has just been explained according to Wiegand (2002b), is illustrated graphically.
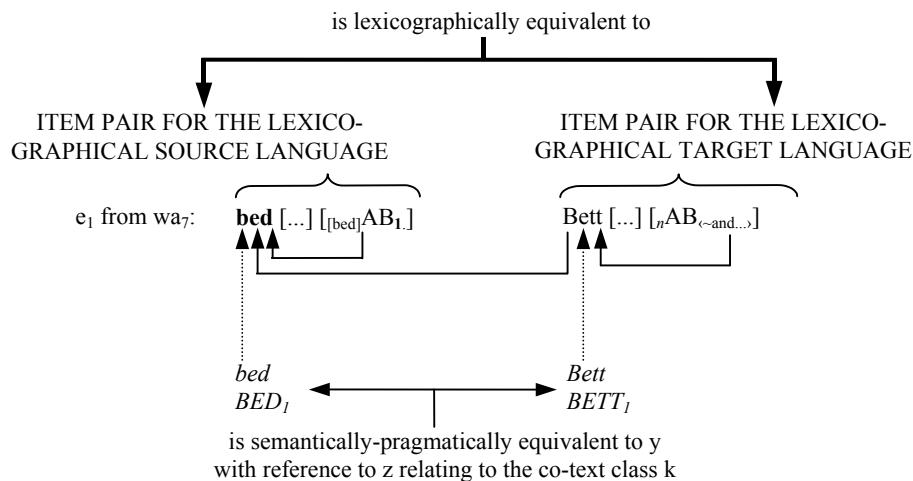
is lexicographically equivalent to

ITEM PAIR FOR THE LEXICO-
GRAPHICAL SOURCE LANGUAGE

ITEM PAIR FOR THE LEXICO-
GRAPHICAL TARGET LANGUAGE

e_1 from wa_7:   **bed** [...] $[_{[bed]}AB_1.]$ Bett [...] $[_nAB_{‹~and…›}]$

bed
BED₁

Bett
BETT₁

is semantically-pragmatically equivalent to y
with reference to z relating to the co-text class k

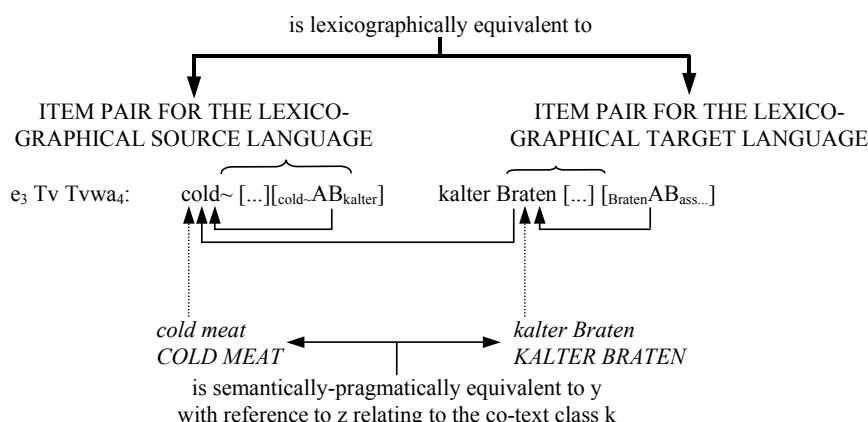**Fig. 2-1:**   Illustration of a lexicographical equivalence relation which is specific for open class expressions and based on semantic-pragmatic word equivalence; *Abbreviations and notation conventions*: AB = blank item; "x ⟶ y" means *x is addressed to y;* "u ┈┈► z" means *u is mentioned by z.*

Now the following statement is valid:

(**bed**, $[_{[bed]}AB_1.]$) is lexicographically equivalent to (Bett, $[_nAB_{‹~and…›}]$).

Accordingly, the following is also valid:

$R_{lexÄ}(e_1) = \{‹\{\textbf{bed}, [_{[bed]}AB_1.])\}, \{(Bett, [_nAB_{‹~and…›})]\}›\}.$

The appropriate reference texts for the analysis of lexicographical equivalence relations are bilingual dictionary articles. Accordingly, $R_{lexÄ}(e_1)$ is only a partial relation (of the cardinality 1) of the lexicographical equivalence relation $R_{lexÄ}(wa_7)$ which belongs to wa_7, so that $R_{lexÄ}(e_1) \subseteq R_{lexÄ}(wa_7)$.[5] The cardinality of an article-specific lexicographical equivalence relation $R_{lexÄ}(wa_x)$ has precisely the same numerical value as the number of equivalence relations which are lexicographically treated in wa_x and thus has the same numerical value as the number of target language items giving the equivalent in wa_x; or, put differently: $R_{lexÄ}(wa_x)$ is equipotent to that subset of $R_{BeAd}(wa_x)$ of which the elements only exhibit equivalence-relevant items as components. As one can easily count in Fig 1-3, the cardinality of $R_{lexÄ}(wa_5)$ is, for example, 5, because wa_5 exhibits the following five items giving the equivalent: *desafilado, romo, opaco, obtuso*, and *apático*.

In order to make the formal connections and the connections of contents even more comprehensible, let us look at another simple example below. We isolate the following part (T) from (v) wa_4:

Tvwa_4: **meat** [mi:t] *s* Fleisch *n* ‹cold ~ kalter Braten; assorted […]› […]

In Tvwa$_4$, two relationships of the semantic-pragmatic equivalence which are specific for open class expressions are textualised lexicographically: one of word equivalence and one of equivalence of syntagmas. The result of the textualisation for word equivalence can be presented as follows, taking into account the zero items for which the blank items are presented:

e$_2$: Tv Tvwa$_4$: **meat** […] [$_{[miːt]}$AB$_s$] […] Fleisch [$_n$AB$_{‹cold…›}$] […]

For the equivalence of syntagmas, there is the following form:

e$_3$ Tv Tvwa$_4$: cold ~ [$_{cold~}$AB$_{kalter}$] kalter Braten [$_{Braten}$AB$_{ass…}$] […]

An illustration of e$_3$ can be found in Fig. 2-2.

is lexicographically equivalent to

ITEM PAIR FOR THE LEXICO-GRAPHICAL SOURCE LANGUAGE

ITEM PAIR FOR THE LEXICO-GRAPHICAL TARGET LANGUAGE

e$_3$ Tv Tvwa$_4$:   cold~ [...][$_{cold~}$AB$_{kalter}$]   kalter Braten [...] [$_{Braten}$AB$_{ass…}$]

cold meat
COLD MEAT

kalter Braten
KALTER BRATEN

is semantically-pragmatically equivalent to y
with reference to z relating to the co-text class k

Fig. 2-2:   Illustration of a lexicographical equivalence relation which is specific for open class expressions and based on semantic-pragmatic equivalence of syntagmas; *Notation conventions*: as in Fig. 2-1.

Next, we construct a relation of the type of lexicographical equivalence relations which belongs to Tvwa$_4$ and which is determined by the proposition schema AF$_{lexÄ}$(Tvwa$_4$) = "x is lexicographically equivalent to y"; this relation is called R$_{lexÄ}$(Tvwa$_4$). Firstly, we constitute the base set for the variable "x" in AF$_{lexÄ}$(Tvwa$_4$) — this is A(Tvwa$_4$) — and then the base set B(Tvwa$_4$) for the variable "y". For the source language range of the variables "x" are then needed the following equivalence-relevant items, which are named here with their item classes and accompanying class symbols:

—   **meat** ∈ LZGA (item giving the form of the lemma sign)

—   [$_{[miːt]}$AB$_s$] ∈ A-pragNM (item giving the pragmatic zero marking [in the item position between "[miːt]" and "$_s$"])

—   cold ~ ∈ v.KollA (condensed item giving the collocation)

— $[_{\text{cold}\sim}AB_{\text{kalter}}] \in$ A-pragNM (item giving the pragmatic zero marking [in the item position between "cold ~" and "kalter Braten"]).

For the target language range of the variables of "y", the following four equivalence-relevant items come to the fore from the reference text $Tvwa_4$:

—   Fleisch $\in$ ÄA.W (item giving the word equivalent)

— $[_{n}AB_{\langle\text{cold}...\rangle}] \in$ A-pragNM (item giving the pragmatic zero marking [in the item position between "n" and "cold ~"])

—   kalter Braten $\in$ ÄA.Synt (item giving the syntagmatic equivalent)

— $[_{\text{Braten}}AB_{\text{ass}...}] \in$ A-pragNM (item giving the pragmatic zero marking [in the item position between "kalter Braten" and "assorted cold ~s"]).

With the eight above-mentioned items we now have all the first components (k_1) and all second components (k_2) of the 2-tuple, which appear as elements of those sets which are elements of $A(Tvwa_4)$ and $B(Tvwa_4)$, so that both these base sets for $AF_{\text{lexÄ}}(Tvwa_4)$ can be presented as follows:

$A(Tvwa_4) = \{\{(\textbf{meat}, [_{[\text{mi:t}]}AB_s])\}, \{(\text{cold} \sim, [_{\text{cold} \sim}AB_{\text{kalter}}])\}\}.$

$B(Tvwa_4) = \{\{(\text{Fleisch}, [_{n}AB_{\langle\text{cold}...\rangle}])\}, \{(\text{kalter Braten}, [_{\text{Braten}}AB_{\text{ass}...}])\}\}.$

Next, we construct the Cartesian product of $A(Tvwa_4)$ and $B(Tvwa_4)$, thus:

$A(Tvwa_4) \times B(Tvwa_4) = \{(k_1, k_2) \mid k_1 \in A(Tvwa_4) \wedge k_2 \in B(Tvwa_4)\}.$

Accordingly, the following set of four elements is found:

$A(Tvwa_4) \times B(Tvwa_4) = \{\langle\{(\textbf{meat}, [_{[\text{mi:t}]}AB_s])\},$
$\{(\text{Fleisch}, [_{n}AB_{\langle\text{cold}...\rangle}])\}\rangle, \langle\{(\textbf{meat}, [_{[\text{mi:t}]}AB_s])\},$
$\{(\text{kalter Braten}, [_{\text{Braten}}AB_{\text{ass}...}])\}\rangle, \langle\{(\text{cold} \sim, [_{\text{cold} \sim}AB_{\text{kalter}}])\},$
$\{(\text{Fleisch}, [_{n}AB_{\langle\text{cold}...\rangle}])\}\rangle, \langle\{(\text{cold} \sim, [_{\text{cold} \sim}AB_{\text{kalter}}])\},$
$\{(\text{kalter Braten}, [_{\text{Braten}}AB_{\text{ass}...}])\}\rangle\}.$

Next, we select from the set $A(Tvwa_4) \times B(Tvwa_4)$ all the elements which exhibit a true proposition schema when the components of their 2-tuple are substituted in the proposition schema "x is lexicographically equivalent to y", in this case k_1 for "x" and the component k_2 for "y". Then we get the subset of $A(Tvwa_4) \times B(Tvwa_4)$. This set is the lexicographical equivalence relation $R_{\text{lexÄ}}(Tvwa_4)$ to the completely condensed lexicographical reference text $Tvwa_4$, so that the following is valid: $R_{\text{lexÄ}}(Tvwa_4) \subseteq A(Tvwa_4) \times B(Tvwa_4)$.

   $R_{\text{lexÄ}}(Tvwa_4)$ is a subset of $R_{\text{lexÄ}}(wa_4)$, exhibits the cardinality 2 and can be presented (extensionally complete) as follows:

$R_{\text{lexÄ}}(Tvwa_4) = \{\langle\{(\textbf{meat}, [_{[\text{mi:t}]}AB_s])\}, \{([\text{Fleisch}, [_{n}AB_{\langle\text{cold}...\rangle}])\}\rangle,$
$\langle\{(\text{cold} \sim, [_{\text{cold} \sim}AB_{\text{kalter}}])\}, \{(\text{kalter Braten}, [_{\text{Braten}}AB_{\text{ass}...}])\}\rangle\}.$

The following propositions about the reference text $Tvwa_4$ are accordingly valid:

(**meat**, [$_{[mi:t]}AB_s$]) is lexicographically equivalent to (Fleisch, [$_nAB_{‹cold…›}$]).

(cold ~, [$_{cold}$ ⌣AB_{kalter}]) is lexicographically equivalent to (kalter Braten, [$_{Braten}AB_{ass…}$]).

By means of the lexicographical equivalence relation $R_{lexÄ}(Tvwa_4)$ one explicitly observes:

(i)   how a case of semantic-pragmatic word equivalence, and

(ii)   how a case of equivalence of syntagmas which are specific to open class expressions are textualised lexicographically in a section of a completely condensed bilingual dictionary article.

With that one also sees

(iii) how a four-place relation of the type of lexical-semantic relations which is specific for open class expressions with the cardinality of 2, and

(iv)   how a four-place relation of the type of lexical-pragmatic relation with the cardinality of 2 have been textualised.

Implicitly one also learns — although only from a specific point of view — how a user-in-action infers the equivalence relationships in $Tvwa_4$ by means of $Tvwa_4$, because it is clear which items, appearing as components in the 2-tuples of the elements in the domain and the range of the lexicographical equivalence relations, are addressed to which other items within the reference text. This information on the addressing relationships can also be made explicit in the notation for a lexicographical equivalence relation and also by providing all equivalence-relevant items which appear as components in a 2-tuple, with their address symbol (which has been worked out in more detail in Wiegand 2002b; cf. also below).

A formal and at the same time graphic representation of $R_{lexÄ}(Tvwa_4)$ can be given in the form of an arrow diagram (cf. Fig. 2-3).
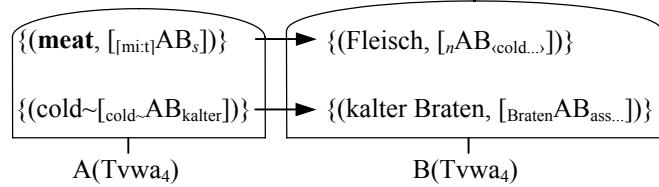

Fig. 2-3:   Arrow diagram for the lexicographical equivalence relation $R_{lexÄ}(Tvwa_4)$.

If one defines on the set of all equivalence-relevant items of a lexicographical reference text (which is Tvwa$_4$ here), a two-place relation of the type of lexicographical equivalence relations, one gets the concrete lexicographical equivalence structure of the reference text. A formal representation of this structure for Tvwa$_4$ can be found in Fig. 2-4.
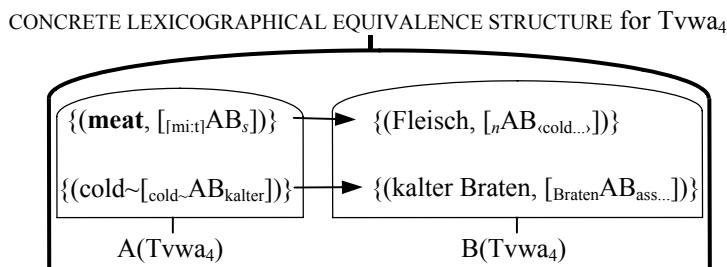
CONCRETE LEXICOGRAPHICAL EQUIVALENCE STRUCTURE for Tvwa$_4$

Fig. 2-4:   Structural graph for the concrete lexicographical equivalence structure of Tvwa$_4$

After the examples and explanations given so far, it should be clear what is meant by a specific concrete lexicographical equivalence relation. Of course, one might also like to have at one's disposal the concept of lexicographical equivalence in such a way that one could make generalisations, and not only make statements about single (excerpts from) dictionary articles. This can happen in two ways. Firstly, one could transfer from all similar concrete cases for a lexicographical equivalence relation to an abstract lexicographical equivalence relation, which represents all the concrete ones. Secondly, the generalisation can be obtained by covering all concrete cases by means of a general form for lexicographical equivalence relations. Below, an explanation is given of how both these generalisations can be obtained.

We already know the following two equivalence relations:

(i) $R_{lex\ddot{A}}(e_1) = \{\langle\{(\textbf{bed}, [_{[bed]}AB_{1.}])\}, \{Bett, [_nAB_{\langle\sim and...\rangle}])\}\rangle\}$.

(ii) $R_{lex\ddot{A}}(e_2) = \{\langle\{\textbf{meat}, [_{[mi:t]}AB_s])\}, \{(Fleisch, [_nAB_{\langle cold...\rangle}])\}\rangle\}$.

Valid are: $R_{lex\ddot{A}}(e_2) \subseteq R_{lex\ddot{A}}(Tvwa_4)$; $R_{lex\ddot{A}}(e_1) \subseteq R_{lex\ddot{A}}(Tvwa_7)$.

In addition, the following dictionary articles wa$_8$ to wa$_{12}$ in Fig. 2-5 from Neubert and Gröger 1991 are given.

wa$_8$:   **fringe** [frɪndʒ] **1.** *s* Franse *f* | Saum *m*, Rand *m* | Besatz *m* | Ponyfrisur *f* | *übertr* Grenze *f*, Rand(zone, -gebiet) *m(f, n)* ⟨the outer ~s die Randbezirke *pl*⟩ | Randgruppe *f*, (besondere) Klasse *f* ⟨the criminal ~ Gruppe *f* der Kriminellen⟩ | *Zool* Haarfranse *f*; **2.** *vt* mit Fransen versehen *od* schmücken | besetzen | einfassen ⟨to ~ with trees⟩ | um-, einzäunen; *vi* Fransen machen; '~ ˌ**ar·e·a** *s* Rundf, *Ferns*

52   Herbert Ernst Wiegand

Randbezirk *m* (mit schlechtem Empfang); ˈ~ ˌ**ben·e·fit** *s*
(*meist pl*) *Wirtsch* (zusätzliche) Sozialleistung *f* ⟨≈s of a
job mit einer Stelle verbundene Extras⟩; **fringed** *adj* ge-
franst; ˈ**~ group** *s* Randgruppe *f*; ˈ~ ˌ**the·at·re** *s Brit* unkon-
ventionelles Theater; ˈ~ **time** *s Ferns* Randzeit *f*; ˈ**fring·y**
adj fransig, Fransen-

wa₉: **gawk** [gɔːk] **1.** s Tölpel *m* Schlacks *m*; **2.** *vi urspr Am umg*
dumm glotzen *od* starren (**at** auf); ˈ**~y 1.** *adj* einfältig,
dumm | tölpelhaft, linkisch; **2.** *s* Tölpel *m*

wa₁₀: **gem** [dʒem] **1.** *s* Edelstein *m* | Gemme *f* | *übertr* Pracht-
stück *n*, Perle *f* | *Am* Brötchen *n*; **2.** *vt* (ˈ**gemmed**,
ˈ**gemmed**) mit Edelsteinen schmücken *od* besetzen

wa₁₁: **hand·sel** [ˈhænsl] **1.** *s* Neujahrsgeschenk *n* | Handgeld *n* |
(Geschäft) erste Einnahme | *übertr* Vorgeschmack *m*; **2.** *vt*
ein Neujahrsgeschenk machen | ein Handgeld geben | ein-
weihen | zum ersten Mal versuchen *od* ausprobieren

wa₁₂: **yacht** [jɔt] **1.** *s* Jacht *f* | (Sport) Segel-, Rennboot *n*; **2.** *vi*
auf einer Jacht fahren | (Sport) segeln; ˈ~ **club** *s* Jachtklub
m; ˈ**~ie** *s umg* Segelboots-, *bes* Jachtbesitzer(in) *m(f)* | Segel-
ler(in) *m(f)*; ˈ**~ing** *s* Jachtsport *m*, Segelsport *m*; **3.** *adj* Se-
gel-; ˈ**~er**, ˈ**yachts·man** (*pl* ˈ**yachts·men**) *s* Jachtfahrer *m* |
Sportsegler *m*; ˈ**yachts·man·ship** *s* Jacht-, Segelkunst *f*;
ˈ**yachtsˌwom·an** (*pl* ˈ**yachtsˌwom·en**) *s* Jachtfahrerin *f* |
Sportseglerin *f*

Fig. 2-5:   Dictionary articles wa₈ to wa₁₂ from Neubert and Gröger 1991

Below, a concrete partial relation of the lexicographical equivalence relation
which belongs to the reference texts, is given for each of the reference texts wa₈
to wa₁₂.

(iii) $R_{lexÄ(1)}(wa_8) = \{⟨\{(\textbf{fringe}, [_{[frɪndʒ]}AB_1.])\}, \{(Franse, [_fAB_{Saum}])\}⟩\}$.

(iv) $R_{lexÄ(1)}(wa_9) = \{⟨\{(\textbf{gawk}, [_{[gɔːk]}AB_1.])\}, \{(Tölpel, [_mAB_{Schlacks}])\}⟩\}$.

(v) $R_{lexÄ(1)}(wa_{10}) = \{⟨\{(\textbf{gem}, [_{[dʒem]}AB_1.])\}, \{(Edelstein, [_mAB_{Gemme}])\}⟩\}$.

(vi) $R_{lexÄ(1)}(wa_{11}) = \{⟨\{\textbf{hand·sel}, [_{[ˈhænsl]}AB_1.])\}, \{(Neujahrsgeschenk,$
        $[_nAB_{Handgeld}])\}⟩\}$.

(vii) $R_{lexÄ(1)}(wa_{12}) = \{⟨\{(\textbf{yacht}, [_{[jɔt]}AB_1.])\}, \{(Jacht, [_fAB_{⟨Sport⟩}])\}⟩\}$.

The lexicographical relations (i) to (vii) are equipotent; the seven relations are,
in addition, of the same kind in the following sense:

(a)   As first components of all 2-tuples of all sets in the domain of the rela-
        tions appear items giving the form of the lemma sign; accordingly, the

following propositions are valid:

bed $\in$ LZGA; **meat** $\in$ LZGA; **fringe** $\in$ LZGA; **gawk** $\in$ LZGA;
  gem $\in$ LZGA; **hand·sel** $\in$ LZGA; **yacht** $\in$ LZGA.

(b)   As second components of all 2-tuples of all sets in the domain of the rela-
tions appear items giving the pragmatic zero marking; accordingly,
seven propositions of the same kind are valid, of which I only name the
one following (i):

$[_{[bed]}AB_{1.}] \in$ A-pragNM.

(c)   As first components of all 2-tuples of all sets in the range of the relations
appear items giving the word equivalent; accordingly, seven proposi-
tions of the same kind are valid in this regard, of which only the one
belonging to (i) is named here:

Bett $\in$ ÄA.W.

(d)   As second components of all 2-tuples of all sets in the range appear items
giving the pragmatic zero marking; of the seven propositions of the
same kind which are valid, only the one for (i) is named:

$[_{n}AB_{\langle\sim and...\rangle}] \in$ A-pragNM.

(e)   In addition, the following is valid: all 2-tuples of all sets in the domains
and ranges of the relations (i)–(vii) are elements of the reference address
relation, which belongs to the reference text at hand. For instance, the
following propositions are valid for (iv):

(**gawk**, $[_{[gɔ:k]}AB_{1.}]) \in R_{BeAd}(wa_9)$.
(**Tölpel**, $[_{m}AB_{Schlacks}]) \in R_{BeAd}(wa_9)$.

Finally, the following is valid:

(f)   All addressing relationships containing those items which appear at any
time as equal components of a 2-tuple, are equal. Consequently, the
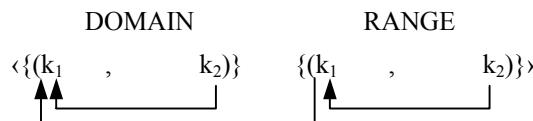illustration in Fig. 2-6 is valid for all examples (i) to (vii).

Fig. 2-6:   Illustration of the addressing relationships in the reference texts (i) to (vii);
Notation convention: "x⟶y" means *x is addressed to y*

The abstract lexicographical equivalence relation which represents the concrete relations (i) to (vii) (and many others from Neubert and Gröger 1991) can now be stated because, instead of naming items as components, one names those class symbols by means of which the classes of items with equal genuine purpose are denoted and to which the particular items belong. Accordingly, the following abstract (A) lexicographical equivalence relation (LEXÄ) belongs to the seven relations (i) to (vii):

$$R_{ALEXÄ}(WA_1) = \{\{(LZGA, A\text{-}pragNM)\}, \{ÄA.W, A\text{-}pragNM)\}\},$$

in which WA_1 is the set of all the involved dictionary articles appearing as reference texts, so that the following is valid: $WA_1 = \{wa_4, wa_7, wa_8, wa_9, wa_{10}, wa_{11}, wa_{12}\}$.

In the description so far, the addressing relationships have been taken into account as prerequisites for a correct formation of the 2-tuples as the elements of the sets in the domain and the range of the lexicographical equivalence relation; however, in the formal notation of the concrete and abstract relations they have not explicitly been taken into account. This can only be changed if one expands the notation with address symbols and consequently follows the formal representation of item structures by means of structural graphs in which the reference address of the particular items in the node labelling is systematically taken into account because one adds an address symbol to each symbol of an item class by keeping to specific notation prescriptions (cf., for example, Wiegand 1990 and 1991: 103ff). Therefore, one can now state: an expression of the form $a_i \cdot b_j$ should be read as "the item a_i with the reference address b_j". The expression "$[_{[bed]}AB_1.] \cdot \textbf{bed}$" should accordingly be read as: "the item '$[_{[bed]}AB_1.]$' with the reference address $\textbf{'bed'}$ ". For the relation (i), this results in the following notation in which the particular address symbols were taken into account:

(i') $R_{lexÄ}(e_1) = \{\langle\{(\textbf{bed}, [_{[bed]}AB_1.] \cdot \textbf{bed})\}, \{(Bett \cdot \textbf{bed}, [_{n}AB_{\langle \sim and \dots \rangle}] \cdot Bett)\}\rangle\}.$

In accordance, for the relation (ii), the notation has the following form:

(ii') $R_{lexÄ}(e_2) = \{\langle\{(\textbf{meat}, [_{[mi:t]}AB_s] \cdot \textbf{meat})\}, \{(Fleisch \cdot \textbf{meat}, [_{n}AB_{\langle cold \dots \rangle}] \cdot Fleisch)\}\rangle\}.$

For the accompanying abstract lexicographical equivalence relation, the following notation will result accordingly:

$$R_{ALEXÄ}(WA_1) = \{\langle\{(LZGA, A\text{-}pragNM \cdot LZGA)\},$$
$$\{(ÄA.W \cdot LZGA, A\text{-}pragNM \cdot ÄA.W)\}\rangle\}.$$

Next, we look at the second possibility for generalisation. In the examples so far, the following has always been the case: (i) to each source language item giving the form (in the examples it was always the item giving the form of the lemma sign), precisely one equivalence-relevant item was addressed source language-internally (in the examples it was always a A-pragNM); (ii) to the item giving the word equivalent, which is addressed to a source language item

giving the form, precisely one equivalence-relevant item was addressed target language-internally (cf. also Fig. 2-6). In many completely condensed bilingual dictionary articles, however, not only one, but various equivalence-relevant items can be addressed to the source language item giving the form as well as to the target language item giving the form. This has as a result that not only one set appears in the domain and the range of a lexicographical equivalence relation which only exhibits one 2-tuple as element (as in all the examples so far), but also sets with m 2-tuple (m ≥ 2; m ∈ NI). Such dictionary articles are, for example, wa_{13} to wa_{16} in Fig. 2-7.

wa_{13}: **stattlich** [ˈʃtatlɪç] *adj* (*beeindruckend*) im-
ponente; [*Betrag*] considerable

wa_{14}: **straff** [ʃtraf] *adj* (*gespannt*) tenso; (*Diszi-
plin)* riguroso; **etw ~ ziehen** tensar algo

wa_{15}: **streng** [ʃtrɛŋ] *adj* severo; (*hart*) duro; (*Ge-
ruch*) acre; (*schmucklos*) austero; **~ ge-
nommen** en rigor; **das ist ~ verboten!**
¡eso está terminantemente prohibido!; **~
geheim** bajo absoluta discreción

wa_{16}: **stürmisch** *adj* **1.** (*Wetter*) tempestuoso;
(*Meer*) agitado **2.** (*ungestüm*) impetuoso;
(*heftig*) violento; (*Liebhaber*) apasionado;
(*Entwicklung*) rápido; (*Beifall*) frenético

Fig. 2-7:   Dictionary articles wa_{13}–wa_{16} from DE. Dt.–Span./Es.–Al. 1999

In the articles wa_{13} to wa_{16}, if-then relationships given in the dictionary subject domain are textualised lexicographically as conditions for equivalence in such a way that various equivalence-relevant items are addressed to the item giving the form of the lemma sign and in most cases also to the target language items giving the equivalent. Firstly, let us look at wa_{13}. According to the accompanying metatext of the dictionary, a zero item should be added between the item giving the word class "*adj*" and the item giving the synonym functioning as the item discriminating the equivalent, "*beeindruckend*", in the form of an item giving pragmatic zero marking (which can be named with "[$_{adj}$AB$_{(be…)}$]", of which the textual scope is the totality of the rest of the article, so that it is addressed within its textual scope to all items giving the form with which language forms that could be pragmatically zero marked, are named. The excerpt

e_4: Tv wa_{13}: **stattlich** [...] *adj* (*beeindruckend*) imponente [...]

should then, with a view to the dictionary subject domain, be read as follows: If the lexical-semantic unit *STATTLICH$_1$* in one of its realised German forms is used in standard language and in its habitual sense so that it means *beein-*

druckend, then it is semantically-pragmatically equivalent to a standard language Spanish form which is a realisation of the lexical-semantic unit *IMPONENTE* and is used in its habitual sense. With a view to the dictionary form, the following statement is accordingly valid: If in the first place the item giving the pragmatic zero label (represented by "[*adj*AB(be...)]") and if in the second place the item discriminating the equivalent, "*beeindruckend*", is addressed to the item giving the form of the lemma sign "**stattlich**" and, finally, in the third place the item giving the pragmatic zero marking is also addressed to the item giving the word equivalent "imponente", then "imponente" is also addressed to the item giving the form of the lemma sign. An illustration of the addressing relationships can be found in Fig. 2-8.
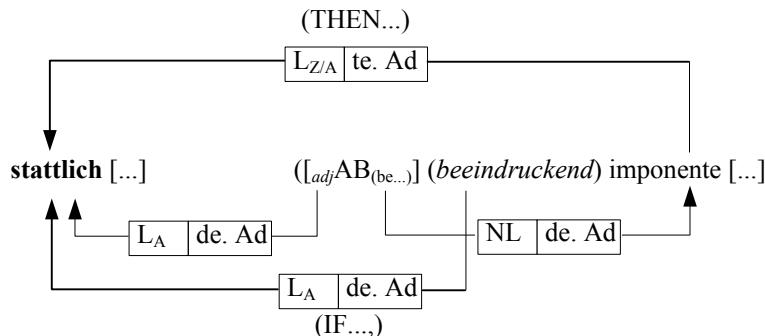
Fig. 2-8: Illustration of the addressing in e_4; *Abbreviations and notation conventions*: A = lexicographical source language (here: German); Z = lexicographical target language (here: Spanish); $L_{Z/A}$ = addressed lemmatically from Z to A; L_A = A-internally lemmatically addressed; NL = non-lemmatically addressed; te.Ad = qualified addressing; de. Ad = required addressing; "x —□u→ y" means *x is article-internally u-addressed to y* with "u" as variable for L_A, $L_{Z/A}$, amongst others.

In the domain (d) of the lexicographical equivalence relation to the reference text e_4 now appears the following set A_d of 2-tuples: A_d = {(**stattlich**, [*adj*AB(be...)]), (**stattlich**, *beeindruckend*)}; in the range (r) appears the following set: B_r = {(imponente), [*adj*AB(be...)])}.

Accordingly, the lexicographical equivalence relation to e_4 can now be stated as follows:

$R_{lexÄ}(e_4)$ = {‹{(**stattlich**, [*adj*AB(be...)], (**stattlich**, *beeindruckend*)},
     {(imponente], [*adj*AB(be...)])}›}.

Next, we look at the entry:

e_5: Tvwa$_{13}$: **stattlich** […] (*Betrag*) considerable […]

Firstly, it should be taken into account that the item giving the word equivalent

"considerable" also lies within the textual scope of "$[_{adj}AB_{(be...)}]$". Accordingly, the addressing relationships in e_5 are formed analogous to e_4, as they are illustrated in Fig. 2-9.
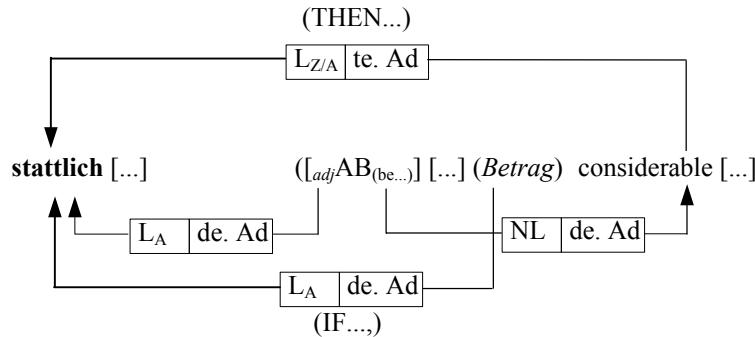

Fig. 2-9:   Illustration of the addressing in e_5; *Abbreviations and notation conventions*: as in Fig. 2-8.

The lexicographical equivalence relation for e_5 thus has the following form:

$R_{lexÄ}(e_5) = \{\langle\{(\textbf{stattlich}, [_{adj}AB_{(be...)}]), (\textbf{stattlich}, \textit{Betrag})\},$
$\{(\text{considerable}, [_{adj}AB_{(be...)}])\}\rangle\}.$

Because e_4 and e_5 are parts of wa_{13}, the following subset relationships are valid:

$R_{lexÄ}(e_4) \subseteq R_{lexÄ}(wa_{13}); R_{lexÄ}(e_5) \subseteq R_{lexÄ}(wa_{13}).$

The lexicographical equivalence relation $R_{lexÄ}(wa_{13})$ to the reference text wa_{13} can be stated (extensionally complete) as follows:

$R_{lexÄ}(wa_{13}) \{\langle\{(\textbf{stattlich}, [_{adj}AB_{(be...)}]), (\textbf{stattlich}, \textit{beeindruckend})\},$
$\{(\text{imponente}, [_{adj}AB_{(be...)}])\}\rangle, \langle\{(\textbf{stattlich}, [_{adj}AB_{(be...)}]), (\textbf{stattlich}, \textit{Betrag})\},$
$\{(\text{considerable}, [_{adj}AB_{(be...)}])\}\rangle\}.$

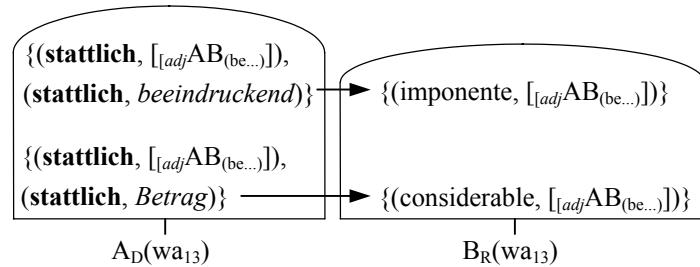Fig. 2-10 shows an arrow diagram for $R_{lexÄ}(wa_{13})$:

Fig. 2-10:   Arrow diagram for the lexicographical equivalence relation $R_{lexÄ}(wa_{13})$

It is clear that sets which present themselves as elements n 2-tuples (with $n \geq 2$; $n \in NI$) can also appear in the range of a lexicographical equivalence relation. The discussion of a similar example should be left out for reasons of space.

The abstract lexicographical equivalence relation which belongs to $R_{lexÄ}(wa_{13})$ can now be stated as follows:

$R_{ALEXÄ}(WA_{13})$ = {‹{(LZGA, A-pragNM), (LZGA, ÄUntA)},
    {(ÄA.W, A-pragNM)})›, ‹{(LZGA, A-pragNM)}, (LZGA, ÄUntA)},
    {(ÄA.W, A-pragNM)}›}$_{wa_{13}}$.

If one wants to follow through the process right to the end to get to the general form of a lexicographical equivalence relation, the last step can be carried out in different ways. In the following procedure a partial analogy is given for the treatment of e_4 and e_5, because this makes it more understandable. One proceeds from the fact that, in every completely condensed bilingual dictionary article, a set of equivalence-relevant items appears, and in addition, that these items belong to classes of equivalence-relevant items; the set could be named as $M_{ÄrelA}$. This set can be divided into four disjunct subsets (where specific subsets could be empty). The following subsets are involved:

— ÄA, the class of all target language items giving the equivalent,

— BeAd, the class of all source language items giving the form which are reference addresses of target language items giving the equivalent,

— ÄrelA·ÄA, the class of all equivalence-relevant items which are addressed to target language items giving the equivalent,

— ÄrelA· BeAd, the class of all equivalence-relevant items addressed to the source language items giving the form which are the reference addresses of the target language items giving the equivalent.

Accordingly, the following propositions are valid:

ÄA $\subseteq M_{ÄrelA}$; BeAd $\subseteq M_{ÄrelA}$; ÄrelA·ÄA $\subseteq M_{ÄrelA}$;
    Ärel·BeAd $\subseteq M_{ÄrelA}$.

ÄA $\cap$ BeAd = Ø; ÄA $\cap$ ÄrelA·ÄA = Ø;
    ÄA $\cap$ ÄrelA·BeAd = Ø; BeAd $\cap$ ÄrelA·ÄA = Ø;
    BeAd $\cap$ ÄrelA·BeAd = Ø; ÄrelA·ÄA $\cap$ ÄrelA·BeAd = Ø.

In the following notation, which is stated for the general form of the lexicographical equivalence relation, one should interpret expressions of the form X^n and Y_m as "the nth element which has been selected from X" and "the mth element which has been selected from Y" respectively, so that therefore, for example, $BeAd^i$ should be read as "the ith element which has been selected from the class BeAd". The general form of a lexicographical equivalence relation can then be stated as follows:

$$\{\langle\{(BeAd^i, \ddot{A}relA_j \cdot BeAd^i)_{j \in J_i}\}, \{(\ddot{A}A^r \cdot BeAd^i, \ddot{A}relA_k \cdot \ddot{A}A^r)_{r \in R_i, k \in K_i}\}\rangle_{i \in I}\}$$

with $I = \{1, ..., n\}, n \in IN$

$J_i = \subset IN, R_i \subset IN, K_i \subset IN.$

Consequently, the following is valid: If j equivalence-relevant items ($\ddot{A}relA_j \cdot BeAd^i$) are addressed to the ith reference address ($BeAd^i)_j$ and if equivalence-relevant items ($\ddot{A}relA_k \cdot \ddot{A}A^r$) are addressed to the rth item giving the equivalent ($\ddot{A}A^r)_k$ and if in addition the rth item giving the equivalent is addressed to the ith reference address ($\ddot{A}A^r \cdot BeAd^i)_j$, then the rth item giving the equivalent is lexicographically equivalent to the ith reference address, whereby the number of items giving the equivalent depends on the odds of the ith reference address.

Finally, an appropriate definition is given for the stated general form of a lexicographical equivalence relation. In this definition, the term *insertable nomination expression* is used as generic term in such a way that all insertable expressions in source and target language sentence constructions by means of which one makes references and predications, are valid as insertable nomination expressions. The definition reads as follows:

(D 2-1: lexicographical equivalence for expressions of open classes)

A two-place relation of lexicographical equivalence exists exactly when, with regard to a bilingual condensed dictionary article, the following conditions are met:

(1)   By means of the items giving the form, appearing in the source language domain of a lexicographical equivalence relation, as well as by means of the items giving the form appearing as items giving the equivalent in the target language range, insertable nomination expressions are mentioned.

(2)   The target language items giving the equivalent in the range of the lexicographical relation are addressed to source language items giving the form in the domain.

(3)   The addressing relationships of the items giving the equivalent in the range of the relation to the items giving the form in the domain of the relation are qualified by the fact that the source language forms mentioned by the items giving the form in the domain are realisations of a lexical-semantic unit which is semantically-pragmatically equivalent to those particular lexical-semantic units which are realisations of those forms which are mentioned by the items giving the equivalent in the range of the relation.

(4)   n Equivalence-relevant items are addressed to the items giving the form in the domain of the relation as well as to the items giving the equivalent in the range of the relation (with $n \geq 1$), amongst which appears at least

one pragmatic item.

(5)   The reference relationship in which the items stand to their respective reference addresses, i.e. the source language items giving the form and those particular items giving the equivalent addressed to them, is qualified by the habitual usage of the source and target language forms named by the items giving the form and the equivalent, as well as by the habitual usage of all forms which are realisations of the accompanying lexical-semantic units.

(6)   The function of the pragmatic language items addressed to the source language items giving the form has to be equal to the function of the pragmatic items addressed to the item giving the equivalent.

3.   Brief view on the possibilities of optimising the practice of lexicographical description in the area of equivalent open class expressions

Especially in more comprehensive bilingual dictionaries of the widely-used languages of the civilized world, there are thousands of completely condensed dictionary articles which are considerably more extensive than the longest article examples used in this discussion, namely, wa_4 and wa_7, which deal with bilingual articles of medium length. Dictionary articles dealing lexicographically with dozens of equivalence relationships are not unusual, but occur all the time. One should, in fact, not suspect in principle that practicing lexicographers do not at least master the practice that they have acquired at their particular dictionary project. There are, however, enough clear indications that the particular practice has not been thought through consistently. At any rate, it has, in my opinion, not been completely understood from a theoretical point of view. I do not think that one can expect an optimisation of similar practice-based results for the description of equivalence relations which stem from practice itself. Because of the specific conditions in dictionary offices, practitioners are — that is demonstrated by the entire history of lexicography — not in a position fundamentally to reform their own practice on their own.

The optimisation of the lexicographical description of equivalent open class expressions has two different aspects, which have to be distinguished absolutely and examined separately. The first aspect deals with the subject and subject domain of the dictionary. In the last decade, the possibilities for optimisation have considerably increased in this aspect because of the steadily improving investigation possibilities in electronic corpora and because of the rapid development of corpus linguistics. This article does not deal with this aspect. An optimisation of the existing lexicographical description of equivalence relationships which relates to the dictionary form and consequently especially to the lexicographical textualisation of equivalence relationships in completely condensed bilingual dictionary articles as well as to the explanation of the textualisation in the accompanying metatexts, cannot be determined in the

recent general bilingual dictionaries, but is urgently needed for the sake of users. The concept of lexicographical equivalence presented in this contribution and also in Wiegand (2002b), with different emphasis and further aspects, constitute in my opinion a firm basis for a far-reaching optimisation of the presentation of equivalents and the accompanying user-friendly explanations. In this way, different types of equivalence structures which belong to the article and the correlating distinct types of bilingual dictionaries can be distinguished (cf. Wiegand 2002c and 2003). Only when the textual features of different types of bilingual dictionary articles are better known, will one be successful in obtaining a well-established, purposeful and systematically teachable optimisation of the lexicographical description of equivalent open class expressions. In completely condensed bilingual dictionary articles, the relation of the article form and the article contents is very complex. Because a user can only manage to get to the article contents via the article form, it is worthwhile to investigate the form, and in future to fashion it more adequately.

Notes

1. For the distinction between text condensation and text compression as two different possibilities to condense a text, cf. Wiegand 1998a: 31f; on the perception of the text form, cf. Wiegand 1999. On micro-architectures, cf. for example, Wiegand 2001: 191ff.
2. The abbreviations which are used further on are class symbols for items with identical general genuine purpose, so that one could, for example, write Bett ∈ ÄA.W.
3. For the expression *habitual usage*, cf. Wiegand 1996.
4. The fact that a set of which the element is a 2-tuple, and not the 2-tuple itself, appears in the domain of the relation, cannot be motivated sufficiently by means of example e_1. It could easily make sense if one understood that an item giving the form of the lemma sign such as "**bed**" could be addressed to various equivalence-relevant items, so that a set of various 2-tuples appear in the domain; with that the equivalence relation to e_1 becomes recognizable as a special case, because the set in the domain (and also in the range) only exhibits one element.
5. In order to draw attention to the status of the partial relation, I added (or omitted) the braces which stand directly in front of and behind the round brackets; because they are needed in the illustration of lexicographical equivalence relations with a cardinality of >1.

Bibliography

I thank Klaus-Peter Konerding for useful suggestions.

Dictionaries

DE. Dt.–Span./Es.–Al. 1999 = *Diccionario Esencial. Deutsch–Spanisch/Español–Alemán. Taschenwörterbuch.* 1999. Barcelona: Biblograf.

Kinne and Strube-Edelmann 1980 = Kinne, Michael and Birgit Strube-Edelmann. 1980. *Kleines Wörterbuch des DDR-Wortschatzes*. Düsseldorf: Schwann.

Lang. Twb. Port.–Dt./Dt.–Port. 2001 = Langenscheidt-Redaktion. 2001. *Langenscheidt Taschen-wörterbuch Portugiesisch–Deutsch/Deutsch–Portugiesisch*. Completely revised edition. Berlin/Munich/Vienna/Zürich/New York: Langenscheidt.

Müller 1999 = Müller, Martin. 1999. *Goethes merkwürdige Wörter. Ein Lexikon*. Darmstadt: Wissen-schaftliche Buchgesellschaft.

Müller-Thurau 1983 = Müller-Thurau, Claus Peter. 1983. *Laß uns mal 'ne Schnecke angraben. Sprache und Sprüche der Jugendszene*. Third edition. Düsseldorf/Vienna: Econ-Verlag.

Neubert and Gröger 1991 = Neubert, Albrecht and Erika Gröger. 1991. *Großes Handwörterbuch Eng-lisch–Deutsch*. Leipzig/Berlin/Munich: Langenscheidt.

Pons 2002 = *Universalwörterbuch Deutsch–Bulgarisch*. 2002. Stuttgart: Ernst Klett.

Schlosser 2000 = Schlosser, Horst Dieter. 2000. *Lexikon der Unwörter*. Gütersloh: Bertelsmann.

Secondary literature

Cruse, D. Alan. 1987. *Lexical Semantics*. Cambridge Textbooks in Linguistics. Cambridge/Lon-don/New York: Cambridge University Press.

Engelberg, Stefan and Lothar Lemnitzer. 2001. *Lexikographie und Wörterbuchbenutzung*. Tübingen: Stauffenburg.

Wiegand, Herbert Ernst. 1983. Was ist eigentlich ein Lemma? Ein Beitrag zur Theorie der lexiko-graphischen Sprachbeschreibung. Wiegand, Herbert Ernst (Ed.). 1983. *Studien zur neuhoch-deutschen Lexikographie III. Germanistische Linguistik* 1–4/82: 401-474. Hildesheim/New York: Georg Olms.

Wiegand, Herbert Ernst. 1990 [1991]. Printed Dictionaries and their Parts as Texts. An Overview on Recent Research as Introduction. *Lexicographica* 6: 1-26.

Wiegand, Herbert Ernst. 1996. Über usuelle und nichtusuelle Benennungskontexte in Alltag und Wissenschaft. Knobloch, Clemens and Burkhard Schaeder (Eds.) 1996. *Nomination — fach-sprachlich und gemeinsprachlich*: 55-103. Opladen: Westdeutscher Verlag.

Wiegand, Herbert Ernst. 1998. *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie. 1. Teilband.* With 159 illustra-tions in the text. Berlin/New York: Walter de Gruyter.

Wiegand, Herbert Ernst. 1998a. Lexikographische Textverdichtung. Entwurf einer vollständigen Konzeption. Zettersten, Arne, Viggo Hjørnager Pedersen and Jens Erik Mogensen (Eds.). 1998. *Symposium on Lexicography VIII. Proceedings of the Eighth International Symposium on Lexi-cography May 2–4, 1996, at the University of Copenhagen*: 1-35. Lexicographica. Series Maior 90. Tübingen: Max Niemeyer.

Wiegand, Herbert Ernst. 1999. Artikel einsprachiger Lernerwörterbücher, Textgestaltwahrneh-mung und Suchbereichsstrukturen. Plädoyer für übersichtliche Printwörterbücher im Zeital-ter der Neuen Medien. Skibitzki, Bernd and Barbara Wotjak (Eds.). 1999. *Linguistik und Deutsch als Fremdsprache*: 259-281. Tübingen: Max Niemeyer.

Wiegand, Herbert Ernst. 2001. Sprachkontaktwörterbücher. Typen, Funktionen, Strukturen. Igla, Birgit, Pavel Petkov and Herbert Ernst Wiegand (Eds.). 2001. *Theoretische und praktische Pro-bleme der Lexikographie. 1. Internationales Kolloquium zur Wörterbuchforschung am Institut Ger-*

manicum der St. Kliment Ohridski-Universität Sofia, 7. bis 8. Juli 2000: 115- 224. *Germanistische Linguistik* 161–162. Hildesheim/New York: Georg Olms.

Wiegand, Herbert Ernst. 2002. Adressierung in zweisprachigen Printwörterbüchern. Wiegand, Herbert Ernst (Ed.). 2002. *Studien zur zweisprachigen Lexikographie mit Deutsch VIII*: 112-175. *Germanistische Linguistik* 166. Hildesheim/New York: Georg Olms.

Wiegand, Herbert Ernst. 2002a. Äquivalenz in zweisprachigen Printwörterbüchern. Kritik und Vorschläge. Wiegand, Herbert Ernst (Ed.). 2002. *Studien zur zweisprachigen Lexikographie mit Deutsch VIII*: 93-110. *Germanistische Linguistik* 166. Hildesheim/New York: Georg Olms. [English version: Equivalence in Bilingual Lexicography: Criticism and Suggestions. *Lexikos* 12: 239-255.]

Wiegand, Herbert Ernst. 2002b. *Äquivalenz, Äquivalentdifferenzierung und Äquivalentpräsentation in zweisprachigen Wörterbüchern. Eine neue einheitliche Konzeption*. [Paper presented at the Eleventh International Symposium on Lexicography, May 2002, at the University of Copenhagen, Copenhagen; in the press.]

Wiegand, Herbert Ernst. 2002c. *Äquivalentpräsentation und Wörterbuchfunktionen in zweisprachigen Printwörterbüchern*. [Paper presented at the Second International Symposium on Dictionary Research, 18 October 2002, in Sofia; to be published in the Proceedings.]

Wiegand, Herbert Ernst. 2002d. Über textuelle Strukturen der Wörterbuchartikel und Artikelnischen im *de Gruyter Wörterbuch Deutsch als Fremdsprache*. Zugleich ein Beitrag zur Weiterentwicklung einer Theorie der Wörterbuchform. Wiegand, Herbert Ernst (Ed.). 2002. *Perspektiven der pädagogischen Lexikographie des Deutschen II. Untersuchungen anhand des* de Gruyter Wörterbuch Deutsch als Fremdsprache: 497-595. Lexicographica. Series Maior 110. Tübingen: Max Niemeyer.

Wiegand, Herbert Ernst. 2003. Überlegungen zur Typologie von Wörterbuchartikeln in Printwörterbüchern. Ein Beitrag zur Theorie der Wörterbuchform. *Lexicographica* 19.

Abbreviations

A	=	source language (Ausgangssprache)
ÄA.Synt	=	item giving the equivalent of syntagmas (Syntagmenäquivalentangabe)
ÄA.W	=	item giving the word equivalent (Wortäquivalentangabe)
AB	=	blank item (Angabeblank)
A-pragNM	=	item giving the pragmatic zero marking (Angabe zur pragmatischen Nullmarkierung)
ÄUntA	=	item giving the distinction for the equivalent (Äquivalentunterscheidungsangabe)
AusA	=	item giving the pronunciation (Ausspracheangabe)
de.Ad	=	required addressing (bedingende Adressierung)
d	=	domain (Vorbereich)
k	=	component (Komponente)
LA	=	source language-internally lemmatically addressed (ausgangssprachenintern lemmatisch adressiert)
lp	=	lexical-pragmatic equivalence relation (lexikalpragmatische Äquivalenzrelation)
ls	=	lexical-semantic equivalence relation (lexikalsematische Äquivalenzrelation)
$L_{Z/A}$	=	addressed lemmatically from Z to A (von Z nach A lemmatisch adressiert)
LZGA	=	item giving the form of the lemma sign (Lemmazeichengestaltangabe)
NL	=	non-lemmatically addressed (nicht lemmatisch adressiert)
r	=	range (Nachbereich)
R_{Ad}	=	addressing relation (Adressierungsrelation)
R_{An}	=	item relation (Angaberelation)
R_{Bea}	=	treatment relation (Bearbeitungsrelation)
R_{BeAd}	=	reference addressing relation (Bezugsadressenrelation)
R_{Erw}	=	relation of mentioning (Erwähnungsrelation)
$R_{lexÄ}$	=	lexicographical equivalence relation (lexikographische Äquivalenzrelation)
RT	=	two-place relation term (zweistelliger Relationsterm)
T	=	part, excerpt (Teil)
TC	=	typology criterion (Typologiekriterium)
te.Ad	=	qualified addressing (bedingte Adressierung)
v	=	of (von)
v.KollA	=	condensed item giving the collocation (verdichtete Kollokationsangabe)
wa	=	dictionary article (Wörterbuchartikel)
WAA	=	item giving the word class (Wortartangabe)
WUntA	=	item giving the word class distinction (Wortunterscheidungsangabe)
Z	=	target language (Zielsprache)

User-oriented Understanding of Descriptive, Proscriptive and Prescriptive Lexicography[*]

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

Abstract: There is much uncertainty and confusion as to the real differences between prescriptive and descriptive dictionaries. In general, the majority of existing accounts can be summarised as follows: Descriptive relates to the empirical basis; accordance between the empirical data and the dictionary is required. Prescriptive relates to the genuine purpose of the dictionary; the dictionary is meant to help with problems concerning text production and will thus affect usage. This asymmetrical understanding would imply prescriptive and descriptive in practice being false contrasts. In this article, a more consistent terminology is suggested which allows for both the function of the dictionary and the relation of the dictionary to the empirical basis.

Keywords: DESCRIPTION, DICTIONARY FUNCTION, EMPIRICAL BASIS, EXACTLY DESCRIPTIVE DICTIONARY, EXACTLY PROSCRIPTIVE DICTIONARY, EXPLICITLY PRESCRIPTIVE DICTIONARY, INTROSPECTION, LINGUISTIC SURVEY, NOT EXACTLY DESCRIPTIVE DICTIONARY, NOT EXACTLY PROSCRIPTIVE DICTIONARY, PRESCRIPTION, PROSCRIPTION, STRONGLY DESCRIPTIVE DICTIONARY, STRONGLY PRESCRIPTIVE DICTIONARY, TEXT CORPUS, USER-ORIENTED, WEAKLY DESCRIPTIVE DICTIONARY, WEAKLY PRESCRIPTIVE DICTIONARY, WEAKLY PROSCRIPTIVE DICTIONARY

Opsomming: Gebruikersgerigte siening van deskriptiewe, proskriptiewe en preskriptiewe woordeboeke. Daar is baie onsekerheid en verwarring oor die werklike verskille tussen preskriptiewe en deskriptiewe woordeboeke. Oor die algemeen kan die meerderheid bestaande verklarings soos volg saamgevat word: Deskriptief hou verband met die empiriese basis; ooreenstemming tussen die empiriese gegewens en die woordeboek is nodig. Preskriptief hou verband met die werklike doel van die woordeboek; die woordeboek is bedoel om te help met probleme betreffende teksproduksie en sal dus gebruik beïnvloed. Hierdie asimmetriese siening sal impliseer dat preskriptief en deskriptief in die praktyk onjuiste teenstellings is. In hierdie artikel word 'n meer konsekwente terminologie aan die hand gedoen wat rekening hou met sowel die funksie van die woordeboek as die verhouding van die woordeboek met die empiriese basis.

Sleutelwoorde: DESKRIPSIE, WOORDEBOEKFUNKSIE, EMPIRIESE BASIS, PRESIES DESKRIPTIEWE WOORDEBOEK, PRESIES PROSKRIPTIEWE WOORDEBOEK, DUIDELIK PRE-

---

SKRIPTIEWE WOORDEBOEK, INTROSPEKSIE, LINGUISTIESE ONDERSOEK, NIE-PRESIES DESKRIPTIEWE WOORDEBOEK, NIE-PRESIES PROSKRIPTIEWE WOORDEBOEK, PRESKRIP-SIE, PROSKRIPSIE, STERK DESKRIPTIEWE WOORDEBOEK, STERK PRESKRIPTIEWE WOOR-DEBOEK, TEKSKORPUS, GEBRUIKERSGERIG, SWAK DESKRIPTIEWE WOORDEBOEK, SWAK PRESKRIPTIEWE WOORDEBOEK, SWAK PROSKRIPTIEWE WOORDEBOEK

1.   Selection

When you explain something to a layman, it is necessary to make some short-cuts or simplifications. So the following situation is no exception:

Peter (a 12 year old boy)**:** What do you do when you write a dictionary? Of course you write it, but how do you find the words for your dictionary and how do you find the information about these words?

Dr. Claes (a lexicographer working on a large monolingual dictionary)**:** There are, as you know, many words in the language, more than you can include in one dictionary. And you can say very much about each word. If you want to, you can write a whole book about one word. It has, by the way, been done several times already. But I want to make just one book with many words. I do it as follows: First I select the words to be explained in the dictionary. Then I explain the meaning of these words. I also want to show how the words are used. It can be done by examples containing the described word, either short examples with a few words or complete sentences.

Peter: I see. But where do you get all these words, meanings and examples from? And how do you select the right words and explanations? I mean, if you have many more words than you want in your dictionary. And if you have many more meanings and examples. How do you decide what to exclude from or to include in the dictionary? Do you have it all in your head before you start working?

Dr. Claes: Of course I have a lot of knowledge about all the words. That is what a lexicographer has to know. But he can't know everything. Therefore he makes investigations and undertakes searches to find out which words are used and how they are used. Nowadays you have a lot of texts in your computer and you have computer programs to help you. You can produce lists with words or with examples, and the computer can help you find out what kind of meanings and uses of the word are most common.

Peter: I see. Isn't that very exhausting and also quite difficult? If you have more words and meanings in your head or in your computer than you want to include the dictionary, how can you be sure that you make the right choice? To me it looks like visiting my aunt. She always makes wonder-

ful cakes, but I am not allowed to take more than one. But perhaps it makes enough sense, because if I were allowed to eat all the cakes, I might get a very painful stomach-ache.

In my opinion, Peter has understood one of the main problems in lexicography. What is the best or the most convenient empirical basis for a certain lexicographical project? And what methods can be chosen in the different selection processes (lemma selection, selection of the main orthographical variant, selection of the grammatical variant, selection of collocations, etc.)? In many ways the lexicographer has the same problem as the boy who wants to eat all the cakes. You might say your problem is bigger because the boy did not bake the cakes himself and did not make the choice himself between baking this or another kind of cake. If you do not have a certain type of text in your text corpus (if you do have a corpus as empirical basis), you will of course not have examples of this specific kind in your dictionary (Bergenholtz and Mugdan 1989). You could then say the dictionary is not correctly descriptive, or you could also say it is prescriptive.

2.   Dictionary functions

It makes no sense however to discuss description contra prescription, before we take a look at what Tarp (2000) calls "the heart of lexicography". The heart of lexicography is the understanding that a dictionary is a tool. And like every tool it has certain functions. A lot of tools have more than one function; they are, like most dictionaries, polyfunctional. The sum of all functions of a given dictionary is called the genuine purpose of the dictionary (Bergenholtz and Tarp 2002). Modern functional lexicographical theory can briefly be described by the following thesis[1]:

—   Lexicography is a discipline in its own right; that means e.g. that lexicography is not a linguistic discipline and not applied lexicology either. On the other hand, specialised lexicography, with its synonym terminography, is part of the discipline lexicography.

Of course, we need linguistic knowledge for a lot of dictionary types, but in the same way we need knowledge about molecular biology preparing an English-Spanish biotechnological dictionary and knowledge about music preparing a monolingual music dictionary.

—   The object of lexicography is dictionaries. That means the object is not the language, but those tools developed by mankind for certain purposes.

Discussing dictionary functions, you need in the first place to consider the following topics:

—   different types of user situations where the consultation of dictionaries can offer help,

—   different types of users distinguished in terms of native language, knowledge of language, general knowledge, specialized knowledge, etc., and

—   different types of problems relevant to certain user types in certain types of user situations.

The most important division of functions is one between communication-oriented and knowledge-oriented functions. Through communication-oriented functions the dictionary is a tool helping to solve problems in an ongoing communication. Through knowledge-oriented functions the dictionary is a tool helping the user to extend his knowledge. Each of these sets of functions can be divided into smaller sets or into single functions, e.g. for solving communicative problems through reception or production of text in the mother tongue or in a foreign language. A further division can be made with regard to different user types and their linguistic or encyclopaedic knowledge. Such a typology of dictionaries, e.g. reception dictionaries and production dictionaries for native language users, or translation dictionaries from native language to foreign language for experienced translators, is much more relevant than descriptive terms such as linguistic dictionaries, encyclopaedic dictionaries or mono- and bilingual dictionaries (Bergenholtz 1998).

But let us return to dictionary functions. The considerations mentioned above are also necessary if you aim at a relevant discussion on what kind of empirical basis you need for a specific dictionary and when you decide whether it is better to compile a descriptive or a prescriptive dictionary.

3.   Empirical basis and dictionary functions

In linguistics there is some but no universal agreement about the use of the terms description and prescription: Normally the first term explains how a language element is used, the second tells you how to deal with a certain language element. This means that description has a clear relation to the empirical basis, prescription to the function of information about language. But besides this very unclear and asymmetric agreement, there are quite different explanations (Bergenholtz 2001). This is not the main topic of this article. Instead of a terminological discussion, I will begin with some data which can be used as part of the empirical basis for a concrete lexicographical project (Bergenholtz 1995). In the following table, are given some results from two linguistic surveys of the Danish language, one with 795, the other with 341 informants. You also find the number of attestations from searching the Internet. In the last column the items from the official Danish dictionary are given. I have chosen orthographical and inflectional variants as examples. But this does not mean that my

arguments are about lexicographical problems regarding inflectional morphology and orthography only. The same arguments could be applied to all kinds of item types in all kinds of dictionary types.

	linguistic survey		Google		official variant
lin**ie** *(line)*	231	68.0%	102,886	71.2%	-
lin**je** *(line)*	109	32.0%	41,572	28.8%	+
lu**dd**er *(prostitute)*	56	16.6%	233	5.8%	+
lu**d**er *(prostitute)*	281	83.4%	3,793	94.2%	+
frå**ds**e *(gorge)*	217	66.2%	412	95.2%	-
frå**s**e *(gorge)*	111	33.8%	21	4.8%	+
kraftvarmeværk *(combined power and heating plant station)*	284	85.0%	5,495	98.3%	-
kraftvarme**-**værk	30	9.0%	90	1.6%	-
kraft**-**varmeværk	6	1.8%	2	0.04%	-
kraft**-**varme**-**værk	14	4.2%	6	0.1%	+
all**ee**n *(the avenue)*	166	20.9%	1,172	69.3%	+
all**ée**n *(the avenue)*	594	74.7%	519	30.7%	+
both	28	3.5%			
don't know	7	0.9%			
grundl**i**ggende *(fundamental)*	249	31.3%	3,220	6.0%	+
grundl**æ**ggende *(fundamental)*	529	66.5%	50,200	94.0%	+
both	7	0.9%			
don't know	6	0.8%			
vani**lle** *(vanilla)*	138	40.8%	2,609	79.3%	-
vani**lje** *(vanilla)*	107	31.7%	381	11.6%	+
vani**lie** *(vanilla)*	93	27.5%	298	9.1%	-

Table 1

In these cases, you have different orthographical variants. Sometimes the variants are used with a clear preference in the linguistic survey and in the Internet texts, e.g. variants *linie* contra *linje* (Eng. *line*) and *ludder* contra *luder* (Eng. *prostitute*). But even in such clear cases, you can see that the normative official Danish Language Council does not decide according to real use: The frequent variant *linje* (Eng. *line*) was prohibited in November 2001, but the infrequent form *ludder* (Eng. *prostitute*) is still allowed. This tendency is especially clear from the very frequent, but prohibited orthographical variant with -d- in *frådse* (Eng. *gorge*) and also from the variant without hyphens *kraftvarmeværk* (Eng. *combined power and heating plant station*). In other cases, you see that the informants decided in a way which differs from the real use in texts, e.g. the variant with or without an accent *alléen* contra *alleen* (Eng. *the avenue*), and with -æ- or

-i- in *grundlæggende* contra *grundliggende* (Eng. *fundamental*). In all these cases, both variants are allowed by the Danish Language Council. And finally you find differences between the answers in the linguistic survey and the real use in texts as shown by the three variants for Eng. *vanilla*, but only one of these variants is allowed by the Danish Language Council. These Language Council decisions are quite variable, in that they change with every new edition of the official dictionary. The Danish Language Council published a new edition in 1996 with quite a number of changes and another in 2001 with even more changes, e.g. from one form to another: *frådse → fråse* (Eng. *gorge*), *kraftvarmeværk → kraft-varme-værk* (Eng. *combined power and heating plant station*), or from two variants to only one: *linie* or *linje → linje* (Eng. *line*), or from one variant to two: *for resten → for resten* or *forresten* (Eng. *incidentally*). This is in contrast with the English-speaking world which does not have such normative traditions.

At least two interesting points can be made with regard to these variants: (a) It is not very clear whether we can call the decisions made by the Danish Language Council prescriptive, because very often they are quite liberal and allow for two or more variants, even if one of these variants is quite infrequent. (b) There is almost no general public discussion pro or contra these decisions, although every pupil, student or government employee is legally bound to use only the allowed orthographical variants.

The same situation applies to the inflection of verbs, nouns and adjectives. In the following, you see a verb with a clear preference for one inflectional variant (*vejledte*, Eng. *instructed*), but the infrequent variant is also allowed together with the frequent one. In the last case, you will find a preference in the linguistic survey for the very frequent variant with the inflection *-det* for singular definite (*postbudet*, Eng. *the postman*), but this variant is prohibited by the Danish Language Council.

	linguistic survey		Google		official variant
vejlede**de** (*instructed*)	73	9.2%	41	9.9%	+
vejled**te** (*instructed*)	710	89.3%	373	90.1%	+
both	6	0.8%			
don't know	6	0.8%			
postbu**ddet** (*the postman*)	96	28.2%	69	6.5%	+
postbu**det** (*the postman*)	244	71.8%	1,434	93.5%	-

Table 2

4.   Description

Collecting and using such data from a linguistic survey and from a text investigation in practical lexicographical work is normally called descriptive lexicography. It can be description, but it need not be. And although the lexico-

graphical results can be called description, this is not exact. You have to distinguish between different ways and different extents of description. It depends on the empirical basis and its use:

(a)   introspection,
(b)   analysis of a linguistic survey,
(c)   involvement of descriptions in existing dictionaries, grammars, monographs, articles, etc.,
(d)   analysis of a number of examples which have been randomly chosen from random texts (corresponding with the practice of dictionary making before the age of computers),
(e)   analysis of a specifically constructed text corpus, and
(f)   analysis of usage found in texts in the examined language in all available web-sites on the Internet.

These possibilities for an empirical basis can be used in different ways and to different extents. In order to deal with these possibilities, we need suitable terms. They are necessary for drawing a distinction between different kinds of lexicographical work:

open description if the outside matter informs about the empirical basis, or
hidden description if the outside matter does not inform about the empirical basis.

Most dictionaries do not exactly inform about the used empirical basis, therefore they use a hidden description. It is possible that a good part of the dictionary users do not read the outside matter. But interested users, especially critical ones and also linguists, really need an open description. Only then is it possible for them in difficult cases to compare the items with their own considerations, particularly if they encounter the lexicographical treatment of variants. As part of an open description the lexicographer has to explain what kind of description he has used, be it his own competence as an empirical basis, or also the investigation of a text corpus etc.:

total description using a combination of description possibilities (a), (b), (c), (d), (e) and (f), or
partial description using only some of the possibilities for description.

You do not often find that all possibilities have been used. The *American Heritage Dictionary* and the *Danske Netordbog* are two of the few examples based on a linguistic survey. (For a theoretical discussion of user survey methods and the representation in dictionaries see Andersen and Bergenholtz 2001.) In the following dictionary articles (see (1)–(10) below), a partial description is given by using possibilities (a), (b), (c) and (f) as empirical basis.

A description, either total or partial, can be given in different ways, depending on the quality of the used empirical basis:

strong description if the items are given after an analysis of a broad empirical basis, e.g. for description possibility (b) a representative sample, or for description (e) an exemplary corpus (a large and broad text corpus), or

weak description if the items are given after an analysis of a narrow empirical basis, e.g. for description possibility (b) a non-representative sample, or for (e) a small corpus, or if the items are given without any kind of references to a broad empirical basis.

All the following dictionary articles are based on a broad empirical basis, but only articles (1)–(9) can be considered as strongly descriptive; the remainder (10)–(12) would be found in a weakly descriptive dictionary. But there are differences between (1)–(9), which can be called different gradations of descriptivity.

As a last differentiation, a distinction must be made between the exclusion of parts of the empirical results or the inclusion of all the empirical results:

explicit description if all results from every part of the empirical basis are part of the dictionary, also from description possibility (c) obsolete words or meanings, or from description possibility (e) mistakes, or

implicit description if only results from description possibilities (d), (e) or (f) are part of the dictionary, e.g. mistakes are not included as items in the dictionary.

Most dictionaries include some of the more infrequent words or meanings, but normally they do not include mistakes. In both cases such an inclusion could be an advantage for the user who wants to know whether he should or should not use a certain word, word combination or meaning. It could also be advantageous for the user situation, if the user has a text reception problem with an older or a rare word or a misused word, e.g. a frequent spelling mistake.

In dictionary articles (1)–(12) you find different examples, all the result of descriptive lexicographical work, but in quite different ways, depending on the kind of descriptiveness. In dictionary articles (1)–(3) you have descriptive information about the inflection of *vejlede* (Eng. *instruct*), all arranged with the infinitive as lemma, followed by the present, the preterite and the past tense:

(1)   **vejlede** verb ⟨-r, vejledte (710 attestations, 373 informants) or -de (73 attestations and 41 informants), ... ⟩
The Danish Language Council allows all inflections mentioned.

(2)   **vejlede** verb ⟨-r, vejledte (89%) or -de (11%); similarly the informants (90% and 10% respectively), ... ⟩
The Danish Language Council allows all inflections mentioned.

(3)   **vejlede** verb ⟨-r, vejledte (frequent) or -de (rare), vejledt (frequent) or -t (rare); most informants also chose *vejledte* and *vejledt*, ... ⟩
The Danish Language Council allows all inflections mentioned.

Using the same kind of differentiation, the lexicographer can prepare diction-ary articles, all descriptive but quite different from each other, depending on the type of lexicographical method used to present the descriptive results. For *kraftvarmeværk* (Eng. *combined power and heating plant station*) you have the same inflection, but different variants with or without hyphens:

(4)   **kraftvarmeværk** noun ⟨et; -et, -er, -erne⟩ (...)
      This spelling without a hyphen is the usual variant (5,495 attes-tations), besides 90 for *kraftvarme-værk*, 2 for *kraft-varmeværk*, 6 for *kraft-varme-værk*, similarly the choice of the informants with 284 votes for *kraftvarmeværk* and 30, 6 and 14 informants respectively for the variants with a hyphen or hyphens. The Danish Language Council allows only the spelling with two hyphens: *kraft-varme-værk*.

(5)   **kraftvarme-værk** → kraftvarmeværk

(6)   **kraft-varmeværk** → kraftvarmeværk

(7)   **kraft-varme-værk** → kraftvarmeværk

(8)   **kraftvarmeværk** noun ⟨et; -et, -er, -erne⟩ (...)
      This spelling is the most frequent (98%), only between 2% and 0.04% for *kraftvarme-værk*, *kraft-varmeværk* and *kraft-varme-værk*, similarly the informants voted with 85% for *kraftvarmeværk* and 9% for *kraftvarme-værk*, 2% for *kraft-varmeværk* and 4% for *kraft-varme-værk*. The Danish Language Council allows only the spelling with two hyphens: *kraft-varme-værk*.

(9)   **kraftvarmeværk** noun ⟨et; -et, -er, -erne⟩ (...)
      This spelling is the most frequent and is also the preferred infor-mant choice. Other spelling variants with one hyphen are rare and with two hyphens very rare. The Danish Language Council allows only the spelling with two hyphens: *kraft-varme-værk.*

The following dictionary articles are also descriptive, but they are examples of a weak description where the lexicographer did not have a broad empirical basis, or did not use the results from such an investigation in the descriptive representation of dictionary articles (10)–(12) for *vejlede* (Eng. *instruct*) and *kraft-varmeværk* (Eng. *combined power and heating plant station*):

(10) **vejlede** verb ⟨-r, vejledte/vejledede, vejledt/-t, ... ⟩
      The Danish Language Council allows all inflections mentioned.

(11) **vejlede** verb ⟨-r, vejledte/-de, vejledt/-t⟩

(12) **kraftvarmeværk** or **kraft-varmeværk** or **kraftvarme-værk** or **kraft-varme-værk** noun ⟨et; -et, -er, -erne⟩

Let us try to look at these dictionary articles from the viewpoint of an imagi-

nary user. He is a teacher of Danish at a secondary school, and his name is Os-kar. Oskar's knowledge of and interest in Danish is on a high level. If he wants to know more about variation in Danish inflectional morphology and ortho-graphy, the strongly descriptive dictionary articles (1), (2) and (8) and (9) are very informative; the remaining less strongly and especially the weakly de-scriptive articles are not relevant for the knowledge-oriented function of a dic-tionary. If Oskar instead is in doubt about the choice of variants when he encounters text production problems, he will probably choose *vejledte* and *kraft-varmeværk* from dictionary articles (1)–(9), because he wants to follow the nor-mal language usage. But perhaps he will not make his choice from dictionary articles (4)–(9), because he wants to follow the Danish Language Council. The weakly descriptive dictionary articles (11) and (12) are not helpful at all when making a choice of a variant. They only inform Oskar to make the choice himself.

5.   Prescription

In the Scandinavian countries, you find some preference for descriptive lexi-cography, but the normal practice is to be more prescriptive, which is also advocated by some metalexicographers. In the USA, you can even find sup-porters of a purely leave-the-language-alone approach:

> It is not the function of a dictionary-maker to tell you how to speak, any more than it is the function of the mapmaker to move rivers or rearrange mountains or fill in lakes. (Urdang 2000: 37)

Neither in specialised lexicography nor in countries like Iceland will it be pos-sible to find many supporters of this approach. The tradition in Africa, e.g. regarding Swahili, but also regarding many other languages, tends towards prescriptive lexicography. As to prescription you have at least the following three possibilities:

(a)   a specific linguistic variant is explicitly prohibited,
(b)   one or more linguistic variants are explicitly prescribed, thus prohibiting all other non-mentioned variants, and
(c)   a specific linguistic variant is explicitly prescribed (as opposed to pre-scription (b) this involves a new word, new spelling, new pronunciation, new inflection or neologism).

All these possibilities are not used in every dictionary article. For other articles you have the same possibilities using the empirical basis for description men-tioned above. A prescriptive dictionary contains prescriptive dictionary arti-cles, but can also contain articles similar to those in e.g. weakly descriptive dic-tionaries. The point is the user cannot know which article corresponds with language usage and which does not. Therefore it will make no sense to say that

some of the articles are descriptive and some prescriptive.[2] Here you may — similarly to the proposals for description — distinguish between different lexicographical labelling techniques:

open prescription if the outside matter informs about its intention to influence and in the end change the language use up to now, or

hidden prescription if the outside matter does not inform about its intention to influence and change the language use up to now.

In a prescriptive dictionary, the method of a hidden prescription is a way to seduce the user if the outer text does not inform about the intention of the prescription. Therefore only an open prescription should be the practice in scientifically based dictionaries.

In the next distinction, the dictionary user is not given such clear advice or instruction, for it depends on the function of the dictionary:

total prescription using a combination of prescription possibilities (a), (b) and (c), or

partial prescription using only one or two of the possibilities for prescription.

The term partial prescription will also be used if a dictionary is only prescriptive according to one or only some of the item types, e.g. only by orthographical items. But here you have a clear classification. This is not the case concerning the following terms with a typological division:

strong prescription if dictionary articles have items that prohibit and only allow certain variants in cases with a clear difference from the normal language use, or

weak prescription if dictionary articles have items that prohibit and only allow certain variants in cases with some but not very clear differences from the normal language use.

With such a typological division you might have had more gradations such as very strongly prescriptive, strongly prescriptive, quite prescriptive, weakly prescriptive and very weakly prescriptive. Here a very simple division with an extreme prescription has been chosen in order to designate it as strong prescription. In this sense dictionary articles (13), (14) and (16) are strongly prescriptive. The prohibited variant has a frequency of more than 95%, or the only allowed variant has a frequency of less than 5%. In dictionary articles (15) and (17), you have a weak prescription, because the prohibited variant is less frequent than 95%, or the only allowed variant is more frequent than 5%.

The terms explicit and implicit are not used in the same way as for description. This is necessary, because you have quite another intention with giving advice or instruction to the dictionary user:

explicit prescription if one variant is explicitly allowed and another is explic-

itly prohibited, or

implicit prescription if one variant is an item and therefore allowed and all
other not mentioned variant implicitly prohibited.

The inflection for preterite *vejledte* (Eng. *instructed*) is not allowed in dictionar-
ies with dictionary articles (13), (14) and (16), but this variant is the frequent
one (90%), similarly the prohibited *kraftvarmeværk* (Eng. *combined power and
heating plant station*). On the other hand, *vejlod* (Eng. *instructed)* and *kraft-varme-
værk* (Eng. *combined power and heating plant station*) are not used in written Dan-
ish texts. But this presentation is possible in prescriptive dictionaries, so they
may differ from real language usage:

(13)   **vejlede** verb ⟨-r, de, -t; not allowed: vejledte, vejledt⟩

(14)   **vejlede** verb ⟨-r, vejlod, vejledt; not allowed: vejledte/vejledede,
vejledt/vejledet⟩

(15)   **kraft-varme-værk** noun ⟨et; -et, -er, -erne⟩
Other spellings without or with only one hyphen are not allowed:
kraftvarmeværk, kraftvarme-værk, kraft-varmeværk

Dictionary articles (13)–(15) are explicitly prescriptive, (16)–(17) implicitly pre-
scriptive.

(16)   **vejlede** verb ⟨-r, de, -t⟩

(17)   **kraft-varme-værk** noun ⟨et; -et, -er, -erne⟩

When Oskar, our secondary school teacher, is not certain about the choice of
variants to solve his text production problems, dictionary articles (13)–(17)
would suit him well. These articles would of course not appear in the same, but
in different dictionaries, and would give different answers. But if Oskar knows
that the dictionary he consults is prescriptive with some items different from
normal language use, he will find what he wants: clear advice for text produc-
tion problems. If Oskar chooses such a dictionary for reception problems, it is
quite useful too. In dictionary articles (13)–(16), he will find orthographical and
inflectional variants and from there arrive at the correct lemma, and in addition
will find some hints about the quality of the texts he reads. He will then find
that the texts are faulty because they often use prohibited variants. About
knowledge-oriented functions, he will not obtain much information, only about
the decisions of the dictionary makers. From the dictionary articles at least he
cannot learn why one variant is prohibited or why another one is preferred.

6.   Proscription

With regard to proscription, you have the same possibilities for the empirical

basis as with description:

(a)   introspection,
(b)   analysis of a linguistic survey,
(c)   involvement of descriptions in existing dictionaries, grammars, mono-
      graphs, articles, etc.,
(d)   analysis of a number of examples which have been randomly chosen
      from random texts (corresponding with the practice of dictionary mak-
      ing before the age of computers),
(e)   analysis of a specifically constructed text corpus, and
(f)   analysis of usage found in texts in the examined language in all available
      web-sites on the Internet.

If a lexicographer chooses to be proscriptive, the way he deals with the results
of the empirical analysis differs from that of being descriptive. The lexicogra-
pher does not stop after having informed the user about language use, he also
wants to tell the user which variant he should choose if there is more than one
possibility. This lexicographical method of presentation can be called **proscrip-
tion**,[3] in practice it is a selective description. Unlike description where every
single case is invented individually and described on its own, proscription
throughout compares similar cases systematically and makes the same deci-
sions in each case. Proscription does not include all elements from the empiri-
cal basis in the dictionary items. The terminological differentiation is similar to
that of description:

open proscription if the outside matter informs about the used empirical basis,
      or
hidden proscription if the outside matter does not inform about the used em-
      pirical basis;

total proscription using a combination of proscription possibilities (a), (b), (c),
      (d), (e) and (f), or
partial proscription using only some of the possibilities for proscription;

strong proscription if the items are given after an analysis of a broad empirical
      basis, or
weak proscription if the items are given after an analysis of a narrow empirical
      basis, or if the items are given without any kind of references to a broad
      empirical basis; and

exact proscription if only one variant is recommended; other variants can be
      mentioned, or
not exact proscription if more than one variant is recommended; other variants
      can be mentioned.

In the following dictionary articles, you have exact proscription in (18) and (19)
and not exact proscription in (20)–(22) (the examples are the same: *vejlede* (Eng.

According to the function of the different kinds of descriptive, prescriptive and proscriptive dictionaries, this can be explained in the following summary:

Strongly descriptive dictionaries are well suited for knowledge-oriented functions, but also for communication-oriented function reception. They cannot easily be used for text production functions, because they often do not give clear advice, but a lot of confusing information.

Weakly and not exactly descriptive dictionaries are not well suited for any communication- and knowledge-oriented functions.

Exactly the same can be said about **weakly and not exactly proscriptive dictionaries:** they are not well suited for any communication- and knowledge-oriented functions.

Exactly proscriptive dictionaries are well suited for all kinds of communication-oriented functions in text production.

Strongly prescriptive dictionaries are not well suited for any dictionary functions. It is a question of language policy whether the decision-makers in a language society decide to choose such a lexicographical solution.

Explicitly and weakly prescriptive dictionaries can be well suited for text production in connection with specialised languages. They can be used successfully as part of national and international language politics. But for this function, proscriptive dictionaries are better suited, because they do not differ from current language usage.

A more consistent lexicographical terminology is described in the following table, where the labelling "+ −" in the last row means that the analysis of an empirical basis can, but might not have taken place, and that in some articles there are differences between the items and the normal language use, but perhaps not always (although the user cannot know):

	empirical basis	accordance with empirical basis	intention to influence the user
descriptive dictionary	+	+	−
proscriptive dictionary	+	+	+
prescriptive dictionary	+ −	+ −	+

Table 3

Endnotes

1. Modern functional lexicographical theory is quite different from old-fashioned theory, e.g. cf. Wiegand (2001). Such theories are primarily based on the outer type of dictionary, not on the user and the use of dictionaries.

2.   It could only be done for a dictionary with special labelling for descriptive and for prescriptive articles. No such dictionary is known for any language.

3.   The term **proscriptive** is formed according to Latin *proscribere* 'make public'. The word *proscriptive* is already in use in English with the meaning 'forbidden', compare *a proscriptive law* or *a proscriptive statement*. In order to avoid a conflict Gregory James has proposed the use of **praeterscriptive** instead of **proscriptive**. It could indeed be a term for proposal items in lexicography, but **proscriptive** is already used in quite a lot of lexicographical articles. A terminological change could lead to new misunderstandings. Besides, it is quite common that you find a specific use of a certain word in ordinary language and another use of the same orthographical word in language for special purposes. Therefore, I do not follow the proposal from James.

Bibliography

Andersen, Henning and Henning Bergenholtz. 2001. Brug af surveys til lingvistiske undersøgelser. *Hermes* 27: 201-209.

Bergenholtz, Henning. 1995. Material for the Dictionary. Bergenholtz, Henning and Sven Tarp (Eds.). 1995. *Manual of Specialised Lexicography. The Preparation of Specialised Dictionaries*: 90-96. Amsterdam: Benjamins.

Bergenholtz, Henning. 1998. Das Schlaue Buch. Vermittlung von Informationen für textbezogene und textunabhängige Fragestellungen. Zettersten, Arne, Jens Erik Mogensen and Viggo Hjørnager Pedersen (Eds.). 1998. *Symposium on Lexicography VIII. Proceedings of the Eighth International Symposium on Lexicography May 2–5, 1996 at the University of Copenhagen*: 93-110. Tübingen: Niemeyer.

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.). 2002. *Sprache im Alltag. Beiträge zu neuen Perspektiven in der Linguistik. Herbert Ernst Wiegand zum 65. Geburtstag gewidmet*: 499-520. Berlin/New York: De Gruyter.

Bergenholtz, Henning and Joachim Mugdan. 1989. Korpusproblematik in der Computerlinguistik: Konstruktionsprinzipien und Repräsentativität. Bàtori, Istvàn S., Winfried Lenders and Wolfgang Putschke (Eds.). 1989. *Computional Linguistics. An International Handbook on Computer Oriented Language Research and Applications. / Computerlinguistik. Ein internationales Handbuch zur computerunterstützten Sprachforschung und ihrer Anwendungen*: 141-149. Berlin/New York: De Gruyter.

Bergenholtz, Henning and Sven Tarp. 2002. Die moderne lexikographische Funktionslehre. Diskussionsbeitrag zu neuen und alten Paradigmen, die Wörterbücher als Gebrauchsgegenstände verstehen. *Lexicographica* 18: 253-263.

Tarp, Sven. 2000. Theoretical Challenges to Practical Specialised Lexicography. *Lexikos* 10: 189-208.

Urdang, Laurence. 2000. Review of R.R.K. Hartmann and Gregory James. *Dictionary of Lexicography. International Journal of Lexicography* 13(1): 35-42.

Wiegand, Herbert Ernst. 2001. Was eigentlich sind Wörterbuchfunktionen? Kritische Anmerkungen zur neueren und neusten Wörterbuchforschung. *Lexicographica* 17: 217-248.

Linguistic Variation in Shona with Special Reference to Monolingual Dictionaries[*]

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

Abstract: This article focuses on the problems lexicographers of monolingual dictionaries face when dealing with a language characterised by differences associated with geographical variation. The article specifically seeks to explore problems with which Shona lexicographers are confronted when working on monolingual Shona dictionaries. It does so by looking at some lexicographic aspects such as tone marking, sense ordering and treatment of synonyms and variants. Linguistic variation has always been a problem for writers of reference works, especially those normative in nature and function. The challenge is that of representativeness, that is, the production of a work which incorporates all the information it is supposed to provide. If the work fails to represent the language of the total population it is assumed to cover, then it is prone to criticism, usually by those whose dialect is not well represented. The article contains the writer's experience in the African Languages Lexical (ALLEX) Project, at present the African Languages Research Institute (ALRI). The ALRI, accommodated at the University of Zimbabwe, has the compilation of monolingual dictionaries for the indigenous languages of Zimbabwe as one of its major objectives. Although illustrative examples are drawn from Shona alone, it is, however, believed that the problems highlighted in this article may not be peculiar to Shona, but can also apply to other Bantu and world languages.

Keywords: ALLEX PROJECT, ALRI, CORPUS, DIALECT, MONOLINGUAL LEXICOGRAPHY, SHONA, SUBDIALECT, SYNONYM, TONE, VARIANT, VARIATION

Opsomming: Linguistiese variasie in Sjona met spesiale verwysing na eentalige woordeboeke. Hierdie artikel konsentreer op die probleme waarvoor leksikograwe van eentalige woordeboeke te staan kom wanneer hulle te doen kry met 'n taal gekenmerk deur verskille wat verband hou met geografiese variasie. Die artikel wil veral probleme ondersoek waarmee Sjonaleksikograwe te doen kry wanneer hulle aan eentalige Sjonawoordeboeke werk. Dit word gedoen deur na 'n aantal leksikografiese aspekte soos toonaanduiding, betekenisordening en die behandeling van sinonieme en variante te kyk. Linguistiese variasie was nog altyd 'n probleem vir die skrywers van naslaanwerke, veral dié wat normatief van aard en funksie is. Die uitdaging is dié van verteenwoordigendheid, dit wil sê die totstandbrenging van 'n werk wat al die inligting bevat wat dit veronderstel is om te verskaf. As die werk in gebreke bly om die taal van die hele bevolking te verteenwoordig wat dit aangeneem word om te dek, dan is dit vatbaar vir kritiek,

---

gewoonlik deur diegene wie se dialek nie goed verteenwoordig is nie. Die artikel bevat die skrywer se ervaring in die African Languages Lexical (ALLEX) Project, tans die African Languages Research Institute (ALRI). Die ALRI, gehuisves by die Universiteit van Zimbabwe, het as een van sy belangrikste doelwitte die samestelling van eentalige woordeboeke vir die inheemse tale van Zimbabwe. Alhoewel verduidelikende voorbeelde slegs aan Sjona ontleen word, word daar egter geglo dat die probleme wat in hierdie artikel na vore gebring word, nie net eie is aan Sjona nie, maar ook geldig is vir ander Bantoe- en wêreldtale.

Sleutelwoorde: ALLEX PROJECT, ALRI, KORPUS, DIALEK, EENTALIGE LEKSIKOGRAFIE, SJONA, SUBDIALEK, SINONIEM, TOON, VARIANT, VARIASIE

1.   Introduction

Researchers who carry out quantitative and qualitative research in various disciplines often refer to the term *variation*. However, this term means different things in different fields. In this article it is asserted that in the field of language, *variation* refers to systematic differences existing between different languages or between varieties of the same language. No language can be described as homogeneous; every language is characterised by a variation in expression. In fact, a simple observation of any language would show that no two speakers of a language speak in exactly the same way.

The issue of language variation can be looked at from different perspectives and at different levels. At the individual level, such factors as mood, health, age, attitude and context can lead a person to use different forms of language to refer to the same thing at different times. At group level, differences can, among others, be determined by social grouping or geographical location. Some factors such as differences in sex, age, social status and region of origin are critical in explaining the differences existing between a variety used by one group of people and that used by another. Whatever the dominant factor, the differences present can be realised at all levels of linguistic structure, that is, phonological, lexical and grammatical levels.

For a fuller understanding of the forces at work, it is usually more appropriate to discuss the joint factors leading to language variation. However, such a large-scale discussion is beyond the scope of this article which narrowly focuses on regional variation, that is systematic differences among groups of speakers of the same language coming from different geographical locations. According to Francis (1983: 15), regional variation results from differential change. Francis' view is that language change is inevitable and that when changes occur, they characteristically affect the speech of only part of the total language community. For example, different regions of a language community, by virtue of differences in their geographical location, would experience different linguistic changes at different times in the history of their language.

Fromkin and Rodman (1993: 277) also note that dialectal diversity develops when speakers of the same language are separated from each other geographically or socially. In this case, the changes that occur in the language spo-

ken in one area or within one group of people do not spread to another. When some communication barrier, be it a physical barrier such as an ocean or a mountain range, or a social barrier of a political, racial, class or religious kind, separates groups of speakers, linguistic changes are not easily spread and dialectal differences are reinforced. A change that occurs in one region and fails to spread to other regions of the same language community gives rise to dialectal differences which also tend to increase proportionately to the degree of communicative isolation among the groups of people. Fromkin and Rodman (1993: 277) argue that when such differences give the language spoken in a particular region its own flavour, this version of the language is referred to as a regional dialect.

The issue of regional dialects is one of many aspects that lexicographers have to treat with great care. This is because their products have a readership that usually cuts across different varieties or dialects. The challenge for the lexicographer is to produce a work that would accommodate the differences among the respective varieties without a bias towards one or the other variety. In the subsequent sections of this article, we will examine the dialect situation in Shona and see how it has added to the general problems Shona lexicographers of monolingual dictionaries encounter in their work.

2.   Dialect Situation in Shona

Shona is a Bantu language spoken in Zimbabwe by approximately 80 percent of the country's total population. The language is mainly spoken in the central, southern, northern and eastern parts of the country. As in any other language, there is dialectal variation in Shona. In fact, taking geographical distribution as the determining factor, Shona is understood to consist of five major regional dialects: Zezuru, Karanga, Manyika, Ndau and Korekore, each of which can also be divided into numerous subdialects. Zezuru is spoken in the central region of the country, in areas surrounding Harare, the capital city. Karanga is spoken in the southern parts, Manyika in the east, Ndau in the southeastern parts and Korekore in the northeastern parts of the country. Although these dialects have a high degree of mutual intelligibility, they show systematic differences. The varieties differ in pronunciation, vocabulary and, at times, even syntactically.

The regional names under which the respective varieties of Shona are known today are labels that came with the colonisation of Zimbabwe by British settlers at the end of the 19th century. In order to appreciate the present dialectal situation in Shona, one has to understand the history of the Shona language from the arrival of missionaries in Zimbabwe to the present day. Because a more comprehensive discussion of this topic is given in Chimhundu (1992), only a brief overview would suffice here. But before we consider the Shona situation, it may be necessary to note a general observation about Africa made by Chimhundu (1992: 88) who says:

> An exaggerated multi-ethnic, multilingual and multi-tribal picture of the African colonies has been painted through misinterpretation and inadequate study on the part of the early missionaries and through manipulation for administrative convenience on the part of the colonial governments.

This point is important because it tries to capture the processes that brought Africans (including the Shona) into their present sociopolitical and economic situation. The exaggeration of differences rather than similarities among people has tended to separate rather than unite people with a common language, history and destiny. Divisions created during the colonial era have caused people to identify themselves as belonging to particular groups, a development which makes co-operation and compromise among these groups more difficult and at times even impossible. These divisions have also tended to complicate the task of those who strive towards unifying the different groups or those who work on projects cutting across the established boundaries.

To return to the case of Shona: Before the arrival of missionaries, there were no known distinctions based on language in Mashonaland, that is, in the area of Zimbabwe where Shona is spoken. The people of this region defined themselves politically as subjects of particular chiefs coexisting with other chiefdoms in unlimited zones of a common culture (Chimhundu 1992: 91). At that time, they did not identify themselves according to language or dialect because, since they considered themselves as speaking the same language despite the minor differences that existed, the linguistic factor was not important. However, when the missionaries came to Zimbabwe during the last quarter of the 19th century, they tended to view all chiefdoms as tribes and also simplistically equated these to tribes speaking different languages or dialects. As noted by Chimhundu (1992: 91), this identification was more convenient for the missionaries, and later for the colonial government since it meant they had to deal with fewer chiefs with territorial claims over wider areas. For the missionaries, it was easier to have dealings with chiefs who would, in turn, become agents in spreading the gospel to their respective subjects rather than taking their own initiatives without the chiefs' backing. These missionaries, therefore, naturally helped in the creation of wider chiefdoms or groupings of chiefdoms, with tribal labels more or less corresponding with the spheres of influence of different missionary societies on the one hand, and the administrative districts and provinces of the settler government on the other (Chimhundu 1992: 91).

One important point to note about the missionaries who came to Zimbabwe is that they were not a single society operating in one area. According to Kahari (1996: 3), different missionary groups strategically deployed themselves to all four corners of the region: the American Methodists among the Ndau at Chirinda and the Manyika at Umtali (present-day Mutare); the Anglicans among the Manyika at St Augustine; the British Methodists among the Zezuru at Waddilove; and the Roman Catholics among the Zezuru at Chishawasha and Kutama. The Dutch Reformed Church of South Africa established itself among the Karanga at Zimuto, while the Salvation Army settled among the

Zezuru at Howard in the Chiweshe tribal area. Operating from these different mission stations, the missionaries set up presses for printing church literature, Bibles and educational books to create literature for the new converts who had become literate (Kahari 1996: 3).

As in most parts of Africa, formal education was seen as a strategic instrument in spreading the Christian gospel in Zimbabwe. By bringing children to mission schools, missionaries were able to teach their Christian message to them at an early age, away from interference by elders who were more difficult to convince and convert. At most mission stations in different regions of the country, mission schools were established. Students were recruited and converted. Later on, the same students were used as agents in spreading the religious teachings to their parents and other people in their local communities. However, the important point to note is that the missionary efforts did not lead to a co-ordinated educational policy. The result was that each missionary society established a different system of education for its schools. They could not agree on a common system of writing. Each group, therefore, developed its own form of Shona orthography based on the dialect of the region in which it operated. In the end, different orthographies were thus developed in the different regions. The result of these developments was the creation of dialect clusters which we have already identified as Zezuru, Karanga, Manyika, Ndau and Korekore.

The Shona people who, before the coming of the Europeans, identified themselves collectively, began to see themselves as belonging to specific dialect areas or regions. This state of affairs was later reinforced by the colonial government employing the divide-and-rule tactic as one of its strongest tools for separating the Africans who could otherwise be a formidable force were they allowed to continue identifying themselves collectively. The colonial regime strengthened the artificially created differences at the expense of the overriding and self-evident similarities. This situation was allowed to continue up to the late 1920s.

From 1903 to 1928 there was a feeling among different missionary societies that a common Shona orthography, which they all needed, should be developed. This issue was kept on the agenda of the Southern Rhodesia Missionary Conference (S.R.M.C.), an association of the various denominations working in the country. Because they could not reconcile the orthographies their constituent members had developed independently of one another, efforts by members of the S.R.M.C. in effecting a common orthography were not successful. Church representatives participating in the discussions took the stance of provincial and tribal defenders (Chimhundu 1992: 99). Following a request by the S.R.M.C., the colonial regime in Southern Rhodesia (now Zimbabwe), in 1929, invited C.M. Doke, a language expert from South Africa, to come and study the language situation in the country and then to make recommendations for the development of the country's languages, including Shona. After a survey of the language situation, Doke in 1931 published his *Report on the Unification of the Shona Dialects* in which he recommended the creation of Union Shona by com-

bining the varieties identified in the Shona language. Of interest with regard to this article is the fact that some dialects were recognised but the use of others for creating Union Shona were discouraged. For example, one of Doke's recommendations was that Union Shona was to draw its vocabulary from Zezuru, Karanga, Manyika and Ndau. Korekore words were supposed to appear only sparingly whilst the inclusion of words from Budya and all the other smaller dialects were discouraged (Doke 1931: 81). For political, economic and geographical reasons, Zezuru was elevated to become the point of reference or the core when creating Union Shona. It was chosen, for example, as the norm for phonetic analysis. In written Shona, Zezuru, whose subdialects were fewer, was to be preferred above all the other varieties. This development elevated Zezuru to a position where it gained prestige, prominence and eventually dominance over the other dialects. The result was a diglossic situation where Zezuru was the high variety and all other dialects were treated as low varieties.

After Doke's recommendations, a standard Shona orthography based on Zezuru was encouraged and even promoted. It is, however, important to note that this development, paralleled with the discouragement of the use of vocabulary from other dialects, did not go without opposition from speakers of other dialects. Chimhundu (1992: 83) notes that the recommendation by Doke to create Union Shona based on Zezuru was opposed by speakers of other dialects from as early as the 1930s. Even today, some people resent Zezuru dominance which is still evident in literature written in Shona. For reasons of ethnic identity, non-Zezuru speakers would want to see their vocabulary reasonably represented in all forms of literature which are national in nature and character. On the other hand, Zezuru speakers are quite happy to see a continued promotion of their variety to the extent of having it as the only variety for standard Shona. Evidence for such feelings can be found in Kahari (1996: 7), who had this to say soon after the publication of Chimhundu's *Duramazwi ReChiShona* in 1996: "I have fears that the pendulum has swung from one extreme, Zezuru, to the other, Karanga, and the other dialects in between have suffered." Kahari here regrets the attempt by the compilers of *Duramazwi ReChiShona* to try and balance inputs from Zezuru and the other dialects. For Kahari, Zezuru's dominance should stay. Any attempt to put the dialect on the same level with other dialects he considers retrogressive. On the other hand, a field research trip carried out by the ALLEX Project in Manicaland in 1996 showed that speakers of Manyika and Ndau dialects felt the kind of domination that would result from the continued use of Zezuru in written Shona literature at the expense of the other dialects, would not only be linguistic, but also social, economic and even political. Because of this, they strongly resent this domination.

From the discussion above, we can see that the Shona language situation as it exists today is heterogeneous and the social situation is fluid (Chimhundu 1979: 78). It is in this situation that the lexicographer of monolingual Shona dictionaries finds himself. An analysis of this situation would show that Shona lexicographers are bound to have problems, some of which emanate from the structure of the language and some from the social situation described above.

In the next section, we will look at some of the problems the team of lexicographers compiling Shona dictionaries has so far encountered.

3.   Challenges for the Lexicographer of Monolingual Shona Dictionaries

Dictionary users consider dictionaries as authoritative reference works. They are believed to be accurate representations of the languages they describe. They are also seen as instruments for describing languages as wholes and not as parts of wholes. Perhaps this is the reason why they have an authoritative influence on aspects of language growth and development. One of the requirements a dictionary, especially a monolingual one, should strive to fulfil is capturing the way a specific language is used by its total speech community. In this case, the role of the lexicographer of a monolingual dictionary would be that of producing a dictionary that unites rather than separates people. A dictionary which excludes vocabulary from other dialects or which emphasises vocabulary from a selected dialect or dialects cannot serve this very important function. The reason is that speakers from ill-represented varieties would find it difficult to identify with the dictionary. This would also have a negative impact on the marketing or acceptance of the dictionary, as well as on the general development of the language involved. Given the nature of the responsibilities of the lexicographer of a monolingual dictionary, the process of compiling a dictionary which balances variations from different dialects of a language is not an easy task. In this section, we will explore the problems lexicographers face in compiling monolingual Shona dictionaries which are representative of the varieties of the language, especially given the fluid social situation characterising the language.

Monolingual Shona lexicography started with the launch of the ALLEX Project in September 1992 when a group of lexicographers undertook the compilation of monolingual dictionaries for the indigenous languages of Zimbabwe. Their efforts have so far culminated in the publication of *Duramazwi Re-ChiShona* (Chimhundu 1996) and *Duramazwi Guru ReChiShona* (Chimhundu 2001), both of which are medium-sized general monolingual Shona dictionaries. During the compilation process, the team has encountered a number of problems rooted in dialectal variation. One such problem concerns the treatment of variants. Variants are taken here to refer to headwords meaning the same but differing slightly in pronounciation. The pronunciation usually differs with respect to one syllable of the word forms concerned, and is usually correlated with dialectal differences. Examples of Shona variants include the following pairs:

shuro	vs	*tsuro*	(hare)
nzara	vs	*zhara*	(hunger)
-famba	vs	*-hamba*	(walk/travel)
-dya	vs	*-ja*	(eat)
chii	vs	*chinyi*	(what?)

Both words forming a pair of such variant forms would appear as dictionary entries. However, because the word forms basically mean the same, they cannot both be given definitions for this would occupy a lot of space without adding anything to the explanation of the entries involved. A more economic way of dealing with such headwords would be to make one of each pair the main headword entry carrying the sense(s). The second word form would not carry a definition, but would just be cross-referred to the main entry. We can illustrate this with the example of *shuro* and *tsuro* given above. If, for example, we decide to take *tsuro* as the main entry, we would present the definition(s) under it. We would also indicate that *shuro* is a variant form of *tsuro*. When we come to *shuro*, information about tone and word class would be given but for the definition(s), the dictionary user would be referred to *tsuro*. The problem with this kind of presentation is to decide which word form to make the main entry, since by making one of the pair the main entry it would appear as if it has been given some superiority over the other.

The principle so far adopted in Shona dictionaries is to use the more common form as the main headword. However, the immediate question that arises is how to determine the more common form of the two. Relying on one's intuition in making decisions on such cases does not always work, for what may be common to one person or a group of persons may not necessarily be common to another person or group. In such cases corpora would be useful. The form with more occurrences in the corpus would automatically become the main entry. However, this is not as simple as it appears on the surface. The reality is that the usefulness of a corpus depends on a number of factors, for example, on whether the corpus being used can be relied on as truly representative of language use as it applies to all the regional dialects of the language concerned. If the corpus used is biased towards one or more dialects, the picture that would emerge may be misleading.

The challenge of making an informed choice between two word forms that should be cross-referred to each other is also encountered when dealing with synonyms. Synonyms are taken here to refer to two or more different words which mean the same, or rather, which refer to the same object or event. Unlike in the case of variants where differences in terms of form are minimal, synonymous words differ greatly with regard to form. Their relationship lies in their similarity of lexical meaning. Examples of Shona synonyms include the following:

gudo	vs	*dede*	vs	*bveni*	vs	*diro*	(baboon)
-mhanya	vs	*-rumba*	vs	*-gogoma*			(run)
-konza	vs	*-nyenga*	vs	*-pfimba*			(declare love)
hwahwa	vs	*doro*	vs	*mhamba*			(beer)
chibage	vs	*barwe*	vs	*bonore*	vs	*gwere*	(maize cob)

The treatment of synonyms in Shona dictionaries is almost the same as that of variants. The more or most common of the synonymous words would carry the

meaning and would become the main entry. The other(s) would be cross-refer-red to it. However, as noted in the case of variants, the problem is that of choosing the main entry and the ones to be cross-referred. We can take as example the Shona verbs -*mhanya* and -*rumba* given above. Whilst -*rumba* might be common to a Ndau speaker, it may not be common to speakers of the other dialects. Although -*mhanya* seems to be more common to speakers of more than one dialect (also according to the Shona corpus where it appears 874 times as compared to the 29 occurrences of -*rumba*), making it the main entry may meet resistance from speakers who do not use it in their everyday language. In fact, the concept of making the more common word form the main entry is not acceptable to speakers of smaller dialects whose representation in the corpus is usually limited. This can be explained mainly by the fact that a large part of the Shona corpus comes from written materials, and the written literature in these dialects is very small. To speakers of these dialects, a strict following of what is in the corpus would appear to be just a way of consolidating the prestigious position of those dialects that have for a longer time been used in writing. De-spite the fact that their vocabulary is included in the dictionary, they may feel that it is playing a secondary role if most of the headwords are not made main entries.

The arrangement of senses also poses a challenge, especially if the senses come from different dialect areas. Often there are cases where a headword has more than one sense. As far as the arrangement of the senses is concerned, the principle adopted for Shona dictionaries has been to list the more commonly used or the literal sense first and the rarely used or the metaphorical sense(s) subsequently. To illustrate this, we can take the Shona example -*bikira* which has two senses, that is, (a) prepare food for someone and (b) prepare a love potion for a beloved. When the above principle is applied, the first of these two senses would be listed as the first sense in the dictionary because its meaning can easily be deduced by combining the meanings of the main verb -*bik*- (cook) and that of the applied extension -*ir*-. The second sense, which is metaphorical, would consequently be listed as a second sense. It is important, however, to note that some senses tend to be more common in some areas and in some contexts than others. Also, what may be regarded as a metaphorical sense in one region may actually be a general and basic sense in another region. As illustrative example we can use the verb -*mamisa* which can mean (a) cause someone to defecate or (b) beat someone thoroughly. Whilst the second sense of this verb may be regarded as metaphorical by speakers of other dialects, to Ndau speakers it is the general and basic sense. In such cases, lexicographers would end up using their intuition, a practice which may not reflect the language use of most Shona speakers.

Another problem is linked to the marking of tone. One function of a dic-tionary is to provide the correct way the words of a language are pronounced. In fact, researches carried out in several languages have shown that tone mark-ing in dictionaries is not only important to the mother-tongue speakers of a

language but is also important to second-language speakers, since they have to learn how to pronounce individual words correctly. To Shona lexicographers, tone marking is a big challenge because there often are cases where one word form is pronounced differently across regional dialects. The following can be used as illustrative examples:

sekuru	*(HHH/LHL)*	(grandfather/uncle)
-fashaidza	*(L/H)*	(boil water or something else)
barwe	*(HH/HL)*	(maize cob)

As can be seen from these examples, a word form can be pronounced with more than one tone. In trying to accommodate this kind of variation, the Shona team has resorted to the policy of marking the different tones that can be realised with a particular headword. The problem of ignorance, however, sometimes arises. Although a lexicographer is expected to know everything about the language with which he is working, this is not the case. Sometimes he may not be aware of the second or third pronunciation of a certain word. To give such a word a single tonal realisation may not accurately capture the situation that pertains in all dialect areas, resulting in a deficiency in the dictionary.

In this section, we have taken examples from Shona to try and show how linguistic variation within a language can be problematic in the compilation of a dictionary. Issues such as synonymy and tone marking have been explored and some problems Shona lexicographers encounter when dealing with these issues, have been singled out.

4.   Conclusion

In this article, we have tried to show some of the difficulties Shona lexicographers face as a result of working with a language characterised by variation. The major problem that has been highlighted is that of trying to compile a dictionary balancing inputs from different dialects. Besides the technicalities involved in the compilation process, the article has shown that some problems concern the attitude towards one dialect or another. The article has also indicated the need for well-informed principles when making choices between pairs or among groups of words.

Space has also been devoted to the discussion of the dialectal situation in Shona. By using the example of Shona, it has been shown that differences existing between speakers of the same language are not always natural. Some have been imposed from outside. To emphasise such differences can only impede research such as dictionary compilation which cuts across different regions or dialects. Although examples have been drawn from one language, the problems discussed and the arguments put forward in this article can apply to almost all languages since almost all languages show variation.

References

Chimhundu, H. 1979. Some Problems Relating to the Incorporation of Loan Words in the Lexicon. *Zambezia* 7(1): 75-91.

Chimhundu, H. 1992. Early Missionaries and the Ethno-linguistic Factor during the 'Invention of Tribalism' in Zimbabwe. *Journal of African History* 33: 87-109.

Chimhundu, H. (Ed.). 1996. *Duramazwi ReChiShona*. Harare: College Press.

Chimhundu, H. (Ed.). 2001. *Duramazwi Guru ReChiShona*. Harare: College Press.

Doke, C.M. 1931. *Report on the Unification of the Shona Dialects*. Hertford: Stephen Austin.

Francis, W.N. 1983. *Dialectology: An Introduction*. London/New York: Longman.

Fromkin, V. and R. Rodman. 1993. *An Introduction to Language*. New York: Harcourt Brace Jovanovich

Kahari, G.P. 1996. Review of *Duramazwi ReChiShona*. *The Herald*, 28 November 1996 and 5 December 1996, *Features Section*, pp. 5 and 12, and p. 4 respectively.

Aspekte van mikrostrukturele verskeidenheid en inkonsekwentheid in woordeboeke

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

Opsomming: Dit is belangrik in enige woordeboek dat die mikrostrukturele aanbod sistematies en konsekwent moet wees. Gemotiveerde afwykings van die verstekstruktuur moet aan die gebruiker verklaar word. In hierdie artikel word aandag gegee aan verskillende tipes mikrostrukturele modelle wat in die metaleksikografie ontwikkel is, met die klem op die geïntegreerde en die niegeïntegreerde mikrostruktuur. Voorstelle word gedoen vir aanpassings in die niegeïntegreerde mikrostruktuur om 'n beter inligtingsoordrag tot gevolg te hê. Aan die hand van voorbeelde uit enkele Afrikaanse woordeboeke word gewys op die gevarieerde en inkonsekwente mikrostrukturele aanbod wat gevolg word. Die nadele van gemengde primitiewe mikrostrukture word bespreek en daar word aandag gegee aan die implikasies daarvan vir die adresseringstruktuur van 'n woordeboek. In hierdie verband word verskillende prosedures van afstandsadressering bespreek. Die teksverdigting wat dit meebring en die probleme wat gebruikers met die koördinering van koteks-inskrywings en betekenisparafrases het, word beklemtoon. Aanbevelings word gedoen vir 'n mikrostruktuur wat groter erkenning gee aan die eise van 'n gebruikersgedrewe leksikografiese proses.

Sleutelwoorde: ADRESSERINGSTRUKTUUR, AFSTANDSADRESSERING, ARTIKEL-STRUKTUUR, GEÏNTEGREERDE MIKROSTRUKTUUR, GERIGTE AFSTANDSADRESSERING, IMPLISIETE AFSTANDSADRESSERING, KONTEKS, KOTEKS, MIKROSTRUKTUUR, NIE-GEÏNTEGREERDE MIKROSTRUKTUUR, PRIMITIEWE MIKROSTRUKTUUR, RUDIMENTÊRE MIKROSTRUKTUUR, SEMANTIESE SUBKOMMENTAAR, SOEKSONE

Abstract: Aspects of Microstructural Diversity and Inconsistency in Dictionaries. In any dictionary it is important that the microstructural presentation should be systematic and consistent. Motivated deviations from the default structure should be explained to the user. This article looks at different types of microstructural models developed within the field of metalexicography, focusing on the integrated and non-integrated microstructure. Proposals are made for changes to the non-integrated microstructure in order to enhance the quality of information transfer. With the aid of examples from a number of Afrikaans dictionaries the varied and inconsistent nature of the microstructural presentation is indicated. Disadvantages of a mixed primitive microstructure are discussed, with special reference to the implications for the addressing structure. In this regard different procedures of remote addressing receive attention. The resulting textual condensation and problems users experience with the co-ordination of cotext entries and paraphrases of meaning are emphasised. Suggestions are made for a microstructure which gives a better reflection of the demands of a user-driven lexicographic process.

1.   Inleidend

In 'n woordeboek as draer van tekssoorte is die sentrale lys, ook bekend as die woordelysteks of, in algemene woordeboeke, die alfabetiese deel, die teksafdeling wat verreweg die meeste deur teikengebruikers van die betrokke woordeboek geraadpleeg word. Dit is dan veral ook hierdie teks wat 'n wesenlike rol daarin speel dat die werklike doel (vgl. Wiegand 1998) van die betrokke woordeboek verwesenlik word. Binne die teorie van woordeboektekste bestaan die sentrale lys uit 'n aantal artikeltrajekte wat elk die versameling lemmata bevat wat uit een alfabetletter as behandelingseenhede gekies is. Binne enige artikeltrajek kan deeltrajekte onderskei word wat telkens uit 'n versameling opeenvolgende artikels bestaan. Elke artikel vertoon 'n bepaalde artikelstruktuur en dikwels ook 'n mikro-argitektuur (vgl. Bergenholtz, Tarp en Wiegand 1999).

Die aard en omvang van verskeidenheid in die leksikografie is nie net beperk tot woordeboektipologie, die datatipes of die dataverspreidingstruktuur nie maar dit blyk ook uit die artikelstruktuur van 'n woordeboek. Hier gaan dit onder meer daaroor dat verskillende artikeltipes onderskei kan word en alhoewel daar meestal gepoog word om die aanbod in 'n woordeboek so konsekwent en eenvormig moontlik te hou, word daar toenemend erkenning gegee aan die geldigheid en waarde van 'n leksikografiese benadering waar 'n woordeboek gekenmerk word deur heterogene artikeltipes en 'n heterogene artikelstruktuur. Hierdie benadering het onder meer daartoe gelei dat 'n onderskeid gemaak word tussen enkelartikels en sinopsisartikels in gespesialiseerde woordeboeke (vgl. Bergenholtz, Tarp en Wiegand 1999), en tussen enkelartikels en komplekse artikels in algemene woordeboeke (vgl. Gouws 2003, 2003a).

Die artikelstruktuur word in 'n hoë mate bepaal deur die keuse wat die leksikograaf tydens die woordeboekkonseptualiseringsfase maak, veral t.o.v. die mikrostruktuur van die betrokke woordeboek. In hierdie verband is dit belangrik om kennis te dra van die verskillende tipes mikrostruktuur en die implikasies van elk hiervan vir die artikelstruktuur. Die daarstelling van 'n funksionele artikelstruktuur voorvereis ook die nodige beplanning t.o.v. die twee hoofkomponente van 'n artikel, te wete die vormkommentaar en die semantiese kommentaar. Daarbenewens moet die leksikograaf deeglik bewus wees van die funksie van en onderskeid tussen twee belangrike inskrywingstipes, naamlik twee kategorieë van funksionele tekssegmente, te wete mikrostrukturele aanduiders en struktuurmerkers. 'n Mikrostrukturele aanduider is 'n woordeboekinskrywing wat dit vir die gebruiker moontlik maak om uit die

aanbod in 'n artikel leksikografiese inligting te ontsluit oor die onderwerp van die betrokke woordeboek en soms ook oor die woordeboekvorm. Inligting word uit die mikrostrukturele aanduiders ontsluit. Die datatipes wat in die bewerking van 'n lemmateken aangebied word, word per mikrostrukturele aanduider daargestel. Mikrostrukturele struktuurmerkers is inskrywings waardeur inligting oor die vorm van 'n artikel en die aanbod in die artikel ontsluit kan word. Dit ondersteun die waarneming van die gebruiker en wys die gebruiker op dele van die artikelstruktuur sodat sekere aanduiders beter geïdentifiseer en sistematies gevind kan word. Daar word 'n onderskeid gemaak tussen tipografiese en nietipografiese struktuurmerkers. Tipografiese struktuurmerkers is byvoorbeeld die vet, romein, kursief, e.s.m. wat gebruik word terwyl nietipografiese struktuurmerkers simbole soos 'n asterisk of 'n vierkant, e.s.m. insluit waardeur bepaalde soeksones en datatipes gemerk word.

In hierdie artikel word 'n verskeidenheid aspekte van veral die mikrostruktuur aan die hand van voorbeelde uit enkele Afrikaanse woordeboeke bespreek. Daar word ook aandag gegee aan die invloed van die mikrostrukturele keuse op die adresseringstruktuur en die vindbaarheid van die aangebode data. 'n Onderskeid word gemaak tussen verskillende vorme van afstandsadressering wat ter sprake is. Die doel van hierdie artikel is nie 'n geïsoleerde bespreking van hierdie kwessies of 'n negatiewe beoordeling van bepaalde woordeboeke nie. Dit gaan veel eerder om 'n poging om na aanleiding van opgetekende voorbeelde uit die tersaaklike woordeboeke te probeer vasstel of daar 'n tipiese patroon in die betrokke woordeboeke te bespeur is wat mikro- en artikelstruktuur betref. Dit behoort te kan lei tot sekere voorstelle ter verbetering van die Afrikaanse leksikografiepraktyk en tot voorstelle waaruit die nodigheid blyk om sekere werkswyses te handhaaf.

2.   Die werklike doel van 'n woordeboek

Vir die leksikograaf van enige woordeboek is dit gebiedend noodsaaklik om reeds voor die aanvang van die leksikografiese bewerking uitsluitsel te hê oor die werklike doel van die betrokke woordeboek (vgl. Wiegand 1998, Gouws 2001). Die werklike doel van 'n woordeboek lê daarin dat dit geproduseer word sodat die teikengebruiker wat die woordeboek gebruik in 'n tipiese gebruiksituasie 'n instrument sal hê om hom/haar te help om 'n suksesvolle woordeboeknaslaanhandeling uit te voer. Hierdie sukses word daardeur bepaal of die behoefte wat die woordeboekraadpleging gemotiveer het, wel beantwoord word. Die werklike doel van 'n woordeboek is dus om suksesvolle woordeboeknaslaanhandelinge te verseker. Die formulering van die werklike doel van 'n spesifieke woordeboek kan gesien word as 'n regstreekse reaksie op die behoeftes van die geïdentifiseerde teikengebruikers van daardie woordeboek. Dit behoort 'n onmiddellike invloed op die besluite oor onder meer die mikro-, artikel- en dataverspreidingstruktuur te hê.

3.   Oor verskillende mikrostruktuurtipes

In verskeie publikasies (vgl. Hausmann en Wiegand 1989, Wiegand 1989, 1996, Bergenholtz, Tarp en Wiegand 1999) word 'n onderskeid gemaak tussen verskillende tipes mikrostrukture. Dit sluit onder meer 'n niegeïntegreerde, 'n geïntegreerde en 'n semigeïntegreerde mikrostruktuur in (vgl. veral Wiegand 1996). Vir 'n algemene verklarende woordeboek maak Hausmann en Wiegand (1989) 'n onderskeid tussen die verpligte, die absolute verpligte en die volledig verpligte mikrostruktuur. Gouws (1999) vereenvoudig hierdie verdeling tot 'n onderskeid tussen die verpligte en die uitgebreide verpligte mikrostruktuur. Hiervolgens maak 'n woordeboek se verpligte mikrostruktuur voorsiening vir die aanbod van daardie substrukture van 'n mikrostruktuur wat in al die artikels van die woordeboek voorkom. Dit impliseer dat daar voorsiening gemaak word vir al die kategorieë inskrywings en datatipes wat verplig vir elke tipe lemmateken verstrek moet word. Hiervolgens sal die abstrakte mikrostruktuur van 'n artikel wat die kleinste aantal inskrywingstipes bevat, bepalend wees vir die aantal kategorieë wat tot die verpligte mikrostruktuur van die betrokke woordeboek behoort. Hiernaas maak 'n uitgebreide verpligte mikrostruktuur vir hierdie verpligte kategorieë voorsiening asook vir daardie kategorieë wat nodig mag wees vir die uitvoerige leksikografiese bewerking van sekere tipes leksikale items wat as lemmata opgeneem is. Die onderskeid tussen die verpligte en die uitgebreide verpligte mikrostruktuur word bepaal deur die verskeidenheid van die data-aanbod.

Vir die vasstelling van die verpligte mikrostruktuur is dit belangrik dat sogenaamde rudimentêre artikels nie as uitgangspunt geneem word vir die besluit oor die minimum aantal kategorieë wat die verpligte mikrostruktuur daarstel nie. By rudimentêre artikels gaan dit om artikels soos suiwer verwysingsartikels waar daar geen volledige bewerking van die behandelingseenheid gebied word nie.

Die onderskeid tussen 'n niegeïntegreerde, 'n geïntegreerde en 'n semigeïntegreerde mikrostruktuur hang onder meer saam met die nabyheid en regstreeksheid van die verhouding tussen elke inskrywing wat in 'n verklarende woordeboek 'n betekenisparafrase verteenwoordig of elke inskrywing wat in 'n vertalende woordeboek 'n vertaalekwivalent verteenwoordig en die ondersteunende koteksinskrywings wat voorbeeldmateriaal verteenwoordig. Hierdie onderskeid word gemaak op grond van 'n basis van regstreekse of afstandsadresseringsprosedures tussen koteksinskrywings en die tersaaklike kerninskrywings van elke semantiese subkommentaar, dit is die betekenisparafrase of vertaalekwivalent wat verstrek word vir elke polisemiese onderskeiding van die leksikale item wat deur die lemmateken verteenwoordig word.

Semigeïntegreerde mikrostrukture kan beskou word as 'n kombinasieproduk waar die artikels sowel 'n geïntegreerde komponent as 'n niegeïntegreerde komponent bevat. Hierdie tipe mikrostruktuur het tot dusver nog nie in die praktyk wyd neerslag gevind nie — en glad nie in Afrikaanse woordeboeke

nie. In die res van hierdie bespreking sal dié mikrostruktuurtipe nie verder aan die orde kom nie.

4.   Geïntegreerde en niegeïntegreerde mikrostruktuur

Die integraat is die betrokke semantiese subkommentaar (vgl. Wiegand 1996: 5 en Bergenholtz, Tarp en Wiegand 1999: 1770), waarin die betrokke polisemiese waarde verklaar of van vertaalekwivalente voorsien word. 'n Geïntegreerde mikrostruktuur word dus daardeur gekenmerk dat alle artikelinterne aanga-wes, met die uitsondering van die polisemie-aanduiders, wat t.o.v. 'n bepaalde polisemiese waarde van die lemmateken verskaf word, in die integraat verskyn en dus binne die bestek van daardie betrokke polisemiese waarde optree. Die polisemie-aanduiders word hier uitgesonder omdat hulle deel vorm van die interne kitstoegangstruktuur van die woordeboek (vgl. in hierdie verband Wie-gand 1996: 6, 7). Die voordeel van hierdie tipe mikrostruktuur is dat daar 'n regstreekse en onmiddellike aanduiding is van die gerigtheid van elke aan-gawe. Vergelyk die volgende voorbeeld van 'n geïntegreerde mikrostruktuur uit die *Tweetalige Aanleerderswoordeboek/Bilingual Learner's Dictionary* (voortaan afgekort as TAW):

(1)   **taak** ⓵ task *It is my **task** to lay the table for dinner in the evening*. Dit is my **taak** om saans die tafel vir ete te dek. ⓶ duty *He works in a bank and his main **duty** is to receive and pay out money*. Hy werk in 'n bank en sy belangrikste **taak** is om geld te ontvang en uit te betaal. … (TAW)

Tipieserwys gaan dit in 'n geïntegreerde mikrostruktuur om koteksinskry-wings wat saam met die betekenisparafrase of vertaalekwivalent in die inte-graat van elke semantiese subkommentaar optree. In aansluiting by Wiegand se beskrywing van die aard van 'n geïntegreerde mikrostruktuur kan daar geargumenteer word, soos wel in hierdie artikel gebeur, dat koteksinskry-wings wat saam met die betekenisparafrase of vertaalekwivalent in 'n enkele integraat optree, ook 'n geïntegreerde mikrostruktuur daarstel. In afdeling 6 van hierdie artikel word voorbeeld (7) en (8) bespreek ter illustrasie van hierdie soort geïntegreerde mikrostruktuur.

Woordeboeke met 'n niegeïntegreerde mikrostruktuur verskil opsigtelik van woordeboeke met 'n geïntegreerde mikrostruktuur. Die tipe woordeboek met 'n niegeïntegreerde mikrostruktuur, soos beskryf in Wiegand (1996), is nog nie vir Afrikaans saamgestel nie. Volgens Wiegand (1996: 23 e.v.) word 'n woordeboek met 'n niegeïntegreerde mikrostruktuur eerstens daardeur geken-merk dat alle betekenisparafrases wat na links aan die lemmateken geadresseer is, in die eerste semantiese subkommentaar van die semantiese kommentaar verstrek word. Dit word die semantiese subkommentaar vir leksikale betekenis genoem. Tweedens word alle koteksaangawes soos voorbeelde, kollokasies en fraseme, in 'n afsonderlike daaropvolgende semantiese subkommentaar, die semantiese subkommentaar vir die koteks, geplaas. Omdat daar verskillende

moontlikhede is waarvolgens die koteksinskrywings verstrek en georden word, is dit van die grootste belang dat hierdie stelsel in een van die voortekste vir die gebruiker uiteengesit moet word. Dit is veral die polisemiese waardes van die leksikale item wat deur die lemma verteenwoordig word wat die artikel-interne ordening van die semantiese subkommentare in 'n woordeboek met 'n geïntegreerde mikrostruktuur bepaal. Gevolglik speel 'n sintagmaties-gemotiveerde stelsel 'n wesenlike rol in die ooreenstemmende ordening in 'n woordeboek met 'n niegeïntegreerde mikrostruktuur. 'n Tipiese verskynsel in 'n woordeboek met 'n niegeïntegreerde mikrostruktuur is die ordening van die koteksinskrywings volgens die woordsoort van die leksikale item wat kombineer met die leksikale item wat deur die lemma verteenwoordig word. Wiegand (1996: 24) bied die artikel van die lemmateken *Kranz* uit die *Deutsches Wörterbuch* van Wahrig aan as 'n voorbeeld van een soort ordening in 'n woordeboek met 'n niegeïntegreerde mikrostruktuur.

(2)   **Kranz** (m. ...) **1** *kreisförmiges Gewinde aus Blüten* ... **2** ein — **Feigen** ... **3** einem Mädchen einen — **aufsetzen** ... **4** einen — **auf** dem Kopf tragen ...

Die eerste semantiese subkommentaar bevat die betekenisverklaring en in die daaropvolgende semantiese subkommentare vir die koteks word voorbeelde gegee van kombinasies waarin die leksikale item *Kranz* met onderskeidelik naamwoorde (2), werkwoorde (3) en voorsetsels (4) optree. Wat belangrik is, is dat hierdie verskillende tipes koteksinskrywings duidelik van mekaar onderskei word. In 'n ander bewerking wys Wiegand daarop hoe daar in die onderskeie koteksinskrywings van die artikel van 'n lemma wat 'n polisemiese leksikale item verteenwoordig, 'n aanduiding gegee kan word van die tersaaklike polisemiese onderskeiding. Weer vir die naamwoord *Kranz* gee hy 'n konsepartikel (weliswaar met 'n semigeïntegreerde mikrostruktuur) met ses semantiese subkommentare vir leksikale betekenis om vir die verskillende polisemiese onderskeidings voorsiening te maak. Daarna volg die niegeïntegreerde deel van die artikel met opeenvolgende soeksones waarin *Kranz* met onderskeidelik naamwoorde, werkwoorde, adjektiewe en voorsetsels verbind. Elk van hierdie soeksones word voorafgegaan deur 'n struktuurmerker wat die betrokke woordsoort van die leksikale item aandui waarmee *Kranz* verbind. Telkens word daar in die verdigte koteksinskrywing bo 'n tilde wat as plekhouersimbool vir die lemmateken optree, 'n nommer geplaas wat die tersaaklike polisemiese waarde van *Kranz* aandui:

(3)   **Kranz** (m. ...) **1** ... **2** ... **3** ... **4** ... **5** ... **6** ...
        Sub. ein ~ [1] *Feigen* / ein bunter ~ [2] tanzender *Kinder* ...
        V. ... / einen ~ [6] *werfen* ...
        Adj. bunter [grüner, verwelkter] ~ [1]
        Präp. einen ~ [1] *auf* das Grab legen ...

Wiegand se uiteensetting van 'n niegeïntegreerde mikrostruktuur sou aangepas

kon word om ook voorsiening te maak vir artikels waar die subkommentaar vir koteksinskrywings georden word op grond van die betrokke polisemiese waardes van die leksikale item wat deur die lemma verteenwoordig word, sonder inagneming van die woordsoort van die items waarmee dit kombineer. Dit sou meebring dat die semantiese kommentaar steeds in verskillende subkommentare verdeel wat elk in 'n afsonderlike integraat aangebied word, maar alle koteksinskrywings word daarna per afsonderlike subkommentaar, die semantiese subkommentaar vir kotekste, verstrek. Kenmerkend van so 'n niegeïntegreerde mikrostruktuur is steeds 'n duidelike verbandlegging tussen 'n koteksinskrywing en die betrokke semantiese subkommentaar wat die betekenisparafrase bevat wat deur die koteksinskrywing geïllustreer word. Hierdie verbandlegging kan op verskillende maniere gedoen word, maar dit is veral die gebruik van struktuurmerkers wat hier 'n belangrike rol speel. 'n Bepaalde koteksinskrywing bevat dan 'n struktuurmerker wat ooreenstem met die polisemie-aanduider van die semantiese subkommentaar wat die relevante betekenisparafrase bevat. So 'n struktuurmerker vorm ook deel van die interne kitstoegangstruktuur en bevestig 'n bepaalde artikel-interne kohesie. Die volgende konsepartikel illustreer hierdie soort aanbieding.

(4)   **bak** ww. (het gebak)
      1 Gaarmaak deur hitte. **2** Gaar word; warm kry. **3** Hard laat word deur hitte. **4** Hitte afgee.
      Eiers in die pan bak (1). In die son sit en bak (2). Stene bak (3). Die son bak op die stoep (4).

Die gebruik van 'n niegeïntegreerde mikrostruktuur kom tipieserwys voor in die bewerking van lemmas wat polisemiese leksikale items verteenwoordig. Die artikel-interne leksikografiese aanbod maak nie vir 'n regstreekse adressering tussen die koteksinskrywings en die tersaaklike betekenisparafrase of vertaalekwivalent voorsiening nie. 'n Regstreekse of onmiddellike adressering kom voor wanneer die koteksinskrywings in dieselfde integraat as die relevante betekenisparafrase of vertaalekwivalent optree sodat hierdie betekenisparafrase of vertaalekwivalent plus die koteks as geheel binne een semantiese subkommentaar kan funksioneer.

Die daarstelling van 'n niegeïntegreerde mikrostruktuur lei tot 'n versplintering in die tekssamehang van 'n semantiese subkommentaar a.g.v. die afstandsadressering tussen koteksinskrywing en betekenisparafrase of vertaalekwivalent. Alhoewel die gebrek aan onmiddellike adressering die samehang versteur, word 'n mate van kohesie steeds behou deur 'n duidelike markering van die koteksinskrywings ter aanduiding van hulle gerigtheid op 'n bepaalde betekenisparafrase of vertaalekwivalent. Wiegand (1998a: 8) onderskei tussen kohesie en koherensie in leksikografiese tekste. 'n Kohesiewe teks word gekenmerk deur tekssamehang op sintaktiese, proposisionele en referensiële vlak. Daarenteen lei koherensie tot tematiese progressie met 'n bepaalde tema wat die tekssamehang bewerk. In die geval van niegeïntegreerde mikrostrukture

blyk kohesie veral uit die referensiële progressie wat die tekssamehang bevestig.

Die onderskeid van Wiegand (1996) tussen geïntegreerde en niegeïntegreerde mikrostrukture geld nie net die indiwiduele artikel nie maar 'n gegewe woordeboek as sodanig. Die artikelstruktuur van 'n woordeboek maak voorsiening vir 'n bepaalde tipe mikrostruktuur om as verstekwaarde te dien en daardie tipe mikrostruktuur is dan verteenwoordigend van die woordeboek as geheel. Die erkenning aan heterogeniteit in die artikel- en mikrostruktuur van 'n woordeboek lei daartoe dat daar gemotiveerde afwykings van hierdie verstekstruktuur mag wees. In die woordeboek se toeligtingsteks moet sowel die verstekstruktuur as die tipes afwykings en 'n motivering vir hierdie afwykings verstrek word. Die artikel- en mikrostrukturele heterogeniteit mag nie die teikengebruiker van die betrokke woordeboek verhinder om 'n ondubbelsinnige onttrekking van inligting te bewerkstellig nie.

5.   'n Rudimentêre mikrostruktuur

Dit is belangrik om daarop te let dat sowel die geïntegreerde as die niegeïntegreerde mikrostruktuur struktuurtipes is wat aan bepaalde kriteria moet voldoen. 'n Verwysing na 'n niegeïntegreerde mikrostruktuur dui dus op 'n bepaalde tipe struktuur met spesifieke kenmerke en nie maar bloot op 'n woordeboekartikel waarvan die mikrostruktuur nie geïntegreer is nie. Die nienakoming van die voorskrifte wat geld vir 'n geïntegreerde mikrostruktuur impliseer nie noodwendig die bestaan van 'n niegeïntegreerde mikrostruktuur nie. Vergelyk die volgende voorbeelde uit *Groot Woordeboek/Major Dictionary* (voortaan afgekort as GW) en *Verklarende Afrikaanse Woordeboek* (voortaan afgekort as VAW).

(5)   **na'spel[1],** (s) after-play, afterpiece; postlude; sequel; aftermath. (GW)

(6)   **paragon', -e. 1.** Voorbeeld van voortreflikheid. **2.** Iem. wat 'n besondere voorbeeld stel, wat die buitengewoon goeie, grote of mooie bereik het. **3.** Iets buitengewoon deugdelik of deeglik. **4.** Soort drukletter. **5.** Volmaakte diamant van 100 karaat of meer. (VAW)

Nie een van hierdie artikels bevat enige konteks- of koteksleiding nie. In artikel (5) bestaan die semantiese kommentaar uit 'n jukstaposisie van vyf vertaalekwivalente wat in vier verskillende semantiese subkommentare aangebied word. Volgens die stelsel van hierdie woordeboek word verskillende semantiese subkommentare deur 'n kommapunt van mekaar geskei terwyl 'n komma in die vertaalekwivalentparadigma 'n struktuurmerker van leksikale divergensie is wat sinonimiese vertaalekwivalente skei wat dieselfde polisemiese waarde van die lemma verteenwoordig (vgl. Gouws 1989). In voorbeeld (5) is daar sprake van 'n ekwivalentverhouding van polidivergensie (vgl. Gouws 2002:

199), maar die woordeboek laat na om aan sy gebruiker hulp te bied met kon-
teks- en koteksinskrywings om die keuse van die regte vertaalekwivalent vir 'n
bepaalde gebruiksoptrede te vergemaklik. In artikel (6) word daar vir elke
semantiese subkommentaar slegs 'n betekenisparafrase in die integraat gegee.
Vanuit 'n mikrostrukturele perspektief gaan dit in albei hierdie artikels om
onverdeelde integrate maar integrate met 'n enkelvoudige datalading. Hier is
geen sprake van óf 'n geïntegreerde óf 'n niegeïntegreerde mikrostruktuur nie,
maar eerder van 'n rudimentêre mikrostruktuur, d.w.s. 'n basis vir 'n meer
komplekse mikrostruktuur.

6.   Wisselende geïntegreerde mikrostrukture

Naas die voorkoms van artikels wat 'n rudimentêre mikrostruktuur vertoon,
bevat GW en VAW ook artikels wat ander mikrostrukturele tipes daarstel.
Vergelyk die volgende voorbeelde:

(7)   **om'geslaan, (-de, .. slane),** turned over (page); overturned (vehicle); turn-down
        (collar) (GW)

(8)   **demp, ge-. 1.** Bedwing, onderdruk (oproer). **2.** Opvul (sloot). **3.** Swakker maak,
        temper (geluid). **4.** Blus (vuur). **5.** Dowwer maak, flouer raak (lig). **6.** Versag
        (kleure). **7.** Verminder, smoor (vuur); **~baar; ~ing; ~klep; ~laag.** (VAW)

Sowel voorbeeld (7) as voorbeeld (8) vertoon 'n tipe geïntegreerde mikrostruk-
tuur. Die artikels van die lemmas **omgeslaan** en **demp** bestaan elk uit meerdere
semantiese subkommentare omdat die lemmas polisemiese leksikale items ver-
teenwoordig en in elk van hierdie subkommentare word die vertaalekwivalent
of betekenisparafrase ondersteun deur konteksinskrywings. Hierdie konteksin-
skrywings wat pragmatiese verankering ten doel het, word telkens in dieselfde
integraat as die betrokke vertaalekwivalent of betekenisparafrase aangebied.
Dieselfde woordeboeke waarin daar artikels is wat 'n rudimentêre mikrostruk-
tuur vertoon (voorbeelde (5) en (6)), bevat ook artikels met daardie weergawe
van 'n geïntegreerde mikrostruktuur wat konteksinskrywings in die integraat
insluit (voorbeelde (7) en (8)).

    'n Geïntegreerde mikrostruktuur kom veral ook voor in artikels waar die
vertaalekwivalent of betekenisparafrase ondersteun word deur voorbeeldmate-
riaal wat as koteksinskrywings optree met die koteksinskrywings onmiddellik
opeenvolgend aan die vertaalekwivalent of betekenisparafrase, d.w.s. as deel
van dieselfde integraat. Koteksinskrywings het die sintaktiese verankering van
die leksikale item wat deur die lemma verteenwoordig word ten doel. Vergelyk
in hierdie verband die volgende voorbeelde uit onderskeidelik die TAW en die
Verklarende Handwoordeboek van die Afrikaanse Taal (voortaan afgekort as HAT):

(9)   **dom** 1 slow *He is clever at languages but a little **slow** at maths*. Hy is slim in tale

maar 'n bietjie **dom** in wiskunde. 2 foolish, silly, stupid *It is **foolish/silly/ stupid** to go too deep into the sea if you can't swim*. Dit is **dom** om te diep in die see in te gaan as jy nie kan swem nie. … (TAW)

(10) **da'nig** b.nw. en bw. **1** (*bw*.) Baie, erg; in hoë mate: *Danig verlief wees. Dit is nie so danig ver na hulle toe nie. Danig in jou skik wees met iets*. **2** Oorvriendelik: *Danig wees met die nuwe landdros. Danig met mekaar wees*, duidelik verlief. **3** (*ietwat neerhalend*) Van groot belang; waffers: *Jy hou jou verniet so danig. Gaan leen by jou danige vriende*. (HAT)

Die TAW het 'n konsekwente aanbieding van 'n geïntegreerde mikrostruktuur terwyl HAT dit as verstekwaarde het, alhoewel baie artikels 'n rudimentêre mikrostruktuur vertoon, soos blyk uit die volgende:

(11) **dek'spaan 1** Klopper om dekstrooi plat te klop. **2** Spaan aan 'n selfbinder om die koppe van gerwe gelyk te maak. (HAT)

'n Geïntegreerde mikrostruktuur maak voorsiening vir 'n onmiddellike en regstreekse adressering tussen die vertaalekwivalent of betekenisparafrase en daardie bykomende inskrywings in die betrokke semantiese subkommentaar wat konteks- en/of koteksleiding verstrek. In voorbeelde (7) en (8) is dit in sowel GW as VAW telkens die optrede van konteksinskrywings wat as deel van die betrokke integrate aangebied word. Die insluiting van sulke inskrywings dra veel by tot sowel die teksresepsie- as die teksproduksiefunksie van 'n woordeboek (vgl. Tarp en Gouws (om te verskyn)). Die afwesigheid van sulke inskrywings stel heelwat hoër eise aan die gebruiker om uit die aangebode data 'n optimale inligtingsonttrekking te kan doen.

Waar konteksinskrywings in artikels met meerdere semantiese subkommentare in GW en VAW verstrek word, word dit nie altyd in elke semantiese subkommentaar gegee nie. Vergelyk die volgende voorbeelde:

(12) **berei', (~),** prepare (meal); dress (leather, tools); preserve (foodstuffs); cure (tobacco); concoct. (GW)

(13) **beoor'deel, (~),** judge, criticise, adjudicate (merits); evaluate, assess; review (books); rate, value. (GW)

(14) **belig', (~),** lighten up; throw a light on, illuminate; expose (photo); explain. (GW)

(15) **hui'sie, -s. 1.** Klein huis. **2.** Dosie (bril). **3.** Skyfie (lemoen). **4.** Skede (mes). **5.** Skulp (slak); **~slak.** (VAW)

(16) **in'dryf, in'drywe, -ge-. 1.** Drywend na binne gaan. **2.** Saamvoer na binne. **3.** Inja, aandryf (vee). **4.** Inslaan (spyker). (VAW)

(17) **hut, -te. 1.** Eenvoudige woninkie. **2.** Kajuit (op skip). **3.** Loge. **4.** Strooihut; **~bewoner; ~koffer.** (VAW)

In voorbeelde (12) en (15) word daar telkens op een na in elke semantiese subkommentaar konteksleiding gegee. In voorbeelde (13) en (16) word meer as een semantiese subkommentaar wees gelaat t.o.v. van konteksleiding en in voorbeelde (14) en (17) word daar telkens slegs vir een semantiese subkommentaar 'n konteksinskrywing voorsien. Die gebruiker se behoefte, die aard van die lemma wat bewerk word en die funksies van die woordeboek behoort te bepaal wat die omvang van konteksleiding is en dus wat die mate is waarin die semantiese subkommentaar 'n rudimentêre al dan 'n geïntegreerde struktuur vertoon. Vanuit 'n teoretiese perspektief is dit egter belangrik dat enige woordeboek 'n konsekwente en voorspelbare data-aanbod moet hê. Die gebruiker kan verwag om in die toeligtingsteks 'n duidelike uiteensetting te kry van die werkswyse wat gevolg word by die weergawe van konteksleiding, veral as die aanbod so arbitrêr en onvoorspelbaar is. Dit ontbreek ongelukkig in sowel GW as VAW.

7.   Die mikrostrukturele plasing van koteksinskrywings

Die semantiese kommentaar in 'n woordeboekartikel bevat alle aangawes wat, in 'n breë sin, te make het met die betekenis van die lemmateken (vgl. in hierdie verband onder meer Hausmann en Wiegand 1989: 353 en Bergenholtz, Tarp en Wiegand 1999: 1768). In die semantiese subkommentare word die betekenisparafrase of vertaalekwivalent ter wille van onder meer 'n eenduidige interpretasie ondersteun deur die koteks- en die konteksinskrywings. Voorbeeldmateriaal vorm 'n wesenlike deel van die semantiese kommentaar en dit is belangrik dat dit op 'n konsekwente en duidelik identifiseerbare manier aangebied moet word. Binne die integraat van die betrokke semantiese subkommentaar moet die voorbeeldmateriaal in 'n eie soeksone verstrek word waar dit probleemloos deur die teikengebruiker herwin kan word. In die VAW word koteksmateriaal uiters onvoorspelbaar en onbevredigend hanteer. Een van die wesenlike probleme in hierdie verband is die gebrek aan 'n onderskeid tussen voorbeeldmateriaal en vaste uitdrukkings. In hierdie verband ervaar die gebruiker dieselfde tipe probleem met die VAW as met die GW. Vergelyk die volgende voorbeelde uit HAT en VAW onderskeidelik, met HAT wat by die onderskeie semantiese subkommentare konteksleiding gee en daarnaas 'n afsonderlike artikelsone het waarin die uitdrukkings bewerk word, terwyl die VAW geen sodanige onderskeid maak nie:

(18)   **maag** (mae) **1** Belangrikste spysverteringsorgaan van die mens (en sommige diere), waar kos 'n tydlank gehou word: *'n Sterk, swak maag hê. Met 'n vol maag gaan slaap. Moenie jou maag oorlaai nie,* te veel eet. **2** Onderste deel van die romp; buik: *Daardie ou het 'n groot maag.* UITDR.: *Dit sit my **dwars** in die maag,* ek is daarmee verleë, opgeskeep. *Sy **oë** is groter as sy maag,* hy wil meer hê as wat hy kan opeet (behartig). *Jy kan dit op jou maag **skryf** en met jou hemp afvee …,* daar sal niks van kom nie. *Jou maag **vashou** van die lag,* onbedaarlik lag. *'n Maag soos 'n **volstruis** hê,* alles kan eet. *Van jou maag 'n **wolsak** (afgod)*

maak, vraatsugtig wees. **maag: ~aandoening, ~bloeding, ~druppels, ~kanker, ~kramp, ~kwaal, ~lyer, ~operasie, ~siekte, ~streek, ~wand, ~wond.** (HAT)

(19) **maag¹, mae.** Vernaamste spysverteringsorgaan; *van sy ~ 'n AFGOD maak,* vrees-lik baie eet; vraatsugtig lewe; *dit sit hom DWARS in die ~,* met iets opgeskeep of verleë wees; *op sy NUGTER ~,* nog voordat hy geëet het; *met iets in sy ~ SIT,* in die verleentheid wees as gevolg van iets; *jy kan dit op jou ~ SKRYF,* daar sal niks van teregkom nie; *jou ~ VASHOU v.d. lag,* onbedaarlik lag; *'n ~ soos 'n VOLSTRUIS hê,* alles kan verteer; *sy ~ WERK,* sy maag ontlas buitengewoon; *van sy ~ 'n WOLSAK maak,* vraatsugtig wees; **~aandoening; ~operasie.** (VAW)

Uit die VAW-aanbieding is dit geensins duidelik of dit in die semantiese sub-kommentaar vir die koteks om voorbeeldmateriaal al dan om uitdrukkings gaan nie. VAW bied in baie artikels slegs 'n rudimentêre mikrostruktuur (vgl. voorbeeld (6)). Soms lei 'n uitbreiding van hierdie rudimentêre mikrostruktuur tot die geïntegreerde weergawe van konteksleiding (vgl. voorbeeld (8)). Waar die uitbreiding van die mikrostruktuur gerig is op die insluiting van koteks-inskrywings blyk die aanbieding daarvan uiteenlopend van aard en uiters lukraak te wees. Die volgende voorbeelde getuig hiervan:

(20) **reg³,** (bw). **1.** Presies, net, bv. *~ bo, ~ agter, ~ onder.* **2.** Reguit; in orde: *~ buig, ~ knip, ~ deur die see gaan.* **3.** Behoorlik, goed: *~ verstaan, nie ~ weet nie.* (VAW)

(21) **privaat'²,** (b, bw), **(..vate); ..vater** of **meer private, ~ste** of **mees private. 1.** Wat op 'n persoon, 'n indiwidu self en alleen betrekking het of aan hom behoort, per-soonlik: *dis Jan se private sienswyse; die minister se private woning.* **2.** Nie publiek, openbaar en gemeenskaplik nie: *dis 'n suiwer private aangeleentheid rakende die plaasbewoners.* **3.** Nie amptelik, beroepmatig nie: *hy hou ~ skool na sy ontslag; hy het die saak ~ onderneem.* **4.** Nie vir algemene gebruik nie: *ek het 'n private deurgang oor sy plaas na my lande toe.* **5.** Vertroulik, geheim: *hy het my dit ~ meegedeel.* **6.** Nie uit beroepsverdiente verkry nie: *die dokter het 'n private inkomste uit sy wildsplaas;* **~heid.** (VAW)

(22) **me'dia.** Mv. van **medium. 1.** Middele; hulpmiddels; *gemengde ~* (skilderk.), ver-skillende skildermiddele, bv. olieverf, pastel, waterverf, ink, ens. **2.** Alle kommunikasiemiddele saam, bv. pers, rolprent, televisie; *dit word in die ~ bekend gestel; 'n verteenwoordiger v.d. ~;* **~persoonlikheid; ~stryd; ~verteen-woordiger; ~wetenskap.** (VAW)

(23) **pot'jie, -s. 1.** Klein pot. **2.** Spel, bv. *'n ~ dambord, skaak, tennis.* **3.** Holgewrig in die bladbeen of heupbeen waarin die bolgewrig van die boarm of -been draai; *in die KLEINSTE ~s bewaar mens die beste salf,* waardevolle dinge is altyd skaars en word dus in klein dingetjies bewaar; *sy eie ~ KRAP,* vir homself sorg; *met iem. 'n ~ LOOP,* met iem. afreken; *klein ~s het groot ORE,* kinders hoor meer as mens dink; *'n ~ SPEEL,* 'n spel speel. (VAW)

(24)   **aan'bied, -ge-. 1.** Vrywillig tot beskikking stel. **2.** Aanprys om te verkoop. **3.** Op-
lewer, gee. **4.** Vir betaling voorlê. **5.** Beskikbaar wees; *'n gulde GELEENTHEID
het hom aangebied om iets te verdien; iem. 'n GESKENK, HULP, SIMPATIE ~; hy
bied die boer R100 vir die perd aan; 'n TJEK ~;* **~er.** (VAW)

(25)   **termyn', -e. 1.** Tydruimte van 'n bepaalde, gesette tyd. **2.** Tydsverloop waarin
betalings, paaiemente, moet geskied; *in ~ BETAAL*, met paaiemente betaal; *op
KORT ~*, vir 'n kort tyd; *op LANG ~*, vir 'n lang tydperk; **~belegging.** (VAW)

Hierdie voorbeelde gee blyke van 'n verskeidenheid maniere waarop VAW met
koteksinskrywings werk. In voorbeelde (20) en (21) het die artikels 'n geïnte-
greerde mikrostruktuur met die koteksinskrywings as die enigste bykomende
inskrywings naas die betekenisparafrase in die betrokke integrate. In voor-
beelde (22) en (23) is daar 'n gemengde aanbod met telkens slegs een seman-
tiese subkommentaar wat 'n geïntegreerde daarstelling het. Na die laaste bete-
kenisparafrase volg 'n verdere soeksone waarin koteksinskrywings en/of vaste
uitdrukkings gegee word. Die aanbod in voorbeelde (22) en (23) verskil ook
onderling. In voorbeeld (22) tree die koteksinskrywing *gemengde media* in die
eerste semantiese subkommentaar ook op as 'n sekondêre bewerkingseenheid.
Dit is die adres van die etiket (*skilderk.*) en ook van 'n afsonderlike betekenis-
parafrase, te wete *verskillende skildermiddele, bv. olieverf, pastel, waterverf, ink, ens.*
Hierdie inskrywing is gerig op die koteksinskrywing deur 'n proses van
onmiddellike adressering aangesien dit in dieselfde integraat gegee word. Die
koteksinskrywing *'n potjie dambord, skaak, tennis* in voorbeeld (23) kry geen eie
bewerking nie. In voorbeeld (22) word die tweede semantiese subkommentaar
gevolg deur die koteksinskrywings: *dit word in die media bekend gestel; 'n
verteenwoordiger v.d. media.* Hier word geen betekenisparafrase gegee nie maar
in voorbeeld (23) word die inskrywings in die soeksone na die verklaring van
die laaste polisemiese waarde *in die KLEINSTE potjies bewaar mens die beste salf,
sy eie potjie KRAP, 'n potjie LOOP, klein potjies het groot ORE* en *'n potjie SPEEL*
wel verklaar. Van die inskrywings in hierdie gemengde sone is uitdrukkings,
bv. *klein potjies het groot ORE*, wat wel verklaar moet word. Daar word egter
geen onderskeid gemaak deur hierdie uitdrukkings te skei of te onderskei van
'n inskrywing soos *'n potjie SPEEL* wat slegs 'n kollokasie met *potjie* as kern is
nie.

In voorbeelde (24) en (25) is daar geen sprake van 'n geïntegreerde mikro-
struktuur nie. Weer eens word die koteksinskrywings uitgefaseer na 'n soek-
sone wat na die laaste integraat volg. Ook die bewerking van die koteksin-
skrywings in voorbeelde (24) en (25) verskil onderling. In voorbeeld (24) kry
hierdie koteksinskrywings geen betekenisverklaring nie, maar in voorbeeld
(25) word die inskrywings wel verklaar. Daar sou aangevoer kon word dat die
inskrywing *in termyn BETAAL* 'n ondeursigtige geykte uitdrukking is en 'n
verklaring regverdig. Dit geld egter nie die inskrywings *op KORT termyn* en *op
LANG termyn* nie. Hier geld telkens 'n betekeniswaarde van *termyn* wat in die
bewerking van die lemma verklaar is. Die betekenisparafrases *vir 'n lang tyd-*

perk en *vir 'n kort tyd* voeg weinig leksikografiese waarde toe aan die bewerking van die lemma *termyn.* Dit bevestig eerder die inkonsekwente werkswyse in die mikrostrukturele aanbod van hierdie woordeboek.

8.   Afstandsadressering

Waar koteksinskrywings nie in dieselfde integraat as die tersaaklike betekenisparafrase of vertaalekwivalent aangebied word nie maar in 'n afsonderlike soeksone, lei dit noodwendig tot 'n prosedure van afstandsadressering (vgl. Louw en Gouws 1996). Afstandsadressering is 'n indirekte en nie-onmiddellike vorm van adressering en 'n produk van teksverdigting, en die toepassing van teksverdigting voorveronderstel veel meer gevorderde woordeboekgebruiksvaardighede aan die kant van die teikengebruiker. Teenoor die regstreekse en onmiddellike adressering tussen koteksinskrywing en betekenisparafrase in 'n woordeboekartikel met 'n geïntegreerde mikrostruktuur geld die afstands- en indirekte adressering in artikels met 'n niegeïntegreerde mikrostruktuur. Ondanks die teksverdigting wat hier voorkom, is daar in die voorkoms van 'n niegeïntegreerde mikrostruktuur steeds 'n goeie kans op 'n optimale herwinning van inligting uit die aangebode data omdat die aard van 'n niegeïntegreerde mikrostruktuur dit vir die gebruiker moontlik maak om elke koteksinskrywing aan 'n spesifieke betekenisparafrase of vertaalekwivalent te koppel. Die gebruik van struktuurmerkers om die koteksinskrywing aan die regte betekenisparafrase of vertaalekwivalent te koppel, lei tot 'n proses van *gerigte afstandsadressering*, dit is afstandsadressering waar die geadresseerde inskrywing per struktuurmerker gerig word op 'n spesifieke adres. Hierdie struktuurmerkers is deel van die interne kitstoegangstruktuur van die betrokke woordeboek en skep 'n eksplisiete soekroete vir die gebruiker wat vanaf die betekenisparafrase of vertaalekwivalent gelei moet word na koteksmateriaal wat die tipiese gebruik van die lemma of vertaalekwivalent illustreer.

'n Verdere inwerking van teksverdigting op 'n artikel met gerigte afstandsadressering kan die struktuurmerkers laat verdwyn en die gebruiker laat met *implisiete afstandsadressering* en die gevolglike implisiete interne soekroete. Indien dit op 'n konsekwente manier toegepas word en daar 'n een-tot-eenverhouding bestaan tussen die koteksinskrywings en die verskillende betekenisparafrases of vertaalekwivalente in die artikel se integrate kan die kundige gebruiker wel die implisiete adressering korrek interpreteer en die verlangde toegang tot die aangebode data kry. Implisiete afstandsadressering kan net optimaal werk indien daar adresseringsekewivalensie bestaan (vgl. Gouws 2000, 2002), en daar vir elke polisemiese onderskeiding van die leksikale item wat deur die lemmateken verteenwoordig en in 'n afsonderlike integraat as die betrokke semantiese subkommentaar vergestalt word een koteksinskrywing is en die volgorde van die koteksinskrywings parallel loop aan die volgorde van die betekenisparafrases of vertaalekwivalente wat die adresse van die onderskeie koteksinskrywings is.

Implisiete afstandsadressering word ondermyn deur die aanwesigheid van prosedures van onder- en ooradressering (vgl. Gouws 2000, 2002 vir 'n bespreking van oor- en onderadressering). Die volgende voorbeeld wys dat daar in die VAW nie sprake is van óf gerigte afstandsadressering óf implisiete afstandsadressering nie en dat onderadressering die leksikografiese bewerking verder benadeel:

(26)   **hou⁴,** (w), **ge-. 1.** Bewaar, behou. **2.** Vashou (leisels). **3.** Bevat (emmer). **4.** Slaan. **5.** Vervul; nakom. **6.** Besig wees met. **7.** Bly (in die bed). **8.** Beweer; voorgee. **9.** Boer, het sy verblyf (wild). **10.** Nakom (belofte). **11.** Vier (feesdag). **12.** Laat plaasvind. **13.** Duur. **14.** Lewer (rede). **15.** In orde hou (boeke). **16.** Goed bly, vars bly. **17.** Besig wees met (skool). **18.** Heilig (Sondag). **19.** Teëhou, bedwing. **20.** Ophou. **21.** By jou laat bly. **22.** Bewaar (boek). **23.** Sorg dat jy dit nie verloor nie. **24.** Nie geleidelik verloor nie (kleur). **25.** Uithou, verduur. **26.** Jou aangetrokke voel tot. **27.** Herberg. **28.** Laat bly. **29.** Vir jou gerief aanhou. **30.** Jou bestaan vind in. **31.** Nie nalaat nie, in stand hou; *jou AAN iets* ~, iets nakom (bv. afspraak); *die BED* ~, in die bed bly (weens siekte); *'n BEDIENDE daarop na*~, 'n bediende in diens hê; *jou GOED* ~, jou goed gedra; nie treur of mismoedig wees nie; ~ *wat jy HET* (*en kry wat jy kan*), moenie wat jy besit uit jou hande laat gaan nie en probeer sover moontlik nog meer kry; *iets vir JOUSELF* ~, met niemand daaroor praat nie; *die LEWE* ~, bly lewe; *hy* ~ *hom SIEK*, hy maak of hy siek is; ~ *dit VIR jou*, hou dit geheim; ~ *tot die WERKERS v.d. land kom*, goed geëet het en nie gou honger sal word nie; *WAARVOOR* ~ *jy my?* is dit jou mening van my? (VAW)

In hierdie voorbeeld word die koteksinskrywings nie gemerk as gerig op spesifieke integrate in die semantiese kommentaar nie. 'n Poging om die koteksinskrywings aan bepaalde betekenisparafrases te koppel, laat blyk dat daar nóg 'n een-tot-een-afparing nóg 'n volgordeparallel bestaan. Die juistheid van die polisemieverdeling word nie hier bespreek nie. 'n Afparing tussen die inskrywings (koteksinskrywings of uitdrukkings) in die artikelfinale soeksone en die betekenisparafrases in die onderskeie semantiese subkommentare sou 'n resultaat soos dié in voorbeeld (27) kon lewer. In hierdie afparing word die betrokke inskrywings (koteksinskrywing of uitdrukking plus betekenisparafrase) gegee en daarnaas die nommer van die semantiese subkommentaar wat ten beste deur die betrokke inskrywing geïllustreer word. Hierdie afparing word bemoeilik deur 'n onvoldoende onderskeid in die woordeboekartikel tussen betekenisbeskrywing en die verklaring van die gebruik van die betrokke woord (vgl. Gouws 1989 in hierdie verband).

(27)   *jou AAN iets hou*, iets nakom (bv. afspraak)   **10**
        die BED hou, in die bed bly (weens siekte)   **7**
        'n BEDIENDE daarop nahou, 'n bediende in diens hê   **Foutiewe voorbeeld:**
                                                                    nahou **is 'n afsonderlike deeltjiewerkwoord**
        jou GOED hou, jou goed gedra; nie treur of mismoedig wees nie   **16**
        hou wat jy HET (*en kry wat jy kan*), moenie wat jy besit uit jou hande laat gaan nie

en probeer sover moontlik nog meer kry	**?1 UITDR**.
iets vir JOUSELF hou, met niemand daaroor praat nie	**?21**
die LEWE hou, bly lewe	**?25, 31 UITDR**.
hy hou hom SIEK, hy maak of hy siek is	**8**
hou dit VIR jou, hou dit geheim	**21**
hou tot die WERKERS v.d. land kom, goed geëet het en nie gou honger sal word nie	
	UITDR.
WAARVOOR hou jy my? is dit jou mening van my?	**? UITDR**.

Die mate waarin onderadressering plaasvind, blyk daaruit dat uit die 31 bete-kenisonderskeidinge wat aangebied word, daar vir minder as tien koteksin-skrywings voorsien word. Die arbitrêrheid van hierdie aanbod blyk verder daaruit dat die koteksinskrywings geensins die volgorde handhaaf van die integrate waarin die relevante betekenisparafrases aangebied word nie. Die volgorde van koteksinskrywings in hierdie soort soeksone in die VAW word bepaal deur die alfabetiese waarde van een van die woorde wat saam met die woord optree wat met die lemmateken van die artikel ooreenstem. In die boge-noemde soeksone word hierdie woorde wat die ordening bepaal in hoofletters gegee.

In die artikels van die VAW gaan dit nóg om gerigte nóg om implisiete af-standsadressering. Dit gaan veel eerder om 'n lukraak en arbitrêre vorm van afstandsadressering. Na die tipe mikrostruktuur onderliggend aan hierdie af-standsadressering wat in VAW en ook in GW voorkom, verwys Gouws en Wiegand (om te verskyn) as 'n *primitiewe mikrostruktuur*. Die nadele wat die ge-bruiker ervaar in 'n woordeboek met 'n primitiewe mikrostruktuur word verer-ger wanneer die dataverspreidingstruktuur van die woordeboek nie daarvoor voorsiening maak dat elke datatipe in 'n afsonderlike soeksone aangebied word nie. Hierdie probleem is akuut in die VAW met die uitdrukkings en koteks-inskrywings wat op 'n ongedefinieerde manier saam in dieselfde soeksone ver-skyn en nie eers van mekaar onderskei word nie maar bloot op grond van die alfabetiese waarde van nog 'n woord in die kombinasie georden word. In so 'n geval gaan dit om 'n *gemengde primitiewe mikrostruktuur*. Voorbeeld (28) illus-treer so 'n mikrostruktuur.

(28)   **bit'ter²,** (b), **-der, -ste. 1.** Wat 'n skerp, onaangename smaak veroorsaak. **2.** Skerp, griewend, bytend. **3.** In hoë mate; ~ *MIN*, uiters min; ~ *in die MOND maak die maag gesond*, medisyne wat sleg smaak is baie goed; **~agtig; ~amandel; ~heid.** (VAW)

In hierdie voorbeeld is *bitter min* 'n kollokasie waarin *bitter* se tweede polise-miese waarde geld. Daarteenoor is *bitter in die mond maak die maag gesond* 'n uit-drukking. Hierdie statusverskil tussen die twee inskrywings blyk nie uit die aanbod nie, laat die gebruiker in die duister en laat hierdie artikel kwalifiseer as een met 'n gemengde primitiewe mikrostruktuur. Die nie-aanduiding van 'n onderskeid tussen uitdrukkings en voorbeeldmateriaal, dit is kollokasies en

ander koteksinskrywings, bemoeilik ook die pogings van die teikengebruiker om 'n gerigte soektog na 'n datatipe soos uitdrukkings uit te voer. Naas artikels met 'n gemengde aanbod van uitdrukkings en koteksmateriaal is daar sommige artikels waarin die ooreenstemmende soeksone slegs koteksinskrywings bevat en ander artikels waar dit slegs uitdrukkings bevat. Vergelyk in hierdie verband voorbeelde (29) en (30) onderskeidelik.

(29)   **psigede'lies, -e. 1.** Wat die uitwerking het om die sintuie sterker, skerper te laat voorkom as in die werklikheid, waardeur vreemde, opgewerkte sensasies van krag, geluk, hopeloosheid e.d.m. ontstaan. **2.** Wat 'n uitwerking op die verstand het a.g.v. sterk, kragtige patrone van klank, kleur, bewegende ligte, lyne, ens. — van 'n vorm van beeldende kuns gesê; ~ *DWELMMIDDELS;* ~ *KLEURE;* ~ *MUSIEK,* wat deur senu- en breinprikkelende inwerking hallusinêre geestesverskynsels teweeg bring. (VAW)

(30)   **profeet', ..fete. 1.** Iem. wat die heilige roeping van God ontvang om te vermaan en tot inkeer te probeer bring. **2.** Iem. wat voorspel wat gaan gebeur, waarsêer, voorspeller; *'n ~ is nie in sy eie LAND geëerd nie,* 'n man van verdienste word in sy eie land nie so geëer as in die buiteland nie; *die OU profete is dood en die kleintjies* (*jonges*) *eet brood,* 'n mens kan nie op enigeen se voorspelling staatmaak nie; *is SAUL ook onder die profete?* gesê om verbasing uit te druk wanneer 'n persoon hom in geselskap bevind waar ons hom nie verwag nie; **..feties.**

9.   Ten slotte

Vanuit die teoretiese leksikografie word daar gemotiveerde voorstelle gemaak oor 'n verskeidenheid aspekte van die struktuur van woordeboeke. Hierdie voorstelle is nie uit die lug gegryp nie maar word gedoen op grond van die eise van 'n gebruikersgedrewe leksikografiese proses. Die suksesvolle onttrekking van inligting uit 'n woordeboek is daarvan afhanklik dat die data op 'n gesistematiseerde en konsekwente manier aangebied moet word. Dit vereis 'n sekere mate van standaardisering in, onder meer, die mikrostrukturele aanbod. Die eerste uitgawes van heelwat woordeboeke, ook Afrikaanse woordeboeke soos die VAW, het in 'n preteoretiese milieu verskyn en die destydse leksikograwe kon nie raad inwin by metaleksikograwe nie. Hierdie situasie het ingrypend verander. Woordeboekgebruikers kan met reg van leksikograwe verwag dat hulle vertroud moet wees met die belangrikste modelle wat vanuit die teoretiese leksikografie voorgehou word. Waar hulle besluit om van voorgestelde modelle af te wyk, moet so 'n afwyking veral vanuit 'n gebruikersperspektief motiveerbaar wees en dit moet in die toeligtingsteks van die woordeboek uiteengesit word.

    Vir Afrikaanse woordeboeke is dit belangrik dat leksikograwe meer aandag moet gee aan die moontlikhede wat geïntegreerde en niegeïntegreerde mikrostrukture bied. Daar is ook genoeg ruimte vir die ontwikkeling en uit-

breiding van bestaande mikrostruktuurmodelle sodat die spesifieke behoeftes en naslaanvaardighede van die teikengebruikers van 'n bepaalde woordeboek erkenning kry. 'n Lukraak benadering tot die mikrostruktuur met 'n arbitrêre aanbod van data is nie aanvaarbaar nie en benadeel die gehalte van die woordeboek asook sy status as gesaghebbende en toeganklike naslaanbron.

Die gebruik van 'n geïntegreerde mikrostruktuur hou groot voordele in, veral met die oog op die regstreekse adressering wat daarmee gepaard gaan. Niegeïntegreerde mikrostrukture kan op 'n vernuwende manier in die Afrikaanse leksikografie aangewend word. 'n Sinvolle benutting van 'n aangepaste weergawe van die niegeïntegreerde mikrostruktuur kan daartoe lei dat Afrikaanse woordeboeke in die semantiese subkommentaar vir kotekste 'n duidelike aansluiting kan bewerkstellig by die verskillende semantiese subkommentare waarin betekenisparafrases aangebied word. So 'n werkswyse dien ook ter versterking van die artikel-interne kohesie. Daarnaas kan daar, in aansluiting by die model van niegeïntegreerde mikrostrukture wat in Wiegand (1996) voorgehou word, voorsiening gemaak word vir gerigte koteksinskrywings waar die verskillende kotekssoeksones telkens bepaal word volgens die woordsoort van die item waarmee die leksikale item wat deur die lemmateken verteenwoordig word, verbind. Dit sal tot 'n verhoging van die gehalte van Afrikaanse woordeboeke lei.

Bibliografie

Woordeboeke

Du Plessis, M. 1993. *Tweetalige Aanleerderswoordeboek/Bilingual Learner's Dictionary*. Kaapstad: Tafelberg.

Eksteen, L.C. (Red.). 1997[14]. *Groot Woordeboek/Major Dictionary*. Kaapstad: Pharos.

Labuschagne, F.J. en L.C. Eksteen (Reds.). 1993[8]. *Verklarende Afrikaanse Woordeboek.* Pretoria: J.L. van Schaik.

Odendal, F.F. en R.H. Gouws. 2000[4]. *Verklarende Handwoordeboek van die Afrikaanse Taal*. Midrand: Perskor.

Wahrig, G. 1994. *Deutsches Wörterbuch*. Gütersloh: Bertelsmann.

Ander bronne

Bergenholtz, H., S. Tarp en H.E. Wiegand. 1999. Datendistributionsstrukturen, Makro- und Mikrostrukturen in neueren Fachwörterbüchern. Hoffmann, L. et al. (Reds.). 1999. *Fachsprachen. Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft/Languages for Special Purposes. An International Handbook of Special-Language and Terminology Research*: 1762–1832. Berlyn: De Gruyter.

Gouws, R.H. 1989. *Leksikografie*. Kaapstad: Academica.

Gouws, R.H. 1999. *A Theoretically Motivated Model for the Lexicographic Processes of the National Lexi-*

cography Units. Ongepubliseerde verslag voorberei vir die Pan Suid-Afrikaanse Taalraad.

Gouws, R.H. 2000. Strategies in Equivalent Discrimination. Mogensen, J.E. et al. (Reds.). 2000. *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, R.H. 2001. Lexicographic Training: Approaches and Topics. Emejulu, J.D. (Red.). 2001. *Éléments de lexicographie gabonaise. Tome I*: 58-94. New York: Jimacs-Hillman.

Gouws, R.H. 2002. Equivalent Relations, Context and Cotext in Bilingual Dictionaries. *Hermes* 28: 195-209.

Gouws, R.H. 2003. Using a Frame Structure to Accommodate Cultural Data. Emejulu, J.D. (Red.). *Éléments de lexicographie gabonaise. Tome II*: 54-69. New York: Jimacs-Hillman.

Gouws, R.H. 2003a. Issues Regarding the Comment on Semantics in Bilingual Dictionaries Dealing with Closely Related Languages. Petkov, P. (Red.). 2003. *Sammelband: Internationales Kolloquium zur Wörterbuchforschung*. Reihe Germanistische Linguistik.

Gouws, R.H. en H.E. Wiegand. Om te verskyn. Strukturtraditionen in der zweisprachigen Lexikographie mit Afrikaans. Ein Beitrag zur Geschichte der Lexikographie in Südafrika. Wiegand, H.E. (Red.). *Studien zur zweisprachigen Lexikographie mit Deutsch IX*.

Hausmann, F.J. en H. E. Wiegand. 1989. Component Parts and Structures of General Monolingual Dictionaries: A Survey. Hausmann, F.J., et al. (Reds.). 1989–1991: 328-360.

Hausmann, F.J., et al. (Reds.). 1989–1991. *Wörterbücher. Ein internationales Handbuch zur Lexikographie/Dictionaries. An International Encyclopedia of Lexicography/Dictionnaires. Encyclopédie internationale de lexicographie*. Berlyn: De Gruyter.

Louw, P.A. en R.H. Gouws. 1996. Lemmatiese en nielemmatiese adressering in Afrikaanse vertalende woordeboeke. *Suid-Afrikaanse Tydskrif vir Taalkunde* 14(3): 92-100.

Tarp, S. en R.H. Gouws. Om te verskyn. Wie leer wat uit Afrikaanse (aan)leerderwoordeboeke? *Tydskrif vir Geesteswetenskappe.*

Wiegand, H.E. 1989. Der Begriff der Mikrostruktur: Geschichte, Probleme, Perspektiven. Hausmann, F.J. et al. (Reds.). 1989-1991: 409-462.

Wiegand, H.E. 1996. Das Konzept der semiintegrierten Mikrostrukturen. Ein Beitrag zur Theorie zweisprachiger Printwörterbücher. Wiegand, H.E. (Red.). 1996. *Wörterbücher in der Diskussion II. Vorträge aus dem Heidelberger Lexikographischen Kolloquium*: 1-82. Lexicographica. Series Maior 76. Tübingen: Max Niemeyer.

Wiegand, H.E. 1998. *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie*. Berlyn: De Gruyter.

Wiegand, H.E. 1998a. Lexikographische Textverdichtung. Entwurf zu einer vollständigen Konzeption. Zettersten, A. et al. (Reds.). 1998. *Symposium on Lexicography VIII. Proceedings of the Eighth Symposium on Lexicography May 2–4, 1996 at the University of Copenhagen*: 1-35. Lexicographica. Series Maior 90. Tübingen: Max Niemeyer.

New Advances in Corpus-based Lexicography[*]

Arvi Hurskainen, *Institute for Asian and African Studies, University of Helsinki, Helsinki, Finland (arvi.hurskainen@helsinki.fi)*

Abstract: This article presents various approaches used in corpus-based computational lexicography. A claim is made that in order for computational lexicography to be efficient, precise and comprehensive, it should utilize the method where the corpus text is first analysed, and the results of this analysis is then processed further to meet the needs of a dictionary. This method has several advantages, including high precision and recall, as well as the possibility to automate the process much further than with more traditional computational methods. The frequency list obtained by using the lemma (the equivalent of the headword) as basis helps in selecting the words to be included in the dictionary. The approach is demonstrated through various phases by applying SALAMA (the Swahili Language Manager) to the process. Manual work will be needed in the phase when examples of use are selected from the corpus, and possibly modified. However, the list of examples of use, arranged alphabetically according to the corresponding headword, can also be produced automatically. Thus the alphabetical list of headwords with examples of use is the material on which the lexicographer works manually. The article deals with problems encountered in compiling traditional printed dictionaries, and it excludes electronic dictionaries and thesauri.

Keywords: LEXICOGRAPHY, DICTIONARY, LANGUAGE TECHNOLOGY, COMPUTATIONAL LINGUISTICS, AUTOMATIC COMPILATION, DICTIONARY TESTING, INFORMATION RETRIEVAL, MORPHOLOGICAL ANALYSIS, SEMANTIC ANALYSIS, DISAMBIGUATION, HEURISTICS

Opsomming: Nuwe ontwikkelinge in korpusgebaseerde leksikografie. Hierdie artikel beskryf verskillende benaderings wat in korpusgebaseerde rekenaarleksikografie gebruik word. Daar word aangevoer dat vir rekenaarleksikografie om doelmatig, noukeurig en omvattend te wees, dit die metode behoort te gebruik waarby die korpusteks eers ontleed word, en die resultaat van hierdie ontleding dan verder verwerk word om te voldoen aan die behoeftes van 'n woordeboek. Hierdie metode het verskillende voordele, insluitende 'n hoë mate van noukeurigheid en herwinning, sowel as die moontlikheid om die proses baie verder as met meer tradisionele rekenaarmetodes te outomatiseer. Die frekwensielys verkry deur die lemma (die ekwivalent van die trefwoord) as basis te gebruik, help met die keuse van woorde vir insluiting in die woordeboek. Die benadering word geïllustreer deur verskillende fases van die aanwending van SALAMA (die Swahili Language Manager) in die proses. Werk met die hand sal nodig wees gedurende die stadium wanneer gebruiksvoorbeelde uit die korpus gekies en moontlik aangepas word. Die lys ge-

---

bruiksvoorbeelde, alfabeties gerangskik volgens die ooreenstemmende trefwoord, kan egter ook outomaties voortgebring word. Die artikel behandel probleme wat teëgekom word by die samestelling van 'n tradisionele gedrukte woordeboek, en dit sluit elektroniese woordeboeke en tesourusse uit.

Sleutelwoorde: LEKSIKOGRAFIE, WOORDEBOEK, TAALTEGNOLOGIE, REKENAAR-LINGUISTIEK, OUTOMATIESE SAMESTELLING, WOORDEBOEKTOETSING, INLIGTINGS-HERWINNING, MORFOLOGIESE ONTLEDING, SEMANTIESE ONTLEDING, ONDUBBELSIN-NIGMAKING, HEURISTIEK

1.   Introduction

The use of computers in lexicographical work has gone through various phases, where enthusiasm on the one hand and disappointment on the other have alternated. The calculating power and speed of computers were thought to revolutionise the compilation of dictionaries, and high expectations were held for automating the process. It was thought that text corpora could be transformed into dictionaries with minimal human intervention.[1]

In this kind of thinking, two major mistakes were made. It was thought that strings in text would, with minimal modifications, become lexemes and possible dictionary entries. The other mistake was that there was no linguistic insight built into the system.[2] At best this approach resulted in various kinds of concordances where the occurrence of a word or a group of words could be retrieved from text with a needed amount of context, and sorted in selected ways. Much of the usefulness of computers in lexicography was seen just in these terms (Jones and Sondrup 1989; Panyr and Zimmermann 1989). The automatic concordancing was, of course, a huge improvement compared with manual compilation, but there was nothing linguistically intelligent in it. These retrieving programs, often called KWIC (Key Word In Context), continue to be standard tools in dictionary work, but they are suitable only for selected tasks.

Because a good dictionary is much more than a list of words, linguistic sophistication is required from computer-based lexicography. In order for the computer-based lexicographical work to be really meaningful, the computer system used for the work has to acquire and make explicit the linguistic information attached to each of the potential lexemes in the dictionary. These requirements include, *inter alia*

—   the category of each word (part of speech),

—   sufficient information for guiding in the use of a word, such as inflection, concordance, tone pattern, argument structure, etc.[3],

—   semantic information, including glosses in bilingual dictionaries,[4]

—   etymological information,[5] and

—   the commonness of a word (frequency category).

Only fairly recently computational lexicography has come to the level where both realism and know-how make it possible to achieve significant advances (Teubert 2001). Much of the current work is still concentrating on the problems encountered in the lexicography of English and other Western languages. African languages raise different kinds of problems, including complex morphology, tonology, disjoining writing systems, etc., and these have to be faced and solved.

A major problem in the computational analysis of language is ambiguity. The extent of ambiguity varies among languages, but in every language it is a problem and needs to be solved. Ambiguity occurs on the morphological level, as well as on the syntactic and semantic levels. A word in isolation may have more than one morphological interpretation. It may have more than one syntactic function, and more than one semantic role, especially several textual meanings.

The computer system designed for lexicographical work should be able to address each of these problems and solve them. This calls for a full computational description of a language, a description that in great detail makes use of linguistic rules and is lexically comprehensive. In other words, the system should be able to analyse unrestricted text of a particular language.

In order to make the subsequent discussion more comprehensible, a description will be given of SALAMA (the Swahili Language Manager), a computer system designed for Swahili, a major Bantu language. Work on the computer description of this language started in 1985, and by now has reached a phase where almost all the problems have at least been addressed, and most of them solved.[6] The system will be briefly described phase by phase, and then by means of examples it will be shown how the system can be applied for dictionary compilation.

2.   Choice of headwords

Data in language dictionaries are usually arranged under headwords ordered alphabetically. Good dictionaries also have sub-entries for listing such lexical words that are either derivatives of headwords or are in some other way closely related to the headword. Lexicographers consider the choice of headwords fairly difficult.[7] Because the final product of dictionary work has to be limited in size, a choice of headwords has to be carried out. Here we will discuss the choice of entries for a general language dictionary, although methods for semi-automatic compilation of domain-specific dictionaries have also been developed.[8]

We may think that a large enough and balanced corpus of general language text is a base for such a dictionary, and by retrieving the lemmas of words in the corpus we will get a reliable list of dictionary entries. The task is

not so simple, however. We need large amounts of various types of text for the corpus, and we also have to think about its representativeness. A problem with text-based lexicography is that words used mainly in spoken contexts will not be represented in text, and such words need to be considered separately. One method is to use transcriptions of spoken corpora as source for spoken language, but sufficiently large and representative spoken corpora are rarely available.

A systematic and comprehensive analysis of written language starts from the identification and analysis of individual words. More specifically, what we find in text is actually word-forms and not such words we find as dictionary entries. Such word-forms will be analysed morphologically, and each interpretation will be made explicit. Thus the interpretation of many word-forms becomes ambiguous, i.e. the word-form has more than one legitimate interpretation.

The concept of 'word' itself is also not as clear as it seems. In lexicography, we are more interested in grammatical words than orthographic words. Grammatical words fairly closely correspond to concepts, and it is the concepts and their definitions we need to deal with in lexicography. A concept may be represented in text by more than one string of characters. The treatment of such multi-word concepts may already be problematic in counting word frequencies of English (Kilgarriff 1997), but it can be detrimental in languages with a disjoining writing system (Hurskainen and Halme 2001).

Multi-word concepts can be treated as single concepts in automatic processing, especially if their constituent parts do not inflect and if they are adjacent to each other. This can be done by temporarily joining such word clusters together, and in the final version the words can be returned to their original shape. Grammatical words allowing other words between the constituent parts cannot be treated in this simple way, but there are means for treating them too (Tapanainen and Järvinen 1998).

One requirement for a useful system is that it has to be comprehensive. In other words, it should not leave words in text without interpretation, however rare or strange they are. There are two major reasons for this. There should be a 'master dictionary' that contains all the grammatical information of the language, as well as all lexical information. When compiling a smaller dictionary for a specific purpose, it is easier to filter out unnecessary analysed material than to cope with unrecognised (and unanalysed) words. Another reason for comprehensiveness is that in order for a disambiguating program to fulfil the task reliably there should not be unanalysed words in text.

If the text corpus is large and balanced enough, the core vocabulary of the dictionary can be selected on the basis of the lemma list arranged in frequency order. For example, we may think of choosing the 10 000 most frequent lemmas for a dictionary. Except for special purpose dictionaries, it is a good policy to include words in order of frequency in the dictionary. The point where the frequency list will be cut depends on the intended size of the dictionary. This

method ensures that at least all common words will be included.

This statement sounds trivial, but it is not trivial at all. In the comprehensive computer evaluation of five Swahili dictionaries (Hurskainen 1994, 2002), it was found that the two most authoritative dictionaries[9] had serious omissions in core vocabulary, although they had a fairly large percentage of words not found in any texts at all. The tests were made with three different corpora, totalling 4 227 362 words. The results show that the monolingual dictionary *Kamusi ya Kiswahili Sanifu* (KKS) was able to recognize between 89.7 and 91.8% of the words of the three corpora, and *Kamusi ya Kiswahili–Kiingereza* (KKK) recognized 90.7 to 92.9% of the words. At the same time, both dictionaries listed a number of such words not found in the corpus. Only half the nouns (precisely 50%) of classes 1/2, 3/4, 5/6, 7/8, and 9/10 listed in KKS were found in the corpus. The corresponding percentage in KKK was 55, i.e. it had less 'excessive' words. With verbs the situation was better: 78% for KKS and 85% for KKK.

If we compare these results with *Swahili–Suomi–Swahili-sanakirja* (Abdulla et al. 2002), which was also tested, we find interesting differences. This dictionary was produced by using a corpus as base for selecting headwords. Its success rate in recognising the words of the corpus ranges between 91 and 94%. In other words, it covers the vocabulary of the corpora slightly better than KKS and KKK. On the other hand, the percentage of 'excessive' nouns of the classes mentioned above was only 24%, and with verbs it was practically zero. In other words, only such verbs also used in the corpora were listed in the dictionary.

These statistics reveal the possibilities of modern language technology to show in detail weaknesses of existing dictionaries, as well as the improvements technology can bring to dictionary compilation.

This lengthy discussion on the problems of selecting headwords for a dictionary reveals that it is a major issue. The use of a frequency list of corpus lemmas is a safe method of avoiding at least major omissions.

The frequency list is, however, not the final entry list of the dictionary. The corpus is rarely so large and balanced that it alone provides all words needed, even for a fairly modest dictionary. Many words used in everyday life are often missing in the corpus, because such matters are not dealt with in texts. Names of flora and fauna are also insufficiently found in texts.

3.   Format of the corpus

It was pointed out above that for the corpus to be maximally useful in dictionary compilation, the linguistic information of the text must be made explicit. Even the first task, i.e. the production of the lemma list, does not succeed in languages with left-branching (prefixing) inflection without a morphological analysis program capable of returning the correct lemma of each word-form. For automatic inclusion of relevant linguistic information needed in a dictionary, the linguistic analyser is an absolute necessity.

Therefore, it is not a question of whether the corpus should be tagged or not, but how and in what phase the tagging is to be performed. Principally there are two methods of tagging, both of them automatic. In one method, which is more traditional, the raw text is tagged with a computer program, and the tagged version of the corpus is then used by the lexicographer as source text. Queries are made to the tagged version, and tags can be used as search keys.

In another method, which basically performs the same operations as the one described above, the lexicographer works with raw text and uses the whole array of programs and utilities in compiling the dictionary. In this method, the user has the raw material (text) and a comprehensive set of tools (programs, utilities, filters, scripts, etc.), which can be used in a number of ways, depending on the type of task.

The latter method is better than the former for several reasons. The user is free to select or prepare their own texts without resorting to tagged corpora prepared by someone else, often for purposes not ideal for the current task. The user also avoids handling of excessively large files. On average, the analysed Swahili text is 16 times larger than the original text, and even after disambiguation it is still 11 times larger than the original. Any editor has difficulties in handling files of this magnitude.

The size problem can be conveniently solved so that the analysis and disambiguation are carried out 'in flight', which means that the user does not even see the results of these phases, because further processing can be carried out in pipe. In lexicography we do not need to see all occurrences of a word in the corpus. We rather want to know in what senses the word occurs in the corpus, and how many times it occurs in each sense. By condensing the format of the information, we do not lose any lexically important information, but the space required for presenting this is cut to a minimum. The larger the corpus, the bigger is the advantage. This method of lexicography requires a working environment, where piping of processes is possible, such as Linux and Unix.

4.   Searching headwords from the corpus

How can the occurrences of a lexical word be found in the corpus? There are currently at least three methods for doing this. Each of these and their suitability for African languages will be briefly discussed below.

4.1   Direct string search — traditional approach

In languages with right-branching inflection and derivation, direct string search is not a major problem, because the potential headwords and their inflected and derived forms are adjacent to each other in alphabetical listing. In languages with predominantly left-branching inflection, the problem is more

serious, as is demonstrated in (1). Our task is to extract all occurrences of the verb *soma* (to read). As can be seen, the search string cannot be the whole verb stem, but only the root *som*, because the verb may also be ending in e or i, and various types of derivative suffixes can be added. Similarly, a large set of (strings of) prefixes has to be taken into account.

(1) Example of string search

[486] donner$ cat maj1999 | kw-alg 'som'

```
     ambaye ana som a darasa la pili √
  simu akiwa Mu som a, Dk. Mazara
Masatu, Rajab M som a, Elia
  ilikuwa ni ku som a ile barua. √
  Kwa sababu wa som aji ndio wateja
            Wa som ali hao wanadaiwa
 kituo cha Tran som a Mabibo na Bw.
sekondari ya Ka som a wilaya ya Musoma
 ndiye aliyewa som ea mashitaka √
   maelezo yali som eka kuwa anakufa √
   ikiwemo kuwa som esha. √
      Azizi ali som ewa mashitaka hayo √
 viongozi na wa som i ambao wamejaa tele
  huko pia ni m som i kwa kuwa ana
   na baada ya  som o, baadhi ya
aendelee na ma som o. Bw. Hiza
 uamuzi huo ku som wa. √
 ili ushahidi u som we hadharani √
```

With the keyword *som* we are likely to get all the real cases, but also a lot of wrong words.[10] If we try to modify the search string so that wrong hits will be reduced, we run the risk of excluding real cases.

4.2   String search with regular expressions

The search is much more accurate if we use regular expressions in formulating the search key. If language analysis tools are not available in dictionary compilation, this is a valuable alternative. It is far more efficient than direct string search, but it is not even nearly as accurate and efficient as the compilation by employing language analysis tools.

Instead of using *som* as search key we have to approach the problem by also trying to describe other elements of the verb that are distinctive enough for separating them from other word categories. As the verb final vowels may be a, e, i and u, this is not a promising approach, because many word categories have similar endings.

A more promising approach is the description of verb prefixes, because there is usually a longer string of characters typical to verbs only. The problem

is that there are at least tens of thousands of such grammatical character combinations. Regular expressions, however, make the formulation of such queries possible, even practical. In (2), such a query has been used, and as the result shows, all findings now are verbs.

(2) Example of search by using regular expressions

[487] donner$ cat maj1999 | \
kw-alg '(ha)?(ni|u|a|tu|m|wa|i|li|ya|ki|vi|zi|ku|pa|mu)\
(na|li|ta|me|si)(ye|o|yo|lo|cho|vyo|zo|ko|po|mo)?\
(ni|u|m|mw|i|li|ya|ki|vi|zi|ku|pa|mu)?som'

```
        ambaye anasoma darasa la pili √
     ilikuwa ni kusoma ile barua. √
     ndiye aliyewasomea mashitaka √
  na maelezo yalisomeka kuwa anakufa √
     ikiwemo kuwasomesha. √
        Azizi alisomewa mashitaka hayo √
     uamuzi huo kusomwa. √
  ili ushahidi usomwe hadharani √
```

Even this search string is not accurate, because it leaves out the so-called general present tense, subjunctive, present tense negative, infinitive, and several more rare tense/aspect forms. It is difficult, and dangerous, to include such possibilities in the same search key, because the danger of getting unwanted strings will multiply.

Let us modify our previous task, so that instead of searching the verb *soma*, we look for all occurrences of each verb in the corpus. We cannot use the verb stem as part of the search key now, because there are thousands of verbs, and we do not know in advance what they are. We may try to simulate the verb stem by defining its minimum length. With some verb forms of monosyllabic verbs it is as short as two characters. Unfortunately this is also the length of the stem in many independent relative constructions, and in some it is even three characters. Thus it seems impossible to get an unmixed list of verbs only. Examples of found strings are shown in (3). Verb roots are in bold face.

(3) An attempt to retrieve verbs by using regular expressions[11]

[489]$ cat maj1999 | \
kwic -s '(ha)?(ni|u|a|tu|m|wa|i|li|ya|ki|vi|zi|ku|pa|mu)\
(na|li|ta|me|si)(ye|o|yo|lo|cho|vyo|zo|ko|po|mo)?\
(ni|u|m|mw|i|li|ya|ki|vi|zi|ku|pa|mu)?[a–z][a–z]+'

```
     wawakilishi wa CUF alichokiita kuwa √
   mwanachama wa chama alichokuwamo wakati √
       NAFCO, anadaiwa aliitumia hali hiyo √
     Malera aliongeza, aliitwa mtuhumiwa namba √
     jana kuwa Mohamed alikiri kosa hilo lakini √
     Bibi Subira kwani alikufa kutokana na √
```

```
          mfupi baba yake alikuja na kuanza √
              Rais Mkapa aliliambia jopo hilo √
      kufungwa kwa duka, alilipa faini hiyo. √
         nusu, Bw. Kahale alimpa fomu za kukata √
              wa mjadala, alivipa changamoto vyombo √
           kwa shuti kali lililomshinda kipa Masuke √
       Hata hivyo, ilivuta usikivu wa washabiki.
        kwa shule hizo ni usimamizi mbovu wa
          tayari kupoteza utaifa wao na kama
      Magharibi ambayo si utamaduni wa wananchi
    polisi waliambiwa na wanakijiji kuwa baadhi
         jijini jana kuwa wanamichezo hao walifariki
         kutokea Kenya na wanamiliki silaha kali √
      Ilala Boma baada ya wanamuziki hao kudai
        na timu ya Vijana wanaume itashiriki
              Mmoja wa wasimamizi mlangoni
```

The search found 5,770 verb candidates, and as expected, there were independent relatives and also nouns that fulfilled the search criteria. Some of these are shown in (3). The precision was, however, very good: more than 98%. The recall was much worse. The analysis with SALAMA showed there were in addition 2 659 such words that were unambiguously verbs. Thus the recall was as low as 68%. This could be improved considerably by using search strings, which were excluded above and which could not be included in the same search.

The identification of a verb lemma is even more difficult than the identification of a verb. We could think of writing a program that would mark the beginning of a verb lemma for each verb in text. This code could then be used in retrieving the lines. In this way we would get a concordance list where the beginning of each verb lemma is marked. It would then be fairly easy to isolate the correct lemma, although a fairly large amount of manual work would be necessary.

4.3   Advanced approach — analyse text first

Although the use of regular expressions facilitates complicated search strings, it is still far from the precision, recall, and ease of the use of an approach where the text is first analysed linguistically. In this method, the following features are made explicit:

—   The lemma or base form of the word can be defined so that it is identical with the headword of the dictionary. As a consequence, we get a list of words to be included in the dictionary.

—   Part-of-speech information is given by the analysis program.

—   The program produces a detailed list of morphological features of the

word-form found in text.

— Semantic features can be added. For example, the information on ani-
   mality or humanness, may be necessary for defining the correct concor-
   dance pattern. Verbs may also be given information on their argument
   structure (SV, SVO, SVOO, etc.).

— If the dictionary is intended to be bilingual, semantic glosses in another
   language can be automatically produced for each dictionary entry.[12]

— Syntactic features (subject, object, various roles of verbs, dependent con-
   stituents in noun phrases, etc.) can be added. In dictionary compilation,
   such features are usually omitted.

— Information on the etymology of words can be added.

— Variant, or non-standard, orthography can be reported.

5.   The problem of ambiguity

Word-forms often have more than one interpretation. A word-form may be-
long to more than one word class. English is a good example of this kind of
ambiguity. In Bantu languages, ambiguity is often caused by the fact that the
same morpheme is a marker of more than one noun class. Although word-
forms may be ambiguous on the word level, in context they normally have only
one interpretation. A general rule is that the more comprehensive the analyser
is, the more ambiguity the result has.

There are two major approaches for solving ambiguity. One method relies
on probabilities. If a word-form has two interpretations and one of these is
common and the other rare, then the common one is chosen. The result is often
correct, but one is never certain whether it is correct or not, because the choice
was made on the basis of probability. In another method, ambiguity is resolved
with context-sensitive 'linguistic' rules. For the vast majority of cases, context-
sensitive rules fulfil the task.

Heuristic rules are used only for cases where there is no basis for con-
structing a linguistic rule. On the basis of morphological features, such rules try
to guess the correct interpretation of the word. For example, if a word begins
with m- and ends with -aji, the word is very likely a deverbative noun of noun
class 1. It is self-evident that ambiguity can be resolved only in context, i.e. as
part of real text.

(4) An example of ambiguity in Swahili

"<ofisi>"
       "ofisi" N 5a/6-SG ENG 'office'
       "ofisi" N 9/10-0-SG ENG 'office'
       "ofisi" N 9/10-0-PL ENG 'office'

```
"<ya>"
        "a" GEN-CON 3/4-PL
        "a" GEN-CON 9/10-SG
        "a" GEN-CON 5/6-PL
        "a" 5/6-PL-SP
"<kampuni>"
        "kampuni" N 5a/6-SG ENG 'company'
        "kampuni" N 9/10-0-SG ENG 'company'
        "kampuni" N 9/10-0-PL ENG 'company'
"<yake>"
        "ake" PRON POSS 3/4-PL SG3 'his/her/its'
        "ake" PRON POSS 9/10-SG SG3 'his/her/its'
        "ake" PRON POSS 5/6-PL SG3 'his/her/its'
"<iko>"
        "iko" 3/4-PL-SP LOC-17 'be (in place)'
        "iko" 9/10-SG-SP LOC-17 'be (in place)'
"<ghorofa>"
        "ghorofa" N 5a/6-SG AR 'storey, floor'
        "ghorofa" N 9/10-0-SG AR 'storey, floor'
        "ghorofa" N 9/10-0-PL AR 'storey, floor'
"<ya>"
        "a" GEN-CON 3/4-PL
        "a" GEN-CON 9/10-SG
        "a" GEN-CON 5/6-PL
        "a" 5/6-PL-SP
"<tano>"
        "tano" NUM 9/10-PL NUM-INFL CARD 'five'
        "tano" NUM NUM-INFL ORD 'fifth'
```

By using a Constraint Grammar parser (CG2) ambiguity is resolved with the help of context-sensitive rules. The process of resolving ambiguity is also called 'disambiguation'. The result is shown below.

(5) Ambiguity resolved

```
"<ofisi>"
        "ofisi" N 9/10-0-SG ENG 'office'
"<ya>"
        "a" GEN-CON 9/10-SG
"<kampuni>"
        "kampuni" N 9/10-0-SG AR 'company'
"<yake>"
        "ake" PRON POSS 9/10-SG SG3
"<iko>"
        "iko" 9/10-SG-SP LOC-17 'be (in place)'
"<ghorofa>"
        "ghorofa" N 9/10-0-SG AR 'storey, floor'
"<ya>"
        "a" GEN-CON 9/10-SG
```

"<tano>"
    "tano" NUM NUM-INFL ORD 'fifth'

6.   Removing excessive tags

Experience has shown that the more detailed the analysis of words, the better possibilities it offers for linguistically motivated disambiguation. Therefore, all features should be made explicit in morphological and semantic analysis, because they may be needed in writing disambiguation rules. An example of complexity is provided in (6), where a few word-forms of the verb *andika* (to write) have been analysed. Note that morpheme boundaries (+) have been manually added, and ambiguity has been removed by rules, so that each form has only one interpretation.

(6) All tags retained

"<wa+me+mw+andik+i+a>"
"andika" V 1/2-PL3-SP VFIN PERF:me 1/2-SG3-OBJ SV SVO SVOO 'write' APPL
"<wa+me+ji+andik+ish+a>"
"andika" V 1/2-PL3-SP VFIN PERF:me REFL-SG-OBJ SV SVO SVOO 'write' CAUS
"<ni+li+andik+ish+w+a>"
"andika" V 1/2-SG1-SP VFIN PAST SV SVO SVOO 'write' CAUS PASS
"<a+li+andik+ish+w+a>"
"andika" V 1/2-SG3-SP VFIN PAST SV SVO SVOO 'write' CAUS PASS
"<a+li+ye+andik+a>"
"andika" V 1/2-SG3-SP VFIN PAST 1/2-SG-REL SV SVO SVOO 'write'
"<a+li+ye+zi+andik+a>"
"andika" V 1/2-SG3-SP VFIN PAST 1/2-SG-REL 9/10-PL-OBJ SV SVO SVOO 'write'
"<a+li+i+andik+i+a>"
"andika" V 1/2-SG3-SP VFIN PAST 9/10-SG-OBJ SV SVO SVOO 'write' APPL
"<a+li+yo+i+andik+a>"
"andika" V 1/2-SG3-SP VFIN PAST 9/10-SG-REL 9/10-SG-OBJ SV SVO SVOO 'write'
"<a+li+li+andik+i+a>"
"andika" V 1/2-SG3-SP VFIN PAST 5/6-SG-OBJ SV SVO SVOO 'write' APPL
"<a+mekwisha+mw+andik+i+a>"
"andika" V 1/2-SG3-SP VFIN PERF:mekwisha 1/2-SG3-OBJ NON-STD SV SVO SVOO
    'write' APPL

The description in (6) has much such information we do not need in a dictionary. Therefore we remove part of the tags and leave those that are useful. After having removed excessive tags, we get a more readable output as in (7).

(7) Part of tags removed

"<wamemwandikia>"   "andika" V SVOO 'write' APPL
"<wamejiandikisha>"   "andika" V SVOO 'write' CAUS
"<niliandikishwa>"   "andika" V SVOO 'write' CAUS

```
"<aliandikishwa>"           "andika" V SVOO 'write' CAUS
"<aliyeandika>"             "andika" V SVOO 'write'
"<aliyeziandika>"           "andika" V SVOO 'write'
"<aliiandikia>"             "andika" V SVOO 'write' APPL
"<aliyoiandika>"            "andika" V SVOO 'write'
"<aliliandikia>"            "andika" V SVOO 'write' APPL
"<amekwishamwandikia>"  "andika" V SVOO 'write' APPL
```

In (7), the analysis program was used in the mode that returned the basic verb lemma but retained the information on verbal extensions. For finding out verb frequencies in the corpus, this mode is useful, because it returns the base form of the verb regardless of its actual form in text. In dictionaries, we often need listing at least part of the extended forms, especially if their meanings are not directly derivable from linguistic rules. For such purposes, a format shown in (8) is better, because it returns extended forms as lemmas. These extended forms are often alphabetically listed as sub-entries after the headword.

(8) Verbal extensions in verbs retained.

```
"<wamemwandikia>"          "andikia" V SVOO 'write' APPL
"<wamejiandikisha>"        "andikisha" V SVOO 'write' CAUS
"<niliandikishwa>"         "andikisha" V SVOO 'write' CAUS
"<aliandikishwa>"          "andikisha" V SVOO 'write' CAUS
"<aliyeandika>"            "andika" V SVOO 'write'
"<aliyeziandika>"          "andika" V SVOO 'write'
"<aliiandikia>"            "andikia" V SVOO 'write' APPL
"<aliyoiandika>"           "andika" V SVOO 'write'
"<aliliandikia>"           "andikia" V SVOO 'write' APPL
"<amekwishamwandikia>"  "andikia" V SVOO 'write' APPL
```

7.   Post-processing of the analysed corpus

When each word in the corpus is analysed and the ambiguity resolved, the result can be manipulated in a number of ways. In dictionary work, we in fact need several kinds of modifications to the result.

For the selection of dictionary entries, we need a frequency list according to the lemma. In order for the list to be correct, we need to remove the actual word-form and all such tags that describe inflection, as well as the codes of verbal extensions. By doing this, we may collapse the list in (7) above and get a single line as shown in (9).

(9) A format needed for counting frequencies of headwords

10 andika V SVOO 'write'

If verbal extensions are also counted as separate lexical entries as in (8) above, we get a list as shown in (10). Note, however, that if the list is sorted in fre-

The dictionary itself is ordered according to the headword, and for this reason
we have to rearrange the data. We also want to retain information on the fre-
quency of the words. Selected entries from the alphabetically arranged data,
extracted from a small section of the news corpus, are shown in (12).

(12) Selected dictionary entries produced by SALAMA

43	awali	ADV 'first, originally (ar)'
32	awali	N 9/10 '1 first. 2 origin, cause. 3 above (ar)'
7	awamu	N 9/10 'phase'
4	azimio	N 5a/6 'declaration'
6	azma	N 9/10 'intention; desire, purpose'
13	baa	N 9/10 'bar, pub. (eng)'
561	baada ya	PREP 'after'
117	baadaye	ADV 'thereafter, afterwards, then, later (on). (ar)'
103	baba	N 9/10 'HUM father, (zamani) sire.'
28	badala ya	PREP 'in stead of'
20	badala yake	PREP 'in stead of him/her/it'
14	badiliko	N 5a/6 AR 'change'
92	bado	ADV 'not yet, still (ar)'
2	bagua	V SVO '1 separate. 2 discriminate against, segregate'
14	baina ya	PREP 'between'
14	baini	V SVO 'realize, recognize (ar)'
11	baiskeli	N 9/10 'bicycle, (hist) velocipede (eng)'
6	baki	N 5a/6 '1 remainder, residue; balance. 2 (chakula) left-overs (ar)'
36	baki	V SV '1 remain. 2 stay/be left behind (ar)'
1	bakiza	V SV SVO 'leave behind; leave (not taking everything)'
6	banda	N 5a/6 'shed, barrack, barn, hut; hovel'
5	bandia	N 9/10 '1 doll, dummy. 2 imitation (ar)'
6	banja	V SVO '1 crack; break, split (nuts, firewood etc). 2 strike. 3 (ms) bark up the wrong tree'
49	bara	N 9/10 'continent (ar)'
86	barabara	N 9/10 'highway, road, street, turnpike, way, avenue'
7	baraka	N 9/10 '1 blessing, benediction, boon, favour. 2 prosperity, progress, abundance (ar)'
	. . .	
1	plastiki	N 5a/6 'plastic (eng)'
3	plastiki	N 9/10 'plastic (eng)'
20	pombe	N 9/10 'local brew, beer'
21	ponda	V SV 'pound, crush, mash; smash, crash'
17	posho	N 9/10 '1 allowance. 2 food, ration'
13	potea	V SV '1 be lost. 2 be wrong, err'
13	potoa	V SVO '1 twist, make crooked/curved/slanting. 2 ruin, pervert, spoil'
2	potofu	ADJ A-INFL '1 stray; misleading. 2 spoiled'
19	profesa	N 9/6 'AN HUM professor. (eng)'
5	pumziko	N 5a/6 'pause; half-time, interval, break, recess'

5	punde	ADV 'soon, in a short while, shortly. (ms) ~ si ~ suddenly'
9	puuza	V SVO 'disregard, ignore, snub'
69	pya	ADJ A-INFL '1 new, recent, modern. 2 novel, strange'
16	rafiki	N 9/6 'AN HUM friend; comrade. (ar)'
30	raia	N 9/10 'AN HUM citizen. 2 civilian (ar)'
113	rais	N 9/6 'AN HUM president (ar)'
38	rasmi	ADJ A-UNINFL 'official, formal (ar)'

We see that some nouns are used in two different noun classes, and the frequencies of each usage are shown. Inflecting adjectives and non-inflecting adjectives have separate codes, which is necessary information for the dictionary user. Verb types are classified and marked with transitive (SVO) and intransitive (SV) tags. Etymological information, if applicable, is given at the end of the gloss.

In (13), we finally have a form where frequency information has been transformed into classes, the most frequent ones being marked with three dark dots, and the least frequent ones with no dots at all. Some further formatting has also been incorporated, all without manual intervention.

(13) Dictionary entries with frequency classes

awali	*adv* 'first, originally (*ar*)' ● ●
awali	*n* 9/10 '1 first. 2 origin, cause. 3 above (*ar*)' ● ●
awamu	*n* 9/10 'phase'
azimio	*n* 5a/6 'declaration'
azma	*n* 9/10 'intention; desire, purpose'
baa	*n* 9/10 'bar, pub. (*eng*)' ●
baada ya	*prep* 'after' ● ● ●
baadaye	*adv* 'thereafter, afterwards, then, later (on). (*ar*)' ● ● ●
baba	*n* 9/10 'HUM father, (zamani) sire.' ●
badala ya	*prep* 'in stead of' ● ●
badala yake	*prep* 'in stead of him/her/it' ●
badiliko	*n* 5a/6 AR 'change' ●
bado	*adv* 'not yet, still (*ar*)' ● ● ●
bagua	*v* SVO '1 separate. 2 discriminate against, segregate'
baina ya	*prep* 'between' ●
baini	*v* SVO 'realize, recognize (*ar*)' ●
baiskeli	*n* 9/10 'bicycle, (hist) velocipede (*eng*)' ●
baki	*n* 5a/6 '1 remainder, residue; balance. 2 (chakula) left-overs (*ar*)'
baki	*v* SV '1 remain. 2 stay/be left behind (*ar*)' ● ●
bakiza	*v* SV SVO 'leave behind; leave (not taking everything)'
banda	*n* 5a/6 'shed, barrack, barn, hut; hovel'
bandia	*n* 9/10 '1 doll, dummy. 2 imitation (*ar*)'
banja	*v* SVO '1 crack; break, split (nuts, firewood etc). 2 strike. 3 (ms) bark up the wrong tree'
bara	*n* 9/10 'continent (*ar*)' ● ●
barabara	*n* 9/10 'highway, road, street, turnpike, way, avenue' ● ● ●

baraka	*n* 9/10 '1 blessing, benediction, boon, favour. 2 prosperity, progress, abundance (*ar*)'
. . .	
plastiki	*n* 5a/6 'plastic (*eng*)'
plastiki	*n* 9/10 'plastic (*eng*)'
pombe	*n* 9/10 'local brew, beer' •
ponda	*v* SV 'pound, crush, mash; smash, crash' •
posho	*n* 9/10 '1 allowance. 2 food, ration' •
potea	*v* SV '1 be lost. 2 be wrong, err' •
potoa	*v* SVO '1 twist, make crooked/curved/slanting. 2 ruin, pervert, spoil' •
potofu	*adj* A-INFL '1 stray; misleading. 2 spoiled'
profesa	*n* 9/6 'AN HUM professor. (*eng*)' •
pumziko	*n* 5a/6 'pause; half-time, interval, break, recess'
punde	*adv* 'soon, in a short while, shortly. (*ms*) ~ si ~ suddenly'
puuza	*v* SVO 'disregard, ignore, snub'
pya	*adj* A-INFL '1 new, recent, modern. 2 novel, strange' • • •
rafiki	*n* 9/6 'AN HUM friend; comrade. (*ar*)' •
raia	*n* 9/10 'AN HUM citizen. 2 civilian (*ar*)' • •
rais	*n* 9/6 'AN HUM president (*ar*)' • • •
rasmi	*adj* A-UNINFL 'official, formal (*ar*)' • •

If we want to furnish the dictionary with examples of use, as we normally do, we need to retrieve such examples from the corpus. In order to automate the process, we need a third kind of list where the lemmas (i.e. headwords) are attached to the actual word-forms in the corpus. Basically the production of such a list is simple, because it is the default format of the analysis result of SALAMA. The problem is that if we do a selection of lemmas according to frequency, it is not easy to delete the correct lemmas from the original list, because the frequency order there is completely different compared with the lemma list. The solution is to retrieve all such lines from the main list where the lemmas of our selection list occur. As a result, we have a list of only those words we intend to include in the dictionary, and the list also has accurate information on the actual word-forms we can use as key for retrieving examples of use in the corpus.

The search for examples of use can be performed in two ways. One possibility is interactive where the dictionary compiler checks from the corpus the use of each lemma by employing one of several search programs or a more user-friendly interface. The other possibility is to retrieve the needed examples with a program. The resulting file will have all those words in the context, for which we want examples of use. By sorting such lines according to the lemma, we get a list of examples of use in the same order as in the dictionary. It is then fairly simple for the dictionary compiler to select and modify suitable examples of use to be included in the final dictionary. In (14), we have an extract from an alphabetically ordered list of the use of words in context. This list was produced by a program which used the word-form (not lemma) as search key.

(14) Words in context

dai: *barua hiyo imesainiwa na watu 10 <waliodai> kuwawakilisha wenzao.
dai: *habari <zilidai> kuwa hatua hiyo inatokana na kile kilichoelezwa kuwa ni mtindo wa *bw.
dai: *hamad kuwataka wanachama wafanye subira kila wanapotaka kufanya jambo fulani la <kudai> haki.
dai: *hata_hivyo, <alidai> kuwa wafuasi wengine wa chama hicho waliendelea kushikiliwa na polisi na kwamba hadi jana mchana walikuwa hawajaachiwa.
dai: *hata_hivyo, majina ya wafuasi wengine <waliodaiwa> kushikiliwa na polisi hayakuweza kupatikana mara_moja.
dai: *ngawaiya <alidai> kuwa baada_ya yeye kufuatilia suala hilo polisi, alielezwa kuwa gari hilo lilikamatwa kwa_kuwa dereva wake hakuwa na leseni.
dai: *profesa *lipumba alisema chama hicho kitafanya maandamano hayo <kudai> mambo matatu.
dai: <*alidai> kuwa kwa sasa wafuasi hao wamefunguliwa mashitaka ya uzururaji.
dai: <*walidai> kuwa uamuzi wa kuteua nyumba zinazostahili kubomolewa ndani_ya bonde hilo umefanywa bila tathmini ya kitaalamu.
fariki: *gabriel *ngwilulupi alisema jana nyumbani kwa marehemu *ukonga *staki *shari, kwamba marehemu <alifariki> juzi usiku katika hospitali ya *taifa *muhimbili kwa ugonjwa wa kiharusi.
fariki: *hezron *mhela <kufariki> muda mfupi kabla_ya uchaguzi.
fariki: *mtumishi wa umma na mwanasiasa wa siku nyingi nchini *mzee *brown *ngwilulupi (76) <amefariki> dunia.
fuatilia: *omari pia wamewaagiza wakaguzi wa kahawa wa bodi hio pia <kufuatilia> kwa karibu suala hilo na kutoa taarifa kwake mwisho wa mwezi.
fuatilia: *wiki moja kabla_ya siku kuu ya *krismasi, mwaka jana, walionekana baadhi ya viongozi wa serikali za vijiji katika wilaya ya *rombo, *moshi na *hai <waki-fuatilia> ushuru huo kwenye makampuni hayo bila mafanikio.
hatua: *alisema uamuzi wa serikali wa kununua umeme kutoa nchini *zambia ni <hatua> thabiti kwani inaonekana ni utekelezaji wa dira ya taifa ya mpango wa kuinua uchumi wa *taifa.
hatua: *omari alisema kuwa, ifikapo mwishoni mwa mwezi huu, kama makampuni hayo yatashindwa kulipa ushuru <hatua> za kisheria zitachukuliwa dhidi_yao kwa_mujibu_wa kanuni na sheria za ununuzi wa kahawa kutoka_kwa waku-lima chini_ya mfumo wa soko huru.
hatua: *taarifa hiyo ilisema <hatua> hiyo inatokana na ukweli kwamba ujenzi wa makazi ya watu katika eneo hilo hauruhusiwi na ni kinyume cha sheria.
hatua: *wamelalamika kuwa ujenzi wa nyumba zao ulitokana na hali ngumu ya kuba-na matumizi kutokana_na kipato kidogo wanachokipata lakini *mkurugenzi huyo amefikia <hatua> ya kutoa agizo lenye athari kubwa kwao na familia zao.
ingia: "*yatakuwa maandamano ya amani, lakini kwa kadri tunavyowajua polisi wetu <watatuingilia> kwa lengo la kuvuruga amani ... wakija na magari yao msiwakimbie na muwe imara kukabiliana nao", alisema *profesa *lipumba ali-pokuwa akiwahutubia wanachama wa chama hicho katika ukumbi wa *diamond *jubilee, *dar_es_*salaam jana.
ingia: *aidha, baada_ya kustaafu shughuli za utumishi, *mzee *ngwilulupi <aliingia> kwenye siasa, ambapo alikuwa miongoni_mwa watu waliopigania mfumo wa

vyama vingi nchini na kufanikiwa.

ingia: *hata_hivyo, wakazi hao wamemuomba *rais *benjamin *mkapa <aingilie> katika hatua hiyo kwa madai kuwa ni ya uonevu.

ingia: *mwenyekiti wa *chama cha *wananchi (*cuf) *profesa *ibrahim *lipumba, amewahimiza wafuasi wa chama hicho kujitokeza kwa wingi kwenye maandamano yaliyopangwa kufanyika nchi nzima *jumamosi ijayo na kwamba wawe imara kukabiliana na polisi pindi <watakapoingilia> maandamano hayo.

8.   Conclusion

After a fairly long period of research and testing, computational lexicography has reached a stage where computers and corpora can be put into effective use. For many years, computers have been used for producing word lists with frequencies from a corpus, as well as for retrieving concordances of word use. This article has shown that the use of regular expressions can significantly increase the precision and recall of search. However, the inclusion of the full linguistic analysis in dictionary work brings the work to a level where precision and recall meet high standards. SALAMA, the working environment developed for Swahili, facilitates the testing of various phases in dictionary compilation based on extensive use of the computer. This article demonstrates that computer-based lexicography does not only greatly benefit from the described approach; it is in fact a necessity in working with highly inflectional left-branching languages.

The system brings the automation of dictionary compilation to the point where the benefits of further automation become questionable. It accurately describes what can safely be described, and leaves ambiguous cases for human checking. Its great advantages are morphological accuracy and coverage, great speed, and ease of use.

The system can be developed still further, especially in the area of semantic disambiguation, so that correct senses of words in each context can also automatically be defined. Research is currently concentrating on the problems in this area.

Endnotes

1.   There were also more realistic opinions that reflected the contemporary state-of-the-art in this field (Calzolari 1989; Wegera and Berg 1989).

2.   By linguistic insight we here mean a kind of simulation of linguistic regularities, which a computer system utilizes and translates as 'linguistic rules'.

3.   There has been discussion on the need of sufficient and systematic grammatical information in dictionaries (Salerno 1999). The approach discussed in this article effectively facilitates the inclusion of this feature.

4.   The need of semantic information in dictionaries has increasingly been emphasized, whether

in terms of frame semantics (Fontenelle 2000, 2000a) or in terms of some other semantic theory. Statistical methods have also been used for identifying such word clusters that seem to occur together. On the basis of such clusters it is possible to carry out cluster analysis (Watters 2002).

5.   In SALAMA, the Swahili Language Manager, etymological information on words of non-Bantu origin has been included by means of specific tags (Hurskainen 1999).

6.   SALAMA is based on two-level morphology, and it is implemented by using finate state automata (Koskenniemi 1983; Hurskainen 1992, 1999). The disambiguation is based on the Constraint Grammar formalism (Karlsson 1995; Tapanainen 1996; Hurskainen 1996).

7.   In fact, according to a survey, the choice of headwords was considered the most difficult among the 13 tasks asked from the team working on the third edition of the *Longman Dictionary of Contemporary English* (Kilgarriff 1998).

8.   Based on SALAMA, the Swahili Language Manager, Sewangi (2000) has developed a system that retrieves term candidates from domain-specific text. This method facilitates the extensive use of domain-specific texts, such as educational books, handbooks, and other written materials of the domain, for compiling domain-specific dictionaries.

9.   These two dictionaries are *Kamusi ya Kiswahili Sanifu* (1981) and *Kamusi ya Kiswahili–Kiingereza* (2001), both produced by the Institute for Kiswahili Research, University of Dar es Salaam.

10.   The strings we wanted to find are shown with √.

11.   Alternative strings are separated with a vertical bar and all alternatives are enclosed in parentheses. The question mark (?) stands for optionality, and the plus sign (+) means that the preceding unit may occur one or more times. The set a–z within square brackets means any character. The backslash (\) in the end of the line signifies that for the computer the same line continues.

12.   The accuracy of the semantic glosses depends on how they were acquired in the analysis system. The most obvious way not requiring too much manual work is to use an electronic version of a good normal dictionary and include relevant parts of its entries in the dictionary of the analysis system. This was done in SALAMA, and the glosses produced are largely the same as those in the original dictionary, for good and bad. We should not, however, be content with these glosses, because they are just approximations of the various meanings of the lexemes and they should be checked and amended on the basis of the information available in the corpus. In addition to helping in the selection of headwords, the corpus is useful in identifying various meanings of the lexemes.

References

Calzolari, N. 1989. Computer-Aided Lexicography: Dictionaries and Word Data Bases. Steger, H. and H.E. Wiegand (Eds.). *Handbuch zur Sprach- und Kommunikationswissenschaft, Band 4*: 510-519. Berlin: Walter de Gruyter.

Fontenelle, T. 2000. A Bilingual Lexical Database for Frame Semantics. *International Journal of Lexicography* 13(4): 232-248.

Fontenelle, T. 2000a. Introduction: Dictionaries, Thesauri and Lexical-semantic Relations. *International Journal of Lexicography* 13(4): 229-231.

Hurskainen, A. 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, A. 1994. Kamusi ya Kiswahili Sanifu in Test: A Computer System for Analyzing Dictionaries and for Retrieving Lexical Data. *Afrikanistische Arbeitspapiere* 37 (Swahili Forum I): 169-179.

Hurskainen, A. 1996. Disambiguation of Morphological Analysis in Bantu Languages. *COLING-96, Proceedings of the 16th International Conference on Computational Linguistics, Copenhagen, August 5–9, 1996*: 568-573. Copenhagen: Center for Language Technology.

Hurskainen, A. 1999. SALAMA: Swahili Language Manager. *Nordic Journal of African Studies* 8(2): 139-157.

Hurskainen, A. 2002. Tathmini ya Kamusi Tano za Kiswahili. *Nordic Journal of African Studies* 11(2): 283-301.

Hurskainen, A. and R. Halme. 2001. Mapping between Disjoining and Conjoining Writing Systems in Bantu Languages: Implementation on Kwanyama. *Nordic Journal of African Studies* 10(3): 399-414.

Jones, R.L. and S.P. Sondrup. 1989. Computer-Aided Lexicography: Indexes and Concordances. Steger, H. and H.E. Wiegand (Eds.). *Handbuch zur Sprach- und Kommunikationswissenschaft, Band 4*: 490-509. Berlin: Walter de Gruyter.

Karlsson, F. 1995. Designing a Parser for Unrestricted Text. Karlsson, F. et al. (Eds.). *Constraint Grammar: A Language-Independent System for Parsing Unrestricted Text*: 1-40. Berlin: Mouton de Gruyter.

Kilgarriff, A. 1997. Putting Frequencies in the Dictionary. *International Journal of Lexicography* 10(2): 135-155.

Kilgarriff, A. 1998. The Hard Parts of Lexicography. *International Journal of Lexicography* 11(1): 51-54.

Koskenniemi, K. 1983. *Two-level Morphology: A General Computational Model for Word-form Recognition and Production*. Publications No. 11. Helsinki: Department of General Linguistics, University of Helsinki.

Panyr, J. and H.H. Zimmermann. 1989. Information Retrieval: Überblick über active Systeme und Entwicklungstendenzen. Steger, H. and H.E. Wiegand (Eds.). *Handbuch zur Sprach- und Kommunikationswissenschaft, Band 4*: 696-708. Berlin: Walter de Gruyter.

Salerno, L. 1999. Grammatical Information in the Bilingual Dictionary: A Study of Five Italian–French Dictionaries. *International Journal of Lexicography* 12(3): 209-222.

Sewangi, 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.

Tapanainen, P. 1996. *The Constraint Grammar Parser CG-2*. Publications No. 27. Helsinki: Department of General Linguistics, University of Helsinki.

Tapanainen, P. and T. Järvinen. 1998. Dependency Concordances. *International Journal of Lexicography* 11(3): 187-203.

Teubert, W. 2001. Corpus Linguistics and Lexicography. *International Journal of Corpus Linguistics* 6 (Special Issue): 125-153.

Watters, P.A. 2002. Discriminating English Word Senses Using Cluster Analysis. *Journal of Quantitative Linguistics* 9(1): 77-86.

Wegera, K.-P. and E. Berg. 1989. Computergestützte Grammatikographie: Eine Fallstudie. Steger,

H. and H.E. Wiegand (Eds.). *Handbuch zur Sprach- und Kommunikationswissenschaft, Band 4*: 519-527. Berlin: Walter de Gruyter.

Dictionaries

Abdulla, A., R. Halme, L. Harjula and M. Pesari-Pajunen (Eds.). 2002. *Swahili–Suomi–Swahili-sanakirja*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Kamusi ya Kiswahili–Kiingereza **(Swahili–English Dictionary).** 2001. Dar es Salaam: Taasisi ya Uchunguzi wa Kiswahili.

Kamusi ya Kiswahili Sanifu. 1981. Dar es Salaam: Oxford University Press.

The Reproduction of Cultural Aspects in Dictionaries in French and the Gabonese Languages[*]

P.A. Mavoungou *(moudika2@yahoo.fr)*, Thierry Afane Otsaga
(thiafane@yahoo.fr) and Guy-Roger Mihindou
(guyr_mihindou@hotmail.com), *Department of Afrikaans and Dutch,*
University of Stellenbosch, Stellenbosch, Republic of South Africa

Abstract: The reproduction of culture in dictionaries constitutes one of the fundamental problems of lexicographers today. What is the nature of cultural data in dictionaries? To what extent should cultural aspects be transferred from one language to another? How should this transfer take place? This article attempts to discuss the relevance of the reproduction of Gabonese cultural aspects in dictionaries in French and the Gabonese languages. One of the main problems the compilers of these dictionaries encountered was the transfer and the translation of some cultural aspects.

In order to discuss the nature and extent of cultural information in Gabonese dictionaries, this article will restrict itself to the following points:

— The translation of different realities
— The role of culture in the change of meaning
— Dictionaries and cultural activities
— Dictionaries and cultural ethics
— Dictionaries and language registers
— Dictionaries and culture maintenance
— Dictionaries and the standardization of culture

Prior to a discussion of the above-mentioned points, a brief description of the dictionaries investigated will be given. After a description of the influence of cultural contexts on lexicographers in the choice of macrostructural elements and their treatment, cultural differences between source and target languages in the dictionaries of Gabonese languages are discussed. The article concludes with the observation that the majority of existing lexicographic works tend to survey the full vocabulary of the language. The lexicographic treatment of some words is satisfactory because an account of the underlying worldview of the people is given. For example, in this article, the following themes are considered: dietary practices, sexuality, mythology, traditional pharmacopoeia, the kinship system, hospitality, and respect for traditional authority and elders. However, to be used in the most efficient way, these lexicographic publications need to be revised.

Keywords: CULTURE, DICTIONARIES, TARGET USER, TRANSLATION, CULTURAL

ACTIVITIES, CULTURAL ETHICS, LANGUAGE LEVELS, CULTURE MAINTENANCE, STAN-
DARDIZATION OF CULTURE, GABONESE LANGUAGES

**Résumé: La reproduction des aspects culturels dans les dictionnaires en
français et en langues gabonaises.** La reproduction de la culture dans les dictionnaires
constitue l'un des problèmes fondamentaux des lexicographes aujourd'hui. Quelle est la nature des
données culturelles dans les dictionnaires? Jusqu'à quel point les données culturelles doivent-elles
être transférées d'une langue à une autre? Comment ce transfert doit-il se faire? Cet article tente
d'examiner la pertinence de la reproduction des aspects des cultures gabonaises dans les diction-
naires en français et en langues gabonaises. L'un des principaux problèmes rencontrés par les
rédacteurs de ses dictionnaires était le transfert et la traduction de certains aspects culturels.

Pour discuter de la nature et du degré des informations culturelles dans les dictionnaires en
langues gabonaises, cet article va se limiter aux points suivants:

—   La traduction de différentes réalités
—   Le rôle de la culture dans le changement du sens
—   Les dictionnaires et les activités culturelles
—   Les dictionnaires et la morale culturelle
—   Les dictionnaires et les registres de langue
—   Les dictionnaires et la maintenance de la culture
—   Les dictionnaires et la standardisation de la culture

Avant une discussion des points sus-mentionnés, une brève description des dictionnaires examinés
sera donnée. Après une description de l'influence des contextes culturels dans le choix des élé-
ments macrostructuraux ainsi que de leur traitement, les divergences culturelles entre langues-
sources et langues-cibles dans les dictionnaires en langues gabonaises seront examinés. L'article
conclut par l'observation que la majorité des travaux lexicographiques existants tente de donner un
aperçu du lexique total de la langue. Le traitement lexicographique de certains mots est satisfaisant
parce qu'il prend en compte la vision du monde sous-jacente des populations. Par exemple, dans
cet article, les thèmes suivants sont pris en compte: habitudes alimentaires, sexualité, mythologie,
pharmacopée traditionnelle, système de parenté, hospitalité, et le respect de l'autorité traditionnelle
et des aînés. Cependant, pour être utilisées de la manière la plus efficace, ses publications lexico-
graphiques méritent d'être révisées.

Mots-clés: CULTURE, DICTIONNAIRES, PUBLIC CIBLE, TRADUCTION, ACTIVITÉS CUL-
TURELLES, CULTURE MORALE, NIVEAUX DE LANGUE, MAINTENANCE CULTURELLE,
STANDARDISATION DE LA CULTURE, LANGUES GABONAISES

1.   Introduction

In present-day lexicography, there is a tendency to give an account of the
underlying culture and civilization of the languages being described. However,
reservations can be made with regard to an encyclopaedic bias. Too often lexi-
cographers compiling dictionaries in the languages in question have been
focusing on encyclopaedic or extra-linguistic aspects of the lemma instead of
on linguistic data.

The majority of dictionaries available in Gabonese languages have a common denominator: they are translation dictionaries and were compiled by Catholic and Protestant missionaries or colonial administrators (cf. Mavoungou 2001a: 160). One of the main problems the compilers of these dictionaries encountered, was the transfer and the translation of certain cultural aspects from one language to another. In fact, the lexicographer does not only propose the description of the lexicon, but should also take the context of use into account. The dictionary is a mirror in which the user must recognize himself not only as speaker but also as a member of a cultural community (Dubois and Dubois 1971).

Gabonese dictionaries are investigated here with a view to comparing how the cultural transfer is taking place as well as its limitations. An attempt will be made to answer the following questions: What is the nature and extent of cultural information in existing Gabonese dictionaries? How can the cultural transfer of information in these dictionaries be improved? Moreover, various cultural gaps between source and target languages in the existing dictionaries will be shown as well as the ways cultural and environmental contexts have influenced lexicographers in the choice of macrostructural elements and their treatment.

Before discussing these, a brief description of the dictionaries investigated will be given. The dictionaries will be restricted to the following:

— *Encyclopédie pahouine*, Largeau (1901),

— *Lexique fãn–français*, Martrou (1924),

— *Les plantes utiles du Gabon*, Raponda-Walker and Sillans (1961),

— *Dictionnaire fang–français et français–fang, suivi d'une grammaire fang*, Galley (1964),

— *Dictionnaire français–yipounou, yipounou–français*, Église Évangélique du Sud-Gabon (1966),

— *Dictionnaire Français–Mpongwé*, Raponda-Walker (1995[2]), and

— "Lexique" in Nsuka-Nkutsi (Ed.), Rittaud-Hutinet (1980).

2. A brief description of the investigated dictionaries

2.1 The *Encyclopédie pahouine* of Largeau (1901)

The colonial administrator Largeau published the *Encyclopédie pahouine* in 1901. It is an alphabetically arranged dictionary of considerable size, consisting of ± 4 996 articles and covering 699 pages. The value of the work lies in the useful information regarding the history of the Fang people, their value system as well as other anthropological issues such as rituals and mythology which the

user can find in its introductory section. It also contains special articles focusing on cultural data introduced by the structural marker *Encyclopédie*, as well as the systematic use of literal translations aimed at the translation equivalents and competence examples (cf. Nyangone Assam and Mavoungou 2000: 257 and Mavoungou 2001: 127 respectively).

2.2   The *Lexique fãn–français* of Martrou (1924)

The *Lexique fãn–français* published in 1924 by Archbishop Martrou, is an alphabetically arranged lexicon consisting of 3 431 articles covering 137 pages. Although the title of this publication modestly describes it as a "lexicon", it is in fact a comprehensive and thorough dictionary.

The user is provided with various data categories: paraphrases of meaning, translation equivalents, competence examples, etc. Many typographical structural markers such as roman, italic, bold and small characters are used in the work to fulfil a metacommunicative function. However, the lexicon does not clearly identify its target users or their reference skills. Martrou's lexicon also fails to give an account of its primary and secondary sources (Mavoungou 2001: 127-130).

2.3   *Les plantes utiles du Gabon* of Raponda-Walker and Sillans (1961)

Although entitled *Les plantes utiles du Gabon*, this publication is in fact an encyclopedic dictionary of the names and uses of Gabonese plants, published in 1961 by Raponda-Walker in collaboration with Sillans. It is arranged alphabetically, and consists of 1 497 articles covering 582 pages. It contains some 8 000 indigenous plant names (as well as their scientific names and usages), collected over a period of thirty years by Raponda-Walker. The work gives an account of its primary, secondary and tertiary sources and the front and back matter texts are also sources of valuable information for the users concerning the scientific description (height and diameter, arrangement and texture of the foliage and leaves, etc.), and the distribution and the cultivation requirements of the tree dealt with. As visual aids for users, the book is illustrated with 53 pictures by Sillans (cf. Nyangone Assam and Mavoungou 2000: 271 and Mavoungou 2001a: 169-174).

2.4   The *Dictionnaire fang–français et français–fang, suivi d'une grammaire fang* of Galley (1964)

The *Dictionnaire fang–français et français–fang* by the Genevan pastor Samuel Galley was published in 1964. It is an alphabetically arranged bilingual dictionary consisting of ± 13 925 articles and covering 588 pages. It is based on the translation of the Bible in Fang, another gigantic task Galley undertook for the Fang community of Talagouga.

The value of the work lies in its biscopal character. The microstructure includes a considerable amount of data, e.g. parts of speech, class numbers, translation equivalents, meaning explanations as well as cross-references (cf. Nyangone Assam and Mavoungou 2000: 258).

2.5 The *Dictionnaire français–yipounou, yipounou–français* of the Église Evangélique du Sud-Gabon (1966)

The *Dictionnaire français–yipounou, yipounou–français* of the Christian and Missionary Alliance (CMA) was published in 1966. It is an alphabetically arranged bilingual publication consisting of 8 934 articles and covering 145 pages. In the front matter, the compilers, American missionaries, make it clear that the work, published for a target group of Yipunu mother-tongue speakers, is meant to help them in studying the Bible (cf. Mihindou 2001: 26). Although this publication is described as a "dictionary", it is in fact only a glossary. In the first section, macro- and microstructural data include lemmata in French followed by their Yipunu translation equivalents presented in two columns, whereas in the second section this information appears in the reversed direction, Yipunu followed by French.

2.6 The *Dictionnaire Français–Mpongwé* of Raponda-Walker (1995[2])

The *Dictionnaire Français–Mpongwé* of Raponda-Walker was published for the first time by the Imprimerie de la Libre Lorraine in Metz (1930–1934). It is monoscopal, with Mpongwé as source and French as target language. In 1961, 27 years later, the dictionary was published in the reverse direction, French–Mpongwé, by the Imprimerie Saint Paul in Brazzaville. This version of the dictionary was reprinted in 1995 under the auspices of the Raponda-Walker Foundation (cf. Nyangone Assam and Mavoungou 2000: 260).

This last version of the dictionary (that will be investigated in this study) is an alphabetically arranged bilingual publication consisting of some 8 000 articles and covering 723 pages. The dictionary articles include much data, such as lemmata in French, translation equivalents in Mpongwé, part of speech indicators, labels, illustrative examples, etc.

2.7 "Lexique" in Nsuka-Nkutsi (Ed.) of Rittaud-Hutinet (1980)

The work of Rittaud-Hutinet entitled "Lexique" was published in 1980. It is an alphabetically arranged bilingual Yipunu–French lexicon consisting of 945 articles and covering 52 pages. The publication is based on several sources, its primary source being a questionnaire from Joseph H. Greenberg. Moreover, the lexicographer gives an account of the criteria for the incorporation of all the lexical items (nouns, verbs, adjectives, pronouns, etc.) entered in the lexicon according to the stem tradition.

3.   The nature and extent of cultural information in the existing Gabonese dictionaries

The discussion of the nature and extent of cultural information in existing Gabonese dictionaries will be restricted to the following focus areas:

—   The translation of different realities

—   The role of culture in the change of meaning

—   Dictionaries and cultural activities

—   Dictionaries and cultural ethics

—   Dictionaries and language registers

—   Dictionaries and culture maintenance

—   Dictionaries and the standardization of culture

3.1   The translation of different realities

No matter what its typology is, a dictionary is a description of culture (Dubois and Dubois 1971). The translation of a culture in a language to which it is uncommon or unknown, is never easy. In such cases, the lexicographer has to be careful regarding the context and the environment of the dictionary users.

In some dictionaries of Gabonese languages, the French National Day **14 juillet**, for instance, has been translated as 'emu awom be<u>n</u>in' (14th July) in Fang. This translation is meaningless to a Fang speaker, who will not recognise the relation between July 14 and the French National Day. The best way to translate this concept is to use the meaning paraphrase 'emu France a<u>ng</u>a nyo<u>ng</u> fili' ("the day France got freedom"). In the same way, August 17 is defined by the meaning paraphrase 'emu Gabon a<u>ng</u>a nyo<u>ng</u> fili' ("the day Gabon got freedom"), used by Fang speakers to designate the Gabonese Independence Day.

Numerous examples of the same kind can be found in existing dictionaries of Gabonese languages. The Gabonese, for instance, will not understand the term "crucify", because this practice (the crucifixion) does not exist in Gabonese cultures. In Western cultures, this term primarily designates the death of Jesus Christ. This point is illustrated by the treatment of the article of the lemma **crucifier**.

(1)   Dictionary excerpt from Galley 1964: 447.

>   **CRUCIFIER** quelqu'un, *domele môr éli* (hbh). Crucifié, *ñkelba éli* (bh)

In this case, the concept crucifixion should be substituted by a term designating the same reality in Gabonese languages (e.g. hanging).

3.2   The role of culture in the change of meaning

Cultural gaps between Gabonese and European languages (French in particular) play an important role in the change of meaning of numerous current words. As far as French is concerned, many words do have another meaning in the Gabonese environment as compared to the meaning they have in French society. The term **cadeau** for example, firstly means "present" or "gift" in French. In the context of Gabonese languages, this term also means "free" or "gratis". In existing Gabonese dictionaries, these cultural specificities have not been taken into account as is clearly shown in (2) and (3).

(2)   Dictionary excerpt from Galley 1964: 431.

> **CADEAU** don, *évévé* (h) (exactement *eveévé*). Cadeau empaqueté et envoyé à un ami pour une occasion, *atôm* (h). Cadeau quelconque, n'importe quoi, *ékeélé* (bm). Cadeau à celui qui part, promesse, *meyagha* (h), *magha* (h). Cadeau qu'on apporte à un ami quand on va en visite chez lui, *meyeñ* (h). Cadeau, souvenir d'un ami à son ami, objet qui lui a appartenu, *mvôm* (bb). Cadeau offert par crainte et pour calmer quelqu'un, *ndun* (h), *ntun* (h), *ntãn abi* (hh). Cadeaux de la belle-famille (*abè*) au gendre, *ñgañ* (m), *mevale* (h).

(3)   Dictionary excerpt from Raponda-Walker 1995: 69.

> **cadeau** N. kado, dasi (néol.). *Cadeau donné*, eza zi mpagaga; cadeau reçu, eza zi mpagago.
> *Petit cadeau*, izege. *Cadeau de bienvenue*, asango.

In Gabonese French, as mentioned above, the word "cadeau" is not only used in the sense of "present" or "gift" but also in the sense of "free" or "gratis". Articles (2) and (3) do not account for the sense "free" or "gratis". This is detrimental to users because it is well attested that in any translation dictionary with English as a treated language, the focus should be on one variety of English, e.g. British or American English. However, when including lexical items restricted to a variety exclusively used by the target users of the dictionary, the lexicographer should indicate this restricted usage by employing a system of geographical labels. The same line of argumentation holds true for French. French is nowadays geographically dispersed across a significant number of continents and countries. The geographical dispersal of French has brought the language into daily contact with new usages in French-speaking countries and territories in Europe, North America, Africa and Asia. This has strengthened and enriched French as it is spoken in France. Thus it is important that dictionaries should take account of the usage of French spoken outside France, especially as it influences the form used in France (cf. Mavoungou 2002). As far as the treatment of the article of the lemma **cadeau** is concerned, e.g. in Galley (1964) and Raponda-Walker (1995), both compilers have failed to take cognisance of the differences between Gabonese French and French as it is spoken in France, and

to supply the target users with the necessary data.

Despite the shortcoming of the article on **cadeau** in the *Dictionnaire Français–Mpongwé*, this dictionary with its 8 000 articles contains quite a representative lexis of national and local forms of French. Since these items relate to Gabonese cultures, they deserve special mention. The first illustration in this regard comes from the article of the lemma **levée**:

(4)   Dictionary excerpt from Raponda-Walker 1995: 382.

> **levée** N. (*de terre*) oralyo; (*de deuil*) intsugu. *Danser à cette occasion*, boliz'intsugu.

Moreover, there are, among others, seventeen articles that can be labelled as Gabonese French, viz. **caïman**, **concombre**, **fromager**, **hangar**, **homonyme**, **kapok**, **kapokier**, **kinkéliba**, **levée de terre**, **règle**, **taloche**, **sanglier**, **talocher**, **cassada**, **blinder**, **chat huant**, and **gris-gris**. Of these seventeen, four lexical items, namely *kapok*, *kapokier*, *kinkéliba* (or *quinquéliba*) and *gris-gris*, are widely used in the French-speaking world.

A further six, viz. *caïman*, *levée de terre*, *concombre*, *hangar*, *homonyme*, and *sanglier* contain a great deal of information about fauna and flora and have a special relevance for Gabonese cultures. A brief examination of these reveals that *caïman* "a crocodile found in America" and *sanglier* "a bush-pig mainly found in Europe" are the local forms of French used to refer to *Osteolaemus tetrapis* (or *Crocodylus niloticus*) and *Potamochoerus porcus* respectively. Similarly, the lexical item *concombre* "cucumber" does not refer to the same extra-linguistic reality in France and in Gabon. In the former, people easily assimilate *concombre* to the French *courge* because the two belong to the same family *Cucurbitaceae*. In the Gabonese context, the lexical item *hangar* encompasses a slightly different meaning compared to that in France. In Gabon, it does not only refer to a warehouse but it is primarily a building where the local population goes to worship ancestral spirits or tutelary spirits (cf. Mavoungou 2002). In Africa like elsewhere in the world, the naming of a child is a matter of great importance. In the Gabonese context, *homonyme* "homonym" is a member of the community whose name has been chosen to be given to a new-born baby. Kwenzi-Mikala (1990: 114) comments as follows on this lexeme:

> Un "homonyme" est une personne de la communauté dont on a choisi le nom pour l'attribuer au nouveau-né. Une fois le choix fait, il s'établit des liens spéciaux entre l'enfant et la personne qui a accepté de donner le nom. Cette espèce de parrainnage entraîne l'utilisation réciproque comme terme d'adresse du mot **dînə** qui signifie par ailleurs "nom". Le véritable but recherché à travers cette institution semble être de trouver un remplaçant à la personne sollicitée en assurant la perpétuation de son nom. Le choix peut porter aussi sur le nom d'un défunt, surtout si celui-ci n'a laissé aucun descendant. Il est encore plus clair, dans ce cas, qu'il s'agit d'honorer et d'immortaliser la mémoire du défunt.

Lexical items such as *levée de terre*, *port de deuil*, and *levée* or *retrait de deuil* are

important funeral rites. After undergoing purificatory rites, members of the family usually wear mourning for the deceased (referred to as "port de deuil" in Gabonese French). The rising from the earth (referred to as "levée de terre" in Gabonese French) generally follows this ceremony. Finally, after a year, a celebration is officially held to end the mourning period ("retrait" or "levée de deuil" in Gabonese French) (cf. Mavoungou 2002). From these local forms of French, some seem to have been omitted. Readers will therefore miss items such as *safou* (the edible fruit produced by the *safoutier*), *atanga* (the edible fruit produced by the *atangatier*) as well as more recent lexical items such as *fétichisme* (fetishism), *tresseuse* (a female who plaits), *marabout* (a sangoma), *maraboutisme* (the craft of the sangoma), *marabouter* (to act as a sangoma), *se saper* (to dress elegantly), *sapeur* (an elegant person), *absenter* (to miss someone), *grever* (to be on strike), *siester* (to take a nap), *régler* (to have menstrual periods), *tôler* (to cover a house with corrugated iron), *têter* (to play a header with a soccer ball), *tchatcher* (to chat, to converse), *tchatcheur* (a good speaker), etc.

To summarise: Raponda-Walker's dictionary substantially adds to our knowledge of Gabon and the forms of French used in this country.

3.3   Dictionaries and cultural activities

As will be seen later in this article, any dictionary should reflect the lexicon of the language being treated. Because of traditional tasks or activities (fishing, hunting, agriculture, etc.), the vocabulary most commonly used in Gabonese speech communities encompasses names of plants, insects and animals as well as names of other natural phenomena (birth, death, traditional pharmacopoeia, traditional beliefs, etc.). In the majority of the dictionaries under discussion, an attempt has been made to include a reasonable number of words meeting these needs. In addition, where available, the scientific names of plants, insects and animals are supplied.

Gabon covers 267 667 square kilometers, most of which is dense tropical forest, interspersed with savannah, and fed by a network of rivers of which the Ogooué is the most important. All seven lexicographic publications studied deal, in greater or lesser detail, with the names of the following fully protected animals of Gabon given in Table 1:

Animal names	Scientific names
Aigle couronné	Stephanoaetus coronatus
Aigle pêcheur	Haliaeetus vocifer
Céphalophe à pattes blanches	Cephalophus ogilbyi crusalbaum
Cercopithèque à queue de Soleil	Cercopithecus solatus
Chimpanzé	Pan troglodytes
Cobe des roseaux	Redunca arundinum
Cobe onctueux	Kobus defassa
Daman des arbres	Dendrohyrax arboreus
Galago d'Allen	Galago alleni

Galago de Demidoff	Galago demidovi
Galago elegant	Galago elegantus
Gorille[1]	Gorilla gorilla
Hippopotame	Hippopotamus amphibious
Lamantin	Trichechus senegalensis
Lion	Panthera leo
Oryctérope	Orycteropus afer
Pangolin géant	Manis gigantean
Panthère	Panthera pardus
Pélican gris	Pelecanus rufescens
Potto de Bosman	Perodicticus potto

Table 1: List of the fully protected animals of Gabon

Of all the dictionaries investigated, it is perhaps the *Dictionnaire Français–Mpongwé* (1995) that best illustrates Gabon's wildlife as it exists today. If one looks at the macrostructural elements of this dictionary, it encompasses to a greater or lesser degree the names of the following partially protected animals of Gabon given in Table 2:

Animal names	Scientific names
Bongo	Boocerus euryceros
Buffle	Syncerus caffer nanus
Céphalophe à dos jaune	Cephalophus silvicultor
Chat doré	Felis aurata
Crocodile du Nil	Crocodylus niloticus
Crocodile nain	Osteolaemus tetraspis tetraspis
Drill	Mandrillus (Papio) leucophaeus
Éléphant	Loxodonta Africana
Faux gavial	Crocodylus cataphractus
Guib harnaché	Tragelaphus scriptus
Hylochère	Hylochoerus meinertzhageni
Ibis sacré	Threskiomis aethiopica
Jaribu du Sénégal	Ephippiorhynchus senegalensis
Mandrill	Mandrillus (Papio) sphinx
Perroquet gris à queue rouge	Psittacus erithacus
Potamochère	Potamochoerus porcus
Python de Seba	Python sebae
Serval	Felis serval
Situtunga	Tragelaphus spekei gratus
Spatule d'Afrique	Platalea alba
Tantale ibis	Ibis ibis
Tortue luth	Dermochelys coriacea
Varan	Varanus niloticus
Vautour palmiste	Cypohierax angolensis

Table 2: List of the partially protected animals of Gabon

The dictionary also contains a fairly large percentage of the names of fish found in rivers and streams as well as the names of sea fish. Compare Table 3 and Table 4 in this regard.

Fish names	Scientific names
Brochet	Hepsetus odoe
Carpe de l'Ogooué	Oreochromis schwebischi
mâchoiron	Chrysichthys nigrodigitatus
Mormyre	Brienomyrus longicandatus
ngotu	Heterobranchus longifilis
oyara	Schilbe grenfelli
Petite anguille épineuse	Aethiomastacembelus marchei
Poisson à tête de serpent	Parachanna obscura
Protoptère	Protopterus dolloi
Sardines des lacs	Pellonula vorax
Silure	Clarias gariepinus
Synodontis	Synodontis batesii
Tetra africain	Brycinus macrolepidotus
Tilapia	Tilapia cabrae

Table 3: List of the names of fish found in rivers and streams of Gabon

Fish names	Scientific names
Bar	Pseudotolithus typus
Bécune	Sphyraena afra
Bossu	Pseudotolithus elongatus
Carangue	Caranx hippos
Dorade grise	Pomodasys peroteti
Engil	Lutjanus dentatus
Espadon voilier	Istiophorus albicans
Grand capitaine	Polydactylus quadrifilis
Mâchoiron de mer	Arius latiscutatus
Mérou noir	Epinepphelus marginatus
Mullet	Mugil cephalus
Plexiglass	Galeoides decadactylus
Raie Pastenague	Dasyatis ukpam
Requin bouledogue	Carcharhinus leucas
Sardine des estuaires	Ethmalosa fimbriata
Tarpon	Tarpon atlanticus

Table 4: List of the names of sea fish of Gabon.

In spite of this, there are some shortcomings. Although dialectal differences have been clearly established, many dictionaries of Gabonese languages are biased toward one dialect. This is detrimental to users of the speech community. When macrostructural elements in a particular dictionary are from one

dialect, users of the other dialects too often do not recognize themselves in the dictionary. In existing dictionaries compiled with Fang-Atsi as a treated language, the lexical stock mainly comprises fishing terms (e.g. several kinds of fish), because Fang-Atsi people are predominantly fishermen. As a result, these dictionaries do not have many terms relative to hunting, harvesting and agricultural activities common to other Fang communities. The point in question is illustrated in the treatment of the following lemmata: **aba**, **abò** and **poisson**.

(5)   Dictionary excerpt from Martrou 1924: 21 and 23.

> **Aba** Petit poisson de ruisseau.
>
> **Abò** Variété de grenouille brune et tâchetée.
> — Espèce de crabe. *Abo-kara*.
> — *evus*, araignée crabe dont la piqûre est venimeuse.

(6)   Dictionary excerpt from Galley 1964: 520-521.

> **Poisson**   *ko* (m), *kos* (m). Noms de poissons (environ septante): *abem* (b) petit poisson; *adua* (bm), famille des *ntetom* (b), *akom* (h), famille des *ñgol* (b), syn.: *mvè* (h); *akweghe* (h), petit poisson: *alap ékôni* (bh); grosse carpe, syn.: *ñkãn ékôni* (hh); *amokh* (h), petit poisson noir à écaille, syn. *ñgoméésa* (mb); *andôndôe* (m), petit poisson; *anduma* (h), grand poisson; *anyia* (b), poisson (b) poisson torpille ou silure électrique, voir *ala anyia (mb), añgeñgeme* (h), petit poisson; *añgwañ* (b), petit poisson; *añgwekh ékôni* (hh), carpe encore petite; *aveña* (bm), hénioque, poisson plat en forme de cercle, surtout dans l'eau de mer, syn.: *évulababa* (bm); *ayaol* (h), poisson sole qui ressemble à une limace (*ayol*); *bè* (h), petit poisson de la famille *mvagha* (b);
> … *ñgoña* (h) ou *mvôkh* (h), petit poisson, variété de *keme* (h), *ñken* (h), petit poisson qui se cache dans le sable sous l'eau, la tête seule se voit; *ñkôghli* (b), poisson raie, voir ce mot; *ôbar* (b), poisson qui nage sur le dos, famille *ntetom* (b).

The same situation is true of the vocabulary denoting spiritual activities (traditional initiation rites). Dictionaries of Fang-Atsi or Fang-Mekè, for instance, give more data regarding the *Bieri* (worshipping of a person's ancestral skull) which forms part of their culture, than about the *Melane* (a masculine initiation and secret society) that speakers of these two dialects do not practice.

3.4   Dictionaries and cultural ethics

Dubois and Dubois (1971) hold that terms do not only refer to the words of a language, they are not only objects of the metalanguage of linguistics, but they also refer to cultural statements, a worldview. In an article dealing with cultural practices of English-speaking North Americans, French Canadians and the French as they are conveyed through advertising, Quillard (1999: 46) points out: "La nudité ou semi nudité ne choque pas les Français. Considérée comme

naturelle, elle n'a pas lieu de d'être cachée ou censurée" (Nudity or semi-nudity do not shock the French. Regarded as natural, it does not have to be hidden or censured.) This is not the case in Gabon. Under normal circumstances, Gabonese are extremely decorous. The private parts of the body are taboo and are spoken of only in metaphors, euphemisms and other rhetorical expressions.

It is a well-attested fact that any dictionary should reflect the lexicon of the language being treated. The lexicon in question must encompass all its aspects, including taboos. In the existing dictionaries of the Gabonese languages, various terms referring to some cultural taboos (particularly about sex and some parts of the body) are found. Compare the following articles as examples:

(7) Dictionary excerpt from Largeau 1901: 388.

> INCESTE. *N'sem nden* (grand péché). Celui qui le commet. *Môr a bo' nsem ndem*. — *Encyclopédie*: L'inceste, chez les pahouins, est limité à ces définitions: 1 Rapports sexuels entre un homme, sa mère, sa sœur ou sa fille; 2 Mêmes rapports entre une femme mariée et les fils des autres femmes de son mari …

(8) Dictionary excerpt from Martrou 1924: 27.

> **Akan** Aine, hanche, cuisse.

(9) Dictionary excerpt from Galley 1964: 423.

> AVORTER (femme), *kwé abmum* (hb). Celui qui fait avorter une femme, *ñkwé abmum* (hb).

(10) Dictionary excerpt from Raponda-Walker 1995: 17.

> **anus** N. esulu, orolo.

(11) Dictionary excerpt from CMA 1966: 44.

> fesse — dirangi

In (10) and (11), it is rather inconceivable to supply a direct translation equivalent for the lemmata **anus** and **fesse**, since in Gabonese cultures, reference to these parts of the body require euphemisms.

It is part of the responsibility of the lexicographer to identify taboo terms and to warn the user against their indecorous nature. In this regard, J. Dubois (1970: 43) mentions that the lexicographic speech is not only a pedagogical statement about the language; it is also a cultural training whose propositions must be culturally accepted. As pedagogical aids dictionaries have to make users culturally aware of the underlying worldview of the language or languages being described. That is why dictionaries also aim to constitute themselves as an explicit cultural norm for the linguistic community. The compilation of dictionaries is useless if the target group they are compiled for, rejects them.

In (8), (10) and (11), objections should not be made with regard to the inclusion of **akan**, **anus** and **fesse** as macrostructural elements but rather with regard to the absence of contextual data addressed to the lemmas, given their taboo nature. Compared to the above-mentioned articles, the treatment of the article of the lemma **inceste** is quite satisfactory. In fact, as long as the explanation of meaning (*n'sem nden* = grand péché) continues, the user is warned against the unmentionable nature of the lemma.

The aim of dictionaries to constitute themselves as an explicit cultural norm for the speech community goes a bit further in the article on the lemma **conseil** which implicitly contains the following theme: respect for traditional authority and elders. The encyclopedic section of this article in both source language and target language is worth quoting *in extenso*:

(12) Dictionary excerpt from Largeau 1901: 190-191.

> CONSEIL 1° avis sur la manière d'agir, de se conduire, *Alôa*, Pl. *me lôa* (b); *Alôgha*, Pl. *me logha* ... — *Encyclopédie*. Les Pahouins ont leur Code de morale qu'ils transmettent religieusement à leurs enfants. Nous croyons devoir donner ci-après, avec le texte original tel qu'il nous a été dicté, un exemple des conseils donnés par les vieillards aux jeunes gens après la cérémonie de la circoncision. La traduction en est aussi littérale que possible.

ME LÔA	LES CONSEILS
Mi yôṁ mia zô bonghe na: ngue wa yi tabe ne ntaṁ wouïa, mvoghi; o dighe ʒu éba boṅga, wona kagha ʒu éba boṅga, wona o wôghā ndzough, o oboā okôkôr. O kagha bo' ke wôghe azô me zô we di, wona wé-mieṅ o yènā dô. E ma zô we va, ke tabe ési; kan' ghi wa dighi boṅga e mi laṁ; o lough. O kagha ke wa dighi boṅga, o bômi, o soli ne ñe, be mani vegh. Wona, wa tabe ési, o bôme ne é moṅga wouïa; a non' ghi aberṁ; a biê moṅ-a-faṁ, o yô we éso; wona o tab' ési nôṁ-o-môr. O kagha yèṅ, o boā ne nló mvê; wona o ñegheïa. Nina: me zô me nôṁ me nga' zô we, o wôghā mô? — Ana o bo' ke mô wôghā, ve wa bo' okôkôr.	Les vieillards disent aux jeunes gens que: si tu veux rester avec ton bien, sois sage; ne débauche pas les femmes des autres. Si tu débauche les femmes des autres, alors tu auras des désagréments, tu seras malheureux. Si tu n'écoutes pas ce que je te dis, alors toi-même verras cela: Quand je te parle ici, ne reste pas inactif; va voir les femmes dans les villages; tu (en) épouse (une); si tu vas voir les femmes, tu (en) amènes (une), on la paie. Alors tu te fixes, tu couche avec ta femme; elle devient enceinte, elle accouche d'un garçon auquel tu donnes le nom de ton père; puis tu restes vieillard. Si tu vois (cela), tu as été un homme sensé; alors tu es aimé. Il dit: les sages paroles que je t'ai adressées, les as-tu comprises? — Si tu ne les a pas comprises, tu seras malheureux.

With regard to the *Encyclopédie pahouine*, Mavoungou (2001: 125) has mentioned the existence of special articles focusing on cultural data presented in both source language (to the left) and target language (to the right) in two columns set apart typographically, and therefore clearly perceived by the user. In

the encyclopedic section of the above article, introduced by the structural marker *Encyclopédie*, Largeau states that the Fang take great pride in their morality, and the youth are instructed from early childhood to develop a sense of uprightness in all their undertakings. He also mentions that he thought that more data regarding the moral codicil in the Fang community was necessary for a better understanding of the treated lexical item. To fulfil these anticipated expectations on the side of the users, Largeau proposes a text containing some advice given by elderly people to young people undergoing circumcision rites. The French version of the text gives a word-for-word translation of the Fang text. The main purpose here is to describe the value system, which is reflected in the use of the source language. The text riddle instructs young people not to sleep around with married women, but to assume adult duties and responsibilities as early as possible.

3.5   Dictionaries and language registers

According to Dubois and Dubois (1971), language registers are indications that define the attitude of the community regarding verbal behaviour and the way oral and written speech are considered. Terms and expressions become indices and social symptoms: they show that the speech which contains them, belongs to a specific social group. They add that the dictionary informs about different language levels and shows how to ascertain these levels, to delineate them and culturally to censor them: one who uses popular or familiar words or expressions does not belong to the so-called "good society". The word is a mark of recognition. In French society for instance, the way people speak (the language level or register they use) can indicate to which social milieu they belong or what their educational background is. It is a commonly used practice in dictionaries to mark, e.g. familiar, popular and slangy words and expressions. This is done to show that these words and expressions are not part of the standard language. Dictionaries of Gabonese languages hardly contain markings signalling language levels. There are two explanations for this lack of level markings:

—   The first explanation is the fact that Gabonese languages are not standardized yet. In this context, it could be difficult to distinguish the standard level from the others (popular, familiar or slangy).

—   The second explanation is related to the first one. There is hardly ever a relation between someone's linguistic proficiency and his/her social status. A person from a very modest background could have a high proficiency in the use of his/her mother tongue and *vice versa*.

3.6   Dictionaries and culture maintenance

Dictionaries try to satisfy the curiosity of the community. The community

wants to know more about its past. By including cultural customs (even those which are no longer utilized or practiced) in dictionaries, lexicographers can help the society to know about the past and to keep it alive for future generations. In many dictionaries of Gabonese languages, terms referring to old customs and activities are included. The *Encyclopédie Pahouine* by Largeau, is a good example in this regard. An important introductory part, presenting several Fang traditions, is included in this book. Aspects related to marriage, hunting, fishing, agriculture, resolution of problems, accommodation, social status, initiation rites, etc. in Fang society are broadly presented. Even if some of the described customs are obsolete nowadays, their existence in dictionaries give younger generations the opportunity to learn more about their culture. The dowry, for instance, was not practiced in the same way in the past as it is today. Dictionaries can teach their users about the real meaning of the dowry in traditional Fang society as illustrated in the following article from Galley's *Dictionnaire fang–français et français–fang*:

(13) Dictionary excerpt from Galley 1964: 445.

> **DOT** de mariage qui se payait autrefois avec de la monnaie de fer, *nsua bikî* (bmb), *bikî* (b), *évegha* (b), *alo kama* (hh), *bilî bi kama* (bh), *mevekh* (b). Actuellement on paie en marchandise et en argent. Compte des choses données en dot d'une part et en cadeaux (ñgañ) d'autre part, *fa* (h). On fait le compte, *ba zia fa* (hh). Paiement de la dot (la plus grosse part), *mbîkìne* (h), *mbukîne* (h). Premier versement de la dot, petit versement de début, *ôlum* (h), *bem ôlum* (bh). Choses que le beau-père donne à son gendre, *mevale* (h), *ñgañ* (m). Tout ce qu'on ajoute à la dot en plus des vieux fusils (couteaux, cuillères, chat, boubou), *ntsim* (b).

Although this dictionary article provides the user with a good deal of information about the bride price in Fang culture, reservations can be made with regard to the absence of contextual entries addressed at the lemma sign **dot** (dowry). Concerning this point, Busane (1990: 33-34) has pointed out that the word for the English *bride price* has been treated unsatisfactorily in several bilingual dictionaries of African languages — unsatisfactorily because the word in African languages meaning *bride price* is usually associated with the English *dowry* on the one hand and with the French *dot* on the other. However, *dowry* and *dot* in the English and French traditions convey the following meaning: Possessions that a woman brings to her husband under a marriage contract (Les biens qu'une femme apporte à son mari lorsqu'elle rentre dans un contrat de mariage). This meaning is different from that which the term *bride price* conveys. In an attempt to co-ordinate source and target language forms, lexicographers in the English-speaking world compiling dictionaries of African languages generally use the lexeme *bride price* as a translation equivalent for the word meaning "a payment of money or cattle made to the family of a woman when she is married".

In the French-speaking world, apart for "dot", there is no suitable transla-

tion equivalent for the word meaning "bride price" in the African languages. In keeping with the user-perspective and given the fact that *dot* does not capture the dominant meaning of the Fang word *nswá* (bride price), the lemma sign **dot** as given in (13) should be accompanied by contextual entries as it occurs in (14). In spite of all these shortcomings, the treatment of the article **dot** given in (13) does not only explain how the bride price was paid but it also accounts for the different forms of payment. The payment was usually made in metallic bars (the traditional currency before the arrival of the metrical colonial currency of *Franc CFA*, Central African Francs) together with various goods and gifts. In addition, paying of the bride price was hardly ever a single, once only transaction. It was not about buying a wife as goods are purchased. Marriage was not the union between two individuals but a contract between different families. This union was performed by a kinship that could not be dissolved by the divorce of the married couple.

(14) **DOT** (au sens que ce mot a dans les langues gabonaises) de mariage qui se payait autrefois avec de la monnaie de fer, …

In *Les plantes utiles du Gabon*, the lexicographic treatment offered for plant names makes provision for the inclusion of a significant amount of cultural data regarding, e.g. traditional pharmacopoeia together with dietary practices and traditional beliefs. Cultural information on traditional pharmacopoeia, for example, can be retrieved by users at **Irvingia gabonensis**:

(15) Dictionary excerpt from Raponda-Walker and Sillans 1961: 207-208.

> 3. IRVINGIA GABONENSIS Baill. — *Manguier sauvage* (colons); *Oba*. Arbre de grande taille, muni de forts accotements aliformes à la base. Ecorce cendrée roussâtre. Feuillage très dense. Feuilles alternes, simples et entières, luisantes à face inférieure. Stipules courtes, promptement caduques. Fruit drupiforme, à pulpe charnue, plus ou moins amère et très fibreuse. Noyau dur renfermant une graine unique, aplatie, très oléagineuse. Bois très dur, d'un blanc-rosé, lavé de jaune clair, putrescible dès qu'il est exposé à la pluie. Très répandu dans tout le Gabon. *Cf*. PL. XXI.
>
> *Us*. — Le bois, difficile à travailler, mais très résistant, est utilisable pour la grosse charpente, à l'abri de la pluie. Les indigènes l'emploient pour la confection des pilons à mortiers ou auges. Les ménagères recueillent les amandes des fruits pour préparer le pain d'odika (mpongwè, galoa, nkomi, orungu), ndòc (fang), mudika (éshira, bavili, bapunu, masango, banzabi), d'un usage courant dans la cuisine gabonaise. La pulpe amère se mange aussi. Le chocolat indigène, réduit en poudre, peut s'appliquer sur les brûlures. Il intervient encore dans la composition de certains remèdes astringents. En outre, on consomme les râpures de l'écorce dans une grosse banane verte cuite sous la braise, pour enrayer la diarrhée ou la dysenterie; ou bien on la fait macérer pour s'en servir en lavement. À cause de sa fécondité, cette essence est considérée comme un porte-bonheur pour les plantations vivrières. Avec son écorce on fait des talismans pour donner ou rendre la fertilité au sol.

> *Noms vernac.*: Oba (mpongwè, galoa, ikomi, orungu); uba (mitsogo, bavové); wéba (apindji); wiba (ivéa); bwibè (benga); mubè (bakota); mwiba (éshira, bavarama, bavungu, bapunu, balumbu, loango, masangu, ngowé); andòc (fang); vidòkò (béséki); béképé (bakèlè); ondimba (mindumu); mundjiku (bavili); mundjiku (baduma); mundjika (banzabi).

This dictionary article is divided into three sections distinguished typographically and therefore clearly perceived by the user. The first section deals with the description of the plant listed. The second section introduced by an italicised structural indicator (*Us.* = Uses), deals with uses. In the third section, the user is provided with a list of indigenous names in some 21 Gabonese languages. An italicised structural marker (*Noms vernac.* = Vernacular names) also introduces this last section of the article. Section 2 of the text can be translated as follows:

> The wood, difficult to carve but very resistant, is used to make frames for huts. The indigenous population uses it to make pestles and mortars. Housewives collect seeds of the fruits in order to prepare the *dika bread* (mpongwè, galoa, nkomi, orungu), *ndòc* (fang), *mudika* (éshira, bavili, bapunu, masango, banzabi), widely used in Gabonese *cuisine*. The bitter pulp (of the fruit) is also edible. The powder obtained from the seeds can be applied to burns. It is also used in the preparation of some astringent medicines. Moreover, raspings of the bark together with a big banana (plantain) that is cooked under hot ashes are used as a remedy for diarrhoea or dysentery. However, a decoction of the bark can also be used for enemas. Because of its fecundity, this tree is believed to bring luck to farms (alimentary plantations). Talismans that give or restore the fertility of the soil are made from its bark.

With regard to the second topic (traditional beliefs), the treatment of the article of the lemma **Strychnos icaja**, for example, offers the user a great deal of cultural information as can be seen in (16):

(16) Dictionary excerpt from Raponda-Walker and Sillans 1961: 269-270.

> 7. STRYCHNOS ICAJA L. — *Poison d'épreuve*. Arbuste sarmenteux de la forêt dont toutes les parties sont amères. Feuilles opposées, d'un vert terne, coriaces et parcheminées, avec de grandes nervures en demi-cercle. Fleurs en cymes axillaires. Fruits de la grosseur du pouce, rouge à maturité. Racines à écorces rouge vif.
>   *Us.* — L'écorce de la racine est fréquemment employée comme poison d'épreuve, dans les ordalies, pour prouver l'innocence ou la culpabilité des accusés. On fait macérer des râpures de la racine dans une grande tasse ou une calebasse. La fermentation a lieu aussitôt et l'eau prend bientôt une teinte rougeâtre. Une fois l'effervescence apaisée, l'accusé saisit la tasse ou la calebasse et en avale le contenu. Il faut noter qu'à faible dose, cette drogue ne détermine pas la mort; elle est simplement enivrante et diurétique. Comme contre-poison, on préconise les excréments humains avec des oeufs de poule et du jus de canne à sucre.

Noms vernac.: Ikaza (mpongwè); ikadja, kwa (benga); kasè (bakèlè, béséki); mbundu (galoa, nkomi, orungu, ngowé, éshira, bavarama, bavungu, bapunu, balumbu, bavili, baduma, banzabi, loango, masangu, mindumu); mbondo (ivéa, bavové, bakota); moléla (apindji); mwéya (mitsogo); bilon (fang).

Section 2 of the article dealing with uses of the plant can be translated as follows:

> The bark of the root is commonly used in trials by ordeal in order to prove the innocence or the culpability of the accused. Bark from the root has to be macerated in a large bowl or in a calabash. The fermentation takes place immediately and the water soon turns into a reddish colour. After the effervescence, the accused takes the bowl or the calabash and drinks from it. It should be noted that a small dosage of this drug does not kill; it is simply intoxicating and diuretic. As an antidote, human faeces together with chicken eggs and sugar cane juice are recommended.

3.7   Dictionaries and the standardization of culture

For a language with several dialects, dictionaries are usually compiled in one dialect. The gaps between dialects from the same language are not only about linguistic forms, but also about culture, activities and habits. The compilation of a dictionary in a certain dialect means that the lifestyle of the speakers of that dialect will also be presented. Unconsciously, dictionary users will be influenced by that dialect. In other words, they will tend to use the underlying customs of the dialect that is described in the dictionary instead of following their own way of life conveyed by their own dialects. Therefore dictionaries somehow impose the use of the included dialect and favour its customs as the standard ones. By standardizing the language, dictionaries also standardize the culture. In the Gabonese context, this situation has not applied yet, either because, in many Gabonese languages, dictionaries do not have a long existence, or because the existing dictionaries are not available to the public at large. As a result, they cannot really influence the use of these languages.

4.   Conclusion

In this article dealing with the reproduction of Gabonese cultural aspects in existing dictionaries of Gabonese languages, the primary focus has been on the rendering of words of a traditional nature, e.g. the names of plants, insects and animals, as well as the names of other natural phenomena. If the articles are looked at along encyclopedic lines, then all the dictionaries under discussion give a wealth of information about Gabon: its place-names, flora and fauna. Moreover, most of these dictionaries contain introductions encompassing a description of a wide variety of customs and traditions typical of Gabonese cul-

tures. To a larger or lesser degree, all these dictionaries encompass essential Gabonese cultural aspects such as dietary practices, sexuality, mythology, traditional pharmacopoeia, the kinship system, hospitality, and respect for traditional authority and elders.

Urbanisation, the destruction of the environment, the westernisation of communities and the change in life-style have already been responsible for a serious decrease in the proficiency of mother-tongue speakers of Gabonese languages. This is why the cultural dimension is becoming one of the central factors in Gabonese lexicography. Because of the widely recognized importance of the user in dictionary making, modern-day lexicographers are striving to put user-friendly products on the market. These products also encompass the cultural dimension.

Endnote

1. Lexical items in this table are differentiated by topographic means. Items in bold type refer to lemmata as they appear in the dictionaries consulted, whereas words in roman characters either occur as microstructural elements or are insertions by the authors.

References

Dictionaries and encyclopaedias

Église Évangelique du Sud-Gabon. 1966. *Dictionnaire français–yipounou/yipounou–français*. Mouila.

Galley, S. 1964. *Dictionnaire fang–français et français–fang, suivi d'une grammaire fang*. Neuchâtel: Henri Messeiller.

Largeau, V. 1901. *Encyclopédie Pahouine. Éléments de grammaire et dictionnaire français–pahouin*. Paris: E. Leroux.

Martrou, L. 1924. *Lexique FÃN–FRANÇAIS*. Paris: Procure Générale (des Pères du St Esprit)/Abbeville: Imp. Paillard.

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.

Raponda-Walker, A. 1995². *Dictionnaire Français–Mpongwé*. Éditions Raponda-Walker. Classiques Africains. Libreville: Imprimerie St Paul.

Other

Busane, M. 1990. Lexicography in Central Africa: the User Perspective, with Special Reference to Zaïre. Hartmann, R.R.K. (Ed.). 1990. *Lexicography in Africa*: 19-35. Exeter: University of Exeter Press.

Dubois, J. 1970. Dictionnaire et dicours didactique. Rey-Debove, J. (Ed.). 1970: 34-47.

Dubois, J. and C. Dubois. 1971. *Introduction à la Lexicographie: "Le Dictionnaire"*. Paris: Larousse.

Kwenzi-Mikala, J.T. 1990. L'anthroponymie chez les Bapunu du Sud-Gabon. *Pholia. Revue du Laboratoire de Phonétique et Linguistique Africaine* 5: 113-120.

Mavoungou, P.A. 2001. Macro- and Microstructural Issues in Mazuna Lexicography. *Lexikos* 11: 122-138.

Mavoungou, P.A. 2001a. La mondialisation et la lexicographie trilingue ou plurilingue au Gabon. Emejulu J.D. (Ed.). 2001: *Éléments de lexicographie gabonaise, Tome I*: 160-183. New York: Jimacs-Hillman.

Mavoungou, P.A. 2002. Synopsis Articles in the Planning of a Trilingual Dictionary: Yilumbu–French–English. *Lexikos* 12: 181-200.

Mihindou, G-R. 2001. Apports des missionnaires à la lexicographie gabonaise: dictionnaires bilingues Fang–Français/Français–Fang; Français–Yipounou/Yipounou–Français; Français/Mpongwé. Emejulu J.D. (Ed.). 2001. *Éléments de lexicographie gabonaise, Tome I*: 7-37. New York: Jimacs-Hillman.

Nyangone Assam, B. and P.A. Mavoungou. 2000. Lexicography in Gabon: A Survey. *Lexikos* 10: 252-274.

Quillard, G. 1999. Publicité, traduction et reproduction de la culture. *Babel* 45 (1): 39-52.

Rey-Debove, J. (Ed.). 1970. *La lexicographie. Langages* 19.

Rittaud-Hutinet, C. 1980. Lexique. Nsuka-Nkutsi, F. (Ed.). 1980. *Éléments de description du Punu*: 193-245. Lyons: Presses Universitaires de Lyon.

The Lexicographic Treatment of Sublexical and Multilexical Items in a Northern Sotho Monolingual Dictionary: A Challenge for Lexicographers

Motlokwe Clifford Mphahlele, *Sesotho sa Leboa National Lexicography Unit, University of the North, Polokwane, Republic of South Africa* (*mojelav@unorth.ac.za*)

Abstract: Dictionaries have in the past used a word-based approach in which sublexical and multilexical items were not regarded as lemmata. Metalexicography as the theoretical component of lexicography requires that sublexical and multilexical items be lemmatized and treated as independent lemmata in the macrostructure of dictionaries. One of the greater challenges for compiling a better and user-oriented Northern Sotho monolingual dictionary is to treat sublexical and multilexical items as macrostructural elements. Treating these items, the lexicographer faces quite a number of challenges. This article proposes possible ways in which sublexical and multilexical elements could be successfully treated in a Northern Sotho monolingual dictionary. Taking stock of these challenges, the writer comes with suggestions that would assist lexicographers in the compilation of a user-friendly, lexicon-based monolingual dictionary that would lead users to successful information retrieval.

Keywords: SUBLEXICAL ITEMS, MULTILEXICAL ITEMS, AFFIXES, PREFIXAL MORPHEMES, SUFFIXAL MORPHEMES, INTEGRATED MICROSTRUCTURE, WORD-BASED APPROACH, LEMMATIZATION, COLLOCATIONS, COMPOUNDS, COMPLEXES, GROUP PREPOSITIONS, FIXED EXPRESSIONS, MORPHEMES, METALEXICOGRAPHICAL ASPECTS, WORD-FORMATION PROCESSES, WORD-INTERNAL FUNCTION, USER-ORIENTED MONOLINGUAL DICTIONARY, GRAMMAR, HOMONYMOUS LEMMA, POLYSEMOUS LEMMA, DECODING, ENCODING, SEMANTIC COMMENT, SEMANTIC TRANSPARENCY

Opsomming: Die leksikografiese behandeling van subleksikale en multileksikale items in 'n Noord-Sotho- eentalige woordeboek: 'n Uitdaging vir leksikograwe. Woordeboeke het in die verlede 'n woordgebaseerde benadering gevolg waarby subleksikale en multileksikale items nie as lemmas beskou is nie. Die metaleksikografie as die teoretiese komponent van die leksikografie vereis dat subleksikale en multileksikale items gelemmatiseer word en as onafhanklike lemmas in die makrostuktuur van woordeboeke behandel word. Een van die groter uitdagings in die samestelling van 'n beter en gebruikersgerigte Noord-Sotho-eentalige woordeboek is om subleksikale en multileksikale items as makrostrukturele elemente te behandel. By die behandeling van hierdie items word die leksikograaf met 'n hele aantal uitdagings

gekonfronteer. Hierdie artikel stel moontlike maniere voor waarop subleksikale en multileksikale elemente suksesvol in 'n Noord-Sotho- eentalige woordeboek behandel kan word. Deur hierdie uitdagings in oënskou te neem, kom die skrywer met voorstelle wat leksikograwe sal help met die samestelling van 'n gebruikersvriendelike, leksikongebaseerde eentalige woordeboek wat gebruikers tot suksesvolle inligtingsherwinning sal lei.

Sleutelwoorde: SUBLEKSIKALE ITEMS, MULTILEKSIKALE ITEMS, AFFIKSE, PREFI-
GALE MORFEME, SUFFIGALE MORFEME, GEÏNTEGREERDE MIKROSTRUKTUUR, WOORD-
GEBASEERDE BENADERING, LEMMATISERING, KOLLOKASIES, SAMESTELLINGS, KOM-
PLEKSE, GROEPVOORSETSELS, VASTE UITDRUKKINGS, MORFEME, METALEKSIKOGRA-
FIESE ASPEKTE, WOORDVORMINGSPROSESSE, WOORDINTERNE FUNKSIE, GEBRUIKERS-
GERIGTE EENTALIGE WOORDEBOEK, GRAMMATIKA, HOMONIMIESE LEMMA, POLISE-
MIESE LEMMA, DEKODERING, ENKODERING, SEMANTIESE KOMMENTAAR, SEMANTIE-
SE DEURSIGTIGHEID

1.   Introduction

During the last decades, before the publication of Zgusta's *Manual of Lexicography* in 1971 and F.J. Hausmann et al.'s *Wörterbücher Dictionaries Dictionnaires* during 1989–1991, dictionaries were regarded as containers of words. This perception encouraged lexicographers to compile word-based dictionaries that dominated practical lexicography for many years. Like other dictionaries, Northern Sotho dictionaries also consisted of word-based macrostructures. This approach has not assisted dictionary users to retrieve the required information. Due to the fact that dictionaries contained only words as macrostructural elements, users were failed because they could not find the meanings of sublexical and multilexical items. The fact that sublexical items are bound morphemes does not necessarily mean that they must be excluded as independent lemmata from the macrostructure of dictionaries. As part of the lexicon, these items should be included in the macrostructure of dictionaries and be accorded comprehensive treatment.

The Northern Sotho lexicon can be divided into three categories, namely, lexical, sublexical and multilexical elements. Compiling user-friendly dictionaries of a high lexicographic standard for African languages poses a great challenge for prospective lexicographers. One of the greater challenges to compiling a better and user-oriented Northern Sotho monolingual dictionary is to treat sublexical and multilexical items as macrostructural elements. Given that there has never been a Northern Sotho monolingual descriptive dictionary before, the lexicographer faces a difficult and challenging task of treating these items in a monolingual dictionary. Sublexical and multilexical items are fully-fledged members of the lexicon. Like many languages, Northern Sotho has a substantial number of sublexical and multilexical items that should be treated in the articles of a dictionary (cf. Mphahlele 2001). Northern Sotho sublexical items play an important role in word-formation processes while multilexical

items convey meanings like any other lexical unit (word). As lexical items (words) require comprehensive treatment in a dictionary, sublexical and multilexical items also do.

This article proposes possible ways in which sublexical and multilexical items could be treated in a Northern Sotho monolingual dictionary. Taking stock of these challenges, the writer wishes to make suggestions that could assist lexicographers in the compilation of user-friendly, lexicon-based monolingual dictionaries that would lead users to successful information retrieval. A number of suggestions are made that could assist in the enhancement of the linguistic quality of a Northern Sotho monolingual dictionary.

2.   Sublexical items

Sublexical items are items that are smaller than words. Most of these items are productive in word-formation processes and they constitute part of the active lexicon of a language. Sublexical items are commonly known as stems and affixes. With regard to sublexical items, Gouws (1991: 76) writes that the lexicon contains a substantial number of items smaller than words including affixes, combination forms and stems. He stresses that many sublexical items are productive in the word-formation process and constitute part of the active lexicon of a standard language and that dictionaries have to give an account of these items. Although limited, their occurrence as components of complexes and compounds must be recorded. Just like any other language, Northern Sotho has a number of sublexical items that are active in the word-formation process. These items should be included in the macrostructure of dictionaries as treatment units (lemmata). This section will only deal with the treatment of affixes as part of the sublexical items in a Northern Sotho monolingual dictionary. Prefixal and suffixal morphemes are examples of affixes. They are added to the roots or stems during word-formation.

Fromkin and Rodman (1993: 42) define a morpheme as the minimal linguistic sign, a grammatical unit in which there is an arbitrary union of a sound and meaning that cannot be further analyzed. According to Van Wyk (1969: 40-41), a morpheme is a structural characteristic of a group of words, which can be associated with a particular semantic aspect or grammatical function of that group of words. If morphemes can be associated with a particular semantic aspect of words, this therefore means that they convey particular meanings associated with words. Because of this, they should be accorded an independent lemmatic status as treatment units in dictionaries.

2.1   The treatment of prefixal morphemes

According to Lombard et al. (1985: 22), prefixal morphemes are morphemes that precede the root of a word. This means that prefixal morphemes can be

found on the left-hand side of the root of a word. For these morphemes to receive proper treatment in a monolingual dictionary, lexicographers must have knowledge of their occurrence or behaviour in words. For example, the treatment of noun class prefixes in a monolingual dictionary requires the lexicographer to have a thorough knowledge of their occurrence as noun class prefixes in a lexicon. The reason for this is that the definition of a sublexical item should indicate how a particular morpheme behaves in a word. For the lexicographer of a Northern Sotho monolingual dictionary, this therefore becomes a challenging if not a difficult task.

In dealing with this challenge, the lexicographer must first of all know the noun class system in Northern Sotho. Some of the noun classes share the same prefixal morphemes. In this case, the lexicographer has to make a differentiation in the microstructure so that the user can retrieve the required information. This means that all the instances of the occurrence or behaviour of a particular prefixal morpheme should be recorded. How then should this prefix as a polysemous or homonymous lemma be recorded? The lexicographer of a Northern Sotho monolingual dictionary does not know whether to record the meanings of one prefixal morpheme in one article or whether to accord all the occurrences an independent status as lemmata. With his/her linguistic sensitivity and knowledge of complicated morphological structures, the lexicographer would be able to solve this problem. What is important here is that sublexical items can function as polysemous or homonymous lexical items. In other words, if a morpheme has polysemic distinctions, the lexicographer should accord such a morphemic lemma an integrated microstructure and if a morpheme has two unrelated occurrences such a morpheme should be accorded a twofold independent lemmatic status in the macrostructure.

When treating the noun class system in the dictionary, the lexicographer should bear in mind that the morphology of the noun is different from that of the verb. With regard to this, Van Wyk (1995: 43) writes: "The morphology of the noun differs in crucial ways from that of the verb and the noun prefix is not mobile or freely exchangeable." Without this knowledge, the lexicographer of a monolingual dictionary cannot succeed in the treatment of the prefixal morphemes.

There are eighteen noun classes in Northern Sotho. By virtue of the form of its prefix, every noun belongs to what has been traditionally called a noun class. Poulos and Louwrens (1994: 11) point out that each noun class is numbered according to the numbering system customary in the comparative study of the family of languages to which Northern Sotho belongs.

Table 1: The noun class system in Northern Sotho

Class	Prefix	Example	Translation
1	mo-	*motho*	person
1a	Ø	*rramogolo*	paternal uncle
2	ba-	*batho*	people

2a	bo-	*borramogolo*	paternal uncles
3	mo-	*molete*	hole
4	me-	*melete*	holes
5	le-	*lefene*	cockroach
6	ma-	*mafene*	cockroaches
7	se-	*sefatanaga*	car
8	di-	*difatanaga*	cars
9	n-/Ø	*nko*	nose
10	di-	*dinko*	noses
14	bo-	*bogobe*	porridge
15	go	*go ngwala*	to write
16	fa-	*fase*	below/down
17	go-	*godimo*	up/above
18	mo-	*morago*	behind

The noun **monna** (man) belongs to noun class 1 because it has the prefix **mo-** and has a [+human] content. The noun **motse** (village) happens to have the same prefix **mo-**, but it belongs to class 3, because it also has a [–human] content. The noun **dilepe** (axes) belongs to class 8 because it has the prefix **di-**. The noun **dikgomo** (cattle) also has the prefix **di-** and it belongs to class 10. This means that noun classes 1 and 3 share the same prefix **mo-** while noun classes 8 and 10 share the same prefix **di-**. In the treatment of these morphemes, the lexicographer must give a clear indication of these occurrences in the definition. What is important is how he/she would make his/her treatment accurate and clear. What the lexicographer needs to do is to establish whether the prefixal morpheme **mo-** can be classified as either a polysemous or a homonymous sublexical item. After having determined this, he/she will therefore be able to give a clear and accurate presentation of the prefixal morpheme. The following presentation would be an accurate treatment of the sublexical item **mo-**:

> **mo-**   hlogo ya maina a legoro la 1
>   (class 1 noun prefix)
>   hlogo ya maina a legoro la 3
>   (class 3 noun prefix)

The integrated microstructure above shows that the sublexical item **mo-** can function as a prefix of noun classes 1 and 3. In this case, the dictionary user is able to see the dual function of the sublexical item **mo-** in a noun class system. This presentation is apt because the polysemic distinctions of the lemma are clearly indicated. Awarding **mo-** an independent lemmatic status is also a fit lexicographic procedure because this morpheme is a fully-fledged member of the Northern Sotho lexicon. With this approach, the lexicographer will be able to assist the dictionary user to retrieve the required information. For an accurate presentation of prefixal morphemes, lexicographers of a Northern Sotho

monolingual dictionary should cooperate with linguists so that the behaviour or occurrence of morphemes in words is clearly indicated. According to Gouws (1989), not only words but also sublexical and multilexical elements should be included as macrostructural elements in a dictionary. With regard to this, he stresses that a lexicon-based approach to lexicography requires a new selection of macrostructural elements as well as a change in the structuring of dictionary articles. This means that dictionary articles should be restructured so that sublexical and multilexical elements could form part of the macrostructure of a dictionary.

The treatment of the sublexical item **le-** poses some problems for the lexicographer of a Northern Sotho monolingual dictionary. The following treatment of the sublexical item **le-** and lexical item **le** could make the dictionary user retrieve the required information:

> **le-** hlogo ya maina a legoro la 5
> (class 5 noun prefix)
> *mohlala:* **le**gapu
> *(example: watermelon)*

> **le-** hlogo ya lesala la motho wa bobedi ka botšing
> (the pronoun prefix of the second person plural)
> *mohlala:* **le**na
> *(example: you)*

> **le** lekgokedi la maina a legoro la 5
> (the concord of noun class 5)
> *mohlala: lefene* **le** *hwile*
> *(example: the cockroach is dead)*

> **le** lekopanyi la maina
> (a noun conjunction)
> *mohlala: monna* **le** *mosadi*
> *(example: the man and the woman)*

> **le** lešupi la maina a legoro la 5
> (the demonstrative of noun class 5)
> *mohlala: lefene* **le**
> *(example: this cockroach)*

From the above presentation, the dictionary user will be able to deduce that the sublexical item **le-** is different from the lexical item **le**. In the above articles, the sublexical item **le-** has been awarded a double lemmatic status which means that it can function twice in different occurrences in the lexicon. In the first article, **le-** is defined as a class 5 noun prefix whilst in the second article it is defined as the pronoun prefix of the second person plural. This is a fitter presentation of these sublexical items which would allow the user to see the different occurrences of the morpheme **le-** in the lexicon. The first two articles are

followed by articles where the lexical item **le** functions as homonymous lemmata. Although this article does not deal with the treatment of lexical items, the above presentation is to show how lexicographers can differentiate sublexical items from lexical items in the macrostructure of a dictionary. Provided with examples, the dictionary user would be able to use the presented morphemes correctly. Lexicographers should provide each article of a sublexical item with an appropriate example to assist the user to comprehend the presented semantic information. The structural resemblance that often exists between sublexical and lexical items should not mislead the lexicographer to include sublexical items in the microstructure of lexical lemmata. This approach or lemmatization procedure is detrimental to the lemmatic status of sublexical items. That is, the presentation of the sublexical item **le-** in the microstructure of the lexical lemma **le** could in no way assist the dictionary user to retrieve the meaning of this prefixal morpheme. This lemmatization procedure used by traditional lexicographers has not assisted dictionary users to retrieve the meanings of sublexical items. Abandoning a word-based approach in favour of a sublexical approach, lexicographers of a Northern Sotho monolingual dictionary need to give an apt account of sublexical items in their dictionary.

The following presentation of the sublexical item **di-** and lexical item **di** could assist the dictionary user to retrieve the presented semantic information successfully:

> **di-** hlogo ya maina a legoro la 8
> (class 8 noun prefix)
> *mohlala: **di**lepe*
> *(example: axes)*
>
> hlogo ya maina a legoro la 10
> (class 10 noun prefix)
> *mohlala: **di**pudi*
> *(example: goats)*
>
> **di** lekgokasediri goba lekgokasedirwa la maina a legoro la 8
> (the subjectival or objectival concord of noun class 8)
> *mohlala: dilepe **di** timetše*
> *(example: the axes are lost)*
>
> lekgokasediri goba lekgokasedirwa la maina a legoro la 10
> (the subjectival or objectival concord of noun class 10)
> *mohlala: dikgomo **di** a fula*
> *(example: the cattle are grazing)*

This treatment is user-friendly because dictionary users will be able to deduce that the sublexical item **di-** can function as a class prefix of classes 8 and 10. This has been aptly presented in an integrated microstructure. The inclusion of the sublexical item **di-** as a lemma in the macrostructure would allow users to

see that there is no semantic relationship between sublexical items and lexical items. If the first article had consisted of one definition only, the user would not have been able to retrieve the required information because one occurrence of the morpheme **di-** would have been lacking. In other words, the dictionary user would have thought that the morpheme **di-** can only function as class prefix of class 8 or as class prefix of class 10. To solve this problem, the lexicographer must do some morphological research so that the articles of sublexical items reflect their occurrences in the lexicon. That is, the way in which the prefixal morphemes are treated in dictionaries should reflect their linguistic status in the lexicon. With regard to this, Gouws and Prinsloo (1997: 46) stress that the first step towards the improvement of the lexicographic standard of dictionaries for African languages must be to do proper groundwork. According to them, dictionaries are instruments of linguistic and communicative empowerment and therefore lexicographers have to make sure that their intended target users receive optimal linguistic information. Lexicographers of a Northern Sotho monolingual dictionary must also cooperate with linguists to find a treatment of sublexical items suitable for a lexicon-based dictionary.

2.2   The treatment of suffixal morphemes

Suffixal morphemes are also known as suffixes. These morphemes, according to Lombard et al. (1985: 22), occur after or behind word roots. Fromkin and Rodman (1993: 518) define a suffix as a bound morpheme which occurs after the root or stem of a word, an affix which is attached to the end of a morpheme or a word. This means that suffixal morphemes have a predominantly word-internal function.

Most dictionaries have not effectively treated suffixal morphemes. In some dictionaries, suffixal morphemes have not been given lemmatic status in the macrostructure. This approach, according to Gouws (1991: 88), is an invalid lexicographic account for it does not regard sublexical items such as suffixes as fully-fledged lexical items. Because they are fully-fledged lexical items, suffixes should be included as lemmata in the macrostructure of dictionaries. The treatment of suffixes in a Northern Sotho monolingual dictionary poses some problems and challenges to the lexicographer.

The following are some of the suffixes that should be treated in the macrostructure of a Northern Sotho monolingual dictionary: **-ana, -ile, -itše, -iša, -ega, -olla, -eng, -ela**. In Northern Sotho, these suffixal morphemes are productive in word-formation processes. The treatment of these suffixal morphemes in the articles of other lexical items would not assist dictionary users to obtain the required information from the dictionary. In some cases, dictionary users would regard these morphemes as extra-linguistic information in the microstructure. With regard to the treatment of sublexical items, Gouws (1991: 88) says: "To the detriment of the user, dictionaries include sublexical items as sublexical lemmas and these lemmas then receive erroneously a complete lexi-

cographical treatment, another invalid lexicographical account." In other words, the inclusion of prefixal and suffixal morphemes in the articles of some lexical lemmata is not an accepted lexicographic procedure. Because morphemes may appear complicated to dictionary users, it is important for these sublexical items to be accorded lemmatic status so that their treatment could be as comprehensive as possible.

The following example may assist lexicographers to treat suffixal morphemes in an accurate and acceptable way:

> **-ana** moselana wa nyenyefatšo woo gantši o hlomesetšwago maineng
> (diminutive suffix which is usually attached to nouns)
> *mohlala: mosela + **ana** = mosel**ana***

> **-ana** moselana wa ledirani
> (the reciprocal suffix)
> *mohlala: nyala + **ana** = nyal**ana***

> **-ile** moselana wa lephethi
> (the perfect suffix)
> *mohlala: rema + **ile** = rem**ile***

> **-iša** moselana wa lediriši
> (the causative suffix)
> *mohlala: lema + **iša** = lem**iša***

> **-ela** moselana wa lediredi
> (the applicative suffix)
> *mohlala: nyaka + **ela** = nyak**ela***

Because the suffix **-ana** receives double lemmatic status in the macrostructure, users of a Northern Sotho monolingual dictionary will be able to see that the suffixal morpheme **-ana** has two unrelated meanings. The given examples after the definition play an important role in assisting users to contextualise the treated sublexical lemmata. Without these examples, dictionary users may find it difficult to apply these suffixes to nouns and verbs. In the first article of **-ana**, users are able to deduce that this suffixal morpheme is attached to nouns to indicate the diminutive of objects. Giving this sublexical item double independent lemmatic status is an apt lemmatization procedure because if the lexicographer accords it single lemmatic status, users will only see **-ana** as either the diminutive or the reciprocal suffix. That is, single lemmatic status for this morpheme would only partly represent its occurrence in the lexicon. Even if the two definitions for **-ana** were included in a single lemma, this would also be problematic because an integrated microstructure would mean that the meanings are related to each other. Although this lemmatization procedure is challenging, lexicographers should always try to be accurate in their treatment so that the decoding and encoding needs of their target users can be met.

Sublexical items should not appear in the microstructure or semantic com-

ment of other lexical lemmata. These items should be regarded as fully-fledged members of the lexicon of a language. The placing of sublexical items in the microstructure or semantic comment of other lexical lemmata disregards the status of these items as members of the lexicon. This lemmatization procedure is also detrimental to the optimal retrieval of semantic information in dictionaries.

3.   Multilexical items

Multilexical items are lexical items that consist of more than one word. This combination of words is always a unit and should be treated likewise in a dictionary. Although multiword lexical items consist of more than one word, they should, according to Gouws (1991: 78), be regarded as single lexical items. These items should therefore be included as multilexical lemmata in the macrostructure of dictionaries.

Just like other languages, Northern Sotho possesses multilexical items such as collocations, fixed expressions, complexes, compounds, group prepositions, etc. As independent and fully-fledged members of the lexicon, these items should be included in the macrostructure of a dictionary as multilexical lemmata (cf. Mphahlele 2001a: 103-104). When treating multilexical items, the lexicographer of a Northern Sotho monolingual dictionary is confronted with quite a number of problems and challenges. The lexicographer does not know firstly, whether compound words should be treated in the articles of their headwords or in the macrostructure as multilexical lemmata, secondly, whether a collocation should be included in full or whether one component of a collocation should be given in the macrostructure as a headword, and, thirdly, whether fixed expressions should be included in full in the macrostructure of a standard, general or special dictionary or whether these should form part of a glossary in the back matter of the dictionary. This section tries to formulate proposals on how to treat multilexical items in a Northern Sotho monolingual dictionary.

With regard to the treatment of multilexical items, Gouws (1991: 82) writes: "Treating multiword lexical items in the entries of lexical lemmas diminishes the linguistic soundness of a dictionary. The conventional treatment of the majority of multiword lexical items results directly from a twofold bias dominating lexicographical practice." The implied semantic relationship between multilexical elements and lexical elements should not compel the lexicographer to include multilexical items in the microstructure of lexical lemmata. This approach which is word-biased, is detrimental to the effective retrieval of the semantic information of multilexical items. That is, if multiword lexical items such as compounds, collocations and fixed expressions are included in the microstructure or semantic comment of other lexical items, dictionary users would not be able to retrieve their meanings because they would regard any information presented after the definition of a lexical lemma as extralinguistic.

In other words, users would not expect to find the meaning of a compound word in the microstructure of a lexical lemma.

Multilexical items are independent members of the lexicon of Northern Sotho and they must be granted an independent lemmatic status in the macro-structure of a dictionary. The following are some of the different multilexical items in Northern Sotho:

Table 2: Different types of multilexical items in Northern Sotho

Multilexical item	Type
mothomogolo	compound
tselapedi	compound
ka ntle ga	group preposition
kgauswi le	group preposition
ga mmogo le	group preposition
sebakeng sa	group preposition
ka nnete	collocation
ka mme	collocation
pudi ya leleme le letala	fixed expression (proverb)
ntsho ya mathudi	fixed expression

Jackson (1988: 98) refers to a collocation as a combination of words that has a certain mutual expectancy. According to him, the combination in a collocation is not a fixed expression but there is a greater than chance likelihood the words will cooccur. The cooccurrence of words in a collocation compels the lexicographer to lemmatize a collocation as a multilexical lemma. Failure to lemmatize a collocation would mean that users would not be able to retrieve the meaning associated with a particular collocation. For example, the collocation **ka nnete** does not have any meaning relationship with the lexical item **nnete**. Therefore, the lexicographer of a Northern Sotho monolingual dictionary should not include a collocation **ka nnete** in the microstructure of the lexical lemma **nnete**. This is an unacceptable lexicographic procedure.

The following treatment could assist dictionary users to retrieve the semantic information regarding multilexical items:

> **mothomogolo** Ke motho yo a gotšego yo a nago le mengwaga ye mentši
> (an elderly person)
> Motho wa maemo setšhabeng go swana le kgoši, bjalobjalo
> (a community leader like a chief, etc.)

> **tselapedi** Ke mokgwa wo mongwe wa go tšwelela ntle le wo o tlwae-legilego
> (an option to succeed or survive)

> **ka ntle ga** Ke letlema leo le šomišwago go šupa selo seo se sego gona goba seo se sa šomišwego

(a group preposition meaning "on the side of", "beyond", "not having", etc.)

kgauswi le Ke letlema leo le šomišwago go šupa go batamela ga selo go se sengwe
(a group preposition meaning "next to", "nearer to", etc.)

ga mmogo le Ke letlema leo le šupago go ba gona goba go ba gotee ga dilo tše pedi goba go feta
(a group preposition meaning "together", "with", etc.)

sebakeng sa Ke letlema leo le šupago ge selo se sengwe se se sa kgethwa
(a group preposition meaning "instead of", "in spite of", "in the place of")

ka nnete Ke letlema leo le šupago go dumela, go makala, bjalobjalo
(a collocation meaning "indeed")

ka mme Ke letlema leo le šupago kgonthišišo
(a collocation meaning "sure!")

pudi ya leleme le letala Ke selo sa go ithuša goba go intšha bohloking go go swana le diruiwa go ba tšhelete
(a proverb meaning something that relieves a person from poverty, i.e. money or livestock)

ntsho ya mathudi Morogo
(leafy plant cooked and eaten as a vegetable)

From the presented compound words above, it is obvious that the meaning of all compound words cannot be deduced from their component parts. With regard to compounds, Gouws (1999: 25) writes: "Complex lexical items are also included in dictionaries but all complex items do not get a similar treatment. In linguistics a distinction is made between transparent and intransparent complex items. The meaning of a transparent complex item can be deduced from the meaning of its components but the meaning of an intransparent complex cannot be deduced from the meaning of its components."

The lexicographer of a Northern Sotho monolingual dictionary should know that there are self-explanatory and intransparent compounds. In the case of intransparent compounds, where the meaning cannot be deduced from their constituent parts, lexicographers should give a comprehensive definition in the microstructure of such multilexical lemmata. In the article of the multilexical lemma **mothomogolo**, for example, the second polysemic distinction does not have any meaning relation with the constituent parts of the lemma. In this case, it is important for the lexicographer to give a comprehensive definition regarding the intransparent compound. Self-explanatory lemmata display 'semantic transparency' and in this case, the lexicographer may give a brief definition to assist the user to retrieve the required information.

With regard to the treatment of fixed expressions (idioms and proverbs) in dictionaries, lexicographers may choose to include these items as multilexical lemmata in the macrostructures of general or special dictionaries.

4.   Conclusions and recommendations

Obviously, the treatment of sublexical and multilexical items in a Northern Sotho monolingual dictionary poses some challenges to the prospective lexicographer. In treating these items some useful guidelines need to be taken into consideration. These are:

— Sublexical items are productive in the word-formation process.

— Sublexical and multilexical items are fully-fledged members of the lexicon of a language and therefore need to be accorded independent lemmatic status.

— Sublexical items can be divided into two types, i.e. prefixal and suffixal morphemes that should be presented as lemmata in the macrostructure.

— There are different kinds of multilexical items in Northern Sotho, i.e. group prepositions, compounds, complexes, collocations, fixed expressions, etc., that should be accorded independent status in the macrostructure of a dictionary.

— The implied semantic relationship between multilexical items and lexical items should not compel the lexicographer to include multilexical items in the microstructure of other lexical items. This approach is in disregard of the lemmatic status of multilexical items.

— To be accurate in their definitions, lexicographers should observe the behaviour or occurrence of sublexical items in a word and how these items affect the meanings of words. In other words, lexicographers should study the morphology of the words with which they are dealing.

— The treatment of sublexical and multilexical items requires lexicographers to be sensitive to language so that they can be able to see the different occurrences of one affix in a lexicon.

— Without a lexicon-based approach, no dictionary will ever present itself as a linguistically-matured instrument.

— The morphology of the noun differs in crucial ways from that of the verb and the noun prefix is not mobile or freely exchangeable like the verb prefix.

— Like some lexical items, sublexical and multilexical items can function as polysemous and homonymous lemmata. In other words, if a morpheme has polysemic distinctions, the lexicographer should accord such a mor-

phemic lemma an integrated microstructure and if a morpheme has two or more unrelated occurrences, such a morphemic lemma should be accorded a twofold or multiple independent lemmatic status in the macrostructure.

As mediators between complicated grammatical structures and the decoding and encoding needs of their target users, the lexicographers of African languages should always face lexicographic challenges so that their dictionaries can display a high lexicographic standard. Traditional lexicography that regards dictionaries as containers of words, should be replaced by a lexicon-based approach to lexicography.

References

Fromkin, V. and R. Rodman. 1993⁵. *An Introduction to Language.* New York: Harcourt Brace Jovanovich College Publishers.

Gouws, R.H. 1989. Setsels as makrostruktuurelemente van Afrikaanse woordeboeke. *Suid-Afrikaanse Tydskrif vir Taalkunde* 7(2): 59-67.

Gouws, R.H. 1991. Toward a Lexicon-based Lexicography. *Dictionaries* 13: 75-90.

Gouws, R.H. 1999. *General Theory of Lexicography. An Introduction.* M.Phil. Study Guide. Stellenbosch: University of Stellenbosch.

Gouws, R.H. and D.J. Prinsloo. 1997. Lemmatisation of Adjectives in Sepedi. *Lexikos* 7: 45-57.

Jackson, H. 1988. *Words and their Meanings.* New York: Longman Inc.

Lombard, D.P., E.B. van Wyk and P.C. Mokgokong. 1985. *Introduction to the Grammar of Northern Sotho.* Pretoria: Sigma Press.

Mphahlele, M.C. 2001. *A Monolingual Descriptive Dictionary of Sesotho sa Leboa: A Lexicon-based Dictionary.* Unpublished Article.

Mphahlele, M.C. 2001a. *A Model to Achieve Communicative Equivalence in Translation Dictionaries.* Unpublished M.Phil. Thesis. Stellenbosch: University of Stellenbosch.

Poulos, G. and L.J. Louwrens. 1994. *A Linguistic Analysis of Northern Sotho.* Pretoria: Via Afrika.

Van Wyk, E.B. 1969. 'n Kritiese ontleding van enkele aspekte van die mees algemene morfeemopvatting. *Klasgids* 4(3): 35-43.

Van Wyk, E.B. 1995. Linguistic Assumptions and Lexicographical Traditions in the African Languages. *Lexikos* 5: 82-96.

The Implications of Culture for Dictionaries of the African Languages[*]

A.C. Nkabinde, *Thornville, Republic of South Africa*

Abstract: This article attempts to show how culture or aspects thereof can be used to complement linguistic and other information in the compilation of dictionaries of African languages. Some obstacles in the way of achieving this goal are identified and proposals made on how to deal with them. Although only some cultural aspects of a single language are examined, the conclusions are valid for cultural aspects of all African languages.

Keywords: CONTEXT, CULTURE, CORPUS, EUPHEMISM, FIGURATIVE LANGUAGE, HLONIPHA, KINSHIP TERMS, LEXICAL BORROWING, MULTILINGUALISM, SPEECH COMMUNITY, STANDARD ZULU, TABOO

Opsomming: Die implikasies van kultuur vir woordeboeke van die Afrikatale. Hierdie artikel probeer om aan te toon hoe kultuur of aspekte daarvan gebruik kan word om taalkundige en ander inligting aan te vul by die samestelling van woordeboeke van die Afrikatale. 'n Aantal struikelblokke op die weg om hierdie doel te bereik, word geïdentifiseer en voorstelle gemaak oor hoe om hulle te hanteer. Alhoewel slegs sommige kulturele aspekte van 'n enkele taal ondersoek word, is die gevolgtrekkings geldig vir kulturele aspekte van alle Afrikatale.

Sleutelwoorde: EUFEMISME, FIGUURLIKE TAAL, HLONIPHA, KONTEKS, KORPUS, KULTUUR, LEKSIKALE ONTLENING, TAALGEMEENSKAP, TABOE, STANDAARDZOELOE, VEELTALIGHEID, VERWANTSKAPSTERME

1. Introduction

The size, content, format, purpose, style and "depth" of a dictionary may vary from dictionary to dictionary according to the compiler's perceived needs of the users of the dictionary. A dictionary may also be largely influenced by the tradition of dictionary making adopted as well as the language described or the level of linguistic description of the language concerned. Other factors which may have a bearing on the language in dictionary making comprise the history and traditions of the specific speech community, including its origins, mythology, legends, exploits, rituals, ceremonies, wisdom, world-view, arts and crafts and other activities together with phenomena in nature and the environment as a whole — indeed the entire universe.

---

[*] This article was presented as a keynote address at the Seventh International Conference of the African Association for Lexicography, organized by the Dictionary Unit for South African English, Rhodes University, Grahamstown, 8–10 July 2002.

In South Africa there is a preponderance of translation dictionaries over explanatory dictionaries of the African languages belonging to the Bantu language family. Most of these were produced by Christian missionaries. In addition to the indigenous lexical items they contain, there is also a fair proportion of lexical borrowings from Indo-Germanic languages adapted for school, religious and other uses. The borrowings and adaptation of existing words in the language are mainly used to accommodate new ideas brought about by the contact of languages. They signal the onset of multilingualism in society and the decline of monolingual societies in the various countries in which colonialism or trade or both have occurred. They are a consequence or evidence of the co-existence of legal, social, cultural, religious, economic and other systems over a considerable period of time. A dictionary of "fanakalo" chronicles multilingualism in industry and on the mines in South Africa. A few monolingual dictionaries in the African languages have been produced or are in the process of production.

A notable feature of translation and explanatory dictionaries in the African languages is their relatively limited description of various aspects of traditional life or culture of the African. *The Greater Dictionary of Xhosa* can be considered an exception to this. Various factors may have contributed to this situation, *inter alia*, urbanisation, industrialisation, social and political organisation, the adoption of Christianity, a new value system, and the decline of the influence of tribal and family life. Many of the traditional rites and practices have either disappeared or continue to exist in a drastically reduced or fragmented fashion.

2.1   What is Culture?

Herskovits (1960: 17) defines "culture" as "the man-made part of the environment". Culture is also often defined in an hierarchical fashion to indicate various levels of social development in order to distinguish between what is regarded as "primitive" on the one hand and "civilised" on the other. Culture is often regarded by some Africans as a form of national identity. Writing in the *Echo*, supplement to the *Natal Witness* (6 June 2002: 8), Bongani Mthethwa decries the fact that black people are "no longer proud of our Africanness". According to this, the communality of African life is contrasted with the individualism of the West. For purposes of this article, an examination of the interconnection between culture and language in general and lexicography in particular is undertaken.

2.2   The Interconnection between Language and Culture

The interconnection between language and culture is widely recognised. Sapir (1921: 233) indicates it as follows: "Culture may be defined as *what* a society does and thinks. Language is a particular *how* of thought."

Thipa (1989: 43-44) stresses the systematic ordering or patterning of culture: "Culture has order or pattern ... comprising a system. This system or unit is composed of interrelated parts which are mutually influencing." He concludes that "culture refers to the *total* way of life of a particular group of people".

The interconnection between language and culture is, however, not always direct or straightforward. Words denoting culture or aspects thereof do not necessarily have special markers or some special features in the grammar. Sapir (1921: 234) says: "In the sense that the vocabulary of a language more or less faithfully reflects the culture whose purposes it serves it is perfectly true that the history of language and the history of culture move along parallel lines." This makes it difficult to distinguish cultural words from the rest of the vocabulary of a language. Context appears to be the only means of establishing the use of words denoting culture.[1] We accordingly want to investigate the identification of words in cultural context as a means of determining the interconnection between culture and language.

One of the methods of identifying words in cultural context is to determine the various cultural aspects peculiar to a speech community rather than address the entire domain of culture as such. It is also more illuminating to concentrate on a single speech community rather than to deal with the broad subject of the interconnection between culture and language in general.

Zulu culture can roughly be classified according to about twenty aspects, e.g. kinship and social organization, communal activities, recreation, customs and beliefs, food and beverages, sorcery and witchcraft, mourning, burial and death, agriculture and animal husbandry, mythology and legends, music, dancing and song, folklore, national ceremonies, etc. Only one of these aspects, viz. kinship terms, will be discussed briefly.

Some Zulu kinship terms show peculiar linguistic behaviour. The singular noun forms, for example, have two or more meanings, e.g. *ubaba* (father/paternal uncle/paternal aunt) and *umama* (mother/maternal aunt). In addition, the terms *ubaba* and *umama* denote plurality. Some kinship terms also accommodate one's siblings. They may further be used to refer to any senior adult male or female. They are sufficiently inclusive to be used without the possessive/genitive form *wami* or *wethu*.[2]

Another set of singular nouns contains singular forms with compound morphemes having plural genitive forms indicating plurality, viz. *udadewethu* (my/our (older) sister), *umfowethu* (my/our (older) brother), *uzakwethu* (my/our fellow bride/companion), *umnewethu* (my/our elder brother), etc. These kinship terms may also apply beyond the ambit of the clan. In spite of the plural genitive morpheme used in these nouns, they retain their singular forms because of the singular prefix they use. This shows clearly that these nouns coincide in meaning with those divested of this morpheme, viz. *ubaba*, *umama*, etc.

There are kinship terms that are contrastive. These are:

(a) Age:
  (i) according to Seniority:
    umnewethu (elder brother)
    udadewethu (older sister)
    ubabamkhulu (older than parents — grandfather)
    ugogo/ukhulu (grandmother)
  (ii) according to Juniority:
    umnawami (younger brother)

(b) Sib:
  (i) by Marriage
    umlanda/umlamu (brother-/sister-in-law on wife's side — her brother or sister)
    umfumbesi/umnakwethu (fellow son-in-law — married to wife's sister)
    umalume (uncle — mother's brother)
    umukhwe (wife's father)
    umkhwekazi (wife's mother/wife's mother's sister)
    ubabekazi (father's sister)
    ubabezala (husband's father/husband's paternal uncle)
    abalandakazi (in-laws on wife's side)
    umkhozi/umlingani/usebele (fellow parent-in-law)
    umamezala (husband's mother)
  (ii) by Descent
    ukanina (cousin on mother's side — child of mother's sister)
    umzala (cousin on father's side — aunt's/uncle's son or daughter).
  (iii) as Offspring
    umshana (sister's son or daughter)

Some kinship terms are borrowed from Afrikaans or English. These are:

    usisi (sister/suster — elder sister)
    ubhuti (boetie — elder brother)
    uanti (aunt/tante — father's or mother's sister/any adult woman older than oneself/any adult woman of the same age as one's mother)
    ubhululu (broer — friend/mate)

It is not clear why these borrowings took place. They may have resulted from the following factors:

(a) Redundancy: *Ubaba* which has a long series of meanings (father/paternal uncle/paternal aunt) often requires some distinguishing qualification, e.g. *ubaba omncane* (father's younger brother) and *ubaba omdala/omkhulu* (father's older brother).

(b) Compensation: *Usisi* (sister/suster) probably reinforces *udadewethu* which is the only term for sister in contrast with *umfowethu*, *umnewethu* and *umnawami* the masculine counterparts.

(c) Harmonisation: *Ubhuti/Ubhululu* (boetie/broer) may be the result of harmonisation of the vocabulary in a multilingual environment. *Ubhululu* could also be evidence of the decline or disuse of indigenous words such as *unkabimalanga*, *untanga* or *utate*.[3]

The foregoing nouns demonstrate forcibly the inclusivity peculiar to kinship terms in Zulu. They are classificatory by nature and appear to conform to the social organisation of the people where the clan or sib plays an important role. The layout of the *umuzi* (kraal) also conforms to the hierarchical structure of the family and the various positions of the nuclear forms of the households.

The lexical borrowings of kinship terms are testimony of a steady change in the culture of the people. They indicate the influence of multilingualism on the culture and language of the people. The lexical borrowing found in kinship terms illustrates the need and importance of viewing culture as a whole without confining it to a traditional setting. Furthermore, the examination of lexical items without also taking cognizance of their cultural background is likely to result in superficial definitions of entries in a dictionary.

The rest of the aspects we have identified in Zulu culture is likely to confirm our findings in the examination of the interconnection between language and culture demonstrated by kinship terms.

3.   Some Problems

The accommodation of Zulu culture, like that of the entire material used in the compilation of a dictionary, presents very complex and often controversial problems. These are discussed below.

3.1.1 Standard Zulu

Standard language is a form of language prescribed and recognised for all official communication (both written and spoken) in a language. The vocabulary and usage of such a form of language are usually based on a prestigious variant or dialect of a language. Standard language is often elitist. It does not accommodate regional or dialectal forms of a language.

Ideally, each of the six Zulu dialects identified by Kubheka (1979: 90), for example, ought to have its own literature and dictionaries where all its individual linguistic idiosyncrasies could be fully catered for. Basically, each dialect is a language in its own right (see Mokgokong 1966: 32). This is, however, impractical for literary purposes. It is unviable to cater for the literary needs of small populations of dialects, more than 50% of whose members are illiterate.

The practical solution for such speakers is to provide jointly for the needs of the Zulu people as a whole. The implication of this is to find a compromise by declaring one of the dialects a standard language. However, such a plan creates scope for strife and conflict.

Earlier efforts to recognise a speech form that could pass for a standard Zulu language have not been successful. Suter's (n.d.) proposal that *isiNtungwa* be made the standard language did not receive support. Kubheka's (1979) suggestion that the Central Zululand dialect be adopted as standard language has for historical and other reasons been a failure. Ndlovu (1963) advocated what he calls "standard educated speech" as the norm for standard Zulu.

Some reasons for the failure to have proposals on standard Zulu recognised include the following:

1. The proposals are based solely on written language. On his own admission, Suter says that even educated Zulus speak a mixture of dialects. It is not clear how such speakers would acquire and use *isiNtungwa* in their writing without adequate written sources in this language.

   It seems unrealistic to determine a standard language on the basis of a language with limited literary sources and a very small percentage of the population literate. Despite the declaration of compulsory education, thousands of children in KwaZulu-Natal still do not attend school or leave school before acquiring functional literacy.

2. The adoption of a standard language purely on the historical accident that a certain dialect had a headstart over the other dialects of having been written, could cause speakers of other dialects to resist or reject it.

3. It seems odd artificially to constrain speakers of a dialect of the same language not to use certain forms of speech under certain conditions.

4. The concept of "standard educated speech" does not help much without an indication of the characteristics of such speech. Except for some distinct phonological, tonological and vocabulary characteristics of the various dialects, the Zulu language is fairly homogeneous. The different dialects can be accommodated in a single grammar with fairly identical syntactic and morphological features.

Economic, social, educational, religious and political influences during the past century, and more particularly during the past fifty years, have drawn Kwa-Zulu-Natal citizens irrevocably closer together. This provides a platform on which a more dynamic and uniform language can flourish.

The foregoing statement has serious implications for dictionary makers, writers and other linguists. The promotion of multilingualism enshrined in the

South African Constitution offers scope for the resurgence of dialects in addition to the acquisition of other languages. A vigorous development of multilingualism has the capacity to enrich the existing written African languages.

There is a need for a more intensive study of the cultures of speakers of the various Zulu dialects and a recording of lexical items of these dialects in dictionaries. Forms drawn from dialects could be entered alongside conventional Zulu forms as variants and indicated accordingly. The peculiarities of the various dialects could be discussed in the introductory section of the dictionary.

3.1.2   Use of a Corpus

A corpus is a useful tool in dictionary making. It enables the lexicographer to determine the various usages and frequencies of different words with the least effort. It is generally assumed that the corpus contains all or most words in a language together with their different usages. It is, however, doubtful if the use of a corpus is equally effective in a language like Zulu with a relatively small literature.

Due to market forces, the output of Zulu literature has been largely confined to educational and religious material. Traditional customs and beliefs are not fully documented in the language itself. The work environment, trade and industry are still dominated by English. Zulu is not used for recording in the law-courts. Advertisements, bill-boards and road signs are still found in English even in predominantly Black areas. Politicians still address their constituents in English in largely monolingual constituencies. The electronic media also tend to use Zulu interspersed with English terms even where such terms are available in Zulu. The language is still under siege. A corpus should be used to supplement the usages obtained through fieldwork in the compilation of a dictionary.

3.2.1   Figurative Use of Language

The use of metaphor, simile, hyperbole and other figures of speech presents some difficulty. Understanding figures of speech requires some knowledge of the history, mythology, legends, or culture of a people.

As in the case of the examination of kinship terms, some of the aspects of culture that may throw light upon certain proverbs and idioms will be discussed briefly.

(a) Parts of a Hut
umsamo (inside back of the hut), *umnyango* (door)
Lithatha osemsamo limphose emnyango, lithathe osemnyango limphose emsamo (Lightning moves a person from the inside back of a hut (favourable place) to the door (unfavourable place) and *vice versa*, i.e. the first shall be last and the last first).

iziko (hearth)
Ikati lilala eziko (The cat lies at the hearth, i.e. there is famine because there is nothing to cook).

(b) Social Organisation
Recognition of status
Umlomo ongathethi manga (the King, i.e. one whose mouth utters no lies/falsehoods).

(c) Communal Activities
Akudlulwa ngendlu yakhiwa (It is a good practice to give others a hand, i.e. one must not forbear to do communal work).

(d) Artifacts
igula (calabash, milk-vessel)
Igula lendlebe aligcwali (The gourd of the ear does not get filled, i.e. the ear does not tire of hearing).
ithunga (wooden milking-pail)
Selidumela emasumpeni (The milking-pail is resounding near the handles, i.e. the matter is coming to fruition).
umkhombe (meat tray)
Imikhombe iyenanana (Meat trays are exchanged, i.e. hospitality is expressed between homes).

In a study of the colour names of *Sanga-Nguni* cattle, Poland (1996) shows the important part played by imagery. She views the similarities established between the different colours of cattle with other objects, e.g. birds, vegetation, etc. as iterative rather than simply comparative or substitutive. This view appears to have a high validity in the use of imagery in cultural material. It is important to know how to deal with this in the compilation of a dictionary.

There seems to be three ways of dealing with this, viz. by

1. identifying a metaphor, simile or other figure of speech,
2. establishing the context in which it occurs, and
3. determining the way it could be entered in a dictionary.

— **Some Characteristics of Proverbs and Idioms**

1. Proverbs occur as independent sentences in speech. They are found individually or interspersed among other sentences, much depending on the context. The predicate of the proverb comprises a simple, complex or compound verb or a copulative (in the positive or negative conjugation). It can also be used in various tenses. Some proverbs are commonly used with the subject of the sentence unexpressed. Similarly, the subject may occur in post-verbal position in the imperative.

Examples:
Umendo awuthunyelwa gundane (A girl cannot spy on conditions that will bring her happiness in marriage, i.e. they are unpredictable).
Sobohla Manyosi (Your prosperity will decline, like Manyosi's).
Ayikhulunyelwa eziko (Do not discuss your plans or secrets in public).

2. An idiom generally constitutes a clause, a word or a phrase. When used, the idiom is usually integrated into a sentence. It does not exist independently. For purposes of linguistic discussion or classification, the idiom with a verbal stem is used in the infinitive form.

Examples:
ukuphuma umuzi (to move away from the main homestead to establish one's own — said of a married son), e.g. *UGebhuza usephume umuzi.*
ukudliwa yinkatho (to be chosen by means of casting lots), e.g. *UJabulile udliwe inkatho* (Jabulile was chosen).
ukudlala umkhosi (to celebrate the first fruits ceremony), e.g. *Umkhosi uzodlalwa ngolwesine* (The first fruits ceremony will be held on Thursday).
ukudlala ngegeja kuziliwe (to do something held in taboo during mourning), e.g. *Le ndoda idlala ngegeja kuziliwe* (This man is impudent).
selimathunzi (towards sunset), e.g. *Bafike selimathunzi* (They arrived towards sunset).
abaphansi (the departed spirits), e.g. *Bayasibona abaphansi* (The departed spirits are with us).
kwelenyoni (at the highest point of a tree), e.g. *Sebekhuphuke baze bafika kwelenyoni* (They have climbed to the highest point of the tree).

— **Context of Proverbs and Idioms**

Proverbs and idioms occur in various contexts. They are also found in elements of culture.

— **Entry of Proverbs and Idioms in a Dictionary**

The Proverb

The proverb occurs as a sentence in Zulu. The subject of the proverb may be expressed or deleted.

Historically, the subject of the sentence (i.e. the proverb) is collocated with a headword in the dictionary. However, where such a subject is not expressed or where it occurs after the predicate, the latter is collocated with a headword in the dictionary.

Examples:
Unyawo alunampumulo (The foot has no rest/does not suspect/smell trouble, i.e. one's travels may take one to people one has ill-treated).
Lala lulaza bakwengule (Let the cream set so that it can be creamed off — said of someone taking advantage of another).
Ithi ingalamba iphenduke inkentshane (A starving dog can go wild, i.e. need can change one's behaviour).

The entry of proverbs is usually arbitrary and inconsistent in Zulu dictionaries.

Proposal
The proverb can be entered by collocating the predicate with a headword in the dictionary.

The Idiom

Zulu idioms with a verbal stem are currently entered in the infinitive form in dictionaries. Other forms are entered without change.

If the infinitive is followed by a noun, the verb-stem of the infinitive is collocated with a headword. If the verb-stem occurs alone without a following noun, its stem is collocated with a headword in the dictionary. An idiom comprising a single noun has its stem collocated with a headword, and one with one or more following nouns has the stem of the first noun collocated with a headword in the dictionary.

Examples:
ukudla umuntu izithende (to chew up a person's heels, i.e. to backbite)
ukuthela ngehlazo (to pour shame onto someone, i.e. to disgrace someone)
ukuvaleka (to be broke)
umthakathi wendaba (a witch of a story, i.e. something on everyone's lips)
uwafawafa (a stiff contest)

The entry of an idiom in the dictionary purely on the basis of the linear order of the words contained in the idiom is unreliable. It leads to unnecessary repetition when idioms entered under the infinitive also occur under the definition of the complement of the infinitive. This is evidence of uncertainty on the part of the lexicographers.

Proposals
1. The repetition of the definition of idioms in a dictionary can be avoided by entering idioms comprising an infinitive and a complement under the noun (or first noun) of the complement.

Examples:
indlebe (as headword)
(a) *isitho somzimba sokuzwa* (an ear)
(b) *ukuluma indlebe (ssh)*, izwi noma umsindo wokuvusa umuntu
       ngento ebucayi ephuthumayo (to inform someone of impend-
       ing trouble/danger)
       ukungabi nandlebe (ssh), ukungalaleli (insolence)
       ukushaywa yindlebe (ssh), ukuzwa okuthile kube kungakhu-
       lunywa' nawe (to overhear something)
       ukubeka indlebe ngomuntu (ssh), ukwalusa umuntu (to spy on
       someone)

imali (as headword)
(a) *uhlamvu noma amaphepha okuthenga* (money)
(b) *ukuchitha imali (ssh)*, ukuhlaphaza imali (to waste money)
       ukwenza imali (ssh), ukungenisa imali (to make money)
       ukudla imali (ssh), ukuchithiza imali (to waste money)
       insumansumane imali yamakhanda (ssh), indida yentela (bewilder-
       ment, head/poll tax)

2. Idioms occurring as single words in the infinitive form can be collo-
    cated with a headword in the dictionary.

    Examples:
    ukubaleka (as headword)
    (a) *ukusuka ngejubane ugwema ingozi* (to run)
    (b) *ukweqa kwentombazane iyogana* (to marry a man by fleeing to his
           home to claim lobolo)

    ukuvaleka (as headword)
    (a) *ukuvimbeka*; *ukusitheka* (to be closed)
    (b) *ukuphelelwa yimali*; *ukushona* (to be hard up)

    ukuqina (as headword)
    (a) *ukuba lukhuni* (to be strong)
    (b) *ukujiya noma ukuhlangana kokuthile kube luzica* (to be firm)
    (c) *ukuphapha noma ukuba nobuqha* (to be precocious)

3. Nouns used as idioms can be collocated with a headword in the dic-
    tionary.

    Examples:
    inkomo (as headword)
    (a) *isilwane esikhulu esifuywayo esinezimpondo nezi* (a cow, a beast)
    (b) *umuntu ohluleka ukudlala ibhola* (a poor football player)

    ingwenya (as headword)
    (a) *isilwane esinomzimba omaholoholo esidla ezinye izilwane esihlala*

emanzini (a crocodile)
(b) *umuntu oyisichwensi oyisichokolozi* (a bully)
(c) *insimbi eqoba amatshe* (a stone crusher)

inkabi (as headword)
(a) *inkomo yeduna etheniwe* (an ox)
(b) *umuntu oqashelwe ukubulala abantu kwezemibango* (a hired execu-
    tioner in violent conflict)

3.2.2 Euphemism

Euphemism is used to ameliorate the force or impact of something unpleasant
said to someone. It is a form of avoidance of using a word or speech that is
harsh or unpleasant.

Examples:
amasimba/uthuvi: indle (faeces)
ukushona/ukuhamba emhlabeni: ukufa (to die)
ukuzala: ukubeletha (to give birth)
ukumitha: ukukhulelwa (to be pregnant)
ukuphunza: ukuphuphuma kwesisu esilwaneni (to have a miscarriage
    — said of an animal)
ukubhebha: ukulala owesifazane (to have sexual intercourse)

The equivalents of the abusive words could be used as definitions of the words
in the dictionary.

3.2.3 Taboo

Vulgarisms are common in Zulu. The use of some anatomical terms, particu-
larly those referring to private parts of human beings or sexual organs of ani-
mals are avoided in polite speech. Vulgarisms also include the use of swear-
words or the description of acts which are not acceptable in polite speech.

The question is whether or not to enter vulgarisms in a dictionary. As vul-
garisms are part and parcel of the language, they should be included in a dic-
tionary. It is important, however, to describe vulgarisms in a way that is inof-
fensive and that does not reinforce the hurt or cruelty with which they are gen-
erally associated.

Examples:
ubolo/ufenu: iphobana (penis)
umnqundu: inhlangano yamathambo emilenze (crotch)
umsunu: isitho sobulili sowesifazane (clitoris)

The use of anatomical terms by themselves is innocuous. The sting of a vulgar-
ism appears to be the genitive noun following it. The latter seems to equate the

person addressed with the preceding noun. To equate a person with something so private and personal is the height of insult and disrespect.

Examples:
msunu kanyoko: ukubiza umuntu ngesitho sobulili sikanina (your mother's clitoris, i.e. you are the clitoris of your mother)
golo likadadewenu: ukubiza umuntu ngesitho sobulili sikadadewabo (your sister's vulva, i.e. you are your sister's vulva)
mdidi wakho: ukubiza umuntu ngendunu yakhe (you are your own anus)

Lexical borrowing of swear-words is also found in Zulu. These are, however, used without the accompanying genitive form.

3.2.4 *Hlonipha*

Much has been written about *hlonipha* used by women (see among others Mncube (1950)). *Hlonipha* is a form of avoidance of using words resembling the names of male persons and others of the extended family of a married woman's in-laws. Except forms of behaviour that avoid eye contact between the woman and those she *hlonipha*s, forms of dress that cover parts of her body such as the breasts and the head, and the adoption of certain forms of posture, such avoidance is also expressed through the substitution of certain sounds contained in words that resemble the names which refer to her in-laws.

Examples:
amada: amanzi (water)
ukucuya: ukufuya (to rear stock)
isigosho: isivalo (door)
umjaqanga: umbhaqanga (thick porridge)

It is not clear whether certain *hlonipha* words are used across the entire Zulu-speaking region. It is, however, common to find *hlonipha* words which occur in certain families.

Hlonipha words should specially be marked in a dictionary in order to avoid confusing them with the rest of the vocabulary.

4.   Conclusion

Culture is indispensable in the description of a language in general and in the compilation of a dictionary in particular. It is, however, important not to confine culture to traditional culture. Culture is dynamic, so is language. The new world culture brought about by globalisation, for example, can no longer be ignored by linguists. The evidence of lexical borrowing is testimony of ongoing changes and adaptation of the Zulu language to new situations.

The new political dispensation is likely to accelerate multilingualism and lower the barriers and inhibitions about other languages. Dialects are likely to blossom and groups of similar languages may come closer and closer together. The over-sensitivity about the differences between dialects may progressively diminish. A good example of this is the benefit of the grouping together of the Nguni languages on the national television network. Speakers of Nguni languages have not clamoured for separate services. This sense of pragmatism needs to be complemented by the implementation of the work of authors, terminographers and language planners, especially the rationalisation of the orthographies of cognate African languages.

Lexicographers have an important task to perform in the preservation and transfer of cultural aspects contained in language through explanatory and translation dictionaries.

Endnotes

1. Cf. *ukuthwala* in *ukuthwala intombi* (to marry a girl by abducting her, as opposed to *ukuthwala* (to carry) or *ukuthwala* (to seek a fortune through magical means).
2. Cf. the Lord's Prayer which uses the genitive form *wethu* instead of *ubaba*.
3. Note that the Afrikaans word *broer* is a kinship term of relationship but *ubhululu* is not.

References

Fischer, A. 1985. *English/Xhosa Dictionary*. Cape Town: Oxford University Press.

Hamel, L. 1965. *English–Southern Sotho Dictionary*. Mazenod: The Catholic Centre.

Hannan, M. 1974. *Standard Shona Dictionary*. Harare: The College Press.

Herskovits, M.J. 1960. *Man and his Works*. New York: A. Knopf.

Kubheka, I.S. 1979. *A Preliminary Survey of Zulu Dialects in Natal and Zululand*. Unpublished M.A. Thesis. Pietermaritzburg: University of Natal.

Mabille, A. and H. Dieterlen. 1983. *Southern Sotho–English Dictionary*. Reclassified, revised and enlarged by R.A. Paroz. Morija: Sesuto Book Depot.

Mncube, F.S. 1950. *Hlonipha Language as Found among Zulu-Xhosa Women*. Unpublished M.A. Thesis. Johannesburg: University of the Witwatersrand.

Mokgokong, P.C. 1966. *A Dialect-Geographical Survey of the Phonology of the Northern-Sotho Area*. Unpublished Ph.D. Thesis. Pretoria: University of South Africa.

Ndlovu, R.S. 1963. *A Short Study of Slang*. Unpublished M.A. Thesis. Durban: University of Natal.

Nkabinde, A.C. 1982. *Isichazamazwi 1.* Pietermaritzburg: Shuter and Shooter.

Nkabinde, A.C. 1985. *Isichazamazwi 2.* Cape Town: Oxford University Press.

Nyembezi, S. 1992. *Isichazimazwi Sanamuhla Nangomuso*. Pietermaritzburg: Reach-out Publishers.

Pahl, H.W., A.M. Pienaar and T.A. Ndungane. 1989. *The Greater Dictionary of Xhosa. Volume 3. Q to Z*. Alice: University of Fort Hare.

Poland, M. 1996. *UChibidolo: The Abundant Herds. A Descriptive Study of the Sanga-Nguni Cattle of the Zulus; With Special Reference to Colour-Pattern Terminology and Naming Practice*. Unpublished Ph.D. Thesis. Pietermaritzburg: University of Natal.

Sapir, E. 1921. *Language*. New York: Harcourt, Brace.

Shabangu, S.S. 1987. *Isichazamazwi Samagama Amqondo Ofanayo*. Pietermaritzburg: Shuter and Shooter.

Snyman, H.H., J.S. Shole and J.C. le Roux. 1990. *Setswana–English–Afrikaans Dictionary/Woordeboek*. Pretoria: Via Afrika.

Suter, F. n.d. *Lessons in Zulu*. London: McDougall's Education Co.

Thipa, H.M. 1989. *The Difference Between Rural and Urban Xhosa Varieties: A Sociolinguistic Study*. Unpublished D.Phil. Thesis. Pietermaritzburg: University of Natal.

Wentzel, P.J. and T.W. Muloiwa. 1982. *Venda–Afrikaans–English*. Pretoria: University of South Africa.

White Fathers. 1954. *Bemba–English Dictionary*. Cape Town: Longmans, Green and Co. for Northern Rhodesia and Nyasaland Joint Publications Bureau.

Ziervogel, D. and P.C. Mokgokong. 1975. *Comprehensive Northern Sotho Dictionary*. Pretoria: J.L. van Schaik.

Compiling a Dictionary of an Unwritten Language: A Non-corpus-based Approach

Jacques van Keymeulen, *Department of Dutch Linguistics, Ghent University, Belgium (jacques.vankeymeulen@ugent.be)*

Abstract: In this article an account is given of the experience in fieldwork by the Dictionary of the Flemish Dialects (Woordenboek van de Vlaamse Dialecten, WVD), Ghent University, Belgium. The focus is on the practical aspects with regard to methods of lexicographic fieldwork. It is maintained that the analysis of 'metalinguistical conversations' with groups of respondents in which their lexicographic competence is explored, is a suitable way of collecting lexicographic data. Fieldwork by correspondence (questionnaires) can amplify and verify the data collected through interviews.

Keywords: LEXICOGRAPHY, UNWRITTEN LANGUAGE, DIALECT, REGIONAL DICTIONARY, FIELDWORK, GENERAL VOCABULARY, DUTCH, SOUTHERN DUTCH, FLEMISH, BRABANT DIALECT, LIMBURG DIALECT, THE NETHERLANDS, BELGIUM, SYSTEMATIC ARRANGEMENT, METHODOLOGY, QUESTIONNAIRE, INTERVIEW, WORD ATLAS, LANGUAGE VARIATION

Samenvatting: Een woordenboek samenstellen van een ongeschreven taal: een niet-corpusgebaseerde benadering. In dit artikel wordt een verslag gegeven van de ervaringen met veldwerk van het Woordenboek van de Vlaamse Dialecten (WVD), Universiteit Gent, België. De klemtoon ligt op de praktische aspecten met betrekking tot lexicografische veldwerkmethodes. Er wordt verdedigd dat de analyse van 'metalinguistische conversaties' met informantengroepen, waarbij de lexicale competentie wordt geëxploreerd, een goede manier is om lexicografische gegevens te verzamelen. Veldwerk per correspondentie (vragenlijsten) kan de gegevens die mondeling verzameld zijn, vervolledigen en controleren.

Sleutelwoorden: LEXICOGRAFIE, ONGESCHREVEN TAAL, DIALECT, REGIONAAL WOORDENBOEK, VELDWERK, ALGEMENE WOORDENSCHAT, NEDERLANDS, ZUID-NEDERLANDS, VLAAMS, BRABANTS, LIMBURGS, NEDERLAND, BELGIË, SYSTEMATISCHE ORDENING, METHODOLOGIE, VRAGENLIJST, INTERVIEW, WOORDATLAS, TAALVARIATIE

Introduction

Unwritten languages exist everywhere. Many standard languages in Europe span a wide variety of purely oral — unwritten — dialects. The 'traditional'

dialects are not regional 'accents' of a standard language, but are ancient language varieties that for historical, political and sociological reasons did not reach the status of a 'cultural language'; they are linguistically characterized by the presence of geographical differentiation. They are sometimes not mutually intelligible and do not have, with a few exceptions, a tradition of written texts.

The aim of this article is to share the experiences with regard to the methods of collecting dialect words/dialect meanings used by the editorial board of the Dictionary of the Flemish Dialects (Woordenboek van de Vlaamse Dialecten, WVD) at Ghent University, Belgium. The WVD is also counselling local dictionary projects, for which a manual has been written (Van Keymeulen 2003a). We hope that some of our experiences will be of interest to other projects aiming at making dictionaries for oral languages. The focus of the article is on the practical aspects of the methodology of data collection, more particularly on the fieldwork to be performed (par. 1). We will briefly comment on the presentation of the macro- and microstructure of the WVD and its two cognate projects (par. 2). In par. 3 we present the conclusions.

There is a vast literature on the methodology of fieldwork, especially for sciences such as anthropology, ethnology, sociology and oral history. For a theoretical background with regard to participant observation, interview techniques, questionnaires, etc., we refer the reader to the relevant handbooks. In what follows, we restrict ourselves to the methods used by the WVD, the result of 30 years of experience.

Flemish dialects

Flemish dialects are spoken in the south-western corner of the Dutch language area, i.e. two provinces of Dutch-speaking Belgium (West- and East-Flanders), plus two smaller areas in France (French-Flanders) and the Netherlands (Zealand-Flanders). The traditional Flemish dialect vocabulary is rapidly disappearing under the pressure of the Dutch standard language, the official language of the Flemings. In France, where the dialect is not 'roofed' by Dutch, but by French, its disappearance is imminent. It should be noted that the term 'Flemish' is used here in its restricted — dialectological — sense. In popular speech, both Flemings and foreigners use the term to denote the 'Belgian' accent (and other particularities) of the Dutch standard language.

Dictionaries of the Southern Dutch Dialects

The set-up of the WVD, which started in 1972, is largely parallel with that of two cognate and older projects: the Dictionary of the Brabant Dialects (Woordenboek van de Brabantse Dialecten, WBD) and the Dictionary of the Limburg Dialects (Woordenboek van de Limburgse Dialecten, WLD). Both dictionaries started in 1960 at the Catholic University of Nijmegen in the Netherlands; from 1991 onwards, an additional editorial board for WBD and WLD has been estab-

lished at the Catholic University of Louvain in Belgium. The three dictionaries together cover the southern Dutch language area.

The methodology of the dictionaries, which was initially developed by prof. A.A. Weijnen (at the Catholic University of Nijmegen), has already been described in Kruijsen en Van Keymeulen (1997), so I limit myself to a few major points. All three dictionaries are arranged systematically (with alphabetical indexes) and combine a dictionary with a word atlas, since for every dictionary the focus is not on one dialect but on a group of related dialects. The dictionaries are divided into three separate parts: I. Agricultural vocabulary; II. Technical vocabularies; III. General vocabulary (i.e. the vocabulary not restricted to professional activities). Every fascicle of the dictionary is devoted to a certain conceptual field (e.g. 'housing', 'wild animals', etc.) and consists of a series of concepts related to it (onomasiological arrangement). For every concept the heteronymy (the different lexemes which can be used to refer to the concept) in the different dialects of the area under investigation is presented, together with general indications as to frequency and location (details with regard to phonetics and location are kept in an automated database). (For further information, see Kruijsen and Van Keymeulen 1997 and the website of the three projects at <http://www.flwi.ugent.be/dialect>.)

The aim of the three dictionaries is to document the vocabulary of the oldest layer of the traditional dialects and its geographical patterns. Since the traditional dialects are not written, the bulk of the data is taken from interviews or consists of answers to questionnaires filled in by hundreds of volunteering elderly dialect speakers. To these data, material taken from older investigations or from written sources such as older dictionaries is added.

The metalexicographical questions with regard to aim ('Why?') and target users ('For whom?') are answered as follows. The ultimate aim is to document a disappearing vocabulary and its lexico-geographical patterns for scientific reasons. This geographical aim explains the neglect of collocations (expressions, proverbs, etc.). The dictionaries are strictly descriptive with the future scientist as target user. Practical usage of the dictionaries is considered marginal. Although the dictionaries enjoy a wide popular support today, in future the scientific user will come more to the fore as the dialect vocabulary becomes extinct. The dictionaries provide data for the future study of the lexical history of the Dutch language; they are complementary to the dictionaries of both the historical and present-day periods of Dutch, which are based on written text corpora. The metalanguage of the three dictionaries is standard Dutch.

In what follows, we will not dwell on all the possible answers to the questions 'Why?' and 'For whom?', which determine both the macro- and microstructure of any dictionary, since they may vary widely according to the different sociological and linguistic contexts in which a dictionary has to function. In case a purely oral language has to be turned into an official and written language, a 'norm' has to be chosen and codified and an orthography has to be agreed upon. The problems with regard to codifying a dominant dialect, of

which all the other dialects are to be considered as 'variants', may engender heated discussions. Solutions to questions pertaining to codification are in fact too important to be left to lexicographers. It is the political authorities and the language policy makers/financing bodies that should provide the metalexicographical answers in this respect. The introduction of standard Dutch in the north of Belgium, incidentally, provides good examples of linguistic frustration, linguistic struggle, linguistic planning, etc. (see Van Keymeulen 2003b).

1.   Data collection

By definition, it is hardly possible to make a dictionary of an *oral* language on the basis of a text corpus. For some 'unwritten' languages or language varieties texts are available, but mostly they are scarce, if not absent altogether. Although it is certainly worthwhile always to collect all the available texts, a corpus-based approach is not the best way to success, given the fact that a collection of texts will at best form an unbalanced and unrepresentative corpus. Any bit of lexical information, however, taken from whatever type of source, may prove interesting when it comes to preparing the fieldwork. As we shall see later, fieldwork needs an input in order to maximize its succes.

In what follows, we focus on the different stages of the lexicographic fieldwork and on the way the five key questions pertaining to data collection may be answered. These questions are: 'What?', 'Who?', 'Where?', 'How?', and 'How much?'.

1.1   What?

The question 'What to collect?' pertains to the macro- and microstructure of the dictionary: what words/collocations are to be collected, and what information about the words has to be gathered? The answers to the 'What?' are very central ones, which should, after careful reflection, be answered in the very first stage of the project.

Oral language traditions may be 'roofed' by a linguistically related standard language, as is the case for all dialects both in the Netherlands and Flanders. This has the effect that the non-standard lexicon coincides partly with the lexicon of the standard language: some differences are restricted to phonology, while others are lexical proper. In a case like this, one has to decide between a confrontative and a contrastive dictionary (the terms are Wiegand's). In a confrontative dictionary, the aim is to present the totality of the lexicon of the non-standard language; in a contrastive dictionary the focus is on the lexical and semantic differences (neglecting the purely phonological ones) between dialect and standard language.

Every language varies both geographically and socially (age, class, level of education and gender being the typical parameters). A clear position has to be taken with regard to the inclusion of the horizontal (geographical) and vertical

(social) variation. Another important decision pertains to the inclusion or exclusion of terminologies of professions or other socially restricted domains. Some occupations, e.g. farming, may be so omnipresent in the culture that the vocabulary accompanying it can be considered essential for the language as a whole, whereas other, more restricted, terminologies, e.g. the vocabulary of the tailor or the smith, may be excluded. In principle, all variables with regard to linguistic variation should be kept uniform, and if not, the words/meanings should be labelled in the dictionary.

Finally, decisions have to be made with regard to the microstructure: What elements of e.g. the phonology, morphology and syntactical valency of the words have to be collected? What labels should be introduced?

Generally speaking, we advise the lexicographer not to undertake more than can be handled. It is better to succeed in reaching a modest aim, than to fail trying to adapt an unmanagable mass of data. Most local dialect dictionaries in Flanders, for instance, are of the contrastive type, and restrict themselves to the 'general vocabulary' (excluding dialectal terminologies) of the traditional dialect of the oldest generation. Their microstructure mostly contains only pronunciation, meaning, collocations and example sentences. The three regional dialect dictionaries mentioned above arrange concepts, together with the accompanying dialectal heteronymy in the area of investigation. Only the location and pronunciation of the lexical items are accounted for.

The answers to the question 'What?' determine the answers to the question 'Who?'

1.2   Who?

The answers to the question 'What?' determine the answers to the question 'Who?'

In the absence of a corpus of written texts, the word material for dictionaries of oral languages has to be collected by way of fieldwork. The quality of the data depends heavily on the choice of the volunteering respondents, the profile of whom is determined by the answers to the question 'What?'. Two types of volunteers may be distinguished: intermediaries and dialect respondents proper. The task of the intermediary is to look for suitable dialect respondents (for a specific subject) and to facilitate the fieldwork by introducing the researcher to potential collaborators within the local community.

An intermediary should meet the following requirements:

— take an interest in the dictionary project;

— be highly respected in the local community; and

— know a large number of people at all levels of society.

In practice, intermediaries can be found amongst the clergy, (school)teachers, school directors, chairpersons of organisations, etc. It is these people who have

the necessary educational background to understand the importance of the dictionary project and who may be willing to introduce the researcher to the local community. In the case of the WVD, a campaign directed towards parish priests to find respondents for agrarian vocabulary proved to be highly productive.

A good relationship with the intermediary is very important since one has regularly to revert to him/her. Lexical knowledge, as it happens, is distributed unevenly in the language community, and one has to recruit different respondents for different topics: farmers know more about the words for wild plants than other people; women know more about the kitchen than men, etc. The WVD regularly convenes with the intermediaries/respondents in order to keep them motivated by communicating interim results and by organizing social activities.

The profile of the ideal respondent should be established very carefully, taking into account the options with regard to the macrostructure of the dictionary. A respondent should meet a list of both subjective and objective requirements.

The subjective requirements amount to:

— an interest in the dictionary project;

— intelligence;

— willingness;

— communicability; and

— suitability for the fieldwork according to his/her own opinion.

The objective requirements usually have to do with sociological and geographical parameters, and with the subject matter under investigation. In the case of a dictionary of a traditional local dialect in danger of disappearing under the pressure of a standard language, the objective requirements to be met by the respondents are:

— being of an advanced age;

— being of a lower social class;

— having a low level of schooling;

— having been locality-bound, i.e.

  – having grown up/lived all his/her life in the same locality,

  – having exercised an occupation/a trade in that same locality,

  – having both parents come from that same locality, and

  – being married to someone of that same locality;

— being expert in the topic under investigation.

The question whether men or women are more suitable as respondents in general is a tricky one. It has been observed regularly that women are more sensitive to language change than men, hence they tend to be more influenced by standard varieties. On the other hand, they are often more home-bound, which makes them linguistically more conservative. In practice, many dictionaries of traditional local dialects rely on Trudgill's NORM-respondent, the Non-educated Old Rural Male.

In some cases, a conflict may arise between the subjective and the objective requirements. The editorial board of the WVD experienced the paradox that some very suitable dialect speakers were unfit to cooperate with the project because they could not understand its purpose.

Respondents should be tested. In the case of the WVD, the statements of respondents who could not spontaneously produce the dialect word for 'butterfly' or 'hedgehog', but used the 'dialectized' standard Dutch lexeme instead, were regarded as doubtful. A further requirement is that the biographical data of the respondent and other relevant information pertaining to his profile should be noted down.

1.3   Where?

The geographical scope of a dictionary may vary from a single locality or a region to a whole language community. If the aim of the dictionary is to codify a vocabulary, a choice may have to be made as to the socio-economical dominant region. Regional dictionaries of the WVD-type only aim to describe a geographically differentiated dialect landscape. The word maps, however, indicated that the Brabant dialect is dominant in northern Belgium. Brabantine words turned up in both Flemish and Limburg dialects; the opposite was never the case.

The editorial board of the WVD could rely on a long tradition of dialect investigation and a fairly good knowledge of the lexical geography of the dialects. Although the heteronymy for every concept has its own geographical pattern (the pattern of the dialect words for 'butterfly', for instance, is totally different from the pattern of the words for 'duck'), it was possible, on the basis of the comparison of many hundreds of word maps, to detect recurrent patterns. A number of 'lexical areas' could be distinguished, i.e. areas which regularly have their 'own' word for a given concept. In every lexical area, one rural 'reference point' for an in-depth interview with a group of respondents was selected. To these, six urban points were added to account for the major urban dialects. A number of interviews were conducted by a researcher to obtain sound (orally collected!) lexical information for every lexical area (the minimal goal). In a later phase, questionnaires were sent out to respondents in as many localities as possible to make it possible to draw word maps and to detect the course of the isolexes (the maximal goal). Since one has to depend on volunteers, the maximal goal is hard to meet.

1.4   How?

The lexicon of an unwritten language can be collected by several methods:

— self-observation;

— observation of spontaneous speech;

— recording and transcribing 'free speech';

— purposive systematic fieldwork, i.e.

   – recording and transcribing 'free speech',

   – oral investigation (interviews/questionnaires), and

   – investigation by correspondence (questionnaires).

Many dictionary projects derive their lexical information from a variety of sources. Some methods are complementary; the results of one method may be used as the input for another. The methods differ firstly with respect to the relative validity of the data: lexemes taken down from spontaneous speech are thought to be of higher quality than those collected purposively. Secondly, the methods differ with regard to the possibility of being carried out systematically. Thirdly, methods may vary according to the type of vocabulary one wants to collect. Some elements of the lexicon are more easily collected by purposive questioning than others, because the introspective capacity of a language user differs according to the different lexical types. The main difference in this respect is between the 'open' (e.g. substantives, adjectives, verbs) and the 'closed' word categories (e.g. abverbs, prepositions, conjunctions). A classic dichotomy, finally, is the distinction between direct and indirect methods, i.e. fieldwork conducted orally or by correspondence respectively. In our view, it is the presence of an intermediary, rather than the medium of communication, that makes a method direct or indirect.

In what follows, we will briefly discuss the various methods of data collection and point out the advantages and disadvantages of each of them.

1.4.1   Self-observation

The advantage of writing a dictionary of one's own language is self-evident. A word collection can indeed be expanded greatly when using one's own language competence. There is, however, a risk. A lexicographer may be influenced by the standard variety of the language — more than he is aware. A phonologically adapted borrowing from a standard variety, for instance, may have replaced the traditional dialect word in his linguistic memory. The question is whether the lexicographer's profile as 'respondent' is in accordance with metalexicographical considerations. However, data drawn from one's own language memory may always serve as an input for other methods.

The above-mentioned risk firstly pertains to word collecting. The lexicographer's language competence is of course invaluable when it comes to describing the meaning of the lexemes or adding other microstructural elements.

1.4.2   Observation of spontaneous speech

The word material can be extended by noting down overheard words or expressions. Many a lexicographer of a local dialect dictionary carries a notebook for this purpose. Since the word material is collected in real language situations, the advantage of this method is the high validity of the data. This advantage is even more important for the closed word categories, which are harder than the open ones to inventory by explicit questioning. The disadvantages of the method pertain to the unknown sociological profile of the speakers and the unsystematic nature of the method. It is, however, a good way to collect material for further fieldwork and it certainly is a good method to collect real and lively example sentences. In many dictionaries example sentences sometimes appear too 'construed', and sentences taken from spontaneous speech often add realism and liveliness to a dictionary.

1.4.3   Recording and transcribing 'free speech'

In an attempt to overcome the problem of the availability of few or no written texts, some lexicographers have tried creating the necessary texts themselves by making transcriptions of interviews with respondents. The method amounts to choosing a good respondent and making him/her talk as much as possible about a specific subject. The interview is then fully transcribed. On the basis of the written transcriptions of the language material provided by the respondent, corpus analysis procedures may be carried out.

Although the validity of the lexicographic data is very high, since it is derived from spontaneous speech, the method is seldom used because it is very time-consuming. Moreover, it is hard to collect a vocabulary systematically by this method only, unless the subject matter of the interviews is restricted and the interviews themselves are highly structured.

In the case of the WVD, the existing transcriptions of the 400 odd recordings of 'free' dialect speech collected within the framework of a project on dialect syntax, are as yet not used for lexicographic purposes, since there are more efficient ways for collecting vocabulary. For the open word categories, the advantages of the method were deemed not to counterbalance the effort. For the words of the closed categories, however, which are highly frequent and harder to collect by way of questions, the collection of transcriptions may be used as corpus.

In the next paragraph, we will present a fieldwork procedure which includes relatively unstructured interviews. Through this process, a collection

of recordings which may afterwards be used for a variety of purposes, can be made.

1.4.4   Systematic fieldwork: in search for concepts, words and meanings

In the absence of written texts, fieldwork is the only systematic way of collecting material for a dictionary of an unwritten language or language variety. In what follows, we will discuss a number of fieldwork procedures which may be of use both for a small-sized local dictionary and for a comprehensive geographically-oriented dictionary. We assume that little or nothing is known about the lexicon under investigation and we consider all data collected by other methods, as an input for more purposive actions.

The methods we have discussed so far, all share the characteristic that the vocabulary is collected unsystematically. None of these guarantees completeness. In what follows, we describe the fieldwork procedure carried out by the editorial board of the WVD. It aims at systematically collecting the lexicon of an unwritten language, i.e. exploring the vocabulary topic by topic. We will first pay attention to the 'classification of reality' (par. 1.4.4.1) and then proceed to the successive stages of the fieldwork procedure: preparation (par. 1.4.4.2), group interviews (par.1.4.4.3), and large-scale investigation by correspondence (par. 1.4.4.4.). In a separate paragraph (par. 1.4.4.5.), we discuss a number of question types.

1.4.4.1   A classification of 'reality'

The fieldwork is conducted systematically, i.e. subject by subject. For this purpose, human experience and the surrounding world have to be classified in a coherent system. It has to be stressed that this system is not a linguistic but an ontologic one, i.e. it is a classification of reality, not of the lexicon. The system only serves as framework for a thematic lexicographic investigation.

The classification used by the three editorial boards of the regional dialect dictionaries of southern Dutch (WBD, WLD, WVD) for the collection of the 'General vocabulary' is based on the 'classical' classification of Hallig–Von Wartburg (1952), adapted by Frissen (1981) and Van Keymeulen (1992), who was inspired by the classification of the WALD. The classification tries to account for what Weijnen and Van Bakel (1967: 40) called 'the concrete coherence of things in daily life' as experienced by the dialect-speaking community. A detailed discussion of the principles according to which the classification below has been made is not possible here. Generally speaking, man is placed at the centre of things, and reality is assigned to him in ever broadening circles: domestic life, society and the surrounding world. Departing from the four main sections, the classification forms an ever finer hierarchy of sections, indicating a particular subject matter (e.g. 4.2 'Plants and animals' > 4.2.2. 'Animals' > 4.2.2.3 'Reptiles and amphibians' > 4.2.2.3.2. 'Amphibians').

Classification

1. Man
    1.1 Man as a physical being (parts of the body, movements, …)
    1.2 Clothing
    1.3 Internal reality of man (emotions, character, …)

2. Domestic life
    2.1 The house
    2.2 Family life/phases of life (birth, childhood, marriage, …)
    2.3 Eating and drinking

3. Society
    3.1 Social behaviour
    3.2 School
    3.3 Religion
    3.4 Amusement, festivities, art

4. The world around us
    4.1 The material world (weather, …)
    4.2 Plants and animals
    4.3 The abstract world (time, space, form, …)

Since the classification is made with 'the concrete coherence of things in daily life' of the dialect-speaking community in mind, is it a classification of the *culture* rather than the lexicon of that community. The main framework of the classification has universal value, and adaptations with regard to specific cultural contexts are only needed at its lower levels.

In the case of the three dictionaries of the southern Dutch dialects, the classification has been adapted during fieldwork and has consequently been used for the presentation of the macrostructure, which is issued in a series of thematic fascicles, each of which is devoted to a specific subject. In our view, however, a systematic (theme by theme) exploration of the lexicon is also recommendable for alphabetically arranged dictionaries.

1.4.4.2 Preparing for the fieldwork

Preparing for the lexicographic fieldwork has two aspects: (a) studying the subject matter itself; and (b) collecting all lexical data already available concerning the subject matter.

In many cases, good descriptions of the daily-life culture of the language community under investigation are absent, and the lexicographer has to do the necessary ethnological and anthropological research partly or even wholly by himself (see par. 1.4.4.3). In the case of the WVD, a subject such as 'birds' proved to be relatively easy in this respect, since good ornithological handbooks are available. For other subjects, especially the traditional crafts and

trades (such as cooper, basket maker, thatcher, etc.), the documention gathered by the editorial board turned out to be the most extensive existent — often to the amazement of folklorists. An additional advantage of the systematic field-work method is that cooperation between ethnologists and lexicographers is possible and in fact desirable.

During the preparatory phase, the lexicographer should also collect the already available lexical data in written sources (texts if any, already existing dictionaries, scientific studies, etc.), which can afterwards be used as input for the interviews. In the preparatory phase there is also the opportunity to collect items which may stimulate the memory of the respondents (e.g. photographs, drawings, etc.).

The main aim of the preparatory phase which ideally can be restricted to reading specialist (ethnological) literature, is to gather information for structuring the interviews for the fieldwork proper.

1.4.4.3 The group interview: in search of quality

When the preparatory phase is completed, one can turn to the fieldwork proper: the in-depth interviews with respondents. The interviews may be both ethnological and lexical in nature, depending on the amount of ethnological literature one was able to find.

It should be emphasized that from a scientific point of view the source of lexical/ethnological information in fieldwork is *not* the respondent, but the interview itself, which in its oral or written form is the result of an interaction between interviewer and interviewee. An interview is an 'artefact'. The field-worker should be aware that the lexical knowledge in the mind of the respondent may very well be stored and organised quite differently from what has been expected. Cicourel (1988: 907) warns us against 'questions that may not evoke an appropriate recall because of being phrased differently from the storage format of the informants' or respondents' experiences and their use of semantic memory'.

An efficient elicitation of lexical data (words or meanings) in metalinguistical discourse presupposes the establishment of a 'common ground' for meeting between interviewer and interviewee. This common ground should include a common understanding of the aims of the interview and a sound knowledge of the conceptions and motivations of the repondent. The advice of Cicourel (1988: 910) should be taken to heart: 'Elicitation procedures … should seek to maximize the activation of schemata or mental models that have been constructed by informants.'

In the first phase of the lexicographic fieldwork concerning a relatively unknown culture/language variety, the aim of the procedure is threefold: (a) the establishment of an inventory of concepts; (b) the establishment of an inventory of lexically relevant concepts; and (c) the establishment of the way the meaning of the lexemes is stored in the semantic memory of the respondents.

The first aim pertains to an analysis of the culture of the language community and its context, apart from linguistic considerations. In doing this, the lexicographer may be obliged to enter the field of other sciences, anthropology, ethnology, oral history, etc., depending on the level of the already existing knowledge about the culture of the language community. If this knowledge is insufficient, and if good descriptions of the culture are unavailable, the lexicographer may benefit from the existing handbooks on fieldwork in those sciences.

When an inventory of concepts (things, actions, attitudes, etc.) is completed, i.e. when there is enough *encyclopedic information* available, the fieldworker can proceed to the second aim: to try and obtain an insight into the structure of the semantic field related to the subject matter under investigation. In the case of 'birds', for instance, this in practice means an investigation into folk taxonomy, which may differ widely from scientific taxonomies in biology. Different kinds of birds may be denoted by the same lexeme; some species may not be named at all. In practice, the procedure amounts to investigating how the concepts are structured and what concepts are named.

The third aim is that the interview reveal the way meanings are stored in the semantic memory of the respondent. The meaning of substantives for concrete objects (e.g. *table*) is explained in a way different from an explanation of the meaning of an adjective denoting a character trait (e.g. *miserly*). There is a good chance that the respondent will describe the meaning of *table* by pointing to the function of the object, whereas *miserly* could be explained by giving an example of an action by a miser. Such semantic indications are very useful for framing explicit questions in the later stages of the fieldwork, and for the description of the meaning in the dictionary afterwards. Respondents usually describe the meaning of a word by using synonyms and antonyms, translations into another language, analytic definitions with semantic components, showing the thing or pointing to it, drawings, encyclopedic descriptions, prototypical examples, and descriptions of situations in which the word could be used.

For the sake of clarity, the three aims above have been presented as if they could be pursued consecutively. This is, however, seldom possible. A metalinguistical conversation derives its complexity from the fact that the interview constantly provides information of all three types. It is not an easy task to ensure that all necessary information, both encyclopedic and lexical, is elicited from the respondent when conducting the conversation.

In this phase, the actual fieldwork of the WVD is carried out by a researcher by way of a number of semi-structured interviews ('guided conversations') in the different reference points with groups of four to six respondents assembled by an intermediary, all resident in the same locality. The subject matter (e.g. 'birds') is divided into subthemes (on the basis of the above-mentioned classification) which are subsequently used to structure the conversation (e.g. concerning seabirds, birds of prey, forest birds, etc.). The respondents are invited to talk freely about the subthemes. In the course of the conversation, the

data from written sources collected in the preparatory phase are introduced by the fieldworker. In the case of the subject 'birds', for instance, the respondent groups were asked to comment on a series of pictures taken from a handbook on ornithology, and at the same time they were invited to describe the meaning of a number of words already known.

There is a wide variety of interview techniques, each of which has its own merits and limitations, and a multiplicity of uses (see Fontana and Frey 1994). Every technique, however, should take into account the cultural context in which the interview has to be conducted (e.g. politeness rules). It goes without saying that an interviewer should have the necessary personal qualities, of which flexibility and empathy are not the least important.

In our experience, the group interview is a very productive method to elicit ethnological/lexicographic data. Labov's 'observer's paradox' can be overcome, especially when the group is visited regularly. Interviewing a group has the following advantages:

—   the respondents feel self-confident, because the researcher is in the minority;

—   the respondents are used to speak the specific language variety with each other, hence the tendency to adapt to the language of the researcher is low; and

—   the respondents can correct each other, therefore the data are very comprehensive.

Merton et al. (1956) note three specific skills needed by a group interviewer:

—   he has to keep one person or a few persons from dominating the group;

—   he should encourage reluctant respondents to participate; and

—   he must obtain responses from the entire group to ensure the fullest possible coverage of the topic.

A group interview may yield data of a very high quality, provided the interviewer was able to manage the group dynamics efficiently. König (1982/1983: 477-478) distinguishes 'primary' and 'spontaneous' material in oral investigations. Primary material consists of the answers to questions; spontaneous material is the unintentional information given by a respondent when talking. The (partly) unstructured nature of the group interview guarantees that the fieldworker will acquire yet unknown information.

In case of a long-term dictionary project, the participants of the group, being interviewed regularly, obtain in-job training during the process and develop from mere 'respondents' into 'language counsellors', who eventually may even comment on proof versions of the dictionary text.

The WVD tries and conducts group interviews in each of the lexical areas. The data yielded by one group is used as an input for the next interview in another area. In the course of the fieldwork, the interviews become ever more structured, till it is possible to frame questions which can be used for fieldwork by correspondence.

Although the group interview should be loosely structured during this phase of the fieldwork, it is clear that targeted questions should be asked. In par. 1.4.4.5, we present a list of question types which can be used either in interviews or in questionnaires.

1.4.4.4 In search of quantity: questionnaires

On the basis of the results of the group interview(s), the lexicographer is in a position to prepare good questionnaires for large-scale investigation, and move to a more standardized form of research, in which comparable data are produced. Large-scale investigations are inevitable if one wants to document the geographical patterns in a lexicon.

In the case of the WVD, the intermediary takes the place of the researcher and conducts a highy structured interview with a respondent by means of a questionnaire. This is however an ideal situation; in many cases intermediary and respondent are the same person. Research by correspondence has the obvious disadvantage that the researcher himself is not present and that the situational setting of filling in the questionnaire cannot be supervised. Filling in questionnaires of course presupposes that the language has a written form, and that the intermediaries/respondents are capable of using it. Since the Flemish dialects are linguistically related to the Dutch standard, dialect words can be written down in a 'Dutchified' form (see par. 2.2 below).

1.4.4.5 Types of questions (interviews/questionnaires)

In lexicographic interviews/questionnaires, apart from the classical dichotomy between 'open' and 'closed' questions, five main types of questions can be distinguished: exploratory questions, encyclopedic questions, onomasiological questions, semasiological questions and task questions. In 'open' questions, the respondent is entirely free to answer as he pleases. 'Closed' questions, where the respondent has to choose between yes or no, or has to pick an answer out of a series of suggestions, are not frequently used in lexicographic fieldwork. (For a theoretical background on questioning and further reading see Atteslander 1988.)

Exploratory questions

Exploratory questions focus on the external world or on language; usually a certain subject is introduced and the respondent is invited to give as much

information as possible. This type of question is typical of the initial phase of fieldwork, and should be used in the group interview. In questionnaires, the technique is not very productive.

There are techniques which aim to stimulate the 'word-finding' process. When a certain subject is introduced, a number of questions may be asked (e.g. 'What is the opposite?' 'Are there different kinds of it?') in order to expand the word collection.

Encyclopedic questions

Encyclopedic questions in principle deal with the extra-linguistic world; these questions are not aimed at language but the answers will of course contain lexemes. Encyclopedic information should be gathered during the group interview. Lexicographic questions in questionnaires, however, may be masked as encyclopedic ones, as in: 'How do you feel when everything seems to turn round?'. A question like this leads to the answer 'dizzy'. Provided the question is well framed and unequivocal, it is a useful technique to avoid a complex onomasiological definition.

Onomasiological questions

Onomasiological questions describe/define/frame a concept and ask for a lexical expression. The technique is widely used for the open word classes. The definition of an onomasiological question can take a variety of forms: pictures, analytical definitions using semantic/encyclopedic components (e.g. the description of a prototypical example), or framing a context in the object language in which a lexical expression should be used (e.g. by way of sentences in which words should be filled in). The question-frame should lead the respondent unequivocally to the desired answer. The choice of the definition is dependent on the type of lexical item under investigation. Lexemes for concreta can be elicited by using analytical definitions of the concept; the more abstract or grammatical the meaning, the more one has to resort to other definition types, which try to imitate object language.

An old-fashioned onomasiological elicitation technique in Flemish dialect lexicography was to define a concept by one standard Dutch word. The respondent was expected to translate the word into his dialect. This technique, however, proved to be problematic since many respondents merely transposed the Dutch word phonologically in dialect, instead of translating it.

The question whether answers (i.e. lexemes) can be suggested, is still a matter of debate. This technique, which aims at stimulating the memory, is dangerous because it may provoke echo-answers. An all too cooperative respondent may select a lexeme he knows, but does not use — it is difficult to tell the two apart. The WVD, however, has been forced to suggest answers because for some subjects the traditional dialect words were no longer in active usage.

It proved that the oldest words could only still be recognized by the respondents; they were unable to remember them spontaneously any more. Suggestions can be presented in an open or a closed series; in the former, the respondent may add yet another word, in the latter, he has to select an answer.

Semasiological questions

In semasiological questions, a lexical expression is presented, and a 'meaning' (conceptual meaning, connotation, register or use) is asked for. From the respondent's point of view this technique is more demanding than the onomasiological one, where just one word may suffice in many cases. Semasiological questions require more thinking and noting down, which sometimes result in all too short or ill-conceived answers. This type of question calls for a good and reliable respondent.

Semasiological questionnaires presuppose onomasiological ones (one has to collect the lexemes first) and are often used to verify the semantic anticipations underlying onomasiological questions. The aim is to collect data about the meaning of a lexeme. The semantic information given by respondents may take a variety of shapes, which have to be interpreted by the researcher and translated afterwards into 'scientific' lexicographic definitions.

Suggestions as to the meaning or the usage of a lexeme, in an open or closed way, may streamline the answers.

Task questions

The aim of task questions is to collect words in context: collocations and sentences. The respondent is invited to produce expressions, proverbs or example sentences in which a given word is used. The dictionary of the dialects of Drenthe (a province in the north-east of the Netherlands), for instance, was compiled largely on the basis of a collection of example sentences produced by respondents. Cards with words on them were sent out to a large number of respondents, who were asked to illustrate the meaning of each word by constructing a sentence with it.

Cyclicity and feedback

Lexicographic fieldwork carried out in a relatively unknown language area should take advantage of the possibility of using cyclical questioning procedures. The answers to onomasiological questions can be used as an input for semasiological ones, which may provide new words and meanings. The alternation of onomasiological and semasiological points of view thus creates a cyclical questioning procedure until all words and meanings within a given conceptual field are accounted for.

De Schrijver and Prinsloo (2000) introduced the concept of 'simultaneous feedback' in lexicography: 'In a nutshell, this new methodology entails the

release of several small-scale parallel dictionaries which triggers feedback that is channeled back to the compilation process of the main dictionaries' (2000: 4). The concept comprises a procedure in which informal and formal criticism on interim versions of a dictionary is collected and used for compiling ever better dictionaries. The method incorporates the target users' needs at an early stage of the compilation of a dictionary.

In dialect lexicography, a variant of 'simultaneous feedback' is sometimes used in order to expand the word material, apart from gaining a better insight in the target users' needs. Many dialect dictionaries present interim dictionary articles to selected respondents for correction or elaboration. Sometimes these articles are published in periodicals on a regular basis, with a request for comments. In the case of the Supplement to the Dictionary of the Zealand Dialects (2003), for instance, interim versions of the whole dictionary were discussed, word by word, by three groups of respondents. The method proved to be highly effective: apart from many corrections, many new words, meanings, collocations, geographical locations, example sentences, etc. came to light.

1.4.4.6 The structure of the questionnaire

For lexicographic purposes, a thematic arrangement of the questionnaire is advisable because it stimulates the introspection and the memory of the respondent. A questionnaire also has to take the 'filling-in' psychology of the respondent into account. A questionnaire should not be too long; it should contain motivating questions (e.g. easy ones to start with) and should reassure the respondent by pointing out that questions can be skipped if too difficult. Lexemes that easily come to mind should be asked first — these lexemes are not necessarily the ones having the most general meaning. The word for 'very greedy', for instance, is more 'salient' than the word for 'greedy'; hence, the question about the word for 'very greedy' should precede the question about 'greedy', lest the salient words should be given when questions about words with more neutral meanings are asked.

In general, the researcher has to have empathic skills in order to foresee the reactions of the respondents. Filling in a questionnaire should be an agreeable task — although it demands an effort from the respondent's side — and the respondent should have the feeling that the researcher understands the difficulties he may encounter.

1.4.4.7 Fieldwork in practice

The different phases of the fieldwork for the open word classes as — ideally — conducted by the WVD can be summarized as follows:

(a)   Establishment of an inventory of concepts based on
     —   the classification of Hallig–Von Wartburg/Frissen/WALD/Van Keymeulen;

  — thematic re-arrangement of older dialect dictionaries;
  — systematic inventories of the Dutch standard language (e.g. Brouwers 1988); and
  — older questionnaires.

(b) Definition of a section in the systematic classification.

(c) Preparation of the respondent groups by sending a specimen questionnaire.

(d) Processing the answers by the members of the respondent groups in preparation of the group interviews.

(e) Interviews with the respondent groups, aiming at
  — enrichment of the inventory of concepts;
  — detection of the structuring principles/structure itself of the semantic field; and
  — collection of words and meanings.

(f) Processing the results of the group interviews.

(g) Establishment of the definitive questionnaire.

(h) Sending out the questionnaire to all the informants.

(i) Processing the answers of the investigation by correspondence.

1.5 How much?

The question 'How much?', i.e. the question about the saturation of a collection, is difficult to answer. There is, however, always a point where the continuation of fieldwork (more questions, more respondents, and more localities) is felt to be unproportional to the emergence of new data.

In the case of the WVD, the answer to 'How much?' is related to the answer to 'Where?'. The minimal requirement is met when the WVD has one good piece of lexical information for every lexical area (the lexical aim); the maximal objective is met when there is a good piece of lexical information for every locality, so that accurate lexical maps can be drawn (the geographical aim). In practice, the WVD is totally dependent on the cooperation of volunteers; the geographical aim is hardly ever completely met.

2. Presentation

The presentation of the word material in a dictionary, both macrostructurally and microstructurally, is of course dependent on an analysis of the users' needs. A dictionary should maximize the different usages that are possible in a userfriendly way.

2.1   Presentation of the macrostructure

Thematic versus alphabetical

The choice between a thematic or an alphabetical arrangement of the macro-structure of a dictionary is determined by its purpose. An alphabetical arrangement is of course evident if the dictionary is meant to be merely a tool for looking up the meaning/orthography, etc. of a word. If a dictionary is also meant to be a description of the culture of the language community, a thematic arrangement may be considered. In the latter case, the row of dictionary articles with regard to a certain subject matter can be seen as an inventory of the mental constructs (the meanings) accompanying part of the culture under investigation.

Whatever choice one makes, it is possible to combine the advantages of both arrangements by introducing an alphabetical index to a thematic dictionary or *vice versa* (which is rarely done). It speaks for itself that thematic field-work does not prevent a lexicographer from making an alphabetical dictionary. An additional avantage of collecting the lexemes and describing their meanings thematically, is that the lexicographic definitions can be geared to one another, even if they are arranged alphabetically afterwards.

Semasiological versus onomasiological

A semasiologically oriented dictionary tries to describe (among other things) the meaning(s) of lexemes, arranged alphabetically or thematically. The point of departure of an onomasiological dictionary such as the WVD, however, is neither a lexeme, nor a meaning, but the description of a concept, which is defined by a word (or paraphrase) in standard Dutch. Under the so-called standard Dutch 'title' of the dictionary article, all the different dialect words which can be used to refer to the concept are arranged (for an example see Kruijsen and Van Keymeulen 1997: 225-227). It has to be stressed that the description of the concept is not the meaning of all the words listed under it. A word such as *bird*, for instance, may be listed under the concept 'large bird of prey' for those dialects which lack a more specific word.

Classification problems

The assignment of a lexeme to a specific class in a thematic/onomasiological arrangement is not always easy. In the case of the WVD, the principles of functionality and frequency of a concept in the every-day life of the average dialect speaker are used to assign a concept to a particular class (e.g. a tomato is a plant, but its function is to be eaten as a vegetable and it is more frequently thought of as being eaten than grown, hence 'tomato' is in the first place assigned to the class 'food'). In some cases, the same concept is placed in different classes, especially when there are different lexemes corresponding to these

classes. Some Flemish dialects, for instance, use different words for a potato when grown in the garden (*erpel*) or when cooked and eaten (*patat*).

2.2   Presentation of the microstructure

The inclusion of microstructural elements and their presentation is dependent on the overall purpose of the dictionary, based on an analysis of the users' needs. We will only briefly dwell on pronunciation and the written form of the headword.

The extent to which spelling systems for normally unwritten languages exist (and are used in society), vary considerably. If the spelling is not codified, the lexicographer has to aware that the spelling used for the headwords in his dictionary may be looked upon as 'official' by the users, because it meets a societal need. Hence, the problem of the spelling of the headwords should be looked into very carefully. In the case of the dictionary of the Flemish dialects, the dialectal headwords are 'Dutchified'. Since the phonology of the dialect group under investigation is closely related to the phonology of standard Dutch, and since the Dutch spelling is highly phonological, it is possible to replace the dialect phonemes by their Dutch counterparts, thus 'framing' the dialect words as if they existed in standard Dutch. In fact, many respondents use this method spontaneously, when noting down the dialect words in the questionnaires. In doing so, they bear proof to the fact that Dutchification of the dialectal headwords was a good option. A WVD Dutchified headword is to be considered as a kind of 'reference form' under which a variety of different local dialect pronunciations are subsumed.

3.   Conclusions

In the absence of a corpus of written texts, dictionaries of unwritten languages or language varieties are based on lexicographic testimonies of good dialect speakers, collected orally or by correspondence. It is clear that some elements of a possible macro- or microstructure are less conscious in the mind of the speakers than others; most lexical elements of the open word classes can luckily be evoked by introspection and are therefore 'questionable'. The bulk of the lexical data of the major regional dialect dictionaries in the Dutch language area is obtained by questionnaires, especially when they combine a dictionary and a word atlas. Only smaller scale dictionaries can exclusively be based on oral interviews. In many cases word material derived from other sources is only marginal. To my knowledge, no dictionary has been based solely on a corpus of transcriptions of 'free speech'.

Analysing targeted metalinguistical discourse (a series of 'metalinguistical conversations' with groups of trained respondents) in which the linguistic competence of respondents is explored, is in my view a way to gain lexicographic data that are, although not derived from 'spontaneous' speech, adequate for

dictionary making. In such conversations one may proceed from an encyclope-dic analysis of reality itself to the way the world is lexically categorized and further to the collection of lexical data. I think that a thematic lexicographic investigation is very productive, because the researcher can focus on the con-tent of a theme and the completeness of a dictionary with regard to specific subjects can also be assessed. Questionnaires may complete, expand and verify the data gained in interviews. A dictionary based on a corpus of texts is some-times thought to be more scientific than a dictionary based on the introspection of the lexicographer or respondents. Fieldwork can of course be carried out scientifically as well. In the last analysis, it is the user who is the best judge of the quality of a dictionary and his judgement should be systematically enlisted at a very early stage. The method of 'simultaneous feedback' is perhaps even more desirable for fieldwork dictionaries than for corpus-based dictionaries, to overcome the danger of the subjectivity or the incompleteness of collection methods based on linguistic introspection.

In the case of the Dictionary of the Flemish Dialects and its two cognate dictionaries, the thematic approach, both in collecting and presenting the word material, undoubtedly rescued the project financially. Since it was possible to publish a thematic fascicle every year, the editorial board managed to maintain the project financially. Nobody but the lexicographer understands why it takes such a long time to compile a dictionary. The sponsors have to be kept happy.

References

Ammon U, N. Dittmar and K. Mattheier (Eds.). 1988. *Sociolinguistics. An International Handbook of the Science of Language and Society./Soziolinguistik. Ein internationales Handbuch zur Wissen-schaft von Sprache und Gesellschaft*. Berlin/New York: De Gruyter.

Atteslander, P. 1988. Befragung. Ammon U, N. Dittmar and K. Mattheier (Eds.) 1988: 940-951.

Besch, W., U. Knoop, W. Putschke and H.E. Wiegand (Eds.). 1982/1983. *Dialektologie: ein Handbuch zur Deutschen und Allgemeinen Dialektforschung.* Handbücher zur Sprach- und Kommunika-tionswissenschaft 1,1 and 1,2. Berlin/New York: De Gruyter.

Brouwers, L. 1988. *Het juiste woord*. Antwerpen: Standaard Uitgeverij.

Cicourel, A. 1988. Elicitation as a Problem of Discourse. Ammon U, N. Dittmar and K. Mattheier (Eds.). 1988: 903-910.

Denzin N. and Y. Lincoln (Eds.). 1994. *Handbook of Qualitative Research*. Thousand Oaks/London/ New Delhi: Sage Publications.

De Schrijver, Gilles-Maurice and D.J. Prinsloo. 2000. The Concept of "Simultaneous Feedback": Towards a New Methodology for Compiling Dictionaries. *Lexikos* 10: 1-31.

Fontana, A. and J. Frey. 1994. Interviewing. The Art of Science. Denzin, N. and Y. Lincoln (Eds.). 1994: 361-376.

Frissen, C. 1981. *Concept Inventory for WBD/WLD. Result of the Project "Voorbereidend onderzoek in verband met de algemene woordenschat" (Preparatory Investigation Concerning the General Vocabu-lary) 1978–1981*. Unpublished Typescript. Nijmegen: Catholic University.

Hallig, R. and W. von Wartburg. 1952. *Begriffssystem als Grundlage für die Lexikographie*. Abhand-

lungen der deutschen Akademie der Wissenschaften zu Berlin. Klasse der Sprachen, Literatur und Kunst nr. 4. Berlin: Akademie Verlag.

König, W. 1982/1983. Probleme der Repräsentativität in der Dialektologie. Besch, W., U. Knoop, W. Putschke and H.E. Wiegand (Eds.). 1982/1983: 463-485.

Kruijsen, J. en J. van Keymeulen. 1997. The Southern Dutch Dialect Dictionaries. *Lexikos* 7: 207-228.

Merton, R., M. Fiske and P. Kendall. 1956. *The Focused Interview*. Glencoe, IL: Free Press.

Van Keymeulen, J. 1992. *De algemene woordenschat in de grote dialectwoordenboeken (WBD, WLD, WVD). Een methodologische reflectie*. Ph.D. Dissertation. Ghent: Ghent University.

Van Keymeulen, J. 2003a. Dialectwoorden verzamelen. Een praktische handleiding. *Handelingen van de Koninklijke Commissie voor Toponymie en Dialectologie* 75: (forthcoming).

Van Keymeulen, J. 2003b. De introductie van het Nederlands in Vlaanderen: taalstrijd van de middenklasse. Botha, Willem (Ed.). *'n Man wat beur. Huldigingsbundel vir Dirk van Schalkwyk*: 332-339. Stellenbosch: Bureau of the WAT.

WALD = Schaars, A. 1984–. *Woordenboek van de Achterhoekse en Liemerse Dialecten*. Doetinchem: Staring Instituut.

WBD = Weijnen, A. et al. 1967–. *Woordenboek van de Brabantse Dialecten*. Assen: Van Gorcum.

Weijnen, A. en J. van Bakel. 1967. Voorlopige Inleiding op het Woordenboek van de Brabantse Dialecten. WBD. 1967–.

WLD = Weijnen, A. et al. 1983–. *Woordenboek van de Limburgse Dialecten*. Assen: Van Gorcum.

WVD = Devos, M. et al. 1979–. *Woordenboek van de Vlaamse Dialecten*. Tongeren: Michiels.

Capturing Cultural Glossaries: Case-study I[*]

Matete Madiba, *Department: Teaching and Learning Development, Technikon Northern Gauteng, Soshanguve, Republic of South Africa (madiba.m@tng.ac.za)*, Lorna Mphahlele, *Department: Languages, Technikon Northern Gauteng, Soshanguve, Republic of South Africa (mphahlele.l@tng.ac.za)*, and Matlakala Kganyago, *Nkoshilo High School, Mankweng, Republic of South Africa (tetema@mweb.co.za)*

Abstract: This article is a presentation of a brief cultural glossary of Northern Sotho cooking terms. The glossary is mainly composed of names for utensils and ingredients, and action words for the processes involved in the preparation of cultural dishes. It also contains names of dishes tied to some idiomatic expressions in a way eliciting cultural experiences that can lead to an understanding of indigenous knowledge systems. The article seeks to explore ways of capturing cultural glossaries to feed into the national dictionary corpora by using a case-study approach to investigate the processes that led to the generation of this specific school-based project. A number of issues that surfaced in this project, can possibly serve as models for the collection of authentic glossaries that can support dictionary making in African languages.

Keywords: CULTURAL GLOSSARY, INDIGENOUS KNOWLEDGE SYSTEMS, CULTURAL TERMINOLOGY, CORPUS, AUTHENTIC GLOSSARY, CONTEXTUALISATION, MARGINALIZED LANGUAGES, OUTCOMES-BASED EDUCATION, PERFORMANCE INDICATOR, ASSESSMENT CRITERIA, RANGE STATEMENT, TRADITIONAL DISHES, LANGUAGE VALORISATION, SIMULTANEOUS FEEDBACK, COMMUNICATIVE APPROACH, STRUCTURAL APPROACH, METONYMY

Opsomming: Die totstandbrenging van kulturele woordversamelings: Gevallestudie I. Hierdie artikel is 'n aanbieding van 'n kort kulturele woordversameling van Noord-Sothokookterme. Die woordversameling bestaan hoofsaaklik uit die name van gereedskap en bestanddele, en handelingswoorde vir die prosesse betrokke by die voorbereiding van kulturele geregte. Dit bevat ook name van geregte wat verbind is met sekere idiomatiese uitdrukkings wat op 'n manier kulturele ervarings oproep wat kan lei tot die verstaan van inheemse kennisstelsels. Die artikel probeer om maniere te ondersoek waarop kulturele woordversamelings in die nasionale woordeboekkorpusse ingevoer kan word deur 'n gevallestudiebenadering te volg om die prosesse te ondersoek wat tot die ontwikkeling van hierdie spesifieke skoolgebaseerde projek gelei het. 'n Aantal resultate wat uit hierdie projek voortgekom het, kan moontlik as modelle dien vir die tot-

---

[*] 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.

standbrenging van oorspronklike woordversamelings wat woordeboekskepping in die Afrikatale kan ondersteun.

Sleutelwoorde: KULTURELE WOORDVERSAMELING, INHEEMSE KENNISSTELSELS, KULTURELE TERMINOLOGIE, KORPUS, OORSPRONKLIKE WOORDVERSAMELING, KON-TEKSTUALISERING, GEMARGINALISEERDE TALE, UITKOMSGEBASEERDE ONDERWYS, PRESTASIEAANDUIDER, WAARDEBEPALINGSKRITERIA, RANGSTELLING, TRADISIONELE GEREGTE, TAALSTABILISERING, GELYKTYDIGE TERUGVOER, KOMMUNIKATIEWE BENA-DERING, STRUKTURELE BENADERING, METONIMIE

Background

A classroom-based project within an outcomes-based education (OBE) approach is the setting for the work reported here. The considered strength of this project is the creation and use of a meaningful context for the collection of a valuable glossary. The project has demonstrated that contextualisation can be used as a good organising tool for the collection of other glossaries. The school setting, within which the project was carried out, provides a fertile ground for an activity of this nature. The environment (of the school) is dominated by rural settlements, which are even more relevant and useful as authentic sources for cultural embodiments.

Of particular interest to researchers is the fact that projects of this nature have the potential to capture and record cultural words that would otherwise be lost. This task seeks also to investigate how glossaries like these can help to realise and implement innovative methodologies and concepts such as De Schryver and Prinsloo's (2000: 1) proposal of 'simultaneous feedback' to support lexicographic work. As the project had a different target, the compiling of a glossary was a 'secondary' and not 'primary' product thereof. It is this distinctive feature (of being a 'secondary' product), whose further implications have to be investigated.

The case-study approach is found to be more suitable to a project like this as lessons learnt while compiling this brief glossary, will be more clearly spelt out. It is the exploration of these lessons that will eventually be used to propose a possible and authentic model for collecting other glossaries of this nature.

Outcomes-based Education Environment

The rationale behind the project is based on the argument that it will not be easy for the formerly marginalized languages to capture cultural terminology in a corpus built mainly from written texts and frequency lists. It is therefore argued that such an approach to corpus building has the potential of creating gaps in a way that may necessarily exclude cultural terms. The provision of a model for the collection of cultural words and the initiation of similar projects

reported here will address these gaps.

There have been great changes in the education system of the country in recent years. The teaching of Northern Sotho was also affected as the Department of Education brought about changes that lead to the teaching and learning of languages moving away from the structural to the communicative approach, especially for grade 10-12 learners. The first examination within this approach had been written at the end of 2002. In the communicative language-teaching approach the 'greatest challenge in the profession is to move significantly beyond the teaching of rules, patterns, definitions, and other knowledge "about" language to the point that we are teaching our students to communicate genuinely, spontaneously, and meaningfully' (Brown 1994). The adoption of this kind of approach in language teaching, it is believed, will help implement outcomes-based education (OBE).

What all this means for educators is that in order to be successful you, together with your learners, have to take full responsibility for the careful planning and management of your classroom instruction and assessment of learning. Within this OBE approach, the focus is on the final result of each learning process. According to Van der Horst and McDonald (1997), learners must take an active part in learning. Their critical thinking should be developed together with their reasoning and reflection. Learning should be relevant and connected to real life. In the communicative approach to language teaching, learners are therefore encouraged to work together for example in group discussions and projects.

In the structural approach, a learner would be given an idiom like 'legotlo le lefa ka setopo' (an eye for an eye). Questions asked would include:

— What is the meaning of the idiom?

— Construct a sentence using the idiom to show that you understand it.

This type of question has no direct reference to the real life situations of the learner. In the communicative approach, questions should be carefully designed to be meaningful. Learners can be asked to:

— Demonstrate the real meaning of the idiom by contextualising it through storytelling;

— Write a paragraph about their own attitude towards the idiom;

— Relate the idiom to topical issues such as the country's constitution and related global debates;

— Give their viewpoint from their religious background;

— Show a connection towards real life and provide short case-studies — when do people use the idiom?; under what circumstances is the idiom relevant?

The type of questions given in the communicative approach must probe the mind of the learner into more thinking. In order to respond to such questions, the learner has to search for experiences in real life. It is not just the content (as in the wording of the idiom and the terminology used) that is of interest in a learning environment like this, but attitudes and values are also to be taken into consideration.

It is within such a teaching and learning environment that the project with which this article deals, was developed. This article reports on what was initiated as a response to the OBE approach within the teaching of Northern Sotho as a language. The goal of the educator in the broadest sense was to teach semantics. In order to contextualise the content and make it meaningful, identifying various Northern Sotho traditional dishes was used as a starting point. In the structural approach, learners would be expected to name the dishes, or be provided with descriptions of a dish so that the learner can provide the suitable name (the match column A to B type of questions). A demonstration of this kind is given in the following example:

Column A (Description of the dish)	Column B (Name of the dish)
Meetsana ao re šebago ka wona go tšwa nameng yeo e apeilwego (The watery substance coming from cooked meat and used as a relish)	mothotho/moro (gravy)

Within the OBE approach, outcomes as outlined in the policy document from the Department of Education (1997) have to be achieved. For this, Specific Outcome 3 (SO/3) that states learners are to (be made to) respond to the aesthetic, affective, cultural and social values, was used. The performance indicator will be that learners uncover important aspects of style and move towards the ability to discern and describe more subtle features of the language. The assessment criteria will therefore focus on the learner's development of a vocabulary to support impressions. In the range statement, emphasis will be on knowledge, i.e. something related to history, social conditions, human experiences, etc.

For the project described here, learners were divided into groups to prepare various traditional dishes of their choice. All the preparations were done in school. A period during every day of a week was given to complete the project. The group monitored the process of preparation, and its stages were recorded on a daily basis. Some of the learners reported at school as early as 5 a.m. to work on their dishes. It was an experience for both learners and educators to go through this process for a week. There was a lot of excitement, which kept the motivation high. Some of the dishes like 'sekome' were equally new to some of the learners as to some of the educators. With the help of this dish, the following Northern Sotho idiom made more sense to many in the school: 'Šiki-

ša dira le molapo, mphago wa dira ke meetse.' (The literal translation is that soldiers were always to walk next to a stream/river because they use water as their provision for a march.) 'Sekome' was prepared as a special provision for soldiers during war. It is very light to carry and after taking a small amount one will drink a lot of water and then one's hunger is satisfied for a longish period. It can also be used as a snack.

Through this project, a useful glossary of traditional dishes was compiled. Working with high-school learners made this endeavour, which turned into a research project, possible. There was no need for employing field-workers to collect the words. The learners had to gather information from members of their own community and this gave the project such a high level of authenticity. It is within this framework that this article argues for more projects of this nature to collect the wealth of information that would otherwise be lost. The excitement this exercise brought to the learning of the language was evident in the level of motivation that impelled the project. Learners experienced it as an affirmative process to contribute to a knowledge of their language beyond the boundaries of their textbooks and the experience of their teachers.

There are many benefits involved in encouraging and nurturing this type of work which has the potential for research into indigenous knowledge systems. The particular language will enjoy a much-desired growth. Learners are motivated, acquiring a lot of knowledge about a certain topic. With the necessary funding and initiative, more projects can be undertaken and topics such as traditional dress, ceremonies, dances and medicines can be researched in a more focused way to capture cultural glossaries within an authentic environment.

Language Valorisation

A very unique opportunity for the valorization not only of Northern Sotho, but also of other African languages is the transfer of traditional, almost archaic terminology to much-needed names for modern innovations, and as such contributing to terminology work necessary for these languages. This involves the process of metonymy (Carstens 1999). It has already been used in many languages, Northern Sotho included, for language expansion. One such example of the transfer of traditional terminology to modern innovations is 'dish', the concave reflector used especially for radar and radio telescopes. Instead of creating a new term, the word 'dish' has been used metonymically.

In Northern Sotho several terms from traditional initiation processes (as practised in the initiation school) have been applied in this metonymic way. When Western type of schooling was introduced into the Northern Sotho environment, such terms as 'mphato' (which came to mean 'standard' as in standard one, two, etc.) were transferred from their traditional meanings to the new ways of education. The same applies to 'go aloga' or 'dialoga', which refer to graduation or graduates.

Another special feature of metonymy is its potential to reveal the traditional and cultural circumstances under which a proverb or idiom was formulated. Together with this, the etymologies of words and their variations can be better understood.

There is a great variety of cultural terminologies still to be uncovered through processes such as those outlined above. A possible application of these terminologies lies in the expansion of the African languages. The use of a process like metonymy in word creation is undoubtedly preferable to the artificial coining of words.

Glossary

What follows is the presentation of a glossary of traditional dishes, and the utensils involved and processes used in their preparation. It is to be hoped that the terminology will be introduced into the corpora of national dictionaries where it can be given proper lexicographic treatment. The glossary is presented here in a report-based manner, mainly from the recipes prepared in the school project, together with further information from informants. The glossary has not been given any specific lexicographic treatment.

Supplementary work was done to expand the glossary, which involved consulting informants who were able to provide the information needed, although the focus of inquiry was effectuated by the initial ideas and terms originating from the school project.

Bjala bja Sesotho (traditional African beer)

One of the idiomatic expressions associated with traditional beer is 'go wela ka nkgong', which can be translated as 'getting drunk'. In a more literal sense, it also means 'falling into the (traditional) beer container'.

The process of preparing African beer has the following terminology:

Mmela is **mabele** (an African cereal food) that is not smoothly ground; it has to be rough. When the mmela is rough, it is called **gaila**. When the mmela is smoothly ground, it is called **mathume**, and is normally used to make porridge.

Sekhukhu is a mixture of mmela and boiling water.

Riteletša is to mix the mmela and the sekhukhu.

This mixture has to be left for 24 hours in summer or 48 hours in winter for fermentation. After this period has elapsed, the mixture is called **mohlogotse.** This is when fermentation is in process and the water separates from the mmela. The water has foam, which is known as **lehulo.**

After the fermentation process, the water is removed and boiled separately. The fermented mohlogotse left behind is now known as **dintshe** or **pelle.**

When the water has boiled, dintshe or pelle is added and cooked. Once

this has cooked, it is called **lešeleba** or **magebeleta.** Lešeleba or magebeleta is left to stand for a day and thereafter more mmela is added and the container is covered with a **seroto**.

Moroko, an extract that is part of the final process of this beer-making procedure, is a reusable product. It will be dried, ground and mixed with corn cereal to make a different type of porridge. Sometimes this extract will be cooked whilst it is still watery, before it dries, and it will be eaten with salt or sugar as another type of porridge. The cooking is referred to as **go šokwa**.

Other types of beer

(a) **morula** — beer made from the juice of the morula fruit

(b) **makgere** — beer made from prickly pears

Different types of porridge and related dishes

Porridge was the staple food. There were variants of this dish as indicated in the various terminologies used.

(a) **bogobe bja mabele** — porridge made from sorghum cereal

(b) **kgodu** — porridge made from a cooked melon-like fruit called **lerotse** mixed with mealie meal

(c) **lewa** — a dish made from sorghum grain

(d) **lehlahlara** — a type of lewa, prepared from mabele

(e) **lemopa** — a type of lewa made from **ditlhodi** (Chinese beans)

(f) **semotwane / sekgotho** — a dish made from traditional beans and thickened (**loišwa**) with maize meal

The less tasty version of porridge was always accompanied by another dish. Besides meat, it would be eaten with some type of **morogo** or other. Morogo is a traditional vegetable.

Different types of morogo

(a) **leroto** — leaves of a small plant which bears small white flowers when mature, with a bitter taste when cooked

(b) **lehlanye** — a small plant with pointed needle-like leaves with a bitter taste

(c) **monawa** — leaves of the bean plant

(d) **mphodi** — leaves of the pumpkin plant

(e) **monyaku** — leaves of a small plant which bears a hairy fruit that looks like a smaller version (and one that is not edible) of a melon fruit

(f) **motšhatšha —** leaves of a melon plant with a bitter taste when cooked

(g) **mothagaraga —** leaves of a small plant that produces oval-shaped fruit and usually grows on harvested fields

(h) **theepe —** leaves of a small plant that are two-toned, with a reddish colour underneath, and slimy when cooked

Dithotse, pumpkin and watermelon pips that are normally fried without fat (and salt added), were also used as **sešebo**, something to eat together with porridge. These were dried, and could be kept for a considerable time. **Dithu-thupe,** popcorn, was a common seasonal snack. **Mangangale** was another type of a dish made from **marotse** (pumpkins). These melon-like fruits will be sliced (**a hlabelwa**) and dried, then cooked and eaten separately as a single dish.

Wild fruit that was part of the traditional diet

(a) **dithetlwa —** soft wild berries

(b) **mabilo —** wild brownish fruit when fully ripe

(c) **ditoro —** prickly pears

(d) **mabupudu —** wild dates

(e) **mahlatswa —** wild litchi

(f) **matšhidi / ditšhidi —** wild fruit that is yellowish when fully ripe, with a bitter-sweet taste

(g) **dinee —** wild mulberries

(h) **ditlhakola —** big wild berries

(i) **marula —** marula

(j) **magaba —** a wild tuber that has a lot of water

(k) **dikgwane —** hard wild berries

(l) **mago —** figs

Other special dishes

Bokoma was a special dish for those who took long journeys. It was made of ground dried nuts. This came in handy and did not need any preservation. **Sekome** is a variation of this dish that was given to soldiers who spent a long while away from home with no access to fresh food. It consists of groundnuts, maize, sorghum and salt. These ingredients are fried without any fat, ground and mixed.

   Mageu is a drink made of cooked mealie-meal porridge mixed with water and sugar, left for a week or so to ferment.

Traditional utensils

- (a) **pitša —** a three-legged cooking pot, usually black in colour
- (b) **moruswi/thiswane —** a utensil for keeping relish
- (c) **sethebe —** a utensil for keeping porridge
- (d) **sego —** a calabash used for drinking water and for serving porridge
- (e) **seroto —** a lid made from cane
- (f) **leselo —** a sieve made from grass
- (g) **mogopo —** a wooden dish used for serving porridge
- (h) **legopo —** a wooden dish, usually small in size, used for serving relish
- (i) **mokgopu —** a calabash used for drinking traditional beer
- (j) **mothotwana —** a small dish used for mealie meal
- (k) **tšhelwana —** a small lid used for covering such substances as mageu
- (l) **tšhakga —** a sheet of galvanised iron used to make popcorn without adding any fat
- (m) **nkgo —** a clay pot used mainly for the storage of traditional beer

Cleaning materials

- (a) **lešekešedi —** river-sand that is used to wash utensils
- (b) **motaga —** white soil usually found at river-banks and used for polishing washed utensils
- (b) **mogohlo —** white gravel soil used for cleaning utensils

References

Brown, H.D. 1994. *Principles of Language Teaching*. White Plains, NY: Prentice Hall.

Carstens, A. 1999. *Lexicography and Terminography 876*. Lecture Notes. Pretoria: University of Pretoria.

Department of Education. *Senior Phase Grade 7–9*. Policy Document. October 1997.

De Schryver, G.-M. and D.J. Prinsloo. The Concept of 'Simultaneous Feedback': Towards a New Methodology for Compiling Dictionaries. *Lexikos* 12: 1-31.

Van der Horst, H. and R. Macdonald. 1997. *Outcomes-Based Education: Theory and Practice*. Pretoria: Kagiso.

Compiling Dictionaries Using Semantic Domains[*]

Ronald Moe, *Linguistics Consultant, SIL International*[1] *(ron_moe@sil.org)*

Abstract: The task of providing dictionaries for all the world's languages is prodigious, requiring efficient techniques. The text corpus method cannot be used for minority languages lacking texts. To meet the need, the author has constructed a list of 1 600 semantic domains, which he has successfully used to collect words. In a workshop setting, a group of speakers can collect as many as 17 000 words in ten days. This method results in a classified word list that can be efficiently expanded into a full dictionary. The method works because the mental lexicon is a giant web organized around key concepts. A semantic domain can be defined as an important concept together with the words directly related to it by lexical relations. A person can utilize the mental web to quickly jump from word to word within a domain.

The author is developing a template for each domain to aid in collecting words and in describing their semantics. Investigating semantics within the context of a domain yields many insights. The method permits the production of both alphabetically and semantically organized dictionaries. The list of domains is intended to be universal in scope and applicability. Perhaps due to universals of human experience and universals of linguistic competence, there are striking similarities in various lists of semantic domains developed for languages around the world. Using a standardized list of domains to classify multiple dictionaries opens up possibilities for cross-linguistic research into semantic and lexical universals.

Keywords: SEMANTIC DOMAINS, SEMANTIC FIELDS, SEMANTIC CATEGORIES, LEXICAL RELATIONS, SEMANTIC PRIMITIVES, DOMAIN TEMPLATES, MENTAL LEXICON, SEMANTIC UNIVERSALS, MINORITY LANGUAGES, LEXICOGRAPHY

Opsomming: Samestelling van woordeboeke deur gebruikmaking van semantiese domeine. Die taak van die voorsiening van woordeboeke aan al die tale van die wêreld is geweldig en vereis doeltreffende tegnieke. Die tekskorpusmetode kan nie gebruik word vir minderheidstale waarin tekste ontbreek nie. Om in die behoefte te voorsien, het die skrywer 'n lys van 1 600 semantiese domeine opgestel wat hy suksesvol gebruik het om woorde te versamel. In 'n werksessie-omgewing kan 'n groep sprekers tot soveel as 17 000 woorde in tien dae versamel. Hierdie metode lei tot 'n geklassifiseerde woordelys wat doeltreffend uitgebrei kan word tot 'n volledige woordeboek. Die metode werk omdat die mentale leksikon 'n groot web is wat rondom sleutelbegrippe gestruktureer is. 'n Semantiese domein kan gedefinieer word as 'n belangrike konsep saam met die woorde wat direk daarmee verband hou vanweë leksikale verwantskappe. 'n Persoon kan die mentale web gebruik om vinnig van woord tot woord binne 'n domein te spring.

Die skrywer is besig om vir elke domein 'n profiel te ontwikkel om te help met die versameling van woorde en met die beskrywing van hul semantiek. 'n Ondersoek van semantiek binne die konteks van 'n domein lewer baie insigte. Die metode laat die totstandbrenging van sowel alfabeties as semanties gerangskikte woordeboeke toe. Die lys domeine is bedoel om universeel in omvang en toepassing te wees. Moontlik as gevolg van universalia van menslike ervaring en universalia van taalkundige vermoë, is daar treffende ooreenkomste tussen verskillende lyste semantiese domeine wat ontwikkel is vir tale oor die hele wêreld. Die gebruik van 'n gestandaardiseerde lys domeine om veelsoortige woordeboeke te klassifiseer, skep moontlikhede vir kruislinguistiese navorsing oor semantiese en leksikale universalia.

Sleutelwoorde: SEMANTIESE DOMEINE, SEMANTIESE VELDE, SEMANTIESE KATE-GORIEË, LEKSIKALE VERWANTSKAPPE, SEMANTIESE PRIMITIEWES, DOMEINPROFIELE, MENTALE LEKSIKON, SEMANTIESE UNIVERSALIA, MINDERHEIDSTALE, LEKSIKOGRAFIE

The problem (It's going to take forever)

The mental lexicon is far larger than either the grammatical component or the phonological component in a person's linguistic competence. Investigating and describing it is the largest and most time-consuming task in descriptive linguistics. With perhaps 6 000 languages in the world and perhaps 20 000 words in each, we need to collect and describe something on the order of 120 000 000 words.[2] The major languages of the world often have several large published dictionaries available to them. The major publishing companies can afford to hire scores of professional lexicographers to compile massive text corpora and do the research necessary to produce quality dictionaries. But for minority languages the picture is far bleaker. With few or no published texts, few or no professional lexicographers available to them, and little or no funding, the minority languages face a daunting challenge.

I have been involved in the production of dictionaries for minority languages since 1985 and have taught lexicography seminars to train others in the process. I estimate that linguists working in a language development project add words to their lexical database at the average rate of only 650 words per year, or about 2.5 words per working day.[3] At this rate it frequently takes 20 years to produce even a modest dictionary. For many years I have been concerned about this abysmal rate of progress and have attempted to find ways to make the process of compiling a dictionary simpler and more efficient. If we are ever going to finish the task of documenting the world's languages, we need a mass production technique.

The journey (Searching for a solution)

For several years colleagues within SIL, together with other interested scholars, have discussed ways in which we could leverage the linguistic similarities among the Bantu languages to facilitate linguistic investigation and language

development within the Bantu family. We have called this movement the 'Bantu Initiative'. In September 2000 the Bantu Initiative asked me to begin work on a dictionary template, including the production of a list of semantic domains that could be used to classify Bantu language dictionaries. I was a bit sceptical, since I had heard from numerous sources that the semantic category systems of the world's languages were vastly different, and even varied from individual to individual. But since the Bantu languages are closely related, I thought it was worth a try.

In order to construct a list of domains for Bantu languages, I needed to know how Bantu peoples categorized the words of their languages. So in December 2000 and January 2001 I held two workshops[4] for Gikuyu and Lugwere[5] in which I asked 12 speakers of each language to sort and group a list of 1 000 words chosen from a wide variety of semantic domains. I was curious to see how non-westernized peoples would classify the words of their language. My expectation was that they would set up very different domains than an English speaker. They didn't. Their domains were strikingly similar to other lists of semantic domains that I had collected from around the world. As I compared the lists, it became apparent that the universality of human experience and some sort of universal linguistic competence resulted in similar classification systems. The differences came from minor differences of culture and the necessity to squash a multi-dimensional system of relationships into a two dimensional list. So I decided (perhaps presumptuously) to attempt to compile a universal list of semantic domains.

The challenge was to compile an exhaustive list of domains that could be used for any language in the world. None of the lists I had were complete. All were designed for a particular language and purpose. For instance, the *Outline of Cultural Materials* (Murdock et al. 1987) presents a list of anthropological domains, but is missing many lexical domains. *Roget's Thesaurus* (Roget 1958) has 1 000 domains, but due to its purpose it also omits many domains. Newer editions of *Roget's* (e.g. Morehead 1985) contain 600 major domains and thousands of smaller entries. Neither presentation is suitable for our purpose. Louw and Nida (1989: xix) admit that their list is uneven due to the subject matter of the New Testament. Recent semantically organized dictionaries such as the *Longman Language Activator* (Summers 1993) and the *Oxford Learner's Wordfinder Dictionary* (Trappes-Lomax 1997) are highly selective in the domains they include. So I concluded that a new list was needed. I contrasted and compared all the lists at my disposal, ensuring that every domain in every list was covered by a domain in my list. As I studied the organization of the lists, more and more similarities began to emerge. There was a logic to the domains, and a logic to how they were organized.

I knew from the beginning that a list of semantic domains could be used to collect words. Eliciting vocabulary has been a topic of interest for some time, and the literature contains a wealth of practical suggestions, such as using lexical relations (Beekman 1968: 4), concording a text corpus (Naden 1977: 14), and

using semantic domains (Newell 1986: 20).[6] I decided to try it out and see just how easy it would be. I took the semantic domain 'Bodies of water' and started listing words that belong to the domain (e.g. *ocean*, *lake*, *river*, *shore*, *wave*, etc.). In fifteen minutes I had collected and subcategorized 169 words. The rate for collecting words had just jumped from 2.5 words/day to 11 words/minute. I realized that all I needed was a list of semantic domains and I could collect the words of a language in a matter of days rather than years.

As I thought about how the list of domains could be used to collect words, I realized that a simple domain label, such as 'Bodies of water', would not be adequate. Three things were needed: (1) a simple statement of the central idea of the domain, (2) elicitation questions that would prompt a person to think of words that might belong to the domain, and (3) sample words from English.[7]

I have tested the materials and method in three workshops. The first test, held in May 2001, used a beta version of the semantic domains list with a group of fifteen speakers of the Lugwere language. In ten days, the participants collected over 10 000 words and 1 000 example sentences.[8] In January 2002, 30 speakers of Lunyole used version one to collect 17 000 words in ten days. In February 2002, 12 speakers of Kitharaka[9] collected 12 000 words in eight days. In the months since the workshops, speakers of each language have been editing and glossing the word lists. As the result of a few months work, we expect to have a classified dictionary in each language of over 10 000 words, including part of speech, noun class, the plural form of each noun, and a simple gloss. The chart below compares the historical average rate of progress with the results of the three workshops.

Rate of Progress
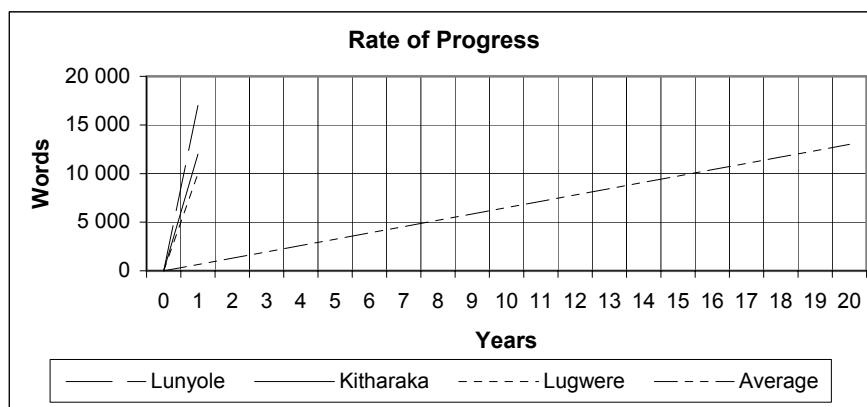
Why does it work? (Semantic domains, lexical relations, and semantic primitives)

The field of semantics has yet to reach a consensus on the nature and validity of semantic domains and semantic primitives. 'Semantic domain' is just another

way of saying 'area of meaning', but the notion that a meaning occupies an area is obviously figurative. Wierzbicka (1996: 170) comes close to endorsing the notion of universal semantic domains when she says: "The idea that words form more or less *natural* groupings, and that at least some of these groupings are *non-arbitrary*, is intuitively appealing, even *irresistible*" (emphasis added). She also indicates that domains vary in their nature from "self-contained fields of semantically related words" to "irregular and open-ended networks of interlacing networks". The question remains — just what is a semantic domain?

I envisioned that the list of domains would serve several purposes. It could be used to collect words, it could serve to classify a dictionary, and it could aid in semantic investigation. In order for it to be an effective tool in collecting words, I felt I should list sample words from English that belong to each domain. As I analyzed the words that I was listing under each domain, and compared them to the words others had included in the same domain, I began to see patterns. Some domains consisted of a generic term, such as 'Game', and a list of specifics: *chess, checkers, charades, monopoly*. Others were based on the Whole-Part lexical relation, such as 'Head' and *eye, nose, mouth*. Other domains included a variety of words related by different lexical relations, such as 'Wave' and *tidal wave, crest, break, roar, surfboard*.

It became apparent that a semantic domain was really some important concept and all the words directly related to it by some lexical relation. The words of a language are all linked together in the mind in a gigantic multidimensional web of relationships. But these mental links tend to cluster around a central nexus. A semantic domain isn't so much an area of the web as it is one of these central hubs. One of the intriguing questions about these hubs is: What is their relationship to semantic primitives? Many domains appear to be based on semantic primitives or a combination of two or three primitives (e.g. 'Bad behavior' = do + bad; 'Parts of things' = part (of) + something). Many are headed by high frequency words which constitute the core vocabulary of a language.

Several recently published dictionaries employ a "defining vocabulary". For instance, the *Longman Language Activator* (Summers 1993) lists the 2 222 words of its defining vocabulary in an appendix. When one excludes the functors (e.g. the, to, of), what is left is very similar to a list of domains. The notions of "semantic domain", "semantic primitive", "core vocabulary", and "defining vocabulary" seem to be converging.

As I developed the list, I began organizing the sample English words into lexical sets. I found that each lexical set was related to the central idea of the domain by a single lexical relation. I have already mentioned that lexicographers recommend that we employ lexical relations in collecting words. This seemed like a very useful idea in the light of what I was discovering. However, lexical relations are very hard to grasp in the abstract (e.g. $Conv_{13}$ (buy) = sell (Grimes 1987: 27)). Grimes (1994) has attempted to make lexical relations more user-friendly. But there are so many of them[10] that it is extremely inefficient to

have to think through the entire list of lexical relations for each new word encountered, in order to determine which ones might be productive. So I felt it was best if I thought through the list and identified which lexical relations were productive for each domain. I worded each productive lexical relation in the form of a simple question. For example, the domain 'Sing' has the following productive lexical relations:

What words refer to singing? *sing, serenade, warble, yodel, burst into song*
What words refer to singing without using words? *hum, whistle*
What words refer to a person who sings? *singer, vocalist, soloist*
What words refer to a group of people singing together? *choir, chorale, singing group, duet, trio, ensemble*
What words refer to something that is sung? *song, singing, tune, melody*
What types of songs are there? *lullaby, hymn, psalm, carol, national anthem, lament, ballad*
What words refer to a part of a song? *verse, chorus, theme, note, melody, harmony*
What words describe how well a person sings? *beautiful singing voice, can't carry a tune in a bucket, sing on/off key, monotone*
What words describe how high or low a person sings? *pitch, soprano, alto, baritone, bass*
What words describe whether or not people are singing the same thing together? *sing in unison, sing in harmony, sing the melody/harmony*

The questions and sample words are not meant to be exhaustive. It doesn't take much effort to think of other words. In practice, it has turned out that the combination of semantic domains and lexical relations is extremely productive. The mind quickly jumps from one word to another along the mental paths formed by lexical relations.

What do we need? (Domain templates)

Atkins (1997) has recommended that lexicographers produce a template for each lexical set they are investigating. She points out that a template enables the lexicographer to gather information faster, prompts the lexicographer to look for common features, and makes the approach to the whole lexical set much more systematic. I believe we could produce universal templates, which would be based on cross-linguistic research and would present features that the lexicographer would be likely to encounter in each domain.

Thus far I have worked to identify the lexical relations that are productive in each domain and have listed sample words from English. My purpose is to produce a tool which can be used to collect words. Here is an example:

2.4.1 See
What words refer to seeing something (in general or without conscious choice)?
     see, behold, come into view

What words refer to consciously looking at something? *look at, view, observe, scan*

What words refer to looking at something in order to learn? *watch, scrutinize*

What words are used of looking at something for a long time or in amazement? *stare, gaze, gape, gawk*

What words are used of looking at something for a short time? *glance, cursory glance, look at briefly, (eyes) flicker over*

What words refer to the sense of sight? *sight, sense of sight, vision*

What words refer to someone who sees? *observer, beholder, witness*

What words refer to a group of people who are watching something? *audience*

What words refer to what is seen? *sight, view*

Once the members of a lexical set are identified, we can identify the semantic features which distinguish them. For instance, the English words which belong to the domain 'Movement' often incorporate a component of direction, such as *advance* (front), *retreat* (back), *step aside* (side), *climb* (up), and *descend* (down). Other components include manner (*walk, run, jump*), beginning or ending point (*leave, arrive*), and medium (*fly, swim*). Once we have investigated the semantics of this domain for several languages, commonly occurring features can be noted. So the template for 'Movement' would prompt the researcher to look for these components. We could also include sample definitions, pragmatic and cultural issues to look out for, possible subcategorizations, and possible variations in the conceptualization of the domain. As each template is enriched, its usefulness will grow.

Where to from here? (Semantic universals and beyond)

Using semantic domains to produce a dictionary has numerous benefits in addition to speeding up the process of collecting words. We can sort our computer databases alphabetically or by domain. Translators and writers need lists of related words to facilitate composition. We can produce semantically organized dictionaries, such as the *Longman Language Activator* (Summers 1993), the *Oxford Learner's Wordfinder Dictionary* (Trappes-Lomax 1997), and the *Greek-English Lexicon of the New Testament, Based on Semantic Domains* (Louw and Nida 1989). Or we can publish alphabetical dictionaries and include an appendix of domains.

It is far more insightful to study the members of a lexical set together than to study them in isolation. As Wierzbicka (1996: 170) has pointed out: "Although the meaning of a word does not depend on the meanings of other words, to establish what the meaning of a word is one has to compare it with the meanings of other, intuitively related words."

Wierzbicka concludes her chapter on semantic primitives and semantic fields by saying: "I think, therefore, that the semantic primitives approach to semantic analysis also offers a necessary firm ground for the study of semantic

fields" (1996: 183). I would agree, and add that the study of semantic fields is necessary for the study of semantic primitives and universals.

The existence of the International Phonetic Alphabet permits cross-linguistic comparisons of phonological systems. The existence of (fairly) standardized grammatical categories allows us to search for universals of grammar. Anthropology has the *Outline of Cultural Materials* (Murdoch 1987). Chemistry has the periodic table. What does semantics have? I suggest that we cooperate to produce a standardized list of semantic domains. Such a list would enable us to do cross-linguistic comparisons and search for linguistic universals in the field of semantics, just as our colleagues are doing in the fields of phonology and grammar. What I have done is only a poor first attempt in this direction, but I hope it will lead to productive avenues of research.[11]

Endnotes

1. SIL International (the Summer Institute of Linguistics International) is an organization of volunteers, devoted to the promotion and development of minority languages. SIL International works in over 50 countries and over 1 000 languages.
2. In the interests of simplicity and naturalness, if not accuracy, this article employs the term 'word' to refer to lexical items of all sorts, including roots, derivatives, compounds, idioms, and phrases.
3. This estimate is based on observation of the number of years it has taken to produce published dictionaries, both within and outside of SIL, and has been confirmed by numerous SIL colleagues.
4. Thanks are due the Bantu Initiative for funding these workshops.
5. Both languages are Bantu. Gikuyu is spoken in Kenya, and Lugwere in Uganda. Dr. Mary Muchiri of Daystar University organized the Gikuyu workshop, and Dr. Ruth Mukama of Makerere University the Lugwere workshop.
6. Ideally lexicographic research should utilize both semantic domains and a concordance. However, unless a computerized text corpus running into the millions of words is available, using a list of domains is the only effective way of collecting words. If no corpus is available, it would be good to begin collecting or producing one.
7. These materials are currently being translated into Swahili, and plans are to have them translated into French, Spanish, Chinese, and other major languages of the world.
8. By comparison many bilingual dictionaries are published with only 3 000–5 000 entries.
9. All three languages are Bantu. Lugwere and Lunyole are spoken in Uganda, and Kitharaka in Kenya.
10. In fact, there are far more than the literature would suggest. It is apparent that lexical relations are not all the same sort of thing. I believe that lexical relations are based on similarities of meaning, and are as varied as the meanings of words.
11. Copies of the author's list of semantic domains and related materials are available from him via email at ron_moe@sil.org. The materials are also available in Swahili.

References

Atkins, Sue. 1997. Template Entries. Unpublished SALEX seminar handout.

Beekman, John. 1968. Eliciting Vocabulary, Meaning, and Collocations. *Notes on Translation* 29:1-11.

Grimes, Charles E. 1994. Mapping semantic relationships in the lexicon using lexical functions. *Notes on Linguistics* 66: 5-25.

Grimes, Joseph E. 1987. Relations and Linkages in the Lexicon. Prepublication Draft. Dallas: Summer Institute of Linguistics.

Louw, Johannes P. and Eugene A. Nida. 1989. *Greek-English Lexicon of the New Testament: Based on Semantic Domains, Vol. 1.* New York: United Bible Societies.

Morehead, Albert H. (Ed.). 1985. *The New American Roget's College Thesaurus in Dictionary Form.* New York: Signet.

Murdock, George P., et al. 1987[5]. *Outline of Cultural Materials.* New Haven: Human Relations Area Files.

Naden, Tony. 1977. *Words and Meanings.* Ghana: Institute of Linguistics.

Newell, Leonard E. 1986. *Lexicography Notes.* Manila: Summer Institute of Linguistics.

Roget, Peter Mark. 1958. *Roget's Thesaurus.* Harmondsworth, Middlesex: Penguin Books.

Summers, Della (Ed.). 1993. *Longman Language Activator.* Essex: Longman.

Trappes-Lomax, Hugh. 1997. *Oxford Learner's Wordfinder Dictionary.* Oxford: Oxford University Press.

Wierzbicka, Anna. 1996. *Semantics, Primes and Universals.* Oxford: Oxford University Press.

Disregarding the Corpus: Head-word and Sense Treatment in Shona Monolingual Lexicography[*]

Webster M. Mavhu, *African Languages Research Institute (ALRI),*
University of Zimbabwe, Harare, Zimbabwe (vhezh2000@yahoo.com)

Abstract: With specific reference to Shona monolingual lexicography, this article discusses how corpus-based lexicographers might, in some instances, decide not strictly to adhere to the corpus when it comes to headword and sense treatment. The writer is a member of the African Languages Research Institute (ALRI), formerly known as the African Languages Lexical (ALLEX) Project. ALRI is a nonfaculty interdisciplinary unit dedicated to research on and the development of African languages in Zimbabwe. The writer is part of the six-member team that compiled the now published Shona monolingual, synchronic, medium-sized and general-purpose dictionary *Duramazwi Guru ReChiShona* (2001). The article originates from the writer's experience of working on this dictionary. The article highlights the fact that being corpus-based does not necessarily imply being corpus-bound.

Keywords: CORPUS, CORPUS-BASED, FREQUENCY, HEADWORD, LEXICOGRAPHY, SENSE, SHONA, SLANG, SYNONYMS

Opsomming: Verontagsaming van die korpus: Trefwoord- en betekenisbe-handeling in die Sjona- eentalige leksikografie. Met spesifieke verwysing na die Sjona- eentalige leksikografie bespreek hierdie artikel hoe korpusgebaseerde leksikograwe in sommige gevalle kan besluit om nie streng by die korpus te bly wanneer dit kom by trefwoord- en betekenisbehandeling nie. Die skrywer is 'n lid van die African Language Research Institute (ALRI), vroeër bekend as die African Languages Lexical (ALLEX) Project. ALRI is 'n niefakulteits-interdissiplinêre eenheid wat hom beywer vir navorsing oor en die ontwikkeling van die Afrikatale in Zimbabwe. Die skrywer is deel van 'n span van ses lede wat die reeds gepubliseerde Sjona- eentalige, sinchroniese, middelgroot en meerdoelige woordeboek *Duramazwi Guru ReChiShona* (2001) saamgestel het. Die artikel het uit die skrywer se ervaring van werk aan hierdie woordeboek ontstaan. Die artikel belig die feit dat korpusgebaseerdheid nie noodwendig korpusgebondenheid impliseer nie.

Sleutelwoorde: BETEKENIS, FREKWENSIE, KORPUS, KORPUSGEBASEER, LEKSIKO-GRAFIE, SINONIEME, SJONA, SLENG, TREFWOORD

---

1.   Introduction

Corpora may be compiled (and used) for many different purposes in language research, including their lexicographic use. The majority of ALRI's research activities are either corpus-based or corpus-aided. In fact, research in corpus work is one of ALRI's basic and essential research areas (Chimhundu 2000: 5). The ALRI team's research activities have so far culminated in the development of corpora for two of Zimbabwe's main languages, Shona and Ndebele. Work is currently under way to develop corpora for four of ALRI's prioritised and Zimbabwe's officially recognised 'minority' languages, Kalanga, Nambya, Tonga and Shangani.

2.   A Brief Discussion of the Shona Corpus

The contents of the Shona corpus came from oral and written data. For oral data collection, undergraduate Shona students were sent out to tape-record interviews on almost all aspects of life, in all Shona-speaking districts of Zimbabwe and from males and females of different age groups. In the process of systematically collecting this oral material, details on the context of the discourse, date of interview, physical location, topic, setting and other relevant details were recorded. Extra-linguistic features such as hesitations, coughs and pauses were also recorded and marked. Some written data from Shona texts was also introduced into the corpus. The material then underwent the processes of transcription, encoding, proofreading, tagging and parsing. These processes are the main stages of corpus design. Transcription is the process of reducing an oral text to writing. Encoding is the keying in of data into a computer. Scanning refers to the process of electronically recognising written material that appear as hard copies and saving them as soft copies. Tagging is the process of assigning a specific code to each word in a text. Parsing involves checking tagging errors.

A discussion of how the above-mentioned processes were employed to produce the Shona corpus must be left to a more detailed report. Suffice it to say that at this point, oral material constitute seventy percent of the 2 600 000 running words that are in the current Shona corpus and written material thirty percent (Chabata 2000: 79). It should be noted that the Shona corpus could be viewed as a monitor corpus, since it is open-ended. Texts are continuously being added to it so that it gets larger and larger as more samples are added. A monitor corpus is important for ALRI, which specialises in dictionary making. In fact, monitor corpora, according to McEnery and Wilson (2001: 30), 'are primarily important in lexicographic work for they enable lexicographers to trawl a stream of new texts looking for occurrence of new words or for changing meanings of old words'.

The Shona corpus was utilised in the production of two Shona dictionaries: *Duramazwi ReChiShona* (DRC) (1996) and *Duramazwi Guru ReChiShona* (DGC) (2001). Whilst the compilation of DRC was corpus-aided, that is, its compilation was assisted with material from the Shona corpus, that of DGC was corpus-based.

3.   The Implications of Being Corpus-Based

Before discussing the degree to which DGC was corpus-based, it is perhaps necessary to survey the debates that have been conducted with regard to the idea of relying on a corpus in linguistic research. Reliance on a corpus would be biased towards an empiricist approach to the study of language that is dominated by the observation of naturally occurring data, typically through the medium of a corpus. Rationalists (notably Chomsky) have maintained that this approach has its limitations. Their main argument is that no one corpus can ever be regarded as a significant record of any language. Perhaps such an argument used to make sense at a time when texts were put on slips of paper and where relevant information could only be accessed manually. Then, there were only very small corpora. However, this is no longer the case. As McEnery and Wilson (2001: 31) put it, 'nowadays, the term "corpus" almost always implies the additional feature, machine-readable'. At present, researchers are coming up with machine-readable corpora that contain several billions of running words that can easily be searched and manipulated.

A corpus has the advantage that corpus-based observations are intrinsically more verifiable than introspectively based judgements. Empiricists observed that the type of sentence typically analysed by the introspective linguist is far removed from the type of evidence we typically tend to see occurring in the corpus. Empiricism maintains that the corpus does not only seem to be a more reliable source of frequency-based data but also provides the basis for a much more systematic approach to the analysis of language. There is, therefore, no doubt that a corpus is an essential linguistic tool. Since DGC was intended to be corpus-based, it meant that all headwords, senses, citations and other relevant linguistic information that would be required in the compilation of the dictionary would come from the Shona corpus. Whilst the Shona corpus was heavily relied upon for the majority of these items, there were instances when the editors of DGC had to disregard this corpus as shall be illustrated in the following sections.

4.   Disregarding the Corpus in the Treatment of Headwords

At times the editors of DGC disregarded the Shona corpus in their treatment of the words that they selected as headwords for the dictionary. This was particularly the case in two areas:

4.1   Headword Selection

Headword selection is one of the most crucial stages in compiling dictionaries because it is during this stage that the contents of a dictionary are determined. Comprehensive criteria defining the process of headword selection has to be set up and should be detailed in the style manual that guides the compilation of any dictionary. If headword selection is corpus-based, as was intended in the compilation of DGC, lexicographers have to rely heavily on frequency, that is, the number of times a word appears in the corpus. Thus, the most frequent words should be selected first, then the less frequent and ultimately the least frequent ones. Since it is not practically possible to include all the words of a language in a dictionary, it follows that some words have to be left out. DGC was intended to contain approximately 50 000 words. It was not possible to go beyond this number to prevent the dictionary from becoming too voluminous in size, too expensive to produce and also too highly priced for its target users.

It was, however, difficult solely to rely on the corpus when deciding on which words to include in or to exclude from the dictionary. The following are the 20 most frequent words in the Shona corpus, listed in descending order according to their frequency.

46 021 *kuti* (that, so that, in order that)
25 272 *kana* (when, although, even, or, if, whether)
10 505 *asi* (but, except)
 9 197 *zvino* (now)
 8 460 *munhu* (person)
 8 259 *saka* (hence, consequently, therefore, for this reason)
 7 840 *here?* (is that so?)
 7 064 *vanhu* (people)
 5 916 *chete* (only)
 5 781 *mwana* (child)
 5 766 *uyu* (this one)
 5 110 *ari* (who is)
 5 018 *ini* (me)
 4 660 *nokuti* (because)
 4 401 *iri* (this one)
 4 280 *iyi* (this one)
 4 137 *sei?* (how?/why?)
 4 093 *izvi* (these)
 4 011 *vana* (children)
 3 997 *iye* (him/her)

The words in the above frequency list are of not much value to a Shona lexicographer, especially a monolingual one. Neither are they of much value to the target audience of DGC that happens to be mother-tongue speaker-writers of the Shona language. The reason is that they are mostly function words. In fact,

the most frequent word in the Shona corpus, *kuti*, which occurs more than 46 000 times, is a conjunctive. No verbs are found in the list and only a couple of nouns such as *munhu*, *vanhu*, *mwana* and *vana*.

As far as headword selection is concerned, it would not make much sense, at least in monolingual Shona lexicography, to prioritise the most frequent word *kuti* over say, for example, either *rufu* (death) which occurs 258 times or *ivhu* (soil) which occurs 254 times in the Shona corpus. Thus, by prioritising certain less frequent lexical items over those that were the most frequent, but were suppletive and function forms, the Shona corpus was disregarded.

4.2   Presentation of Synonyms

Another instance where the editors of DGC did not strictly adhere to what features in the Shona corpus, is in the presentation of synonyms. According to Jackson (1988: 65), two words are said to be synonyms if they have the same meaning. He also notes that since the term 'meaning' can only be understood contextually, synonymy also needs to be defined in terms of contexts of use. He then proceeds to give a rather revised definition of the term 'synonym'. He maintains that two words are synonyms if they can be used interchangeably in all sentence contexts (Jackson 1988: 65). Examples of Shona synonyms would be *-mhanya* and *-rumba* both of which mean 'run'.

As a way of saving space, it had been decided that synonyms were to be defined only when it was deemed necessary. Otherwise, the more commonly used form would carry the definition and the less commonly used one(s) would be cross-referred to the commonly used form. Where in doubt, the strength of the corpus would help to determine the main headword (Mawema 2000: 218). This would be through the use of the frequency counts that have already been mentioned. The frequency counts were, however, disregarded in some cases, for example, when an indigenous word competed with an adoptive.

The general desire of the editors of DGC was to promote indigenous words as much as possible. However, at times indigenous words appear less frequently than adoptives in the corpus as can be seen from the following example. The English noun 'nurse' is rendered by two equivalents in the Shona language: *mukoti* and *nesi*. The former is indigenous whilst the latter is borrowed. Following the principle of prioritising indigenous words over adopted ones, the editors chose *mukoti* to carry the definition whilst they cross-referred *nesi* to *mukoti*. If one looks at the frequency counts in the Shona corpus, one finds that *nesi* appears more frequently (61 times) than *mukoti* (51 times). This example shows that the editors of DGC disregarded the Shona corpus in the presentation of some synonyms.

5.   Disregarding the Corpus in the Treatment of Sense

The editors of DGC at times disregarded the Shona corpus in their treatment of

sense. This is particularly noticeable in two areas:

5.1   Sense Selection

As has already been noted, a corpus is useful in dictionary making since it provides certain senses of words that lexicographers might not think of among themselves. In this regard the Shona corpus was quite useful to the editors of DGC. There were, however, instances when some senses that appear in the Shona corpus were deliberately omitted despite their occurrence in it. This was particularly so with some terms or senses that can be regarded as slang. According to Flexner and Wentworth (1975: vii), 'slang is an ever changing set of colloquial words and phrases that speakers use so as to establish group identity and solidarity'. It was noted earlier on that the Shona corpus comprises oral material that came from different groups of Shona speakers and that focus on various aspects of life. Among these groups of people were youths using Shona slang. Hence some slang found its way into the Shona corpus.

Editors of DGC were quite cautious when dealing with Shona slang. They decided to enter into the dictionary only slang that has become an integral part of the Shona language. They resolved to omit slang that was considered ephemeral. Thus, some senses that can be regarded as Shona slang, and were frequent in the corpus but were considered to be of ephemeral use, were omitted. An example is the term *chitunha* which in typical Shona refers to a corpse, the body of a dead human being. Shona slang extends the term to refer to a slaughtered chicken, the result of metonomy, a type of semantic transfer whereby one entity is taken to stand for another on the basis of some contextual relationship (Bonvillain 1993: 75). Although the 'second' sense appears in the Shona corpus, it was omitted in DGC for fear that it will be short-lived. A practice such as this disregards the Shona corpus.

5.2   Ordering of Senses

During the defining process, in cases where there were two or more meanings for a headword, senses were to be ranked, with the basic meaning appearing first. Where the basic meaning could not be ascertained, usage would determine the ranking of definitions. The literal sense would precede the metaphorical, idiomatic and proverbial senses. Frequency of occurrence would be considered with the aid of the corpus. The corpus was, however, only useful when there were two or three senses being dealt with and when all the senses could be found in it. In the case of some verbs, for example, the senses would sometimes be so many that it was difficult and problematic to handle them.

An example of such a problematic lexical item is the verb *-bata* (lit. touch, hold, catch). In addition to its basic senses, the verb has several other metaphorical and idiomatic ones. In DGC, the senses of the verb are listed as follows:

1.   to hold/touch
2.   to catch
3.   to work somewhere
4.   to do your work wholeheartedly
5.   to be firm (as in a planted seedling)
6.   to attack (as in disease)
7.   to be tight (as in small clothing)
8.   to arrest (as in arresting by the police)
9.   to discover someone doing something bad
10. to understand something
11. to pin (as in pinning a shirt)
12. to be firm and strong (as in something being made/being constructed)
13. to catch (as in catching a bus)
14. to treat (as in treating a subordinate)
15. to face a hindrance
16. to have a lot of money
17. to catch up
18. to be dense (as in a forest)

Altogether, there are eighteen senses listed. Whilst some of them occur in the corpus several times, some do not. However, some of those that occur nil times are also listed in the dictionary and, more so, even before some of those that occur several times. This is because they were found to be more important and closer to the primary meaning of the verb. This example of sense treatment also shows an instance where the corpus was disregarded.

6.   Conclusion

This article has shown that although the corpus is a very useful tool, especially in aiding some lexicographic decisions in corpus-based lexicography, there are times when lexicographers have to disregard it during the compiling process. It has highlighted the fact that being corpus-based does not necessarily have to imply being corpus-bound. This has been shown through focusing on head-word and sense selection with specific reference to corpus-based monolingual Shona lexicography. Most of the considerations in this article could, however, be true of the modus operandi in corpus-based lexicographic projects of other languages of the world.

References

Bonvillain, N. 1993². *Language, Culture and Communication.* Englewood Cliffs, New Jersey: Prentice-Hall.

Chabata, E. 2000. The Shona Corpus and the Problem of Tagging. *Lexikos* 10: 75-85

Chimhundu, H. 2000. *The Agenda for ALRI.* Harare: University of Zimbabwe.

Chimhundu, H. (Ed.). 1996. *Duramazwi ReChiShona*. Harare: College Press.

Chimhundu, H. (Ed.). 2001. *Duramazwi Guru ReChiShona.* Harare: College Press.

Flexner, S. and H. Wentworth. 1975. *Dictionary of American Slang*. New York: Crowell.

Jackson, H. 1988. *Words and their Meaning*. London/New York: Longman.

Mawema, M.B. 2000. Challenges Encountered in the Compilation of an Advanced Shona Diction-
ary. *Lexikos* 10: 209-224.

McEnery, T. and A. Wilson (Eds.). 2001. *Corpus Linguistics*. Edinburgh: Edinburgh University
Press.

Cultural Aspects in the Shona Monolingual Dictionary *Duramazwi Guru reChiShona*[*]

Nomalanga Mpofu, *African Languages Research Institute, University of Zimbabwe, Harare, Zimbabwe (nmpofu@arts.uz.ac.zw)*

Abstract: In the compilation of dictionaries, lexicographers also take cognisance of the culture which is inherent in a language. This article will look at the way the cultural aspect is interwoven in the practice of dictionary making. Language is at the core of culture and it is the major vehicle for the transmission of a people's beliefs and values. Language is also an expression of social structures and attitudes. No culture can exist which does not have a natural language at its centre. A language thus reflects a particular culture. Culture in this article will be taken to mean whatever a person must know in order to function in a particular society (Wardhaugh 1998: 215). The article will look at two aspects: (1) the interrelationship between language and culture and its bearing on lexicography; and (2) the treatment of cultural aspects in *Duramazwi Guru reChiShona* (2001). Examples used in this article are drawn from the advanced Shona monolingual dictionary *Duramazwi Guru reChiShona*, and other Shona dictionaries, both monolingual and bilingual.

Keywords: CULTURAL COMPETENCE, CULTURAL ASPECTS, CULTURAL MEANING, CULTURE, DICTIONARY, EUPHEMISM, EXTENDED MEANING, IDIOM, LEXICOGRAPHY, PROVERBS, SHONA, SHONA SOCIETY, TABOO

Opsomming: Kulturele aspekte in die Sjona- eentalige woordeboek. By die samestelling van woordeboeke neem leksikograwe ook kennis van die kultuur wat inherent aan 'n taal is. Hierdie artikel sal kyk na die manier waarop die kulturele aspek verweef is met die praktyk van woordeboekmaak. Taal lê aan die kern van kultuur en dit is die hoofmedium vir die oordrag van 'n volk se opvattings en waardes. Taal is ook 'n uitdrukking van sosiale strukture en houdings. Geen kultuur kan bestaan wat nie 'n natuurlike taal as middelpunt het nie. 'n Taal weerspieël gevolglik 'n bepaalde kultuur. Kultuur in hierdie artikel sal verstaan word om alles te beteken wat 'n persoon moet weet om in 'n betrokke maatskappy te funksioneer (Wardhaugh 1998: 215). Die artikel sal na twee aspekte kyk: (1) die onderlinge verband tussen taal en kultuur en die betrekking daarvan op die leksikografie; en (2) die behandeling van kulturele aspekte in *Duramazwi Gura reChiShona* (2001). Voorbeelde wat in hierdie artikel gebruik word, is afkomstig uit die gevorderde Sjona- eentalige woordeboek *Duramazwi Guru reChiShona*, en ander Sjonawoordeboeke, sowel eentalig as tweetalig.

Sleutelwoorde: EUFEMISME, IDIOOM, KULTURELE ASPEKTE, KULTURELE BETEKE-

NIS, KULTURELE VAARDIGHEID, KULTUUR, LEKSIKOGRAFIE, SJONA, SJONASAME-LEWING, SPREEKWOORDE, TABOE, UITGEBREIDE BETEKENIS, WOORDEBOEK

1.   Introduction

The purpose of this article is to highlight cultural aspects in *Duramazwi Guru reChiShona* (2001), the advanced monolingual Shona dictionary (hereafter referred to as *Duramazwi*). Culture, according to Wardhaugh (1998: 215), is defined as whatever a person must know in order to function in a particular society. In Shona society, this cultural knowledge entails knowing how to address elders and how to behave in their presence, the difference between being polite and being impolite, what not to say in public, and how to speak about certain subjects. All this knowledge is acquired socially through interaction. Wolff (2000: 302) notes that cultures and societies dictate to a person when to speak or to be silent, and which vocabulary items or formulaic expressions to choose or avoid.

Cultural aspects in this article will be taken to encompass everything that a person has to know in order to function effectively in a particular society. This includes language use which in turn includes veiled and loaded language such as euphemisms and sense extensions. The term will also cover those practices and beliefs characteristic of Shona tradition.

Several aspects found in the Shona language and culture were incorporated in *Duramazwi*. These cultural aspects include sensitive and offensive terms, terms with extended meanings, kinship terms, cultural practices, proverbs and idioms. These depict the Shona way of life, beliefs, norms and values. It is only those cultural aspects which are verbal which can be captured in a dictionary, yet in Shona there are a sizeable number of non-verbal cultural aspects that are part of day-to-day life. For instance, it is rude in Shona society to hand something to someone older than you using only one hand. It is also required that one claps one's hands when accepting something from someone older. There is no way that these cultural aspects can be included in a dictionary. It is the purpose of this article to analyse how cultural aspects were treated in *Duramazwi*. The article will address two aspects: (1) the interrelationship existing between language, culture and lexicography; and (2) the treatment of these cultural aspects in the *Duramazwi*.

2.   The Interrelationship between Language, Lexicography and Culture

Language is at the core of culture and no culture can exist which does not have at its centre a natural language. A language thus reflects a specific culture (Mpofu 2001: 246). Culture is a people's way of life. Two linguists, Edward Sapir (1921) and Benjamin Lee Whorf (1956), wrote extensively on the relationships existing between language and culture. Their findings came to be referred to as the Sapir–Whorf hypothesis which postulates that language and cul-

ture are inextricably related, so that one cannot be understood or appreciated without knowledge of the other. Although this article acknowledges this hypothesis, it will not be a sociological study of language, but an examination of the issues from a lexicographic point of view.

The primary purpose of any dictionary is meaning. But one cannot talk about the meaning of a word without recognising the wider social and cultural reality of the language in question. In other words, society's norms and expectations govern what should be included in a dictionary and how it should be treated. In any language, certain things are not expressed, not because they cannot be, but because people avoid talking about these things (Wardhaugh 1998: 234). According to Svensén (1993: 45), the use of language is influenced not only by linguistic norms but also by social standards. Certain words are socially charged because of being considered by many to be offensive. That is why certain entries in *Duramazwi* were given stylistic markers or labels, e.g. *tuko* (short for *chituko* — offensive words); and *nyadzo* (short for *chinyadzo* — taboo words). Some users would expect words considered by society as offensive or sensitive to be in a dictionary, but because of the linguistic taboo observed in Shona society concerning these words, they were marked or labelled, and defined in as neutral a way as possible. This usage labelling gives information on constraints on the use of these words. According to Kipfer (1984: 6), language is no longer considered as either correct or incorrect, but lexicographers and linguists agree that importance has to be placed on actual usage. In line with this view, Landau (2001[2]: 233) states that the decision of dictionary editors on what to label offensive or disparaging is based on their judgment of society's norm for the limits of reputable public behaviour.

As far as the treatment of these terms are concerned, the compilers of *Duramazwi* were faced with the challenge of how to be precise in defining them, yet at the same time conforming to the cultural norms of the language.

3.   Euphemisms

The use of euphemisms is also prevalent in the Shona language and society. A euphemism is a polite word or expression used to speak about something found unpleasant or embarrassing, such as death or sex. The purpose of euphemisms in language is thus to avoid directly mentioning certain things. In Shona, for example, one does not directly speak about death. When informing someone that a person has died, one does not usually use the more direct word *kufa*. People rarely say, *Nhingi afa* (so and so has died), but rather refer to death as *kupfuura* (to pass on*), kuenda* (to go), *kutisiya* (to leave us behind), *kutungamira* (to go ahead of others), *kufuka rako woga* (to be alone under your bedcovers), *kuenda kwamupfiganebwe* (to go to the place whose entrance is closed by a stone). They may also say, *Nhingi hakusisina* (so and so is no longer there). Such euphemisms are also meant to lessen the pain and fear associated with death. In *Duramazwi*, apart from the words *rufu* (death) or *kufa* (to die), one also finds

-*shaya* (to lose), and -*shayika* (to fail to be located/found). Even terms referring to burial are euphemistic, e.g. *kuchengeta* (to keep), *kuradzika* (to lay down), *kuviga* (to hide something), *kukotsa* (to hide something).

Sex in Shona society, like in other societies, is also one of the least talked about subjects in public. If people have to talk about it, they do so using veiled language. Hence in *Duramazwi* one finds entries such as -*sangana* (to meet), -*rara* (to sleep), and -*ziva* (to know something or someone) which also refer to sexual intimacy. The sexual act itself is commonly referred to as *bonde* (reed mat). Even terms that refer to pregnancy and giving birth are euphemistic. People prefer to use polite expressions such as *kuzvitakura* and *kuzvisenga* (to carry oneself), and *kutsika mwedzi* (to step on the moon) to refer to the state of pregnancy, and instead of the more direct word *kuzvara* (to give birth), *kubatsirwa* (literally, to be helped) and *kusununguka* (literally, to be free) to describe the act of giving birth.

Sex and anything that is associated with it is one of the most tabooed subjects in Shona, both in the public and private domain. The genitals or sexual organs are therefore also not directly referred to. Both the penis and vagina are politely called *sikarudzi* (creator of a clan). Other names euphemistically used for the penis are *mhuka* (animal), *chirema* (cripple), *chombo* (weapon), and *mbonausiku* (that which sees at night), and for the vagina *kunzira kwaamai* (the woman's passageway) or *mukana* (passageway).

Sexually transmitted diseases are also referred to indirectly in Shona. For instance, the general term commonly and politely used for these diseases is *chirwere chepabonde* (the disease of the reed mat) as opposed to direct terms such as *siki* (sexually transmitted disease), and more specifically *gonoriya* (gonorrhea) or *songeya* (syphilis). When one admits to be suffering from a sexually transmitted disease, one usually says *ndakarumwa* (I was bitten). Acquired immunodeficiency syndrome (AIDS) is referred to as *shuramatongo* (that which brings about desolate homes) because of the way it can wipe out whole families. The term *shuramatongo*, before its sense extension, means "a person who is an expert at something, either a sport or a craft" (Hannan 1984: 609, Chimhundu 2001: 932).

4.   Sense extensions

Words that have cultural meanings in Shona are mostly nouns and verbs with extended meanings or sense extensions. In order to understand such terms one needs to have cultural competence in the values of the Shona people. A particular group of people or a specific culture has certain things it reveres and observes. Frake (1964: 260) says of culture:

> Goodenough (1957) has proposed that a description of a culture — an ethnography — should properly specify what it is that a stranger to a society would have to know in order appropriately to perform any role in any scene staged by the society.

For instance, the verb *kukotsira* (to sleep) can also be used to refer to a slow or stupid person. Thus, if one says *Nhingi akakotsira* (so and so is asleep), it not only means that the particular person is sleeping but it can also mean that he/she is dull. *Munhu* literally means a person in Shona, but it also means that the person referred to is morally upright and has a good social standing. It is an insult to say of someone *Nhingi haasi munhu* (so and so is not a person) for this is a deriding and demeaning statement about a person's character. The sense of *munhu* when used in the context *ndinongova munhu* can mean that the person of whom is spoken, has no money at the moment of speaking. *Musikana* (girl) can also refer to a virgin and *mukomana* (boy) to a bachelor. *Bofu* is a blind person, but also refers to a dull student or person. While *chirema* is a cripple, it also refers to a heavily pregnant woman as well as to the male sexual organ. It can also be used to refer to someone who is unable to do anything useful or meaningful in life, or someone who cannot easily do what his/her peers do. *Kufamba-famba* means "to walk about" but it also refers to sexual immorality.

5.   Cultural terms

(a)   Kinship terms

Kinship terms are an integral part of the Shona societal system. There is a sizeable lexicon in Shona that refers to this cultural concept and most, if not all of these words with their different senses were captured in *Duramazwi*. A few can be picked for illustrative purposes. A young girl calls her father's older brother *babamukuru* (literally, older father = uncle) and his younger brother *babamunini* (literally, younger father = uncle) and they in turn call her *mwana* (child/daughter). She calls her mother's brother *sekuru* (uncle) and he in turn calls her *muzaya* (niece). Her father's sister's husband is called *babamukuru* (uncle) and he in turn calls her *mainini* (loosely, young wife), and at times can also refer to her as his wife because in traditional Shona culture when a woman dies, her brother's daughter can be an eligible wife for the widowed husband. In cases of barrenness on the wife's part, the uncle could also marry her. The young Shona girl furthermore calls her mother's older sister's husband *babamukuru* (uncle) and the younger sister's husband *babamunini* (uncle). In *Duramazwi* the entry for *babamukuru* covers all these senses:

> **babamukuru** *DKDKK z 1a. Mubereki kana muramu wechirume mukuru* (parent or sister's husband or father's sister's husband). *1 Uyu mukoma wababa* (father's elder brother). *2 Uyu murume wavatete* (father's sister's husband). *3 Uyu murume wasisi vako, kana uri mukadzi* (sister's husband, if you are female). *4 Uyu murume womukoma womudzimai wako* (wife's elder sister's husband) …

According to Wardhaugh (1998: 226), it is important to remember that when a term is used in a kinship system, it carries with it ideas about how such people

ought to behave toward others in the society using that system. This also holds true of Shona society where the term by which you refer to a member of the family, especially of the opposite sex, governs your behaviour towards that person. For instance, if a person is your *muramu* (wife's sister, wife's brother's daughter, or husband's younger brother), there is linguistic license whereby you may joke with that person regardless of the context. In Shona there is the saying *Muzukuru mukadzi* (Your sister's daughter is your wife) and the proverb *Muzukuru mudonzvo wepwa, ukanzwa nzara unomenya uchidya* (Your sister's daughter is like a sugar-cane walking-stick, if you feel hungry you can peel and eat it, that is, a man can take his sister's daughter as his wife should he so wish). This saying and proverb show that the relationship existing between a man and his sister's daughter does not restrict them from falling in love with each other. The opposite situation prevails when a man refers to a female kin as *ambuya* (wife's brother's wife), while she in turn calls him *mukwasha* (husband's sister's husband). In this case, linguistic and social distance is observed and in strict societies the two are not even allowed to shake hands or sit in the same room for extended periods of time.

(b)   The roora (lobola) custom

Duramazwi also portrays customs considered to be culturally important to the Shona people. One of these is the *roora* (*lobola*) custom. Roora is the money and/or cattle a man pays to his in-laws as a way of seeking permission to live with their daughter as his wife (Mpofu 2001: 247). According to Kavanagh (2000: 103), the vocabulary of a language can provide some evidence of what is considered culturally important, and there may be several words for the most familiar items and concepts. There is a sizeable lexicon on the *roora* custom in the Shona language. These cultural terms were also incorporated in *Duramazwi*, which is an indication of their importance to the Shona people.

One such term is *rusambo* or *rugaba*, money, excluding the cattle, paid as *roora*. Another example is the term *makandinzwanani* (literally, who informed you of me?), which refers to money paid by the son-in-law as a way of introducing himself to his in-laws. *Vhuramuromo* or *zaruramuromo* (open the mouth), another of the roora custom settlements, is money paid by the son-in-law to enable the marriage negotiations to start.

One also finds the terms *danga* (literally, a kraal, but also cattle paid as part of roora), *munyai* (go-between), *masungiro* (goat or money paid to in-laws when the daughter goes to deliver the first child), *matekenyandebvu* (money paid to the father-in-law for rearing the daughter, literally, for the fact that the daughter used to play with her father's beard), and *mapfukudzadumbu* (money paid to the mother-in-law in appreciation for the time she carried her daughter in her womb). In the roora custom, the father-in-law makes many impositions regulated by the culture within which this custom is enshrined.

(c)   Proverbs and idioms

Proverbs were given lemma status in *Duramazwi*. *Duramazwi* is divided into two parts, the first part being the A–Z section of the dictionary including idioms and the second part consisting of proverbs and pithy sayings. Proverbs are important to the Shona people because of their didacticism. They were given lemma status in *Duramazwi* because they form an integral part of Shona discourse. Proverbs can be applied to real life situations as warnings against certain types of behaviour or as words of advice. Among the Shona, proverbs are a versatile medium of communication in that they can be used on any occasion and in any context. According to Pongweni (1989: 1), proverbs derive from, and express the communal view of the world based on firsthand experience.

In *Duramazwi* one thus finds a sizeable number of proverbs, which, when looked at carefully, show similar themes running across them such as communal living and working together, valuing the little you have and discouraging envy. Pongweni (1989: 13) notes that figurative language of all types and from all communities derives its imagery from the environment in which the speakers of the language live. The following sample of proverbs instruct against individualism and being unappreciative of that which you possess:

> *Chara chimwe hachitswanyi inda* (literally, one finger cannot crush a louse, i.e., individualism has no gain)

> *Rume rimwe harikombi churu* (literally, one man cannot surround an anthill, i.e., individualism has no gain)

> *Tsapata rukukwe hazvienzani nekurara pasi* (literally, a worn-out reed mat is better than sleeping on the floor, i.e., half a loaf is better than nothing)

> *Chembere mukadzi, hazvienzani nekurara mugota* (literally, an old woman is a wife, she is better than sleeping alone in a bachelor's hut, i.e., half a loaf is better than nothing)

Idioms, like proverbs, are also a versatile feature of figurative language used on various occasions. Examples are: *kuruma nzeve* (to bite someone's ear = to advise), *kubaya dede nomumukanwa* (to pierce the baboon through the mouth = to do something accurately), *kudya magaka mambishi* (to eat raw cucumbers = to have a fiery temperament), *kubata kumeso* (to touch one's face = to deceive), and *kugara maoko* (to sit on one's hands = to be lazy and unproductive).

6.   Conclusion

This article has attempted to show that one cannot speak of language without referring to the culture embodied in that language. A dictionary is a record of actual usage. It is therefore expected to meet the different needs of its different users. It should also be able to tell the user what the word means and how that

word may be used. The aspect of usage is of relevance to culture because words are used differently in different cultures. What may be deemed proper in one culture may be viewed differently in another.

Cultural items form an integral part of a language, hence they merit inclusion in a dictionary. Their importance in a dictionary hinges on the fact that they mirror what is considered important by the people whose language is being described and who participate in the culture which embodies that language.

References

Chimhundu, H. (Ed.). 2001. *Duramazwi Guru reChiShona*. Harare: College Press.

Frake, C.O. 1964. How to Ask for a Drink in Subanum. Pride, J.B. and J. Holmes. 1972. *Sociolinguistics*. Harmondsworth: Penguin Books.

Goodenough, W.H. 1957. Cultural Anthropology and Linguistics. Garvin, P.L. (Ed.). 1957. *Report of the Seventh Round Table Meeting on Linguistics and Language Study*. Washington, DC: Georgetown University Press.

Hannan, M. 1984. *Standard Shona Dictionary*. Harare: College Press.

Kavanagh, K. 2000. Words in a Cultural Context. *Lexikos* 10: 99-110.

Kipfer, B.A. 1984. *Workbook on Lexicography: A Course for Dictionary Users with a Glossary of English Lexicographical Terms*. Exeter: University of Exeter.

Landau, S.I. 2001². *Dictionaries: The Art and Craft of Lexicography*. Cambridge: Cambridge University Press.

Mpofu, N. 2001. Problems of Equivalence in Shona–English Bilingual Dictionaries. *Lexikos* 11: 242-251.

Pongweni, A.J.C. 1989. *Figurative Language in Shona Discourse*. Gweru: Mambo Press.

Sapir, E. 1921. *Language: An Introduction to the Study of Speech*. New York: Harcourt, Brace and Co.

Svensén, B. 1993. *Practical Lexicography: Principles and Methods of Dictionary Making.* New York: Oxford University Press.

Wardhaugh, R. 1998. *An Introduction to Sociolinguistics*. Oxford: Basil Blackwell.

Whorf, B.L. 1956. *Language, Thought and Reality*. Edited with an introduction by J.B. Carroll. Cambridge, Mass.: Massachusetts Institute of Technology / New York: John Wiley and Sons.

Wolff, H.E. 2000. Language and Society. Heine, B. and D. Nurse. 2000. *African Languages: An Introduction.* Cambridge: Cambridge University Press.

Uitdagings vir die Afrikaanse leksikograaf ten opsigte van Nederlandse lemmata in 'n vertalende Nederlands–Afrikaanse woordeboek[*]

Corlietha Swart, *Departement Afrikaans, Vista Universiteit, Bloemfontein, Republiek van Suid-Afrika (swart-c@blenny.vista.ac.za)*

Opsomming: *Voor Pampus liggen* het 'n oordragtelike betekenis en idiomatiese status in Nederlands verwerf, maar is gewoonlik onbekend aan die Afrikaanse leksikograaf buite die Nederlandse konteks. Vir 'n Afrikaanse leksikograaf met sy unieke verwysingsraamwerk en kulturele en sosiale waardeoordele word dit slegs in 'n Nederlandse milieu duidelik waarom die leksikografiese redigering van Nederlandse lemmata in 'n vertalende Nederlands–Afrikaanse woordeboek so 'n groot uitdaging inhou, ten spyte van die feit dat die twee tale so nou verwant is. Taalversteuring en -beïnvloeding vind makliker plaas tussen nou verwante tale. Taalrelatiwisme impliseer dat elkeen van die betrokke tale 'n unieke soort perspektief op die taaleie werklikheid bied vanweë 'n unieke lewens- en wêreldbeskouing wat 'n inherente deel uitmaak van 'n betrokke moedertaal. Belangrike metaleksikografiese riglyne sluit die oordeelkundige aanwending van pragmatiese leiding (bv. etikette), goedgekose, verhelderende sitate en poëme en relevante gebruiksleiding in. Dit kan die teikengebruiker van 'n vertalende woordeboek 'n insig gee in die unieke en kultuurspesifieke woordeskatelemente wat die tale van mekaar onderskei.

Sleutelwoorde: NOU VERWANTE TALE, VERTALENDE NEDERLANDS–AFRIKAANSE WOORDEBOEK, BRONTAAL, DOELTAAL, SKUIFFUNKSIE, OMKERING, KOMBINASIELEMMA, UNIEKE VERWYSINGSRAAMWERK, SOSIO-KULTURELE WAARDEOORDELE, TAALVERSTEURING EN -BEÏNVLOEDING, TAALRELATIWISME, SAPIR–WHORF-HIPOTESE, LINGUISTIESE RELATIWITEIT EN DETERMINISME, METALEKSIKOGRAFIESE RIGLYNE, KOLLOKASIES, ETIKETTE, PRAGMATIESE LEIDING, SITATE, POËME, GEBRUIKSLEIDING, RAME, METATAAL, TWEETALIGE LEKSIKOGRAFIE

Abstract: Challenges for the Afrikaans Lexicographer with regard to Dutch Lemmata in a Dutch–Afrikaans Translation Dictionary. *Voor Pampus liggen* has acquired a metaphorical meaning and idiomatic status in Dutch, but is generally unknown to the Afrikaans lexicographer outside the Dutch context. For the Afrikaans lexicographer with his unique frame of reference and cultural and social value judgements, it becomes clear only in a Dutch milieu why the lexicographical editing of Dutch lemmata in a Dutch–Afrikaans translation

---

[*] Hierdie artikel is 'n verwerking van 'n lesing wat aangebied is op die LVSA-jaarkongres by die Universiteit van die Vrystaat, Bloemfontein, 2–4 Julie 2001.

dictionary is such a big challenge, despite the fact that the two languages are so closely related. Language interference and influence occur more readily between closely related languages. Language relativity implies that each of the languages concerned gives a unique perspective on its own reality, because of a unique life- and world-view which form an inherent part of a particular mother tongue. Important metalexicographic guidelines include the judicious application of pragmatic signals (e.g. labels), well-chosen clarificatory citations and made-up examples and relevant guidance as to usage. This can give the target user of such a dictionary an insight into the unique and culture-specific lexical elements which distinguish the two languages from each other.

Keywords: CLOSELY-RELATED LANGUAGES, DUTCH–AFRIKAANS TRANSLATION DICTIONARY, SOURCE LANGUAGE, TARGET LANGUAGE, SHIFT FUNCTION, REVERSAL, COMBINATION LEMMA, UNIQUE FRAME OF REFERENCE, SOCIO-CULTURAL VALUE JUDGEMENTS, LINGUISTIC INTERFERENCE AND INFLUENCE, LANGUAGE RELATIVISM, SAPIR–WHORF HYPOTHESIS, LINGUISTIC RELATIVITY AND DETERMINISM, METALEXICOGRAPHICAL GUIDELINES, COLLOCATIONS, LABELS, PRAGMATIC SIGNALS, CITATIONS, MADE-UP EXAMPLES, GUIDANCE FOR USE, FRAMES, METALANGUAGE, BILINGUAL LEXICOGRAPHY

1.   Inleiding

Vertel 'n Afrikaanssprekende aan 'n Nederlandssprekende dat hy " 'n klomp groente" gekoop het, sien die Nederlandssprekende in sy geestesoog 'n groot houtklomp volgepak met groente. *Klomp* verwys in Nederlands eerstens na "skoeisel wat bestaan uit 'n uitgeholde stuk hout" en tweedens na " 'n vormlose massa". In Afrikaans beteken *klomp* egter eerstens " 'n groot versameling, 'n hele aantal of 'n groep" en tweedens verwys dit na " 'n Hollandse houtskoen". Die Afrikaanssprekende verwys dus na " 'n groot versameling groente" en die Nederlandssprekende verstaan " 'n houtskoen vol groente".

Verder vind 'n Afrikaanssprekende ook gou uit dat hy nie in 'n Nederlandse winkel kan sê dat die winkelassistent hom nie hoef te help nie, want hy "loer net 'n bietjie rond". In Afrikaans word *loer* in dié konteks meer neutraal aangevoel. *Loer* in Nederlands impliseer hier egter iets negatiefs: jy wag net jou kans af om iets oneerliks aan te vang, met ander woorde, jy kyk skelmpies rond met die bedoeling om byvoorbeeld te steel as jy die geleentheid kry! Hierdie verwarring kan voorkom word met behulp van 'n Nederlands–Afrikaanse woordeboek. Die grootste gros woorde in Nederlands is egter maklik verstaanbaar vir Afrikaanssprekendes ... Of is hulle werklik?

2.   Die behoefte aan 'n vertalende Nederlands–Afrikaanse woordeboek

Merkwaardig genoeg bestaan daar, ten spyte van die feit dat 'n behoefte aan 'n Nederlands–Afrikaanse woordeboek gereeld geopper word, nog nie so 'n woordeboek nie, miskien omdat die mite van die algehele onderlinge verstaan-

baarheid tot dusver lewendig gehou is. Reeds in 1986 verklaar Jansen en Oli-
vier in *Praktiese Nederlands* (p. 141) dat daar geen Nederlands–Afrikaanse woor-
deboek bestaan nie en dat die enigste alternatief Prisma se *Nederlands–Engelse
sakwoordeboek* is. Die leemte met betrekking tot so 'n woordeboek word in hulle
boek aangevul met 'n bylaag van geselekteerde woorde waarvan sekere beteke-
nisonderskeidinge in Afrikaans óf onbekend óf anders is. Ook Dorsman (1998:
10) wys op hierdie gebrek in die tydskrif vir vertaalwetenskap *Filter*: "Waar het
aan ontbreekt is een goed woordenboek Afrikaans–Nederlands en vice versa."
Een van die weinige resente werke wat wel die belangrikste verskille en afwy-
kinge in die Nederlandse en Afrikaanse woordeskat onder die loep neem, is
Yvette Stoops se werk *Bobbejane of bavianen* (1995). Hierin betreur sy die afwe-
sigheid van vergelykende bronne wat die waarde van so 'n woordeboek weer-
eens belig. Stoops (1995: 8) verklaar: "Wie belangstelling heeft voor het Afri-
kaans als taal, vind er nu in het Nederlands nagenoeg niets over en wat erover
bestaat, is dikwijls verouderd, oppervlakkig of onnauwkeurig, ofwel uiterst ge-
specialiseerd."

3.   Die nut van 'n woordeboek vir nou verwante tale

Die eerste vraag wat by 'n mens opkom en wat ook as vertrekpunt geneem kan
word, is wat die nut van 'n uitgebreide woordeboek vir nou verwante tale is,
aangesien sprekers van die een taal, juis vanweë die noue verwantskap, die
anderstalige spreker met gemak kan verstaan. By nadere leksikografiese onder-
soek blyk dit egter duidelik dat verwarring juis vanweë dieselfde noue ver-
wantskap tussen leksikale items kan insluip, veral as die diverse kontekstuele
faktore waarin leksikale items werksaam is, verreken word. Dit is lank nie
meer so vanselfsprekend dat Afrikaans en Nederlands sonder meer vir respek-
tiewelik die Nederlands- of die Afrikaanssprekende verstaanbaar is nie.

4.   'n Omkeerbare, gerekenariseerde, multifunksionele vertalende Ne-
derlands–Afrikaanse woordeboek in bewerking

Op die oomblik is daar wel so 'n gerekenariseerde, meerfunksionele woorde-
boek in bewerking onder leiding van prof. Willy Martin van die Vrije Universi-
teit, Amsterdam en prof. Rufus Gouws van die Universiteit van Stellenbosch,
tesame met talle medewerkers. Nederlands is die brontaal, die basiswoorde-
skat wat vrylik (gestruktureerd en beredeneerd) kan omkeer om die doeltaal
Afrikaans te verreken, en gevolglik lei tot die ontstaan van 'n omkeerbare data-
basis van die brontaal en doeltaal onderskeidelik.

4.1   Die doel van die beoogde woordeboek

Die Nederlands–Afrikaanse woordeboek het die beskrywing van die leksikon

van twee nou verwante tale deur middel van 'n nuut aangepaste leksikografiese beskrywingsmodel ten doel en hierdeur die stimulering van leksikografiese studies in Suid-Afrika, Nederland en Vlaandere. Die onderskeie tale sal onafhanklik 'n verteenwoordigende weerspieëling bied van die leksikon van hedendaagse Nederlands én Afrikaans.

4.2   'n Verbindingsinstrument tussen tale en kulture

So 'n woordeboek sal op taalkundige en letterkundige vlak wedersydse toeganklikheid en interkulturele kommunikasie bevorder deurdat onverstaanbare woorde en uitdrukkings in Nederlandse en Afrikaanse werke daarin nageslaan kan word. Volgens die haalbaarheids- en definisiestudie van die woordeboek sal die werk as 'n verbindingsweg tussen tale en kulture dien.

4.3   'n Kombinatoriese woordeboekmodel vol moontlikhede in 'n veeltalige konteks en as model vir ander taalpare

Die kern van leksikale aanbieding in die woordeboek is die kombinatoriese data met die verskillende datasones. Dié benadering bied 'n nuwe perspektief as aangepaste leksikografiese infrastruktuur wat selfs wyer as 'n tweetalige konteks is en ook vir 'n veeltalige konteks (soos Suid-Afrika met sy elf amptelike tale) aangewend kan word. Die waarde van die projek is dus meerdoelig, aangesien die visie van so 'n aard is dat die woordeboekmodel ook vir ander taalpare, soos Sesotho en Tswana, as nou verwante tale binne die Suid-Afrikaanse konteks, as model kan dien.

Volgens Renders (en medewerkers in die haalbaarheids- en definisiestudie van die woordeboek wat in 1999 van stapel gestuur is) maak Nederlands intrinsiek deel uit van die verlede (en hede) van die Afrikaanse taal en kultuur in Suid-Afrika. Net so behoort ook die kulturele produkte wat 'n stamverwante taal en gemeenskap voortgebring het, die belangstelling van die Nederlandstalige gemeenskap te wek. Hy noem dat kennisname van die kultuurgeskiedenis en kultuurprodukte van die Afrikaanssprekende die Vlaming en Nederlander die geleentheid bied om hul blik op die wêreld te verruim en hul ervaringshorison uit te brei. Via Afrikaans word dus toegang verkry tot Suid-Afrika en dien die Afrikaanse voedingsbodem en -kultuur vir 'n besonder verrykende ervaring met taalverwantskap as hoofmotivering vir samewerkingsooreenkomste tussen die twee nou verwante tale. Nouer kontak tussen Nederland en Suid-Afrika die afgelope dekade het gelei tot 'n groter wordende openheid en belangstelling in mekaar se kulture. Daarenteen het 'n afname in kennis van die kontaktaal oor die afgelope eeu heen gelei tot 'n behoefte aan 'n Nederlands–Afrikaanse woordeboek as hulpmiddel by uitstek om taalkundige misverstande uit die weg te ruim, insig in en 'n beter begrip van die onderskeie kontekste te verseker en sodoende doeltreffende kommunikasie te bewerkstellig.

5.   Die onverwagse labirint van nou verwante tale se betekenisonder-
skeidings

Alhoewel Afrikaans uit sewentiende-eeuse Nederlands ontwikkel het, het dit daarna 'n eie koers ingeslaan en vind die Nederlands- en Afrikaanssprekende dat hulle soms besig is om te verdwaal in die labirint van die twee nou ver- wante tale se woorde waarvan die betekenisonderskeidings soms wesenlik van mekaar verskil.

5.1   Verskillende ruimtelik-kulturele kontekste

Afrikaans het wél aan die begin van die vorige eeu, wat sy spelling en woor- deskat betref, aansluiting by Nederlands gesoek, maar is ook beïnvloed deur ander tale soos Engels, Frans, Duits en Maleis-Portugees. Mettertyd het in- vloede van ander inheemse tale en van die Suid-Afrikaanse bodem ook inslag begin vind vanweë die konstante taalkontak as gevolg van veeltaligheid en multikulturaliteit. Die resultate daarvan is duidelik in hedendaagse Afrikaans.

Die ruimtelik-kulturele konteks waarin Afrikaans en Nederlands afson- derlik en geïsoleer van mekaar in hul onderskeie lande ontwikkel, is inder- daad kontinente van mekaar verwyderd. Elkeen van die tale het weens maat- skaplike ontwikkeling 'n rykdom kulturele erfgoed opgebou, wat natuurliker- wys in die onderskeie tale se woordeskat neerslag gevind het en wat die dina- miese veelsydigheid en ongebreidelde vitaliteit van die hedendaagse samele- wing en kultuur weerspieël. Beide tale is ontwikkel tot volwaardige kommuni- kasiemiddele op alle gebiede van die maatskaplike lewe.

Dat die twee tale nie so toeganklik is vir onderskeie gebruikers soos met die eerste oogopslag blyk nie, beklemtoon 'n vertaler soos Riet de Jong-Goos- sens (1998: 85): "Omdat Afrikanen en Nederlanders elkaar redelijk verstaan, denken mensen dat het gemakkelijk moet zijn om Afrikaans te lezen. Natuur- lijk komt een groot deel van het lexicon overeen, maar er is ook een groot deel waarvan de uiterlijke verschijningsvorm dezelfde is, maar de betekenis ver- schillend." In dié verband verklaar ook Dorsman (1998: 12) dat Afrikaans 'n ánder taal is, miskien nou verwant aan Nederlands, maar dáárom juis word vertaling bemoeilik. Vanweë ingrypende veranderinge waaraan Afrikaans blootgestel is, beskou hy dit sedert die 1950's geensins meer as 'n aanhangsel van Nederlands nie.

Vanweë politieke, maatskaplike en industriële veranderinge het daar die afgelope dekade 'n verruiming (waaronder onder andere neologismes), ver- nouing en verandering op woordeskatvlak in Afrikaans plaasgevind. Tipiese Afrikaanse lewensituasies vind uiteraard neerslag in talryke woorde, uitdruk- kings en spreekwoorde, wat die behoefte aan 'n volledige en wetenskaplik ver- antwoorde tweetalige Nederlands–Afrikaanse woordeboek soveel dringender maak, aangesien bestaande bronne, woordelyste en internetvertaalwebwerwe óf verouderd óf onvolledig óf ontoereikend is. (Vergelyk byvoorbeeld die inte-

ressante, maar beperkte webwerf *101 Valse Vriende* <http://www.geocities.com/Athens/Acropolis/9756/> waar die verskille tussen Nederlands en Afrikaans bespreek word.)

5.2 Gemeenskaplike en eiesoortige idiomatiese uitdrukkings, kollokasies en woorde in Nederlands en Afrikaans

'n Groot veld wat nog braak lê en wat deur die nuwe woordeboek gedek sal word, is 'n omvattende sinchroniese en diachroniese ondersoek na die omvang van die verskille in die idioomskat van Nederlands en Afrikaans. Beide tale deel 'n groot aantal gemeenskaplike idiomatiese uitdrukkings. Die teendeel is egter ook waar: elkeen van die tale beskik oor 'n groot aantal eiesoortige idiomatiese uitdrukkings wat wedersydse begrip kompliseer. Dit geld natuurlik nie net idiome nie, maar ook die woordeskat self. Vergelyk Nederlandse begrippe soos *berm* vir " 'n strook grond langs dyke of die grondafsaksel in dyke", *strippenkaart* vir " 'n reekskaartjie vir verskillende openbare Nederlandse vervoermiddels", *metro* vir "die ondergrondse spoorweg", *lucifer* vir " 'n vuurhoutjie", *mus* vir " 'n Afrikaanse mossie" (terwyl die Nederlandse *muts* verwys na 'n hoofbedekking wat in Afrikaans as *mus* bekend is), *toetje* vir "nagereg", *zaklantaren* vir " 'n flitslig", *prullebak/vuilnisbak* vir " 'n vullisblik of 'n vuilgoedblik", *nijlpaard* vir " 'n seekoei" en *paraplu* vir " 'n sambreel".

'n Tipies Nederlandse uitdrukking soos *voor Pampus liggen* het oordragtelike betekenis en idiomatiese status in Nederlands verwerf, maar is normaalweg onbekend buite 'n Nederlandse konteks. Alle Nederlanders weet egter dat *Pampus* oorspronklik 'n sandbank in die Ijsselmeer naby Muiden was, wat later ontwikkel is in 'n mensgemaakte militêre forteiland met kanonne vir die beskerming van Amsterdam teen vyandelike skepe in die Noordsee. Die sandbank het soms vir lang tye tydens ebgety skepe verhinder om verder te vaar en die bemanning het hulle tydens dié lang wag oorgegee aan eet en drink, vandaar die betekenis van *voor Pampus liggen*: "jou ooreet en -drink terwyl jy niks (kan) doen om voort te beweeg nie".

Vir die Afrikaanse leksikograaf met sy unieke verwysingsraamwerk en kulturele en sosiale waardeoordele, word dit in 'n Nederlandse milieu duidelik waarom die leksikografiese redigering van Nederlandse lemmata in 'n vertalende Nederlands–Afrikaanse woordeboek so 'n groot uitdaging inhou, ten spyte van die feit dat dit twee nou verwante tale is wat hanteer word. Want hier gaan dit om taalrelatiwisme.

5.3 Taalrelatiwisme (die Sapir–Whorf-hipotese)

Taalrelatiwisme impliseer dat elkeen van die betrokke tale 'n unieke soort perspektief op die taaleie werklikheid bied vanweë 'n unieke lewens- en wêreldbeskouing wat 'n inherente deel uitmaak van 'n betrokke moedertaal, in hierdie geval onderskeidelik Afrikaans en Nederlands. Die Sapir–Whorf-hipotese is 'n

kombinasie van linguistiese relatiwiteit (kultuur beïnvloed taal, vandaar die onbeperktheid van tale se diversiteit) en linguistiese determinisme (taal bepaal die lewens- en wêreldbeskouing).

Volgens die sterk deterministiese weergawe beleef die mens die werklikheid in terme van kategorieë en onderskeidings wat in die betrokke taal geënkodeer is, terwyl die swak deterministiese weergawe, daarteenoor, impliseer dat taal nie 'n persoon se denkprosesse beïnvloed nie, maar wel die menslike geheue, persepsies en gedagtes affekteer. Die swak relatiwistiese weergawe impliseer dat alle taalgebruikers fundamenteel oor ooreenstemmende konseptuele raamwerke beskik, wat betref abstrakte konsepte soos tyd en ruimte.

6.   Die tweetalige Nederlands–Afrikaanse woordeboek

6.1   Infrastruktuur = RBN ↔ OMBI

Die onderliggende vertrekpunt van die woordeboek is op 'n eentalige Nederlandse databasis geskoei. Hieruit vloei die tweetalige Nederlands–Afrikaanse woordeboek se infrastruktuur waarin daar eerstens, as leksikale databank en betroubare referensiepunt van lemmata, 'n Referentiebestand Nederlands, verkort tot RBN, ontstaan. Dit word gekoppel aan 'n teenoorgestelde kant met omkeringsmoontlikhede vir die skep van Omkeerbare bilinguale bestanden, ook verkort tot OMBI, waaraan Afrikaanse lekseme gekoppel word.

Die abstrakte datakategorieë van die tweetalige woordeboek bestaan onder andere uit die trefwoord met sy vorm, betekenisomskrywing/resumé/vertaalprofiel (waaronder ook pragmatiese gebruik ressorteer), kombinatoriek en 'n afdeling vir kommentaar (waaronder kontraste soos "valse vriende", kultuurgebonde en taalspesifieke items en diskoersmerkers ondergebring word). Daar word by elke polisemiese betekenisfaset van 'n woord in die onderskeie tale leksikale kommentaar gegee. Leksikografiese kommentaar word vanuit kontrastiewe oorweginge gegee, veral by "valse vriende" en kultuurgebonde woorde of woorde sonder geleksikaliseerde vertaalekwivalent in die doeltaal, wat ook "leksikale gapings" genoem kan word. Vergelyk in hierdie verband tipies Afrikaanse begrippe (wat in sommige gevalle ook beïnvloed is deur inheemse Suid-Afrikaanse tale), soos *Afrikatyd, biltong, lobola, mahala, slaphakskeentjies, souskluitjies, toyi-toyi,* ensovoorts, teenoor tipies Nederlandse begrippe soos *elfstedentocht* (skaatswedstryd langs elf Friese stede, met Leeuwarden as begin- en eindpunt), of *terp* (heuwel in Friesland waarheen in geval van 'n oorstroming gevlug kan word). Waar daar nie 'n Nederlandse idioom, kollokasie of woord vir 'n Afrikaanse idioom, kollokasie of woord bestaan nie, word die Afrikaanse weergawe ingesleutel as "brontaal". Die skuiffunksie tussen die twee tale word dus geaktiveer tydens elke lemmabewerking en word deur middel van 'n strukturele merker aangedui. Dit is egter nie slegs die verskille tussen lemmata wat aandag geniet nie. Ook die ooreenkomste word wetenskaplik benader waar 'n woord soos *tafel* identies in vorm en betekenis is. Ook

woorde waarvan die betekenis ooreenkom, maar waarby daar sistematiese morfologiese verskille bestaan, word verreken, veral by werkwoorde, soos die Afrikaanse *ontsnap* teenoor die Nederlandse *ontsnappen*.

6.2   Latere praktiese aanpassings in OMBI se infrastruktuur

Tydens 'n besoek aan prof. Willy Martin by die Departement Lexicologie, Vrije Universiteit: Amsterdam, is daar op OMBI gewerk om leemtes in die rekenaar-program, vanuit die Afrikaanse leksikograaf se perspektief, te probeer identi-fiseer. By redigering van die letter J se omkering na Afrikaans, is 'n leemte in die rekenaarprogram geïdentifiseer, naamlik die onderskeiding van die mor-fologiese waarde van die vertaling van die voorbeelde, bekend as *kontrastiewe relevansie* en *niekontrastiewe relevansie.* 'n Verdere behoefte was die duideliker aantoon van die gelyktydige voorkoms van idiome. Dit het gelei tot die invoe-ging van bykomende kodesleutels in die rekenaarprogram: CR en NCR om die kontrastiewe relevansie (of die afwesigheid daarvan) aan te toon en ID-CR en ID-NCR in die geval van idiome. Onder kontrastiewe relevansie word nie bloot "vormlike" afwykings verstaan nie, maar ook leksikale en grammatikale kon-tras.

6.3   Verdere kenmerke van die OMBI-datakategorieë (kodesleutels, leksi-kografiese kommentaar, ensovoorts)

Kodesleutels wat in die rekenaarprogram ingebed is, sluit onder andere in die @ (teen)-teken by "valse vriende", die # (hekkie)-teken wat aandui dat Neder-lands en Afrikaans vormlik absoluut identies is en die % (persentasie)-teken by leksikale gapings of kulturele items. Selfs fynere aspekte soos ortografiese ver-skille tussen woordpare (byvoorbeeld weeksdae wat in Afrikaans met hooflet-ters en in Nederlands daarsonder geskryf word), en konnotatiewe, denotatiewe en kombinatoriese verskille (wat ook die ko-teks en konteks van woorde in-sluit) word verreken. Ook die verskille tussen gegewe voorbeelde van verwan-te tale word op woordvlak en woordkombinasievlak bespreek. Waar kontras-tiewe verskille lei tot gebruiks- en/of interpretasieprobleme, word dit toegelig met leksikografiese kommentaar.

Anders as by klassieke vertaalwoordeboeke en bestaande beperkte Afri-kaans–Nederlandse woordeboeke en -lyste, verg hierdie benadering 'n kon-trastiewe behandeling met die noodsaaklike afdeling vir kommentaar wat lin-guistiese aspekte by sommige woordpare (waarvan die vorm verskil) en ook by uitdrukkings betref. 'n Ander benadering word gevolg by woordpare waar daar nie 'n verskil is nie. Die datakategorieë/beskrywingsmodel bly dieselfde, maar die omvang en aanbieding van die doeltaalbewerking word bepaal deur die ooreenkomste of verskille tussen die morfologiese, semantiese of pragma-tiese eienskappe van die bron- en doeltaalwoorde. Elke woordeboekartikel be-skik dus oor dieselfde datakategorieë wat daarop gemik is om ook betekenis-

en gebruiksverskille van woorde te belig. Die rame verskil egter van woord tot woord, afhangende van die tipe woord, woordpaar of gebruiker. Funksiewoorde word dus anders bewerk as leksikale woorde (sien 6.4 vir 'n waarneming rakende funksiewoorde se omkering).

Linguistiese niveaus wat aandag geniet, sluit in spelling, uitspraak, morfologie, sintaksis, semantiek, pragmatiek en kombinatoriek. Hierdie vooraf gedefinieerde struktuur van die omkeerbare rekenaarprogram lei dus die Afrikaanse leksikograaf doelgerig tydens die omkeringsbewerking van Afrikaanse lemmata. 'n Belangrike aspek van die Nederlands–Afrikaanse woordeboek is die gebruikskontekste en ander talige komponente van die leksikon, wat dui op 'n intensiewe vergelykende studie van onder andere die woordeskat, morfologie, sintaksis en pragmatiese aspekte.

6.4   Enkele waarnemings tydens die omkeringsbewerkings op OMBI

Tydens die bewerking van die kombinasielemma *bluffen/bluf* was een van die betekenisonderskeidinge, eie aan die Nederlandse konteks, 'n verwysing na *Haagse bluf* met die betekenis "nagereg van styfgeklopte eierwit, soms met bessiesap". In Afrikaans verwys *bluf* na "grootpratery, spogtery of oëverblindery". Die Afrikaanse leksikograaf kan dus met 'n mate van sekerheid die afleiding maak dat die Haagse nagereg misleidend weinig bestanddele van substansie bevat, maar indrukwekkend vertoon met 'n bessiesousie.

Die nuanseverskille tussen die twee tale val alleen maar die fyn woordwaarnemer op. Op 'n bussitplek gereserveer vir gestremde persone, is die volgende amelioratief in Nederlands aangebring: *zitplaats bestemd voor **minder valide**.* In Afrikaans bestaan daar (nie sover bekend) 'n eufemisme vir *invalide* wat *nievalide* impliseer nie.

'n Verdere verskynsel wat 'n interessante ondersoek kan regverdig, is woorde uit sewentiende-eeuse Nederlands wat tans in Afrikaans as verouderd aangevoel word, terwyl hulle in Nederlands nog algemeen in gebruik is, byvoorbeeld *laarse* (in Afrikaans het *laarse* die etiket *verouderd*, terwyl *stewels* in Afrikaans algemeen gebruiklik is). Vergelyk ook in dié verband *oortje/oortjie*, *stuiver/stuiwer* (5c), *dubbeltje/dubbeltjie* (10c), *kwartje/kwartjie* (25c), en *rijksdaalder/riksdaalder* (250c) wat tot en met die koms van die Euro in Nederlands gebruik is, maar verouderd is in Afrikaans vanweë 'n "nuwer" geldstelsel. Dan bestaan daar ook verouderde vorme in beide Nederlands en Afrikaans, naamlik *daalder/daler* (150c) en *duit*. In Afrikaans se spreekwoordeskat en verhaleskat is die verouderde vorme egter steeds lewendig, lank nadat die woorde uit die aktiewe woordeskat verdwyn het. Dink byvoorbeeld aan die uitdrukkings: *nie 'n **duit** omgee nie* en *'n **stuiwer** in die armbeurs werp* en die sprokie *Die gelaarsde kat*. Die afdeling vir kommentaar is dus onontbeerlik.

In 6.3 is gemeld dat funksiewoorde anders bewerk word as leksikale woorde. Die leksikograaf kan daarby baat vind deur hom te vergewis van al die relevante voorsetsels (in beide tale) wat moontlik in kollokasies gebruik kan word alvorens werkwoordbewerkings aangepak word. 'n Geheelbegrip

van die kollokasie tree dus duideliker na vore tydens die bewerking van die werkwoord as lemma.

Metaleksikografiese riglyne is baie belangrik. Met metaleksikografiese riglyne word byvoorbeeld bedoel riglyne vir die oordeelkundige aanwending van pragmatiese leiding (onder andere etikette), goedgekose, verhelderende sitate en poëme en relevante gebruiksleiding. Dit kan die teikengebruiker van 'n vertalende woordeboek insig gee in unieke en kultuurspesifieke woordeskatelemente wat die tale van mekaar onderskei.

7.   Slot

Die Afrikaanse leksikograaf kom voor 'n groot uitdaging te staan, aangesien Nederlands en Afrikaans so nou verwant is en die nuanseverskille weinig en gering. Tydens taalkontak van Nederlandse en Afrikaanse sprekers word geringe aanpassings en verstellings op taalvlak gemaak en spoedig is die verskille tussen die twee tale nie meer so opvallend nie, want deur sy aanpasbaarheid skep die mens 'n soort metataal vir suksesvolle kommunikasie. Taalversteuring en -beïnvloeding vind makliker plaas tussen nou verwante tale. En dit is juis onderskeiding van die fyn nuanseverskille tussen Nederlands en Afrikaans wat so belangrik is vir leksikograwe wat aan hierdie woordeboek werk.

Die deskriptiewe, linguistiese, wetenskaplike en gebruikswaarde van so 'n vernuwende woordeboekprojek is onbetwis en die woordeboek (met 'n ingebedde vertaal-, begrips-, produksie- en leerfunksie) waarvoor al soveel jare aan beide kante gewag is, is dus 'n noodsaaklikheid en stimulus vir navorsing op die gebied van die tweetalige leksikografie. Voeg daarby die ideaal dat ander nou verwante taalpare ook volgens hierdie model omkeerbaar is en die betekenisvolheid van so 'n woordeboek blyk duidelik.

Verwysings

De Jong-Goossens, Riet. 1998. Vertalen in de praktijk. *Filter* 5(3): 83-90.

Dorsman, Robert. 1998. Als een (lach)spiegel in raadselen. *Filter* 5(3): 3-12.

Geerts, G. en H Heestermans. 1992[12]. *Van Dale Groot Woordenboek der Nederlandse Taal*. Utrecht/Antwerpen: Van Dale Lexicografie.

Jansen, Ena en Gerrit Olivier. 1986. *Praktiese Nederlands.* Pretoria/Kaapstad: Academica.

Martin, Willy (Red.), Rufus Gouws en Luc Renders. 1999. *Haalbaarheids- en definitiestudie voor een woordenboek Afrikaans–Nederlands/Nederlands–Afrikaans.* Amsterdam: Vrije Universiteit.

Stoops, Yvette. 1995. *Bobbejane of bavianen: Afrikaans versus Nederlands.* Mechelen: CODA.

Swart, Corlietha. 1995. *'n Leksikaal-semantiese ondersoek na verandering in die Afrikaanse woordeskat soos dit weerspieël word in die hedendaagse nuusmedia.* Ongepubliseerde M.A.-skripsie. Bloemfontein: Universiteit van die Oranje-Vrystaat.

Swart, Corlietha. 2000. *Verslag oor OMBI: Besoek aan vakgroep Lexicologie (prof. Willy Martin), Vrije Universiteit: Amsterdam, Sept.–Okt. 2000.* Ongepubliseerde verslag, elektronies versprei aan medewerkers.

Dictionaries of Canadian English

John Considine, *Department of English, University of Alberta, Edmonton, Canada (jc18@ualberta.ca)*

Abstract: The lexicographical record of English in Canada began with wordlists of the late eighteenth, nineteenth, and early twentieth centuries. From the beginning of the twentieth century onwards, the general vocabulary of English in Canada has been represented in bilingual and monolingual dictionaries, often adapted from American or British dictionaries. In the 1950s, several important projects were initiated, resulting in the publication of general dictionaries of English in Canada, and of dictionaries of Canadianisms and of the vocabulary of particular regions of Canada. This article gives an overview of these dictionaries and of their reception, contextualizing them in the larger picture of the lexicography of Canada's other official language, French, and of a number of its non-official languages. It concludes by looking at the future of English-language lexicography in Canada, and by observing that although it has, at its best, reached a high degree of sophistication, there are still major opportunities waiting to be taken.

Keywords: DICTIONARY, LEXICOGRAPHY, CANADIAN ENGLISH, CANADIANISMS, NATIONAL DICTIONARIES, CANADIAN FRENCH, CANADIAN FIRST NATIONS LANGUAGES, BILINGUAL DICTIONARIES, REGIONAL DICTIONARIES, UNFINISHED DICTIONARY PROJECTS

Opsomming: Woordeboeke van Kanadese Engels.   Die leksikografiese optekening van Engels in Kanada begin met woordelyste van die laat agtiende, neëntiende en vroeë twintigste eeue. Van die begin van die twintigste eeu af en verder, is die algemene woordeskat van Engels weergegee in tweetalige en eentalige woordeboeke, dikwels met wysiginge ontleen aan Amerikaanse en Britse woordeboeke. In die 1950's is verskeie belangrike projekte onderneem wat gelei het tot die publikasie van algemene woordeboeke van Engels in Kanada, en van woordeboeke van Kanadeïsmes en van die woordeskat van bepaalde streke van Kanada. Hierdie artikel gee 'n oorsig van dié woordeboeke, en van hul ontvangs, deur hulle in die konteks te plaas van die groter patroon van die leksikografie van Kanada se ander offisiële taal, Frans, en van 'n aantal van sy nieamptelike tale. Dit eindig deur na die toekoms van Engelstalige leksikografie in Kanada te kyk, en deur op te merk dat, alhoewel dit, op sy beste, 'n hoë graad van sofistikasie bereik het, daar nog belangrike geleenthede wag om te benut.

Sleutelwoorde: WOORDEBOEK, LEKSIKOGRAFIE, KANADESE ENGELS, KANADE-ISMES, NASIONALE WOORDEBOEKE, KANADESE FRANS, KANADESE EERSTENASIETALE, TWEETALIGE WOORDEBOEKE, STREEKSWOORDEBOEKE, ONVOLTOOIDE WOORDEBOEK-PROJEKTE

1.   The beginnings of English-language lexicography in Canada

The history of lexicography in Canada goes back more than four and a half cen-

turies, to a list of fifty-four Iroquoian lexical items noted down by Jacques Cartier in 1534 (Cartier 1534/1986: 224-226). More substantial lexicographical work was undertaken in the seventeenth century, when French-speaking missionaries gathered wordlists of the languages of some of the First Nations they encountered; these include vocabularies of Montagnais, Old Algonquin, Ottawa, Eastern Abenaki, Huron, and Onondaga (Goddard 1996). From 1743 onwards, another French-speaking missionary, Pierre-Philippe Potier SJ, made informal notes on the distinctive French usages he encountered in Quebec City and on his travels further west, amounting to just over two thousand lexical items (Halford 1994). These made the first approach to French-Canadian lexicography.

By the end of the eighteenth century, Anglophones in Canada were also making wordlists: in 1792, for instance, Captain G.C. Pulling made a list of 111 items of the now-extinct Beothuk language, obtained from a native speaker called Oubee, a young girl who had been kidnapped by English-speaking Newfoundlanders the year before and would die in captivity in England (Hewson 1978: 11-32). Almost exactly contemporary with this wordlist was one obtained in less criminal circumstances, the glossary of 134 items printed in George Cartwright's *Journal of Transactions and Events During a Residence of Nearly Sixteen Years on the Coast of Labrador* in 1792 (Cartwright 1792: 1. ix-xvi). This marks the beginning of the study of the distinctive English vocabulary of Canada.

For more than a hundred years after Cartwright's glossary, this study remained at a very primitive stage. While the English of the independent United States was coming to be recognized by lexicographers in that country as having its own rules and its own integrity, the English of the Dominion of Canada was seen as aberrant. The lexical item *Canadian English* first occurs as the title of a lecture given in Toronto by the Revd. A.C. Geikie in 1857. Geikie was an immigrant from Scotland and the minister of a Presbyterian church in Berlin (now Kitchener), in the traditionally Anglophone province of Ontario, and the strongly prescriptive and conservative tone of his lecture may betray the linguistic anxiety of someone triply remote from the norms of the metropolitan elite: as a Scot, as a resident of Canada, and as a resident of a town with a very substantial German-speaking population (see Bassler 1991: 110-14). In Canada, he remarked, the English language was in a bad way (Geikie 1857/1980–: 345):

> It can be shown that words, unrecognized by good authors, are daily used; that words duly recognized are used in improper ways; ... that extraordinary creations, and combinations of letters and phrases, are extensively circulated without supplying a recognized want, or contributing in any sense to the enrichment of the language.

This attitude to Canadian English was not idiosyncratic at the time (Chambers 1993). While it was current, it is not surprising that there should have been no dictionaries of the newly-recognized language variety. Wordlists of Canadian

French in the nineteenth century were often compilations of so-called "locu-tions vicieuses" or "barbarismes" (e.g. Dulong 1966: entries 12, 23, 40, 46 etc.), but Canadians did not think their English was important enough even to give this sort of negative attention to it. Wordlists appeared sporadically, but on a very modest scale. So, for instance, Julian Moreton's *Life and Work in Newfound-land* (Moreton 1863: 28-51), has a glossary of 89 lexical items, and other New-foundland collections are noted by Story (1975a). An early mainland glossary, of 32 lexical items, was included in a pamphlet of church news printed in New Brunswick in 1889 ("Brief Glossary" 1889/1980–), but can never have been widely read, and is now so rare as to have been excluded from the standard bibliography of writings on Canadian English because no copy could be found (see Avis and Kinloch 1978: entry 706). The most thorough documentation of nineteenth-century Canadian vocabulary was done abroad, in dictionaries of Americanisms and in what became the *Oxford English Dictionary*, which are outside the scope of this survey.

This situation began to change in the early twentieth century. In 1905, a dictionary published in Canada acknowledged at last that Canadian English shared some features with the English of the United States rather than with that of Great Britain. This was *Nugent's Up-To-Date English–French and French–English Dictionary* (1905), revised by the translator and miscellaneous writer Sylva Clapin from an earlier Montréal publication. It included so-called Americanisms as headwords in the English–French section (such as *sidewalk* rather than British English *pavement*) and among the glosses in the French–English section. Before his revision of *Nugent's Dictionary*, Clapin had already produced a *Dictionnaire canadien-français* (1894), which treated the French of Canada without the prejudices of the previous century, and *A New Dictionary of American-isms: Being a Glossary of Words Supposed to be Peculiar to the United States and the Dominion of Canada* (1902), in which a number of words were identified specifically as Canadian. Clapin was a native of Quebec, and he had lived in the United States: for him, the French of Quebec was a beloved heritage rather than a poor relation of Parisian French; English was an alternative to French rather than a language of unquestionable and overpowering prestige; and British English was an alternative to American English rather than the preferable variety of the language. He was not simply a libertarian in matters of usage — another of his works was an "inventaire de nos fautes les plus usuelles" in the writing of French (Clapin 1913). However, his openness to North American, and especially to Canadian, linguistic difference may be contrasted with Geikie's defensiveness.

The first Canadian dictionary to include Canadian English was, then, a product of Quebec; the first separately-published wordlist of Canadian English was a product of the Canadian West, an area which was largely Anglophone, but which was, like Quebec, more resistant to British linguistic norms than the conservative Anglophone heartland of Ontario. This was the *Western Canadian Dictionary and Phrase-Book* of John Sandilands, published in Winnipeg in edi-

tions of 1912 and 1913, with 853 and c. 1 530 entries respectively. The *Western Canadian Dictionary* was originally produced with new immigrants from England in mind, and the title-page of the first edition boasted "a careful selection of items of general information immediately helpful to the newcomer". Some of the entries in the first edition, such as *irrigation*, do indeed provide encyclopedic information, while others identify lexical items likely to be unfamiliar to a speaker of British English, such as *snow-shoes* or the verb *trek*, whose borrowing from South African English Sandilands must have supposed to be distinctively Canadian, although the word was in fact beginning to be used across the British Empire: *OED* (s.v. *trek* v., sense 1a(b)) has an Australian citation of 1911 and an English one of 1912. Other entries celebrate Western Canadian cities, institutions, or scenery, e.g. the entry *Foothills*, which begins "The natural features of Alberta combine the beauties of prairie and mountain scenery", and continues in the same vein. The second edition was more evidently meant to be read for pleasure: a note at the end recommends "mailing copies to friends in the Old-Country who want to know about Canada", and one issue was sold as the "Stampede Edition", presumably for sale as a souvenir at the Winnipeg Stampede of 1913 (there is a copy in the National Library of Canada). It has more of a jokey quality than the first: for instance, s.v. *Firefly*, after a simple definition, Sandilands tells the story of the two Irishmen who, "tired after a day's travel on foot, sought refuge from the mosquitoes under a haystack. In the darkness Pat espied a fire-fly, and declared to Mike that the mosquitoes were after him again, with lanterns". But Sandilands was not just a humorist; he had a good ear for vocabulary, and the *Western Canadian Dictionary* provides the *Oxford English Dictionary* with its first examples of *movie* "film" (singular, as opposed to *the movies*); *B.S.* as an abbreviation for *bullshit* with the sense "nonsense"; *make the grade* "be successful"; *kale* "money", and other lexical items. It was not such a substantial work as Charles Pettman's coverage of South African English in *Africanderisms*, with which it was almost exactly contemporaneous; nor is it to be compared with E.E. Morris's ambitious historical dictionary of Australian usage, *Austral English*, or with the fine account of Anglo-Indian usage in Henry Yule and A.C. Burnell's *Hobson-Jobson*. In the first years of the twentieth century, English-language lexicography was, in other words, practiced with less sophistication in Canada than in several of the other major dominions.

2.   The first major Canadian dictionaries of English

After the publication of Sandilands's *Western Canadian Dictionary*, there was a long hiatus in English-language lexicography in Canada. This stagnation contrasted with the flourishing state of French-Canadian lexicography: the Société du parler français au Canada, founded in 1902, produced a 700-page dictionary of Franco-Canadianisms, *Glossaire du parler français au Canada*, in 1930, and one of its members, Louis-Alexandre Bélisle, drew on this to make a general dictionary of all the French used in Quebec, whether distinctively Canadian or

common to Canada and France, in the 1950s. This was being sold in fascicles in Quebec supermarkets at a time when interest in Canadian English was still at a decidedly primitive stage. A characteristic attitude is that expressed by a representative of the Canadian Broadcasting Company (CBC) in 1949. A member of an Ontario school board had attacked the English spoken by CBC announcers, complaining for instance that they stressed the first syllable of *research* rather than the second (this is in fact standard North American usage). The reply was that in fact the announcers read their scripts in advance, consulting a dictionary for help with the pronunciation of difficult words — and that "[w]e have an Oxford and a Webster's dictionary" (*Toronto Daily Star* 13 January 1949: 29). The implication was clearly that although Canadian English was neither identical with British nor with American English, it could be regarded by dictionary-users as a mixture of the two.

This principle underlay the making of the first general dictionary aimed at the Canadian market, the 32 000-word *Winston Simplified Dictionary for Canada*, published in 1937. This belonged to a large family of non-scholarly American dictionaries. The first of these, which appeared in 1919 as *The Winston Simplified Dictionary*, claimed ease of use as its selling point: "You never have to look up the meaning of the meaning" (quoted Tebbel 1972–81: 3. 222). It was revised in 1936 as the *New Winston Simplified Dictionary for Young People*, and then repackaged, adapted, or further revised in a long line of dictionaries for children or for schools, and of college editions, advanced editions (these two might be identical), and encyclopedic editions (Tebbel 1972–81: 3. 614; Kister 1977: 166, 325-26). The Canadian edition, prepared by the language historian and dialectologist Henry Alexander of Queen's University, substituted British usages for some of the American usages in the original dictionary, and likewise substituted lists of Canadian premiers and governors-general for the list of American presidents in the original. It was reviewed thoughtfully if a little coolly in the Toronto *Globe and Mail*, which began with an interesting description of Canadian usage as "split between English custom in universities and American in business and newspapers, with most Canadians making their own compromises in no orderly fashion". The review (*Globe and Mail* 11 September 1937: 8) ended by observing:

> Of course it is not true, as the editors assume, that Canadians invariably follow English practice any more than it would be true that we invariably follow American standards. That the problem of a new Canadian dictionary is more subtle and difficult than Dr Alexander supposes will not interfere in the least with the rightful popularity of the new Winston dictionary in Canada. This is the first publisher to be so enterprising as to try to suit our peculiar needs.

By the time of the last reissue of a revised edition of this dictionary in 1974 as *The Winston Canadian Dictionary for School, Home, and Office*, the word count had crept up to 38 550, and the publishers claimed to have sold almost two million copies (*Winston Canadian Dictionary* 1974: iv, back cover). The next generation

of Winston Canadian dictionaries, to be discussed below, had by this time overtaken this work, which was fundamentally an undistinguished piece of lexicography but, as the *Globe and Mail* said, an enterprising publication.

Despite the existence of the Winston dictionary, some Canadians were still, at the end of the 1950s, prepared to dismiss Canadian lexicography as pointless. When the idea of a Canadian dictionary was introduced to the Dean of Arts and Science at Dalhousie University in Nova Scotia, for instance, he remarked that he "had always found British definitions satisfactory". On being told that one British dictionary defined *hockey* as "a game played with a ball and stick" (*hockey* in Canadian English means *ice hockey*, a game of great cultural importance, played with a hard disc called a puck, as opposed to *field hockey*, the game usually called *hockey* in British and South African English), he replied that no Canadian would need to look that word up anyway (*Halifax Chronicle-Herald* 24 April 1959: 28). By that time, however, new projects, much more interesting than Winston's, were under way. In 1954, the founding meeting of the Canadian Linguistic Association was held at the University of Manitoba, in Winnipeg. At this meeting, "some discussion was given to the possibility of preparing dictionaries of Canadian English" (Scargill 1967: vi). The desirability of preparing a bilingual English–French dictionary, which would replace Clapin's long-obsolete work, was also raised. One of the first articles in the association's journal, published in 1955, was called "Lexicographic Challenges of Canadian English", and concluded with the hope that it might help "to give impetus to plans for an eventual Dictionary of Canadian English, based upon historical principles" (Lovell 1955a: 5). It was the work of Charles Lovell, a former member of the staff of the *Dictionary of Americanisms*, who went on to contribute a two-part article on the principles of collecting for the dictionary in 1956 (Lovell 1955b, 1956). In the following year, the association formed a Lexicographical Committee and began to plan three types of English-language dictionaries: a series for use by schoolchildren and undergraduates; a dictionary of Canadianisms; and a larger historical dictionary of the English language in Canada (Scargill 1967: vi). Separate committees were subsequently established with French-Canadian lexicography and Ukrainian lexicography as their concerns; a founding member of the Canadian Linguistic Association, J.B. Rudnyckj, produced a Ukrainian etymological dictionary in fascicles from 1962 onwards.

In 1958, the Toronto publishing house of W.J. Gage agreed to publish a dictionary of Canadianisms, to be edited by Lovell (Scargill 1967). At his death in 1960, Lovell left this project well begun, but far from completion. What followed was an exemplary collaboration between commercial, academic, and governmental sponsors of a dictionary project. Gage purchased Lovell's lexicographical collection from his estate and invited two members of the Canadian Linguistic Association, Matthew Scargill and Walter Avis, to take over the editing of the unfinished dictionary. The University of Alberta at Calgary (now the University of Calgary) provided space and secretarial assistance for a lexi-

cographical centre which gave the dictionary a home from 1960 to 1964, and the University of Victoria did so after 1964. Walter Avis was given study leaves for dictionary editing purposes by the Canada Council and the university at which he taught, the Royal Military College of Canada. A number of academics made materials for the dictionary available to its editors. Gaston Dulong, secretary of the Société du parler français au Canada, acted as consultant for French-Canadian etymologies. The whole undertaking was remarkable for its collegiality and efficiency. The *Dictionary of Canadianisms* (*DC*) was completed in 1967, the centenary of the confederated Dominion of Canada.

The dictionary itself was edited on historical principles, with a wordlist confined to lexical items or senses of lexical items "native to Canada or ... distinctively characteristic of Canadian usage though not necessarily exclusive to Canada" (*DC* 1967: xiii). Sandilands and Clapin, by contrast, had simply noted forms which they had encountered in Canada and knew or supposed to be different from British usage, and many of these were American rather than Canadian. The bibliography included around two thousand books and pamphlets and around five hundred journals, and although this list was far from comprehensive, it was most impressive. No attempt was made to document spoken English, which would have made the lexicographic process very much longer and more expensive. The ten thousand entries were handsomely laid out, with a sprinkling of useful line drawings. The publishers' foreword described the dictionary as "a contribution to Centennial thinking" and hoped that it would contribute not only to the study of language but also to the understanding of Canadian identity (Wees 1967). The American lexicographer Sidney Landau perceived the same double function when he called the dictionary "an indispensable source from a linguistic point of view and a fascinating treasury of cultural information about Canada's past" (Landau 1984/1989: 340). An abridgement was published six years later as *A Concise Dictionary of Canadianisms*. *DC* had been planned as a pilot volume for a major *Dictionary of Canadian English on Historical Principles*, which would presumably have been on the same sort of scale as the *Dictionary of Americanisms* on which Lovell had worked (Scargill 1958: 117). Collecting for this project was apparently still in progress at the Lexicographical Research Centre in Victoria in the late 1970s (McConnell 1978: 127 note). The dictionary was, however, never published, and the files of the Lexicographical Research Centre, now in the University of Victoria Archives, appear to have been closed in 1983 (British Columbia Archival Union List 2003).

The *Dictionary of Canadianisms* was not Gage's only contribution to Canadian lexicography in the 1960s. The series of general dictionaries which had been planned by the lexicographical committee of the Canadian Linguistic Association was already far enough in planning to be discussed in print in 1958 (Scargill 1958: 116f). By this stage, permission had been obtained to base the proposed dictionaries on the graded series of American school dictionaries published by Thorndike-Barnhart, and this made it possible to complete them

expeditiously. They were the work of good scholars — Avis, Scargill, Lovell, and R.J. Gregg of the University of British Columbia — and were rather more thoughtfully developed from their American bases than the Canadian edition of the Winston dictionary had been. The series began with *The Beginning Dictionary*, with about 20 000 entries, which was published in 1962; *The Intermediate Dictionary*, with 64 000 entries, followed in 1963; and *The Advanced Dictionary*, with 90 000, in 1967 (overview in Avis and Kinloch 1978: entries 151-53). Revisions of these dictionaries will be mentioned below.

Two other dictionary projects dealing with Canadian English had been launched by 1960. The first of these takes us back to the highly distinctive English of Newfoundland, which had, as we have seen, been an object of attention since the very beginnings of interest in Canadian English. In 1956, George Story, a professor at Memorial University (the principal university of Newfoundland) with a background in English literature of the early seventeenth century, published a fourteen-page pamphlet called *A Newfoundland Dialect Dictionary: A Survey of the Problems*. In the following year, he reported to the Canadian Linguistic Association on his work compiling such a dictionary. Story stated in his report: "I have on cards what I believe to be most of the words in *general* use in the Island ... [and] a fairly good, though not complete, record of words peculiar to [certain named locations]" (Story 1957: 51-52). He also acknowledged that "the limits of individual collecting are being reached", and that, since the project would almost certainly not be completed for many years, he would publish a preliminary glossary "in about three years" (Story 1957: 53). This preliminary glossary appears not to have materialized, but the collaboration which Story looked for did take place; he was joined by two colleagues at Memorial, William Kirwin and John Widdowson. Their work, whose results will be discussed in the next section, was coordinated with other parts of a wider project concerned with the preservation of Newfoundland heritage. This had been initiated by the head of the department of English at Memorial, Edgar Seary, formerly of Rhodes University, Grahamstown (for him, see Story 1975b). In this project, "the languages (especially the English language), the place- and family-names, and the folklore of Newfoundland were to be subjected to a scrutiny worthy of their importance for Newfoundlanders and their interest to others" (*DNE* 1982[1]: v; cf. Seary 1971 and Seary 1977). This concern to set lexicographical research in a wider cultural context was exemplary, and continues in the work of the English Language Research Centre at Memorial (Memorial University 2003).

The second important project of the late 1950s and early 1960s was the *Canadian Dictionary/Dictionnaire canadien*, prepared at the Lexicographic Research Centre/Centre de Recherches lexicographiques at the Université de Montréal. Its chief editor was Jean-Paul Vinay, a translator and educator who had been born in Paris, founded the linguistics department at Montréal, and would go on to be Matthew Scargill's successor as Chair of the Linguistics Department at the University of Victoria. He was assisted by the distinguished translator

Pierre Daviault, and (until 1957) by Henry Alexander, who had adapted the Winston dictionary for Canadian use a quarter of a century earlier. The *Canadian Dictionary* was a practical, synchronic dictionary for translators rather than a dictionary of record. It was, on its publication, described as a "concise edition"; it only contained about 11 000 English headwords (Avis and Kinloch 1978: item 621). A multi-volume successor was anticipated with interest, and one reviewer even urged readers to note "minor blemishes" and communicate them to the editors: "all such information will feed into the archives for the larger volumes of the dictionary to which we now look forward" (Gregg 1962: 71). These volumes never appeared, although a more substantial *Bilingual Canadian Dictionary/Dictionnaire canadien bilingue* is now in preparation (University of Ottawa 2003). The French text of the *Canadian Dictionary* falls outside the scope of this article, but it is significant that the English it treated was meant to be specifically Canadian English. Two of the editors wrote à propos of the dictionary that "il est temps que nos variétés d'anglais et de français sortent de leur lieu d'incubation, et accèdent à la dignité de particularismes reconnus" (Vinay and Daviault 1958: 110). Here, as in the lexicography of Sylva Clapin, Francophone Canadians can be seen treating the English of Canada with more respect than their Anglophone neighbours.

Finally, two more Canadian editions of American dictionaries appeared in the 1960s. The first was the substantial *Funk and Wagnalls Standard College Dictionary: Canadian Edition*. This was a revision for Canadian readers of a 150 000-entry American dictionary (for which see Landau 1984/1989: 341-42); it had, as did the American original, an essay on Canadian English by Walter Avis. The second was a new *Winston Dictionary of Canadian English, Intermediate Edition* (1969), based on the *Holt Intermediate Dictionary of American English* of 1966: the publishing companies of John Winston and Henry Holt had merged in 1960 (Dzwonkoski 1986). The new dictionary, of about 65 000 entries, was reprinted in paperback as *The Compact Dictionary of Canadian English*, and abridged as *The Winston Dictionary of Canadian English: Elementary Edition*, with about 30 000 entries (Avis and Kinloch 1978: entries 490, 492-93). It was made under the editorship of Thomas M. Paikeday, who had been born in India and had lectured at the University of Delhi before beginning his lexicographical career working for Gage. Thereafter, Paikeday turned his attention to what would become the *New York Times Everyday Dictionary* (1982); a subsequent Canadian dictionary of his will be discussed below.

3.   Regional and general dictionaries in the 1970s and 1980s

The next two decades of English-language lexicography in Canada saw three kinds of activity: the proliferation of general dictionaries, none of them wholly original; the making of two major regional dictionaries; and the making of a number of humorous, local, and technical wordlists of minor importance.

The first area of activity can be summed up briefly. In 1973, a new Cana-

dian edition of *Funk and Wagnalls Standard College Dictionary* was published, this time based on the revised American edition of 1968 (for which see Landau 1984/1989: 341-42). It was followed in 1980 by *The Houghton Mifflin Canadian Dictionary of the English Language*. The title of this book concealed its American origins; it was in fact a Canadian edition of the *American Heritage Dictionary of the English Language*, which made the good features of that dictionary, for instance its ample illustrations and learned appendix of Indo-European roots, available to Canadian readers as well as its bad features, for instance its conservative and authoritarian ideology. It has been widely used in Canadian universities and in public life, being quoted in at least one Canadian legislature, and drawn on in the rulings of Canadian federal agencies (*Hansard* [Nova Scotia] 2001; Canadian International Trade Tribunal 1989; *Federal Court Reports* 1994). Finally, in the 1970s and 1980s, the Gage dictionary series underwent revisions and changes of title: *The Beginning Dictionary* became the *Canadian Junior Dictionary* in 1977 and the *Gage Junior Dictionary* in 1985; *The Intermediate Dictionary* became the *Canadian Intermediate Dictionary* in 1979; *The Senior Dictionary* became the *Gage Canadian Dictionary* in two editions of 1973, then the *Canadian Senior Dictionary* in its fourth edition, of 1979, then *The Gage Canadian Dictionary* again in its fifth, of 1983.

The second area of activity in the period, the making of regional dictionaries, is much more interesting and important. George Story had known in the 1950s that the dictionary of Newfoundland English which he had begun would take many years to finish. It was actually in progress for about a quarter of a century, supported by Memorial University and the Canada Council: here, as in the making of the *Dictionary of Canadianisms*, government money and support from a university came together fruitfully. Articles on the making of the dictionary were published by Story, Kirwin, and Widdowson in 1973, by Story and Kirwin in 1974, and by Kirwin (illustrated with sample entries) in 1975. In 1977 specimens of the dictionary were circulated in draft form. The completed *Dictionary of Newfoundland English* (*DNE*) was published in 1982, followed by a second edition with an extensive supplement in 1990; Memorial University made its entire text available online in 1999.

The completed dictionary ran to about 4 500 entries; the second edition, with 1 500 new or expanded entries, has an entry count of 5 224 (*DNE* 1999: search screen). The selection criteria were that words which appear to have originated in Newfoundland or to have been first recorded in books about Newfoundland should be included, as should "words which are characteristically Newfoundland by having continued in use here after they died out or declined elsewhere, or by having acquired a different form or developed a different meaning, or by having a distinctly higher or more general degree of use" (*DNE* 1982: xii). These criteria called for some careful judgements, one of which (the case of *high-liner*) has subsequently been written up in some detail by Kirwin and Story (1986). *DNE* was constructed on historical principles, with extensive cross-referencing to other dictionaries. The bibliography of the printed and

manuscript sources in the first edition ran to over five hundred items, excluding periodicals. To this written material, which made up 42% of the dictionary's sources, was added data from the extensive collections of the Memorial University of Newfoundland Folklore and Language Archive (including 2 000 tape-recordings), and other field records (*DNE* 1982: xxi-xxvi). The *DNE* was originally expected to be largely of local interest: the publisher sent three thousand copies of the first printing to Newfoundland, and fewer than five hundred to bookstores in the rest of Canada (*Globe and Mail* 13 November 1982: 8). However, it was reviewed with great warmth across Canada. For instance, one of the leading authorities on Canadian English, J.K. Chambers of the University of Toronto, praised it in a national paper as "more than a book of definitions ... a flahoolach repository of the folklore and the folkways and even the sound and the spirit of ... Newfoundland" (*flahoolach* is a Newfoundland word meaning "lavish, generous") and added that it "is a book which, combined with a working fireplace, can use up your whole winter" (*Globe and Mail* 22 December 1982: 19).

In 1979, work began on a dictionary of the English of Prince Edward Island, which, like Newfoundland, is a large island lying off the east coast of Canada, with a long history of settlement by Anglophones. This was undertaken by Terry Pratt, a professor of English at the University of Prince Edward Island. Like the *Dictionary of Canadianisms* and *DNE*, and indeed earlier Canadian dictionaries such as Clapin's *Dictionnaire canadien-français*, this dictionary project had a clear sense of language as heritage. Early in the project, Pratt observed that "[j]ust about everybody is interested in language when it comes right down to it", and that his dictionary appealed to this interest and also to "Island pride and the swelling interest in heritage matters" (*Montreal Gazette* 23 September 1980: 21; see also Considine 2000: 325). Although the resulting *Dictionary of Prince Edward Island English* (*DPEIE*) was not part of a cultural research project as elaborate as that into which *DNE* fitted, it shared the larger dictionary's concern with the contextualization of lexicography in other kinds of verbal culture. So, for instance, it concluded with the remark that a dictionary of Prince Edward Island sayings would be desirable (*DPEIE* 1988: 179); in fact, Pratt subsequently produced such a volume, in collaboration with Scott Burke, who had been his editorial assistant on *DPEIE*.

DPEIE was published in 1988. Like *DNE*, it used historical evidence gathered in a reading programme as well as material gathered directly from informants, the latter being proportionally more important than in *DNE*. This made it possible for Pratt to specify regional and social variation in impressive detail: so, for instance, *alder mud* "mud from alder swamps used as fertilizer" is labelled "Archaic. Occasional in Egmont, infrequent elsewhere but rare in Charlottetown; significantly older, male; especially less educated". This sociolinguistic depth was made possible in part by stringent selection criteria, which excluded, among other classes of vocabulary, most proper nouns, most slang, and most general Canadianisms (*DPEIE* 1988: xii-xiii), bringing the word count

right down to 873 main entries (*DPEIE* 1988: xxii). G.M. Story suggested in a review that "the rather small lexical corpus which the work contains" might have been "restricted by the nature of the collecting instruments and the special interests of the investigators" (Story 1989: 392). Responding to a draft of an article which made a similar point, Pratt argued that a much larger word count would simply have meant unnecessary overlap with other dictionaries (1999; cf. Considine 2000: 325-27). To that, it might be replied at least that the excellence of his dictionary inevitably leads to the wish that it had been larger.

Smaller and much less thorough or scholarly local wordlists were also produced in the period. For instance, Lewis J. Poteet's *The South Shore Phrase Book* documents about six hundred lexical items collected by the author in parts of Nova Scotia, including the old and linguistically interesting German settlement of Lunenburg, and Tom Parkin's *WetCoast Words* collects about seven hundred lexical items from British Columbia (the title alludes humorously to the high rainfall characteristic of the Pacific coast of Canada). These publications are of some interest, not least as sketches of what might be done in the study of the regional vocabulary of Canadian English, but their technical level is that of Sandilands's dictionary. A slightly different kind of wordlist, Mark Orkin's successful *Canajan, Eh?* of 1973, relied for its effect on the re-spelling of lemmata and defining text to approximate informal Canadian speech. This had already been done for Australia in *Let Stalk Strine* (Lauder 1965), for Texas in the *Illustrated Texas Dictionary of the English Language* (Everhart 1968), and for South Africa in *Ah Big Yaws?* (Malong 1972); examples from these and other dictionaries are most readily available in the work of David Crystal (1995: 410, 357; 1998: 23). In the 1980s, specialized dictionaries aimed exclusively at the Canadian market — a *Canadian Dictionary of Safety Terms*, a *Canadian Dictionary of Abbreviations*, and so on — also began to appear. Although these are outside the scope of this survey, they are worth mentioning because they suggest an increasing sense of the distinctiveness of Canadian usage.

4.   The new generation of Canadian dictionaries of English: the 1990s

Something new happened in Canadian lexicography in 1990, when Thomas Paikeday completed his *Penguin Canadian Dictionary*, a general dictionary of about 75 000 entries. This was the fruit of a long-standing interest in the possibilities of computational methods in lexicography. Paikeday assembled a 20 000 000-word corpus, largely from Canadian sources, with a particular reliance on machine-readable text of the *Globe and Mail*, and used this to establish a wordlist. He also used it to ascertain regular collocations, a subject of great interest to him (see e.g. Paikeday 1989, 1992), and provided over 40 000 collocational and idiomatic examples of uses of the words he defined. The result was the first general dictionary wholly made in Canada. It was not entirely successful. "In too many places," one incisive review pointed out, "one short definition ... has to serve as the anchor for a whole slew of context phrases and sen-

tences showing clearly different meanings, with not enough information about those differences to serve the user properly"; moreover, the reviewer continued, the wordlist of the dictionary was sometimes haphazardly chosen, homographs were not differentiated, and statements about Canadian usage were sometimes questionable (*Globe and Mail* 18 August 1990: C16). All of these criticisms could no doubt be referred to Paikeday's reliance on a corpus which was too small and crudely selected to do all that he needed.

The decade which began with the *Penguin Canadian Dictionary* was marked by the vigorous, and sometimes bewildering, development of Canadian dictionaries. Gage developed a particularly elaborate range. The flagship, the *Gage Canadian Dictionary,* was revised in 1996 and again in 2000, and was released in electronic form in 1996. By 2002, it was accompanied by a concise edition (1 024 pages, 95 000 entries), and by no fewer than five graded dictionaries for younger readers, who could start with the *Gage Canadian First Book of Words* (64 pages, 600 one-word entries), move up to the *Gage Canadian Beginner's Dictionary* (128 pages, 1 200 entries) when they started reading properly, then turn to the *Gage Canadian School Dictionary* (384 pages) at the age of eight or nine, before graduating to the 800-page *Gage Canadian Junior Dictionary* a year or two later, and then to the 1 414-page *Gage Canadian Intermediate Dictionary*, which might be expected to see them through to the age of sixteen (details from Gage Learning and National Library of Canada). Trained scholars worked on these dictionaries: the principal editor of the 1996 *Gage Canadian Dictionary* had written and published a doctoral thesis on social and regional variation in Canadian English (De Wolf 1992), and its general consulting editor was Terry Pratt.

A rival to the *Gage Canadian Dictionary* appeared in 1997 as the *ITP Nelson Canadian Dictionary of the English Language*. This dictionary, almost exactly the same length as *Gage* at 1 728 pages, with an entry count of about 150 000, was an adaptation of the third edition of the *American Heritage High School Dictionary*, itself an abridgement of the *American Heritage Dictionary* which had been adapted nearly two decades earlier as the *Houghton Mifflin Canadian Dictionary*. Its treatment of Canadian English was carefully done, and it was adorned with what its editors zealously called "a toboggan-load of useful information" about Canadian geography and institutions (*ITP Nelson Canadian Dictionary* 1997: vii). It was well received, one reviewer going so far as to call it "a brave, elegant, erudite and eloquent gift to this much beleaguered nation" (*Globe and Mail* 21 December 1996: D14).

A rival to the *Gage Canadian Dictionary* and *ITP Nelson Canadian Dictionary* was published in 1998: the *Canadian Oxford Dictionary* (*COD*), very much their length at 1 728 pages and about 130 000 entries (*COD* 1998: dustcover). Its chief editor, Katherine Barber, had worked on the *Bilingual Canadian Dictionary/Dictionnaire canadien bilingue* project at the University of Ottawa before being recruited by Oxford University Press: a reminder of the importance of Canadian bilingualism in English-Canadian lexicography and, like Terry Pratt's work for Gage, a reminder of the contributions of universities to the making of trade

dictionaries. *COD* was, perhaps surprisingly, the first major entry of Oxford University Press into the Canadian dictionary market, although it had been rumoured in the 1960s that the press was preparing "a Canadian English dictionary" (Gregg 1962: 68). It was based on a two-million-word citation database of Canadian quotations selected in the course of a substantial reading programme which drew on 20 000 000 words of Canadian text; it also drew on the *Oxford English Dictionary*'s 20 000 000-word citation database of largely non-Canadian quotations. These databases were used with greater sophistication than that on which Paikeday founded the *Penguin Canadian Dictionary*. One reason for this is that the *COD* reading programme took in a much wider range of material than Paikeday could hope to do; and a reason for that is that *COD* was produced by a full-time staff of five, with the assistance of a number of outside contributors, including several members of the main Dictionary Department of Oxford University Press in Oxford. It is no longer possible for any general dictionary of English to be produced by a single person.

Its publishers claimed as the dictionary was launched that *COD* "establishes a new authoritative standard for dictionaries in Canada" and "serves Canadians like no other dictionary" (Oxford University Press 1997), even publishing a description of it as "the new, ultimate authority" (*COD* 1998: dustcover). No dictionary is an ultimate authority on a living language, but *COD* is rich in Canadianisms and in Canadian encyclopedic information, and its coverage of contemporary world English is generally very good. In my opinion, it is at present the best general dictionary of Canadian English.

COD was reissued with a supplement of new words in 2001, and released on CD-ROM in 2002. It has generated several offshoots: *The Canadian Oxford Spelling Dictionary* in 1999; *The Canadian Oxford Paperback Dictionary* (this title suggests misleadingly that this abridged dictionary is the full *COD* in soft covers) in 2000; *The Canadian Oxford High School Dictionary* in 2001; *The Canadian Oxford Compact Dictionary* in 2002. A couple of these challenge Gage and other companies in the children's market: *My Very First Canadian Oxford Dictionary*, with illustrations, appeared in 2002, and *My First Canadian Oxford Dictionary*, also with illustrations, in 2003 (apparently *my very first* means "my first" in these titles and *my first* means "my second").

5.   Lexicography in Canada in the new millennium

Some of the most dynamic lexicographical work being done in Canada as the new millennium begins is concerned with languages other than English. The *Alberta Elders' Cree Dictionary*, for instance, has made a striking partnership between academic lexicography and community ownership of a language: published by the University of Alberta Press, under the general editorship of a professor at that university, Earle Waugh, it has nevertheless, as the title suggests, been shaped by native speakers of Cree. It is described in its preface as "a gift from the elders to the Cree people of Alberta", and its principal maker,

George Cardinal, calls it "an honoring gift to all my great Cree ancestors" (Le-Claire and Cardinal 1998: xiii, xi). A similar sense of heritage pervades a dictionary of the distinctive variety of Low German spoken in western Canada by the ethno-religious group called Mennonites, Jack Thiessen's *Mennonitisch-Plautdeutsches Wörterbuch/Mennonite Low German Dictionary*, "a labour of love" which, in the spirit of *DNE* and the projects associated with it, and of *DPEIE* and the *Dictionary of Prince Edward Island Sayings*, brings a great deal of folkloric and encyclopedic material into its wordlist, "to embody a time and a culture, and a way of life that is rapidly disappearing" (Thiessen 1999: 24; cf. Considine, forthcoming). Other important dictionaries of languages other than English have recently been completed, for instance a *Canadian Dictionary of American Sign Language* (Bailey and Dolby 2002), which lists 8 700 signs within the lexicon of ASL, many of which are distinctive to Canada or even to particular regions within Canada. The dictionary even records a special sign for the most famous hockey player in recent history, Wayne Gretzky (*National Post* 10 July 2002: A1). Others are still in progress, such as the *Dictionary of Old English* being undertaken at the University of Toronto, whose editors had by the end of 2002 begun drafting entries for words beginning with the letter *I* (Holland 2003). One important dictionary of a non-Canadian variety of English which is at present in progress is that of Lise Winer of McGill University, Montréal, who has since 1980 been at work on a dictionary of the English of Trinidad (Haldane 2003).

The lexicography of modern English in Canada has changed dramatically in the last fifty years. There are still some Canadians who believe that there is no distinctive national standard for formal written English in Canada: "all Canadian dictionaries are rather useless", wrote one in an online posting; "[a]ny Canadian usages you use you obviously already know, and there is no tyranny of proper usage when it comes to the ones you don't" (Anon c2003). But generations of lexicographical work, from the pioneering *Winston Simplified Dictionary for Canada* in 1937 through the Gage dictionaries of the 1960s to their offspring and the numerous competitors of those offspring, have changed educated opinion. So, for instance, the CBC, which had been content to appeal to "an Oxford and a Webster's dictionary" in 1949, responded to a criticism in 2001 (it had transcribed *the gig is up*, meaning roughly "the party's over" as *the jig is up*, meaning roughly "the scheme's been foiled, the game's up" in an online posting) by citing the *Canadian Oxford Dictionary*, a *Webster's* dictionary, the *Oxford English Dictionary*, the *Gage Canadian Dictionary*, and the *Houghton Mifflin Canadian Dictionary* (Shewchuk 2001).

One aspect of the future of English-language lexicography in Canada is predictable: the lucrative school and university market will continue to be tapped. This may not be a matter of making highly original dictionaries: at the time of writing, the educational publisher Harcourt Canada is completing a new edition of the *Winston Canadian Dictionary*, which will be based on Thomas Paikeday's *Winston Dictionary of Canadian English, Intermediate Edition* of 1969,

which was itself, as we have seen, based on an American dictionary of 1966 (Kretchman 2003). The revision has been entrusted to a couple who trade as "Focus Strategic Communications Incorporated: Packagers for Publishers", rather than to full-time lexicographers or academics (Focus Strategic Communications 2003). Their work is unlikely to produce results superior to those achieved in twenty-five person-years by the makers of *COD*. More elaborate projects may subsume Canadian English within general American English, as does the *Microsoft Encarta College Dictionary*, which has consultants for some major varieties of English spoken outside the US (including Margery Fee of the University of British Columbia for Canada, Rajend Mesthrie for South Africa, and Eva Hertel of the Chemnitz University of Technology for East Africa), and has four Canadians, including Terry Pratt, on its College Usage Advisory Board. The lexicography of Canadian English is too important for it to be handed over to "packagers for publishers" or to publishers for the general American market. The only way for it to move forward will be to take the route pioneered by Paikeday and subsequently taken by *COD*: the best dictionaries of Canadian English will make increasing use of carefully constructed databases rich in Canadian and worldwide material in English, and they will be made by the full-time staff of publishers' dedicated dictionary departments, conceivably in partnership with universities or the federal government.

The academic lexicography of Canadian English is clearly capable of being developed in two respects. The first is the making of regional dictionaries like *DNE* and *DPEIE*, and at a similar scholarly level to these two. Such dictionaries could in theory be compiled for each of the maritime provinces, namely Nova Scotia and New Brunswick, for the three prairie provinces as a single unit of inquiry, for the far north of Canada, and for British Columbia. The second is, in effect, the finishing of the great unfinished Canadian dictionary: the *Dictionary of Canadian English on Historical Principles*. The *Dictionary of Canadianisms* was never intended to be, as it has *de facto* become, the sole historical record of Canadian English, and the *Australian National Dictionary* and the *Dictionary of South African English on Historical Principles* show that the age of the making of major historical dictionaries is not yet past. There is an important respect in which Canada is, at the beginning of the twenty-first century, lagging behind Australia and South Africa in the documentation of its own variety of English, just as it was at the beginning of the twentieth.

References

Anon "from Vancouver, Canajah" [online]. c2003. Available: <http://www.amazon.ca>, listing for *Funk & Wagnalls Canadian School Dictionary* [viewed May 23, 2003].

Avis, Walter S. and A.M. Kinloch. 1978. *Writings on Canadian English 1792–1975: An Annotated Bibliography.* Toronto: Fitzhenry and Whiteside.

Bailey, Carol Sue and Kathy Dolby (Eds.). 2002. *Canadian Dictionary of American Sign Language.* Edmonton: University of Alberta Press.

Bassler, Gerhard P. 1991. *The German Canadian Mosaic Today and Yesterday*. Ottawa: German–Canadian Congress.

Bélisle, Louis-Alexandre. 1954–57. *Dictionnaire général de la langue française au Canada.* Québec: Bélisle.

A Brief Glossary of Canadian Expressions Collected by An Enquiring Englishman. 1889/1980–. *St. John's Church Parish Notes*: 39-40. St. John, NB: J. and A. McMillan for St. John's Church. *CIHM/ICMH Microfiche Series* fiche 93580. Ottawa: Canadian Institute for Historical Microreproductions.

British Columbia Archival Union List [online]. 2003. Available: <http://mayne.aabc.bc.ca> [viewed May 28, 2003].

Canadian Dictionary/Dictionnaire canadien. 1962. Vinay, Jean-Paul, Pierre Daviault, and Henry Alexander (Eds.). Toronto: McClelland and Stewart.

Canadian Dictionary of Abbreviations. 1994. Dobroslavic, Thérèse and Gail Edwards (Eds.). Previous edition published as *Abbreviations: A Canadian Handbook*. Toronto: ECW Press.

Canadian Dictionary of Safety Terms. 1987. Ottawa: LARR Enterprises and RAM Consulting.

Canadian International Trade Tribunal. 1989. Appeal 3003: Bowne of Canada, Inc., v. the Minister of National Revenue [online]. Available: <http://www.citt-tcce.gc.ca/appeals/decision/apin89_e.asp#89> [viewed May 28, 2003].

Cartier, Jacques. 1534/1986. *Relations*. Michel Bideaux (Ed.). Bibliothèque du Nouveau Monde. Montréal: Presses de l'Université de Montréal.

Cartwright, George. 1792. *A Journal of Transactions and Events, During a Residence of Nearly Sixteen Years on the Coast of Labrador; Containing Many Interesting Particulars, Both of the Country and its Inhabitants, Not Hitherto Known*. 3 vols. Newark, UK: Allid and Ridge.

Chambers, J.K. 1993. 'Lawless and Vulgar Innovation': Victorian Views of Canadian English. Clarke, Sandra (Ed.). 1993. *Focus on Canada*: 1-26. Amsterdam/Philadelphia: John Benjamins.

Clapin, Sylva. 1913. *Ne pas dire, mais dire: inventaire de nos fautes les plus usuelles contre le bon langage*. Worcester, Mass.: J.A. Jacques.

Clapin, Sylva (Ed.). 1894. *Dictionnaire canadien-français ou lexique-glossaire des mots, expressions et locutions ne se trouvant pas dans les dictionnaires courants et dont l'usage appartient surtout aux Canadiens-français*. Montréal: C. O. Beauchemin et fils.

Clapin, Sylva (Ed.). 1902. *A New Dictionary of Americanisms, Being a Glossary of Words Supposed to be Peculiar to the United States and the Dominion of Canada*. New York: L. Weiss.

Clapin, Sylva (Ed.). 1905. *Nugent's Up-To-Date English–French and French–English Dictionary According to the Best Authorities, and Containing All the Words Generally Used in Both Languages/ Nugent — Nouveau dictionnaire Français–Anglais et Anglais–Français rédigé d'après les meilleures autorités, et contenant tous les mots généralement en usage dans les deux langues*. Montréal: Librairie Beauchemin.

COD (*Canadian Oxford Dictionary*). 1998. Barber, Katherine (Ed.). Toronto: Oxford University Press.

Concise Dictionary of Canadianisms. 1973. Abridgement of *DC*. Toronto: Gage Educational Publishing.

Considine, John. 2000. Fourteen Words for Moose: Cultural and Intercultural Contexts of Four Canadian Dictionaries, 1977–1998. Antor, H. and K. Stierstorfer (Eds.). 2000. *English Literatures in International Contexts*: 317-331. Anglistische Forschungen 283. Heidelberg: Carl Winter Universitätsverlag.

Considine, John. Forthcoming. Dialectology, Storytelling, and Memory: Jack Thiessen's Mennonite Dictionaries. Antor, H., S.M. Brown, J.P. Considine, and K. Stierstorfer (Eds.). *Refractions of Germany in Canadian Literature and Culture*. Berlin/New York: De Gruyter.

Crystal, David. 1995. *Cambridge Encyclopedia of the English Language*. Cambridge: Cambridge University Press.

Crystal, David. 1998. *Language Play*. Harmondsworth, UK: Penguin.

DC (*A Dictionary of Canadianisms on Historical Principles*). 1967. Avis, Walter S. (Ed.). Charles Crate, Patrick Drysdale, Douglas Leechman, and Matthew H. Scargill (Ed. board). Charles J. Lovell (Ed. until 1960). Toronto: W.J. Gage.

De Wolf, Gaelan Dodds. 1992. *Social and Regional Factors in Canadian English: A Study of Phonological Variables and Grammatical Items in Ottawa and Vancouver*. Toronto: Canadian Scholars' Press.

DNE (*Dictionary of Newfoundland English*). 1982[1]. Story, G.M., W.J. Kirwin, and J.D.A. Widdowson (Eds.). Toronto: University of Toronto Press.

DNE (*Dictionary of Newfoundland English*). 1989[2]. Story, G.M., W.J. Kirwin, and J.D.A. Widdowson (Eds.). Toronto: University of Toronto Press.

DNE (*Dictionary of Newfoundland English*) [online]. 1999[2]. Story, G.M., W.J. Kirwin, and J.D.A. Widdowson (Eds.). Available: <http://www.heritage.nf.ca/dictionary/> [viewed May 30, 2003].

DPEIE (*Dictionary of Prince Edward Island English*). 1988. Pratt, T.K. (Ed.). Toronto: University of Toronto Press.

Dulong, Gaston. 1966. *Bibliographie linguistique du Canada français*. Québec: Les Presses de l'Université Laval/Paris: Librairie C. Klincksieck.

Dzwonkoski, David. 1986. Holt, Rinehart and Winston. Dzwonkoski, Peter (Ed.). 1986. *American Literary Publishing Houses, 1900–1980: Trade and Paperback:* 190-91. Dictionary of Literary Biography 46. Detroit: Gale Research.

Everhart, Jim. 1968. *The Illustrated Texas Dictionary of the English Language*. Houston: Creative Books.

Federal Court Reports. 1994. Canada v. Seaboard Lumber Sales Co. [online]. Available: <http://reports.fja.gc.ca/fc/1994/pub/v3/1994fca0256.html> [viewed May 28, 2003].

Focus Strategic Communications [online]. 2003. Available: <http://www.focussc.com/packpub.html> [viewed May 29, 2003].

Funk and Wagnalls Standard College Dictionary: Canadian Edition. 1963[1]. Toronto: Longmans Canada.

Funk and Wagnalls Standard College Dictionary: Canadian Edition. 1973[2]. Toronto: Fitzhenry and Whiteside.

Gage Canadian Dictionary. 1996[6]. De Wolf, Gaelan Dodds (Ed.). Toronto: Gage Educational Publishing.

Gage Canadian Dictionary. 2000[7]. De Wolf, Gaelan Dodds (Ed.). Vancouver/Toronto: Gage Educational Publishing.

Gage Learning [online]. 2003. Available: <http://www.gagelearning.com> [viewed May 28, 2003].

Geikie, A. Constable. 1857/1980–. Canadian English. *Canadian Journal of Science, Literature, and History* 2(11): 344-55. *CIHM/ICMH Microfiche Series* fiche P05122-077308. Ottawa: Canadian Institute for Historical Microreproductions.

Goddard, Ives. 1996. The Description of the Native Languages of North America before Boas. Goddard, Ives (Ed.). 1996. *Handbook of North American Indians vol. 17: Languages*: 17-42. Washington, DC: Smithsonian Institution.

Gregg, R. J. 1962. Canadian Lexicography [review of *The Canadian Dictionary/Dictionnaire canadien*]. *Canadian Literature* 14: 68-71.

Haldane, Maeve. 2003. Compiling Creole in the Caribbean. *McGill Reporter* 35(16) [online]. Available: <http://www.mcgill.ca/reporter/16/winer/> [viewed May 29, 2003].

Halford, Peter W. (Ed.). 1994. *Le français des Canadiens à la veille de la Conquête: Témoignage du père Pierre Philippe Potier, s.j.*. Collection Amérique Française 2. Ottawa: University of Ottawa Press.

Hansard [Nova Scotia]. 2001. Debate in the fifty-eighth general assembly of the Legislature of Nova Scotia, May 11 [online]. Available: <http://www.gov.ns.ca/legislature/hansard/han58-2/h01may11.htm> [viewed May 28, 2003].

Hewson, John. 1978. *Beothuk Vocabularies: A Comparative Study*. Technical Papers of the Newfoundland Museum 2. St John's: Newfoundland Museum.

Holland, Joan. 2003. *Dictionary of Old English: 2002 Progress Report* [online]. Available: <http://www.doe.utoronto.ca/Progress2002-web.pdf> [viewed June 3, 2003].

Houghton Mifflin Canadian Dictionary of the English Language. 1980. Morris, William (Ed.). Markham, ON: Houghton Mifflin Canada.

ITP Nelson Canadian Dictionary of the English Language. 1997. Scarborough, ON: ITP Nelson.

Kirwin, W.J. 1975. Selecting and Presenting the Lexicon. *Regional Language Studies* 6: 5-9.

Kirwin, W.J. and G.M. Story. 1986. The Etymology of High Liner: Problems of Inclusion in the Dictionary of Newfoundland English. *American Speech* 61(3): 281-284.

Kister, Kenneth F. 1977. *Dictionary Buying Guide: A Consumer Guide to General English-language Wordbooks in Print*. New York: R.R. Bowker.

Kretchman, Julie (Editorial Director, Harcourt Canada). 2003. Personal communication, May 28.

Landau, Sidney I. 1984/1989. *Dictionaries: The Art and Craft of Lexicography*. New York: Charles Scribner's Sons. Reprinted Cambridge: Cambridge University Press.

Lauder, Afferbeck [i.e. Alistair Morrison]. 1965. *Let Stalk Strine*. Sydney: Ure Smith.

LeClaire, Nancy and George Cardinal. *Alberta Elders' Cree Dictionary/alperta ohci kehtehayak nehiyaw otwestamâkewasinahikan*. Waugh, Earle (Ed.). Edmonton: University of Alberta Press/Duval House.

Lovell, Charles J. 1955a. Lexicographic Challenges of Canadian English. *Journal of the Canadian Linguistic Association* 1(1): 2-5.

Lovell, Charles J. 1955b. Whys and Hows of Collecting for the Dictionary of Canadian English I: Scope and Source Material. *Journal of the Canadian Linguistic Association* 1(2): 3-8.

Lovell, Charles J. 1956. Whys and Hows of Collecting for the Dictionary of Canadian English II: Excerption of Quotations. *Journal of the Canadian Linguistic Association* 2(1): 23-32.

Malong, Rawbone [i.e. Robin Malan]. 1972. *Ah Big Yaws? A Guard to Sow Theffricun Innglissh*. Cape Town: David Philip.

McConnell, Ruth E. 1978. *Our Own Voice: Canadian English and How it Came to Be*. Toronto: W.J. Gage.

Memorial University. 2003. English Language Research Centre [online]. Available: <http://www.mun.ca/english/elrc.html> [viewed June 3, 2003].

Microsoft Encarta College Dictionary. 2001. New York: St Martin's Press.

Moreton, Julian. 1863. *Life and Work in Newfoundland: Reminiscences of Thirteen Years Spent There*. London: Rivingtons.

National Library of Canada. 2003. *Amicus Web* [online]. Available: <http://www.nlc-bnc.ca> [viewed May 28, 2003].

Orkin, Mark M. 1973[1]. *Canajan, Eh?* Don Mills, ON: General Publishing.

Orkin, Mark M. 1997[3]. *Canajan, Eh?* Toronto: Stoddart.

Oxford University Press. 1997. *The Canadian Oxford Dictionary: The Foremost Authority on Current Canadian English* [advertising brochure]. Toronto: Oxford University Press.

Paikeday, Thomas M. 1989. Revolutionizing Dictionaries [review of Morton Benson, Evelyn Benson, and Robert Ilson, *The BBI Combinatory Dictionary of English: A Guide to Word Combinations*]. *American Speech* 64(4): 354-362.

Paikeday, Thomas M. 1992. O Corpora! *Lexicographica* 8: 307-317.

Penguin Canadian Dictionary. 1990. Paikeday, Thomas M. (Ed.). Markham, ON: Penguin Books.

Parkin, Tom. 1989. *WetCoast Words: A Dictionary of British Columbia Words and Phrases.* Victoria, BC: Orca.

Poteet, Lewis J. 1983. *The South Shore Phrase Book.* Hantsport, NS: Lancelot Press.

Pratt, Terry and Scott Burke (Eds.). 1998. *Prince Edward Island Sayings*. Toronto: University of Toronto Press.

Pratt, Terry. 1999. Personal communication, September 27.

Rudnyckj, J.B. 1962–[1]. *An Etymological Dictionary of the Ukrainian Language*. 22 fascicles. Winnipeg: Ukrainian Free Academy of Sciences.

Rudnyckj, J.B. 1966–[2]. *An Etymological Dictionary of the Ukrainian Language*. 10 fascicles. Winnipeg: Ukrainian Free Academy of Sciences.

Sandilands, John. 1912[1]/1976–. *Western Canadian Dictionary and Phrase-book: Explaining in Plain English for the Special Benefit of Newcomers the Meaning of the Most Common Canadianisms and Colloquialisms*. Winnipeg: printed for the editor. *Peel Bibliography on Microfiche* fiche 2342. Ottawa: National Library of Canada.

Sandilands, John. 1913[2]/1975. *Western Canadian Dictionary and Phrase-book: Things a Newcomer Wants to Know.* Winnipeg: printed for the editor. John Orrell (Ed.). Facsimile reprint. Edmonton: University of Alberta Press.

Scargill, M.H. 1958. Canadian Dictionary Projects II: Canadian English. *Journal des traducteurs* 3(3): 114-121.

Scargill, M.H. 1967. Preface. Avis, Walter S. (Ed.). 1967. *A Dictionary of Canadianisms on Historical Principles*: vi-vii. Toronto: W.J. Gage.

Seary, E.R. 1971. *Place Names of the Avalon Peninsula of the Island of Newfoundland*. Toronto: University of Toronto Press for Memorial University of Newfoundland.

Seary, E.R. 1977. *Family Names of the Island of Newfoundland*. With Sheila M. P. Lynch. St John's: Memorial University of Newfoundland.

Shewchuk, Blair. 2001. Gigs, Jigs, and Jibes [online]. Available: <http://www.cbc.ca/news/indepth/words/gigjig.html> [viewed May 27, 2003].

Société du parler français au Canada. 1930. *Glossaire du parler français au Canada*. Québec: L'Action Sociale.

Story, G.M. 1956. *A Newfoundland Dialect Dictionary: A Survey of the Problems*. St John's: Memorial University of Newfoundland.

Story, G.M. 1957. Research in the Language and Place-Names of Newfoundland. *Journal of the Canadian Linguistic Association* 3(2): 47-55.

Story, G.M. 1975a. A Critical History of Dialect Collecting in Newfoundland. *Regional Language Studies* 6: 1-4.

Story, G.M. 1975b. Edgar Ronald Seary. Macdonald, A.A., P.A. O'Flaherty, and G.M. Story (Eds.). 1975. *A Festschrift for Edgar Ronald Seary*: 1-8. St John's: Memorial University of Newfoundland.

Story, G.M. 1989. Review of *DPEIE*. *Canadian Historical Review* 70(3): 391-393.

Story, G.M. and W.J. Kirwin. 1974. 'The Dictionary of Newfoundland English': Progress and Promise. *Regional Language Studies* 5: 15-17.

Story, G.M., W.J. Kirwin and J.D.A. Widdowson. 1973. Collecting for the Dictionary of Newfoundland English. McDavid Jr., Raven I. and Audrey R. Duckert (Eds.). 1973. *Lexicography in English. Annals of the New York Academy of Sciences* 211: 104-108. New York: New York Academy of Sciences.

Tebbel, John W. 1972–81. *A History of Book Publishing in the United States*. 4 vols. New York: R.R. Bowker.

Thiessen, Jack. 1999. *Mennonitisch-Plautdeutsches Wörterbuch/Mennonite Low German Dictionary*. Steinbach, MB: Hanover Steinbach Historical Society.

University of Ottawa. 2003. Projet de lexicographie comparée du français et de l'anglais au Canada/Comparative Lexicography Project of French and English in Canada [online]. Available: <http://www.dico.uottawa.ca/> [viewed June 3, 2003].

Vinay, Jean-Paul and Pierre Daviault. 1958. Dictionnaires canadiens I: les dictionnaires bilingues. *Journal des traducteurs* 3(3): 109-113.

Wees, W.R. 1967. Foreword. Avis, Walter S. (Ed.). 1967. *A Dictionary of Canadianisms on Historical Principles*: v. Toronto: W.J. Gage.

Winston Canadian Dictionary for School, Home, and Office. 1974. Toronto/Montréal: Holt, Rinehart and Winston of Canada.

Winston Dictionary of Canadian English: Elementary Edition. 1975. Paikeday, Thomas M. (Ed.). Toronto: Holt, Rinehart and Winston of Canada.

Winston Dictionary of Canadian English: Intermediate Edition. 1969. Paikeday, Thomas M. (Ed.). Toronto: Holt, Rinehart and Winston of Canada.

Winston Simplified Dictionary for Canada. 1937. Brown, Thomas Kite and Henry Alexander (Eds.). Toronto: Winston.

The Winston Simplified Dictionary: Including All the Words in Common Use Defined so that they can be Easily Understood. 1919. Lewis, William D. and Edgar A. Singer (Eds.). Philadelphia: John C. Winston.

'n Vertalende vakwoordeboek van politieke terme

Mariëtta Alberts, *Leksikografie- en Terminologie-ontwikkeling, Pan-Suid-Afrikaanse Taalraad, Pretoria, Republiek van Suid-Afrika*
(marietta@pansalb.org.za)

Opsomming: *Staatkundige en Verwante Terminologie* wat in 1990 verskyn het, is aansienlik deur die Sentrum vir Politieke en Verwante Terminologie in Suider-Afrikaanse Tale (SEPTSA) hersien en bygewerk ten einde dit as die *Nuwerwetse Politieke Woordeboek* te publiseer. Die verskillende stadia in die hersiening en uitvoering van die projek word bespreek. Die hersiening bevat heelwat meer terme as die vorige publikasie en verskeie bylaes word ingesluit. Aandag word gegee aan die waarde van die tweetalige vertalende en die beplande meertalige verklarende politiekewetenskappeprojek binne die Suid-Afrikaanse en wêreldkonteks. Nuutskeppings en verskeie probleemterme word bespreek, spelling- en transliterasieprobleme wat by die verafrikaansing van ontleende terme ontstaan het, word behandel, en winste uit sowel die inheemse tale as uit vreemde tale word beskou. 'n Oorsig word gegee van 'n verskeidenheid funksionele agtervoegsels wat baie produktief in die *Nuwerwetse Politieke Woordeboek* gebruik word. Die woordeboek bevat 'n aansienlike aantal afkortings vir volvorme. 'n Aanduiding word gegee van die soort afkortings wat in die woordeboek opgeneem is. Die bylaes bygevoeg by die woordeboek word bespreek, en die artikel sluit af met 'n oorsig oor toekomsplanne.

Sleutelwoorde: AFKORTING, AKRONIEM, MEERTALIGE VAKWOORDEBOEK, NUUT-SKEPPING, PROBLEEMTERM, TERM, TERMINOLOGIE, TERMSKEPPING, TWEETALIGE TWEERIGTING VERTALENDE VAKWOORDEBOEK, TWEETALIGE VAKWOORDEBOEK, VERAFRIKAANSING, VERKLARENDE WOORDEBOEK, VERTALENDE WOORDEBOEK, VOL-VORM

Abstract: A Translating Technical Dictionary of Political Terms. *Political and Related Terminology* that appeared in 1990 was revised and updated considerably by the Centre for Political and Related Terminology in Southern African Languages (CEPTSA) in order to publish it as the bilingual *Modern Political Dictionary*. The various stages of the revision and execution of the project are discussed. The revision contains many more terms than the previous publication and various annexures are included. The value of the bilingual translating and the envisaged multilingual explanatory political sciences project within South African and world context receives attention. Neologisms and several problem terms are discussed, spelling and transliteration problems that developed from the Afrikaansification of borrowed terms are treated, and gains from the indigenous languages as well as from foreign languages are surveyed. An overview is given of a variety of functional suffixes used productively in the *Modern Political Dictionary*. The dictionary contains quite a number of abbreviations for full forms. An indication is given of the type of abbreviations entered in the dictionary. The annexures added to the dictionary are discussed, and the article ends with an overview of future plans.

http://lexikos.journals.ac.za

272 Mariëtta Alberts

Keywords: ABBREVIATION, ACRONYM, AFRIKAANSIFICATION, BILINGUAL BIDIREC-
TIONAL TRANSLATING TECHNICAL DICTIONARY, BILINGUAL TECHNICAL DICTION-
ARY, EXPLANATORY DICTIONARY, FULL FORM, MULTILINGUAL TECHNICAL DICTION-
ARY, NEOLOGISM, PROBLEMATIC TERM, TERM, TERM CREATION, TERMINOLOGY,
TRANSLATING DICTIONARY

1. Inleiding

Die ontwikkeling van internasionale staatkundige en politieke verhoudings gedurende die afgelope tien jaar asook die koms van demokrasie in Suid-Afrika het 'n groot toename in nuwe staatkundige en politieke terme tot gevolg gehad. Die bywerking en hersiening van *Staatkundige en Verwante Terminologie* wat in Januarie 1990 gepubliseer is, was dus noodsaaklik.

Taalgebruikers se kommunikasie- en leervaardighede en studente in die politieke en staatkundige wetenskappe kan baat vind by die tweetalige *Nuwerwetse Politieke Woordeboek.*

2.1 Agtergrond van die vertalende vakwoordeboek van politieke terme

Die vakwoordeboekprojek is in drie fases beplan. Die eerste fase was die publikasie van 'n tweetalige tweerigting vertalende vakwoordeboek. Daarna sou die definiëringsfase volg. Fase 2 was egter 'n hersieningsfase en die derde fase sal nou uit die definiëring en verskaffing van Afrikataalekwivalente bestaan.

Fase 1: Vertalend ⇨ Fase 2: Hersiening ⇨ Fase 3: Verklarend
(Tweetalig) (Tweetalig) (Meertalig)

Die hersiening van die bestaande vakwoordeboek was 'n grondliggende stap vir die volgende fase. Dit sal meebring dat 'n uitvoerige, bygewerkte vakwoordeboek tot stand kom waarvan die begrippe gedefinieer sal word voordat die Afrikataalekwivalente bygevoeg word ten einde 'n meertalige verklarende vakwoordeboek te publiseer.

2.1.1 Fase 1: Ontstaan

Gedurende 1983–84 is die behoefte aan 'n vakwoordeboek bepaal wat die politiek met sy verwante dissiplines dek. Die destydse Nasionale Terminologiediens (NTD) (tans Terminologiekoördineringsafdeling, Nasionale Taaldiens, Departement van Kuns en Kultuur) het daarop vakkundiges genader om aan die projek mee te werk en 'n Komitee vir Staatkundige en Verwante Terminologie is gevestig. Hierdie Komitee het van 1991 tot 1997 onder die NTD gefunksioneer. Toe die NTD en die eerste komiteelede die woordeboek begin beplan het, het hulle ooreengekom dat die komitee se eerste produk 'n tweetalige tweerigting vertalende vakwoordeboek sou wees. In fase 2 sou 'n opvolgpubli-

kasie wat die politieke terme definieer, tot stand gebring word.

Die eerste fase is gedurende 1989 afgehandel en in Januarie 1990 is *Staatkundige en Verwante Terminologie* gepubliseer en amptelik vrygestel.

Die produk van fase 1 het die subgebiede staatsleer, internasionale politiek en staatsreg, publieke administrasie, munisipale regering en administrasie, Afrikapolitiek en ontwikkelingsadministrasie, en strategiese studies ingesluit. Die termkorpus het sowat 10 000 terme (in sowel die Engels–Afrikaanse kant as die Afrikaans–Engelse kant) bevat. Benewens die NTD se taalkundige en terminografiese adviseurs het akademici verbonde aan universiteite en kundiges van die Kommissie vir Administrasie, Nasionale Intelligensiediens, Suid-Afrikaanse Polisie, Suid-Afrikaanse Weermag, Departement Buitelandse Sake, Departement Staatkundige Ontwikkeling en Beplanning, en die Afrika-Instituut, meegewerk.

2.1.2   Fase 2: Hersiening

Die dramatiese politieke omwenteling in Suid-Afrika het tot gevolg gehad dat baie beplanning op staatkundige terrein gedoen is om op nuwe beleid en strategieë te besluit ten einde nuwe staatkundige en politieke doelwitte te stel. Hierdie inisiatiewe het honderde nuwe staatkundige en politieke terme geskep.

Gedurende die finale redigering van die eerste uitgawe het die vakkomiteelede reeds met die tweede beplande fase, naamlik definiëring, begin. Ongelukkig het hierdie werk om verskeie redes stadig gevorder.

Die vakkundiges het van die begin af besef dat 'n verklarende termlys noodsaaklik is, maar dat die inhoud van die woordeboek van tyd tot tyd hersien sou moes word.

Ten tyde van die publikasie van die vakwoordeboek en die aanvang van fase 2 van die projek moes die vakkomitee oor die aanvanklike beplanning besin. Die behoefte aan 'n uitgebreide en hersiene uitgawe van *Staatkundige en Verwante Terminologie* het al hoe sterker geword. Die vakkomitee het derhalwe reeds in 1990 besluit om eers met die hersiening van die eerste woordeboek te begin voordat die terme gedefinieer sou word. Hierdie besluit is versterk deur die nuwe politieke bedeling wat tot 'n groot aantal nuwe terme aanleiding gegee het.

Die hersieningskomitee het die datakorpus van die eerste woordeboek nagenoeg verdriedubbel deur die bywerking van aanvullende terme en het ook die terme wat wel in die eerste woordeboek verskyn het, maar waarvan die fynere verskille ten opsigte van betekenis en nuanses nie genoegsaam ontrafel was nie, deeglik hersien. Die byvoeging van bylaes met inligting oor die parlemente van die wêreld en verwante instellings, intelligensie- en veiligheidsdienste van state, akronieme, e.s.m. is 'n stap wat die waarde van die nuwe publikasie aansienlik verhoog. Hierdie produk was teen die middel van 2002 persklaar en is in Desember 2002 gepubliseer.

Aan die einde van fase 2 is die eerste woordeboek dus wel aansienlik ver-

beter, maar die behoefte aan die verklaring van tersaaklike terme is nog nie bevredig nie. Twee verdere behoeftes het na vore gekom, naamlik die ontwikkeling van vakkundige terminologie vir die Afrikatale, en die gevolglike vergemakliking van die studie van studente wat nie Engels of Afrikaans as moedertaal het nie. Hoewel 'n aantal terme uit die Afrikatale reeds in die tweede fase van die woordeboek opgeneem is, is dit nie voldoende om die nuwe behoefte te bevredig nie en is dit noodsaaklik dat die huidige woordeboek tydens 'n derde fase verder ontwikkel word.

2.1.3   Fase 3: Meertalige verklarende woordeboek

Navorsing wat deur die Terminologiekoördineringsafdeling, Nasionale Taaldiens, Departement van Kuns en Kultuur gedoen is, toon 'n duidelike behoefte aan 'n verklarende woordeboek wat ook die Afrikatale insluit. 'n Snelgroeiende getal studente wat nie Engels as moedertaal het nie, studeer in vakrigtings wat deur hierdie pasverskene woordeboek gedek word. Dit is alreeds moeilik om in 'n tweede taal te studeer en die probleem word vererger deur die tegniese ingewikkeldheid van baie begrippe op hierdie terrein, wat nie deur die gewone woordeboeke ondervang word nie. Ook parlementariërs, vertalers binne en buite die staatsdiens, en skrywers in die media sal besondere baat vind by so 'n produk. Voorts sal dit daadwerklik tot die uitbou van die Afrikatale bydra en terselfdertyd die Afrikaanse vaktaal bevorder.

Die Regering het homself daartoe verbind om die taalregte van alle Suid-Afrikaanse burgers te beskerm. Die land het tans elf amptelike tale wat regmatige behandeling moet ontvang. Indien daar gepoog moet word om die ongelykhede van die verlede uit die weg te ruim, sal duidelike doelwitte gestel moet word vir taalontwikkeling in Suid-Afrika en stategieë sal bedink moet word om die Regering se doelwit van taalgelykheid te bereik.

Een van die plekke waar taalontwikkeling tot sy reg kan kom, is in die onderwyssituasie. As skoliere en studente begrippe in hul onderskeie eerste tale kan bemeester, sal hul eindresultate ook aansienlik beter wees. Nie alleen behoort die standaard van onderwys in Engels as medium van onderrig te verbeter nie, maar die ander amptelike tale behoort ook gemoderniseer en ontwikkel te word om kragtige kommunikasie- en onderrigmiddele te word.

'n Ideale onderrigsituasie sal vereis dat die terminologie van verskillende vakgebiede in die onderskeie amptelike tale beskikbaar moet wees. Meertalige verklarende vakwoordeboeke kan meehelp om die terminologie van bepaalde vakgebiede te verklaar en die leerproses aan te help.

Omdat daar egter 'n tekort aan vakkundiges bestaan wat tegelyk oor die vereiste taalkennis en kennis van die onderskeie subgebiede beskik, is dit nodig dat omskrywings of definisies van sleutelterme in Engels en Afrikaans geformuleer word om as grondslag te dien vir die vakkundig korrekte en aanvaarbare ekwivalent of omskrywing van die terme in die betrokke Afrikatale.

Fase 3 vereis dat die huidige werkspan uitgebrei word deur die byvoeging

van kundiges met van die Afrikatale as eerste taal. Daarna sal die huidige data-korpus waarskynlik ook aangepas moet word om begrippe wat eie is aan die Afrikakultuur by te voeg. Die huidige datakorpus sal uitgebrei word om vir die bykomende doeltaalekwivalente voorsiening te maak.

Die terme in die databasiskorpus is op MultiTerm ingevoer in databasis-formaat. Dit impliseer dat selfs navorsingsnotas wat nie normaalweg in 'n vakwoordeboek verskyn nie, ook opgeteken is. Die terme maak ook deel uit van die Nasionale Termbank van die Nasionale Taaldiens, Departement van Kuns en Kultuur.

Die loodsondersoek wat reeds gedoen is, bevestig die behoefte aan so 'n meertalige verklarende vakwoordeboek. Dit toon ook dat die projek uitvoer-baar is. Daar is egter 'n belangrike faktor wat nie uit die oog verloor moet word nie, naamlik dat die huidige werkspan nie oor genoegsame fondse beskik om die verdere ontwikkelingsfases van die projek te finansier nie.

2.2.   Sentrum vir Politieke en Verwante Terminologie in Suider-Afrikaanse Tale (SEPTSA)

Die Komitee vir Staatkundige en Verwante Terminologie het as gevolg van die staking van staatsteun in 1997 besluit om 'n Sentrum vir Politieke en Verwante Terminologie in Suider-Afrikaanse Tale (SEPTSA) te stig ten einde die hersiene tweetalige woordeboek uit te gee, die meertalige woordeboek te produseer en ander professionele produkte te lewer. Die reglement van SEPTSA is op 27 No-vember 1998 gefinaliseer. Dit funksioneer tans met die welwillende medewer-king van die Randse Afrikaanse Universiteit en is in sy Departement Politieke en Regeerkunde gevestig. SEPTSA word dus met behulp van 'n grondwet wat deur die Universiteitsraad goedgekeur is, bestuur en met 'n sakeplan bedryf.

Die medewerkers van SEPTSA poog om fondse vir woordeboekprojekte op 'n verantwoordelike wyse te bekom, aan te wend en te administreer. Prolin-gua se finansiële steun van 1999 tot 2000 het dit moontlik gemaak om met die werk voort te gaan. SEPTSA het vroeg in 2001 en in 2002 fondse van die Taal-sekretariaat ontvang wat gesorg het vir die voltooiing van die projek. Die Randse Afrikaanse Universiteit het in 2002 'n bydrae gemaak sodat die woor-deboek gepubliseer kon word.

3.   Politieke terminologie in die vertalende vakwoordeboek

Die volgende vakgebiede word in die *Nuwerwetse Politieke Woordeboek* verteen-woordig: politieke wetenskappe, publieke administrasie en regeerkunde, inter-nasionale verhoudinge, ontwikkelingstudies en strategiese studies. Terme is wyd geëkserpeer, nie slegs uit vakkundige bronne nie, maar ook uit populêre bronne soos tydskrifte en koerante. Die Komiteelede moes op grond van hul eie kennis, aanvoeling en inligting uit verwante bronne besluite neem oor die

omvang en vakgebiedgrense van die woordeboek, met die toepaslikheid van elke term vir hierdie woordeboek as riglyn.

3.1   Nuutskeppings

SEPTSA het met baie interessante terme te doen gekry wat van termekwivalente voorsien moes word. Dit het dan ook vir 'n paar goeie nuutskeppinge in Afrikaans gesorg wat in die *Nuwerwetse Politieke Woordeboek* verskyn:

> **brain gain: kundigheidswins**
> **brain train: kundigheidsopleiding**
> **clout: klapkrag (spierkrag)**
> **constituency (voter's district): kiesafdeling (kiesdistrik)**
> **constituency: kiesergevolg**
> **crony: kornuit**
> **cronyism: kornuitbegunstiging**
> **entryism** *{practice of infiltrating a political party in order to influence that party's policy from within}*: **politieke binnedringing** *{praktyk om 'n politieke party te infiltreer ten einde daardie party se beleid van binne af te beïnvloed}*
> **fat cat: roomvraat (geiljan, vetkat** <af te keur>**)**
> **gangsterism: rampokkery**
> **gender: gender**
> **gutter education: flenteronderwys (afskeeponderwys)**
> **impeach: ampsaanklag instel teen**
> **impeachment** *{charge against incumbent of a specific office}*: **ampsaanklag** *{aanklag teen bekleër van 'n bepaalde amp}*
> **impeachement: staat van aanklag**
> **kickback: gunsloon**
> **lumpenproletariat** *{in the Marxist theory}*: **flenterbroekproletariaat** *{in die Marxistiese teorie}* **(lumpenproletariaat)**
> **periphery doctrine** *{Israeli foreign policy}*: **randdoktrine** *{Israelse buitelandse beleid}*
> **shamocracy: foppokrasie (skyndemokrasie)**

3.2   Probleemterme

Daar is verskeie brontaalterme wat probleme skep wanneer hulle van vertaalekwivalente voorsien moet word. In sommige gevalle is die terme bloot net nie vertaalbaar nie omdat bepaalde emotiewe waardes wat reeds in die brontaal aan die terme gekoppel word, in die vertaalproses verlore gaan. In ander gevalle bestaan daar reeds vertaalekwivalente wat nie dieselfde betekeniswaarde van die terme in die brontaal weergee nie. Voorbeelde hiervan wat in die *Nuwerwetse Politieke Woordeboek* opgeneem is, is die volgende:

> **civics** ⇒ civic associations [1]
> **civic association: burgervereniging**
> **civic associations (civics): civics**

dumping *{eg of radioactive waste}***: storting** *{bv van radio-aktiewe afval}*
dumping *{eg of economic goods}***: dumping** *{bv van ekonomiese goedere}*

struggle: stryd
struggle <the struggle> *{in South Africa it refers to the period ca 1960–1990 in which the black freedom movement conducted a violent revolutionary struggle against the NP government}***: struggle** <die struggle> *{in Suid-Afrika verwys dit na die tydperk ca 1960–1990 waarin swart bevrydingsbewegings 'n gewelddadige rewolusionêre stryd teen die NP-regering gevoer het}*

township *{a black residential area, mainly associated with apartheid}***: township** *{'n swart woongebied, hoofsaaklik verbind met apartheid}*
township: lokasie
township: woonbuurt
township *{term denoting various forms of residential areas in other states such as US, Canada and UK}***: dorpsgebied** *{term wat verskeie vorme van woongebiede in ander state soos die VSA, Kanada en VK aandui}*

3.3   Spelling- en transliterasieprobleme

In sekere gevalle waar daar onduidelikheid bestaan het oor die spelling van die brontaalterme, het SEPTSA seker gemaak dat die korrekte spelwyse opgeneem word, veral by terme van vreemde herkoms. Die oorspronklike taal se spelling en uitspraak is ook met die betrokke ambassades uitgeklaar. Die *Afrikaanse Woordelys en Spelreëls* (AWS) is geraadpleeg wanneer terme van vreemde herkoms 'n verafrikaanste termekwivalent gekry het. Indien die AWS nie genoegsame leiding kon verskaf nie, is die Taalkommissie geraadpleeg.

SEPTSA het baie moeite gedoen om die regte uitspraak van terme te bekom en om dan tydens die transliterasieproses seker te maak dat die uitspraak nie tydens die verafrikaansing daarvan verlore gaan nie. Die navorsingsnotas by die winste uit vreemde tale dui op die besondere sorg waarmee die transliterasieproses hanteer is, bv. by **Hasidim** en **mukhtar**:

Hasidim <Hebrew; pl of Hasid> *{ultra-orthodox Jewish sect}***: Gasidim** *{ultra-ortodokse Joodse sekte}* **Navorsingsnota:** Collins gee die Hebreeuse uitspraak as *xasidim*; Webster gee: *hasid/chasid, hassid/chassid* pl. *hasidim/hassidim* — one who is pious. Bosman, Van der Merwe en Hiemstra verwys van *Has(s)idism* na *Chas(s)idism* en gee as Afrikaanse ekwivalent *Chassidisme.* Die Komitee moet versigtig wees wat die verafrikaansing betref en behoort minstens ook ander spelvorme toe te laat soos *chasidim/chassidim.* Hierdie tipe transliterasie is soortgelyk aan **ch**aos, **ch**emie, **ch**arisma. Gesprekke met kundiges en raadpleging van die AWS toon dat die sagte **g** in Afrikaans meestal as **g** geskryf word, maar dat ander vorme wel as **ch**, byvoorbeeld **ch**emie, voorkom en reeds ingeburger is. Die geval behoort na die Taalkommissie verwys te word vir leiding.

mukhtar <Arabic> *{elected headman of a settlement}***: moegtar** *{gekose hoofman van 'n nedersetting}* **Navorsingsnota:** Op grond van ondervinding met die verafrikaansing van die Engelse vorm *-kh-* soos in *mukhtar*, word daar aanbeveel

dat die Komitee in Afrikaans *moechtar* skryf — dis die sagte *g*-klank soos in *ch*emie. Die voorstel is dat dit minstens as wisselvorm opgeneem word en as voorkeurspelling vir Afrikaans aanvaar word. Die geval behoort na die Taalkommissie verwys te word vir leiding.

3.4.1 Winste uit die inheemse tale

Daar is etlike terme uit die inheemse tale wat reeds in Afrikaans ingeburger is. Voorbeelde hiervan is terme soos **shebeen** (Afr. **sjebien**), **stokvel**, **spaza** en **spaza shop** (Afr. **spazawinkel**).

Meer onlangse winste kan nog teruggevoer word na die oorspronklike tale. Voorbeelde hiervan wat ook die Afrikaanse vertalings toon, is die volgende:

(a) Zoeloe, bv. **amabutho, amadlozi, imbizo, imbongi, impi, impimpi, indaba, induna, inkatha, umKhonto weSizwe, mukhuku, inkosi** en **Shenge.**

amabutho <Zulu; pl of ibutho> *{peer group}*: **amabutho** <Zoeloe; mv van ibutho> *{portuurgroep}*

amabutho <Zulu — pl of ibutho; incorrectly used by the media to refer to people in traditional gear purported to belong to Inkatha> *{a group of traditional warriors}*: **amabutho** <Zoeloe — mv van ibutho; word verkeerdelik deur die media gebruik om na mense in tradisionele drag as behorende tot Inkatha te verwys> *{'n groep tradisionele krygers}*

amadlozi <Zulu> *{spirits of the forefathers}*: **amadlozi** <Zoeloe> *{voorvadergeeste}*

imbizo <Zulu> *{a gathering called by the traditional leader of a nation to discuss serious or important issues pertaining to the nation}*: **imbizo** <Zoeloe> *{'n vergadering byeengeroep deur 'n tradisionele leier van 'n nasie om ernstige of belangrike kwessies wat die nasie raak, te bespreek}*

impi <Zulu>: **impi** <Zoeloe> [2]

indaba <Zulu> *{a serious business to be discussed}*: **indaba** <Zoeloe> *{'n belangrike saak wat bespreek moet word}*

induna <Zulu> *{the traditional counsellor of a ruler or king appointed because of his allegiance and who sometimes manages certain of the affairs of the ruler or king}*: **indoena** <Zoeloe> *{die tradisionele raadgewer van 'n heerser of koning weens sy trou en wat soms sekere sake van die heerser of koning behartig}*

(b) Sotho, bv. **difaqane** en **lekgotla.**

difaqane <Sotho> *{displacement of population caused by wars — eg those by Shaka}* **(mfecane** <Nguni>**): bevolkingsverplasing** *{bevolkingsverplasing agv oorloë — bv dié deur Shaka}* [3]

lekgotla <Sotho> *{eg court where men sit; troop of warriors}*: **lekgotla** <Sotho> *{bv hof waar mans sit; troep krygers}*

(c) Setswana, bv. **Kgosi.**

Kgosi <pl diKgosi; Setswana> *{leader of a group in Tswana communities}*: **Kgosi** <mv diKgosi; Setswana> *{leier van 'n groep in Tswana-gemeenskappe}*

3.4.2   Winste uit Afrikaans

Dié terme, bv. **apartheid**, **Hanskakie**, **heemraad**, **kraal**, **kragdadig**, **verkramp**, **verlig**, **Vierkleur**, **volkskapitalisme**, **volkskool** en **volkstaat** is onveranderd in Engels oorgeneem.

> **apartheid** *{former policy of institutionalised separation of or discrimination between races in South Africa}*: **apartheid** *{voormalige beleid van geïnstitusionaliseerde skeiding van of diskriminasie tussen rasse in Suid-Afrika}*
>
> **kragdadig** <Afrikaans — compare forceful > *{with reference to forceful politics; orginated with John Vorster}*: **kragdadig** *{met verwysing na kragdadige politiek; begin met John Vorster}*
>
> **verkramp** *{highly conservative in politics; originated in RSA}*: **verkramp** *{hoogs konserwatief in politiek; oorsprong in RSA}*
>
> **verlig** *{liberal in politics; originated in RSA}*: **verlig** *{liberaal in politiek; oorsprong in RSA}*

3.4.3   Winste uit vreemde tale

Behalwe terme soos **Great Society, Green Berets, green card, Ku Klux Klan** en **marine** uit Amerikaans, **gentleman** en **Master of the House** uit Engels, **grundgesetz, länder citizenship, Luftwaffe, reichskanzler** en **stiftung** uit Duits en **gendarmerie, laisser-aller, laisser-faire** en **laissez-passer** uit Frans, is dit terme uit veral die Oosterse tale wat in *Nuwerwetse Politieke Woordeboek* opval. Ontlenings aan die volgende tale kan met enkele voorbeelde wat ook die Afrikaanse vertalings toon, vermeld word:

(a)   Arabies, bv. **fedayeen, Hizbollah, jihad, jamhuriyah, mukhtar** en **wazir.**

> **fedayeen** <pl; Arabic> *{Muslim commando force, especially one operating against Israel}*: **fedajien** *{Moslem-kommandomag, veral een wat teen Israel optree}*
>
> **jihad** *{holy Islamic struggle}*: **djehad** <meer korrekte vorm in Afrikaans> *{heilige Islamitiese stryd}* **(djihad, jihad)** [4]
>
> **jamhuriyah** <Arabic> *{republic, sometimes people's republic}* **(jumhúriyah): djomhoerija** *{republiek, soms volksrepubliek}*
>
> **wazir** <Arabic> *{minister}*: **visier** *{minister}*

(b)   Hebreeus, bv. **Haganah, Hasidim, Histadrut, Mifliget Po'alei Eretz Yisrael** en **moshav.**

> **Haganah** <Hebrew> *{Israeli pre-state underground army}*: **Hagana** *{Israelse voorstaatse ondergrondse leër}*
>
> **moshav** <Hebrew; sing of moshavim> *{cooperative settlement in Israel, consisting of a number of small farms}*: **mosjav** *{koöperatiewe nedersetting in Israel wat uit 'n groepie klein plasies bestaan}*

(c)   Indiese tale, bv. **Kshatriya, Lok Sabha, kuchitollah, pariah, raj, raja, Rajya Sabha, Sudra, suttee, Vaisya** en **varna.**

> **Kshatriya** <Indian languages> *{the second of the four main Hindu varna, the warrior*

caste}: **Tsjatrija** *{die tweede kaste van die vier hoof Hindoe-warne, die krygerkaste}*

raj <Indian languages> *{rule, especially in India}*: **radj** *{heerskappy, veral in Indië}*

Sudra <Indian languages> *{lowest of the four main Hindu varna, the workers}*: **Sjoedre** *{die laagste van die vier hoof Hindoe-warne, die werkers}*

suttee <Indian languages> *{former custom of widow burning in India}*: **sati** *{voormalige gewoonte van weduweeverbranding in Indië}*

Vaisya <Indian languages> *{the third of the four main Hindu varna, the traders}*: **Waisje** *{die derde van die vier hoof Hindoe-warne, die handelaars}*

varna <Indian languages> *{the degree of pureness determining a caste}*: **warne** *{graad van reinheid wat 'n kaste bepaal}*

(d) Indonesies, bv. **keterbukaan.**

keterbukaan <Indonesian> *{openness — Indonesian political policy}*: **kieterboekaan** *{openheid — Indonesiese politieke beleid}*

(e) Iraans, bv. **bandbazi, grand ayatollah, jasus-khane** en **velayet-e faqih.**

jasus-khane <Iranian> **(house of spies)** *{with reference to the American embassy in Tehran}*: **jasoes-gane** *{met verwysing na die Amerikaanse ambassade in Teheran}* **(huis van spioene)**

velayet-e faqih <Iranian> *{governance by Islamic jurisprudence}*: **welajet-e fakieh** *{regering volgens Islamitiese regsleer}*

(f) Japannees, bv. **chuzaisho, daimyo, Domei, genro, habatsu, Heisei, ie, kaisha, kaizen, kamikaze, kanban, Keidanren, keiretsu, Keizai Dóyukai, kinken seiji, ko, koenkai, Kokkai, Meiji, mikado, Nikkeiren, ringi, samurai, Sei-i Tai Shogun, shogun, shogunate, Sho Wa, Sohyo, yakuza, zaibatsu** en **zaikai.**

daimyo <Japanese> *{feudal lords in shogunate}*: **daimijô** *{feodale lords in sjoghoenaat}*

Domei <Japanese> *{private sector labour union confederation}*: **Domê** <Japannees> *{vakbondfederasie van die privaatsektor}* [5]

genro <Japanese> *{elder statesmen}*: **ghenrô** *{ringkopstaatsmanne}* [6]

Heisei <Japanese> (era of universal peace) *{title of the reign of emperor Akihito}*: **Heisei** <Japannees> *{titel van die bewind van keiser Akihito}* (tydvak van universele vrede)

ie <Japanese> *{Japanese household; building block of the feudal Japanese political system}*: **ijee** *{Japannese huishouding; bousteen van die feodale Japannese politieke stelsel}*

kamikaze <Japanese> *{literally divine wind blowing away the enemy; generally refers to suicidal military operations}*: **kamikazê** *{letterlik goddelike wind wat die vyand wegwaai; verwys gewoonlik na selfmoord militêre operasies}*

Keidanren <Japanese> *{federation of industrial and trading organisations}*: **Kêdanren** *{federasie van nywerheids- en handelsorganisasies}*

koenkai <Japanese> *{informal campaign organisers and supporters for local politicians}*: **kôênkai** *{informele veldtogorganiseerders en ondersteuners van plaaslike politici}*

mikado <Japanese> *{title foreigners use for the Japanese emperor}*: **mikadô** *{titel wat buitelanders vir die Japannese keiser gebruik}*

ringi <Japanese> *{consensus}*: **ringhie** *{konsensus}*

samurai <Japanese> *{warrior caste in feudal Japan}*: **samoeraai** *{krygskaste in feodale*

Japan}

shogun <Japanese> *{emperor's chief general and* de facto *ruler; specific mandate of emperor}***: sjoghoen** *{keiser se oppergeneraal en* de facto *heerser; bepaalde mandaat van keiser}*

(g) Koreaans, bv. **chaebol** en **juche.**

chaebol <Korean> *{government-promoted dominant business conglomerate}***: tjêbil** *{oorheersende sakekonglomeraat wat deur die regering bevorder word}*

juche <Korean> *{autarchical political economic policy of North Korea}***: joetje** <Koreaans> *{outargiese politieke ekonomiese beleid van Noord-Korea}*

(h) Sjinees, bv. **ganbu** en **Kuo-miin ta-hui.**

ganbu <Chinese> *{leading bureaucrats in the People's Republic of China}***: ghanboe** *{vooraanstaande burokrate in die Volksrepubliek Sjina}*

Kuo-min ta-hui <Chinese> *{National Assembly; amends constitution, formerly used to elect president}***: Gwo-min ta-hê** *{Nasionale Vergadering; wysig grondwet, voorheen gebruik om president te kies}*

(i) Turks, bv. **ulema.**

ulema <Turkish> *{Muslim scholars}* (**ulama** <Arabic>)**: oelema** *{Moslemgeleerdes}*

3.5   Funksionele agtervoegsels

Dit blyk dat daar veral sekere agtervoegsels is wat baie funksioneel in *Nuwerwetse Politieke Woordeboek* optree:

(a) **-acy**, bv. in **conspiracy, diplomacy, efficacy, fallacy, literacy, papacy** en **supremacy.** Behalwe **diplomacy** word nie een van dié terme in Afrikaans met **-asie**-vorme vertaal nie. Dié agtervoegsel is dus nie ook met sy Afrikaanse ekwivalent gelemmatiseer nie.

(b) **-archy,** bv. in **anarchy, autarchy, callarchy, diarchy, monarchy, oligarchy** en **patriarchy.** Al dié terme word ook in Afrikaans vertaal met die gebruikmaking van die agtervoegsel **-argie**. Dié agtervoegsel is dus soos volg gelemmatiseer:

-archy *{rule by}***: -argie** *{heerskappy deur}*

(c) **-cracy,** bv. in **aristocracy, autocracy, bureaucracy, democracy, ergotocracy, gerontocracy, hagiocracy, hierocracy, isocracy, kakistocracy, nomocracy, ochlocracy, plutocracy, shamocracy, technocracy** en **theocracy.** Al dié terme word in Afrikaans met **-krasie**-vorme vertaal. Dié agtervoegsel is dus soos volg opgeneem:

-cracy *{exercise of power by}***: -krasie** *{uitoefening van mag deur}*

Die definisie is egter te beperk, want dit dek nie die volledige spektrum van betekenisse wat dié agtervoegsel kan hê nie: " 'n vorm van regering",

" 'n politieke of sosiale klas", of " 'n sosiale of politieke teorie" (vgl. Urdang 1982: 257).

(d)   **-ism,** bv. in **communism, dualism, elitism, fanaticism, globalism, holism**, en ook in gevalle wat van eiename afgelei is soos **Chartism, Confucianism, Fascism, Gaullism, Hertzogism.** Dit is vreemd dat dié agtervoegsel en die Afrikaanse vorm **-isme** wat albei 'n aansienlike gebruiksfrekwensie het, nie apart gelemmatiseer is nie. In die politiek het dit veral drie betekenisse: " 'n handeling of gebruik", " 'n gesteldheid of toestand", en "beginsels, leerstellings of opvattings, of 'n organisasie gestig om hulle te ondersteun" (vgl. Urdang 1982: 203).

3.6   Afkortings

Die *Nuwerwetse Politieke Woordeboek* bevat 'n groot aantal afkortings vir verskeie volvorme. Die volvorm en afkorting word albei opgeneem en die afkorting kruisverwys na die volvorm. Voorbeelde:

(a)   Letterafkortings

AAM ⇒ air to air missile
air to air missile (AAM): lug-na-lug-missiel (LLM, lug-tot-lug-missiel)
AAM ⇒ Anti-Apartheid Movement
Anti-Apartheid Movement (AAM): Anti-Apartheidsbeweging (AAB)

(b)   Akronieme

COSATU ⇒ Congress of South African Trade Unions
Congress of South African Trade Unions (COSATU): Congress of South African Trade Unions (COSATU)
ECOMOG ⇒ ECOWAS Monitoring Group
ECOWAS Monitoring Group *{military grouping of ECOWAS}* **(ECOMOG): ECOWAS Moniteringsgroep** *{militêre groepering van ECOWAS}* **(EGMOG)**
GEAR ⇒ Growth, Employment and Redistribution Strategy
Growth, Employment and Redistribution Strategy (GEAR): Strategie vir Groei, Werkverskaffing en Herverdeling (GEAR)

3.7   Bylaes

Die woordeboek word afgesluit deur 'n agtal tweetalige bylaes wat 'n aantal terme alfabeties of tabellaries bymekaargroepeer vir vinnige en (of) oorsigtelike naslaandoeleindes.

(a)   Die eerste bylae bevat akronieme en afkortings gevolg deur hul volvorme. Al die inskrywings in dié bylae is ook in die hooflys op hul alfabetiese plekke opgeneem waar hulle na die volvorme verwys word wat ook in die hooflys voorkom.

(b)   Die tweede bylae gee die parlemente van die wêreld in 'n tabel bestaande uit vier kolomme: die name van die state in die eerste kolom, dié van die parlemente in die tweede kolom en, waar parlemente twee kamers het, die kamers in die derde en vierde kolomme. Die tabel is alfabeties gerangskik volgens die algemene name van die state in Engels, gevolg deur die formele name van die state, wat op hul beurt weer deur die name van die state in hul oorspronklike tale gevolg word.

(c)   Die terme in die derde bylae wat adellike titels en ampte bevat, is ook alfabeties as lemmas in die hooflys opgeneem. Dié bylae is van groot nut vir die woordeboekgebruiker wat die Afrikaanse ekwivalente vir die vreemde benamings benodig of vir die oningeligte vir wie die verskil tussen verskillende benamings nie altyd duidelik is nie. Die volgende sal onder andere daar aangetref word: **prince consort** (Afr. **prinsgemaal**), **prince regent** (Afr. **prins-regent**), **prince royal** (Afr. **vorsteprins**), **queen consort** (Afr. **koningsgemalin**), **queen dowager** (Afr. **koninginweduwee**), **queen regent** (Afr. **koninginregentes**), **queen regnant** (Afr. **regerende koningin**).

(d)   In die vierde bylae word klassieke Latynse en Griekse terme wat ook in die hooflys opgeneem is, gerieflik saamgevat.

(e)   Die terminologie ten opsigte van ministeriële en administratiewe ampte in die Verenigde Koninkryk en die Verenigde State van Amerika verskil aansienlik van dié in Suid-Afrika. Al dié terme wat ook alfabeties in die hooflys van die woordeboek opgeneem is, word in die vyfde bylae hiërargies met hul vertalings aangebied. Dié oorsigtelike aanbieding wat ook vergelykings tussen die ampte van die verskillende lande moontlik maak, verskaf onontbeerlike hulp, ook ten opsigte van korrekte vertalings.

(f)   Die sesde bylae bevat tipiese militêre range. Waar dié terme as lemmas in die hooflys opgeneem is, word hulle met kleinletters geskryf. In hierdie hiërargiese lys waar die rangbenaminge met hul afkortings onder die hoofde "Leër/Lugmag" en "Vloot" gerangskik is, word hulle met hoofletters geskryf. Die notas wat op hierdie tabel volg en onder andere inligting oor aanspreekvorme verskaf, is van groot gebruikswaarde.

(g)   Gegewens oor inligtings- en veiligheidsdienste word in tabelvorm in die sewende bylae verskaf: in die eerste kolom word die name van die state vermeld; in die tweede kolom volg die afkortings van die dienste; die derde kolom wat die volvorme van die afkortings verskaf, is in drie afdelings verdeel: eers word die name in die oorspronklike tale gegee, daarna in Engels en Afrikaans, en, in die vierde kolom, volg opmerkings oor die aard van die dienste. Die afkortings is nie ook, soos die terme in die ander bylaes, alfabeties in die hooflys opgeneem nie. Dié tabel verskaf 'n maklik raadpleegbare oorsig van moeilik bekombare inligting.

(h)   In die agste bylae verskyn byname van politieke persoonlikhede. Die volgende sal onder andere daar gevind word: **Groot Krokodil** (skeldnaam vir die voormalige Suid-Afrikaanse staatspresident P.W. Botha), **Iron Lady** (Afr. **Ystervrou**) (naam gebruik vir die voormalige Britse eerste minister Margaret Thatcher), **Jimmy the peanut farmer (Jimmy the peanut king)** (Afr. **Jimmy die grondboontjieboer (Jimmy die grondboontjiekoning)**) (skeldnaamverwysing na Jimmy Carter, president van die V.S.A., 1977–1981), **Kortbroek** (naam gebruik vir Marthinus van Schalkwyk, leier van die Nuwe Nasionale Party, 1996–), **Mother of the Nation** (Afr. **Moeder van die Nasie**) (naam gebruik vir Winnie Madikizela-Mandela).

4.   Slotsom

Die resultate van die navorsing wat reeds gedoen is, die taalverryking wat plaasgevind het, ontlenings uit vreemde en inheemse tale, die navorsing wat reeds ten grondslag van die meertalige verklarende uitgawe aan die gang is en die totstandbrenging van 'n unieke produk waarby die vakgebied soveel sal baat, is van onskatbare waarde.

Die vakkomitee wat die hersiene en bygewerkte uitgawe van die woordeboek so vinnig ná die vorige woordeboek die lig laat sien het, kan met reg trots wees op die eindproduk. Die *Nuwerwetse Politieke Woordeboek* is in werklikheid 'n heeltemal nuwe woordeboek as 'n mens in aanmerking neem dat die sowat 5 000 brontaalterme na 16 000 vermeerder is, en dat uitgebreide verklarende voorstukke, verbandsverklarings, transliterasies en bylaes bygekom het.

Die woordeboek bied velerlei gebruiks- en toepassingsmoontlikhede. Dit is 'n bruikbare vakwoordeboek wat met groot vrug deur taalpraktisyns geraadpleeg kan word en deur hulle aangeskaf behoort te word.

Vertalers en redigeerders sal hierdie vakwoordeboek om verskeie redes nuttig vind. Tony Moen, 'n bedrewe vertaler en redigeerder, noem die volgende as belangrike redes waarom vertalers en redigeerders dié vakwoordeboek behoort te gebruik:

—   alle inskrywings is volledig tweetalig, so ook alle aantekeninge, definisies en kontekstuele inligting;
—   die datakorpus bevat resente inligting;
—   die lys bevat allerlei bruikbare nuutskeppings; en
—   dit bied 'n breë oorsig van die vakgebied, insluitend terme wat uitgelaat sou kon word vanweë hul spesifieke militêre of ekonomiese aard.

Sowel die tweetalige vertalende woordeboek as die beplande meertalige woordeboek behoort 'n besondere rol in die vakterrein te speel. Dit behels:

—   die aanvulling en verryking van die Afrikaanse vaktaal, en die gebruik

en standaardisering van die Afrikaanse terme;
— die lewering van 'n betekenisvolle bydrae tot die vergemakliking van die leerproses vir duisende studente wat nie Engels as moedertaal het nie — ook in sommige van Suid-Afrika se buurstate;
— die instaatstelling van parlementariërs en ander openbare figure tot korrekte taalgebruik;
— die vergemakliking van vertalings binne die staatsdiens;
— die bevordering van die korrekte gebruik van begrippe en taal in die media;
— die daadwerklike uitbouing van die "agtergeblewe" tale en die speel van 'n positiewe bemagtigingsrol; en
— die benutting deur ander lede van die Dietse taalgroep.

Die Engelse gedeelte sou ook geredelik op sy eie gepubliseer kon word om 'n goeie verkoper in Engelstalige lande te word.

Ten einde werk aan die derde fase van die vakwoordeboek, naamlik die definiëring en verskaffing van die Afrikatale te bespoedig, moet ander uitweë vir befondsing as die huidige bedeling gevind word. Dit is noodsaaklik dat die werk aan die meertalige projek vinniger vorder. Die vaktaalgemeenskap sien met groot verwagting uit na hierdie meertalige verklarende woordeboek wat deur SEPTSA beoog word.

Aantekeninge

1. Aanvanklik is die meer formele **civic association** gebruik, maar later is oorgeslaan na **civics**.
2. Die term beteken letterlik "oorlog", maar kan ook gebruik word om na **amabutho** (portuurgroep) te verwys.
3. Die Sothowoord is **difaqane**, en die Nguniwoord **mfecane** wat, indien hulle letterlik vertaal word, "verdelgingsoorloë" beteken. Groot groepe mense het met hul families en besittings voor ander groepe uitgevlug wat hulle wou uitwis, en in dié proses het hulle self ook ander groepe aangeval. Die Engelse en Afrikaanse ekwivalente verwys dus streng gesproke eintlik na die resultaat van die volksverdelging.
4. Die *WAT* gee: **djihâd** *(Arab.)* Heilige oorlog van die Mohammedane teen ongelowiges en ketters, soos deur die Koran voorgeskryf. In die konteks van die staatkunde is dit nie soseer 'n **oorlog** nie, maar eerder 'n **stryd**. Die *dj*-klank in Arabies behoort hier uitgespreek te word. Vgl. bv. Engels **Jakarta** wat in Afrikaans **Djakarta** word. Die 1994-uitgawe van die *HAT* gee net **djihad** as verafrikaanste vorm aan, maar navorsing deur die Komitee toon dat die vorm **jihad** ook vir die ekwivalensie behou behoort te word omdat dit al, veral onder studente, inslag gevind het.
5. Die Japannese **ei** word in sommige woorde as /əi/ en in ander as /ɛ/ uitgespreek — vandaar die tweeledige transliterasie in Afrikaans, vgl. **Heisei, Kêdanren.**
6. Die Komitee het besluit om wel die **r**-skryfwyse te behou, hoewel die Japannese die /ɛr/ as /ɛl/ uitspreek, vgl. **Kêdanren, ringhie.**

Verwysings

Bosman, D.B., I.W. van der Merwe en L.W. Hiemstra. 1984[8]. *Tweetalige Woordeboek/Bilingual Dictionary.* Kaapstad: Tafelberg.

Committee for Political and Related Terminology / Komitee vir Staatkundige en Verwante Terminologie. 2002. *Modern Political Dictionary: Terms, Concepts and Usages in Politics and the Political Sciences/Nuwerwetse Politieke Woordeboek: Terme, begrippe en gebruike in die politiek en politieke wetenskappe.* Johannesburg: Centre for Political and Related Terminology in Southern African Languages (CEPTSA), Rand Afrikaans University/Sentrum vir Politieke en Verwante Terminologie in Suider-Afrikaanse Tale (SEPTSA), Randse Afrikaanse Universiteit.

Gove, B.P. (Hoofred.). 1961[3]. *Webster's Third New International Dictionary of the English Language.* Londen: G. Bell/Springfield, Mass.: G. & C. Merriam.

Moen, Tony. 2003. SATI Award for Outstanding Translation. Ongepubliseerde dokument.

Odendal, F.F. et al. 1994[3]. *Verklarende Handwoordeboek van die Afrikaanse Taal.* Midrand: Perskor.

Schoonees, P.C. (Hoofred.). 1956. *Woordeboek van die Afrikaanse Taal. Tweede Deel: D–F.* Pretoria: Die Staatsdrukker.

Taalkommissie van die Suid-Afrikaanse Akademie vir Wetenskap en Kuns. 2002[8]. *Afrikaanse Woordelys en Spelreëls.* Kaapstad: Pharos.

Urdang, Laurence (Red. dir.). 1982. *Suffixes and Other Word-Final Elements of English.* Detroit: Gale Research Co.

Juri Apresjan and the Development of Semantics and Lexicography

Igor Burkhanov, *Section of Applied Linguistics, Institute of English Studies, University of Rzeszów, Rzeszów, Poland (burkhan@univ.rzeszow.pl)*

Abstract: The major aim of this article is to highlight Juri Apresjan's impact on the development of linguistic semantics and theoretical lexicography. In order to achieve this goal, a number of issues of paramount importance, which have always been in the focus of attention in Apresjan's publications, have to be discussed: (a) the notion of "naïve picture of the world", i.e. language-specific folk categorization encoded in the lexical and grammatical semantics of a particular language, as opposed to the supposedly universal and language-independent system of scientific concepts; (b) basic properties of the formal metalanguage of semantic desciption, its explanatory power and applicability in dictionary-making; and (c) representation of synonymy in a bilingual and a monolingual dictionary of synonyms designed within the framework of systematic lexicography. In addition, considerable attention has been given to two basic categories of systematic lexicography, "lexicographic portrait" and "lexicographic type", as well as the zonal structure of dictionary articles.

Keywords: BILINGUAL DICTIONARY, COMMONSENSE (EVERYDAY) KNOWLEDGE, DEFINITION, DICTIONARY OF SYNONYMS, EXPERT KNOWLEDGE, INTEGRATED LEXICOGRAPHIC DESCRIPTION, LEXICOGRAPHIC PORTRAIT, LEXICOGRAPHIC TYPE, METALANGUAGE, NAÏVE PICTURE OF THE WORLD, SCIENTIFIC PICTURE OF THE WORLD, SYNONYM SERIES, SYSTEMATIC LEXICOGRAPHY, TRANSLATION DICTIONARY, ZONAL STRUCTURE (OF A DICTIONARY ENTRY)

Opsomming: Jury Apresjan en die ontwikkeling van die semantiek en leksikografie. Die hoofdoel van hierdie artikel is om Juri Apresjan se invloed op die ontwikkeling van die linguistiese semantiek en teoretiese leksikografie uit te lig. Om hierdie doel te bereik, moet 'n aantal kwessies van die grootste belang wat nog altyd in die brandpunt van aandag in Apresjan se publikasies was, bespreek word: (a) die opvatting van 'n "naïewe beeld van die wêreld", d.w.s. taalspesifieke volkskategorisering wat in die leksikale en grammatikale semantiek van 'n bepaalde taal geënkodeer is, in teenstelling met die veronderstelde universele en taalonafhanklike stelsel wetenskaplike konsepte; (b) die basiese eienskappe van die formele metataal van semantiese beskrywing, die verklarende krag daarvan en die toepasbaarheid daarvan in woordeboeksamestelling; en (c) die aanbieding van sinonimiteit in 'n tweetalige en 'n eentalige sinoniemwoordeboek beplan binne die raamwerk van die sistematiese leksikografie. Hierbenewens is aansienlike aandag gegee aan twee basiese kategorieë van die sistematiese leksikografie, die "leksikografiese portret" en die "leksikografiese soort", asook die sonestruktuur van woordeboekartikels.

Sleutelwoorde: ALGEMENE (ALLEDAAGSE) KENNIS, DEFINISIE, GEÏNTEGREERDE LEKSIKOGRAFIESE BESKRYWING, LEKSIKOGRAFIESE PORTRET, LEKSIKOGRAFIESE SOORT, METATAAL, NAÏEWE BEELD VAN DIE WÊRELD, SINONIEMREEKS, SINONIEM-

WOORDEBOEK, SISTEMATIESE LEKSIKOGRAFIE, SONESTRUKTUUR (VAN 'N WOORDE-
BOEKINSKRYWING), TWEETALIGE WOORDEBOEK, VAKKUNDIGE KENNIS, VERTALENDE
WOORDEBOEK, WETENSKAPLIKE BEELD VAN DIE WÊRELD

1.   Introduction

The book *Systematic Lexicography* published by Oxford University Press is a
very impressive hardcover edition that comprises English translations of a
number of Juri Apresjan's articles written over a timespan of 16 years. More-
over, as has been noted in the Translator's Note, these papers have been revised
by the author specifically for the purposes of the publication at issue.

This selection of papers seems to cover all the peculiarities of the author's
standpoint concerning theoretical semantics and lexicographic representation
of linguistic phenomena. It is subdivided into two major parts titled Problems
of Synonymy and Systematic Lexicography. The book is also provided with an
Index of English Lexemes, an Index of Russian Lexemes, a Subject Index and an
Index of Names. The Index of English words seems to be particularly im-
portant for the reader who is not engaged in the study of Russian or other Slav-
onic languages, but is an expert in the study of meaning or lexicography and
may be, thus, interested in the author's treatment of common English lexical
items.

The publication of a collection of Yuri Apresjan's articles in English may
be called a very important event.[1] For many linguists, who have not read his
works in the original, Apresjan remains a mysterious figure. On the one hand,
he is reputed to be an outstanding personality in the field of semantics and lexi-
cography. For instance, in the Acknowledgements to *Lingua Mentalis* Anna
Wierzbicka, expresses her gratitude to "Jura Apresjan", along with Igor Mel'-
čuk, for stimulating and challenging discussions that lasted for over 15 years
(Wierzbicka 1980: xi). On the other hand, his publications available in English
are scarce, to say the least. Hence, most of the international readership has been
deprived of any first-hand information concerning his views on pertinent sub-
jects. The volume under consideration, at least partly, compensates for this
deficiency.

Though this article is intended to deal with the matters of lexicography,
i.e. applied-linguistic description of the lexicon, we shall repeatedly (and un-
avoidably) touch upon theoretical-linguistic issues discussed in Apresjan's
publications for a two-fold reason.

Firstly, Apresjan is both a theoretical semanticist and a practicing lexico-
grapher, which is no exception in Eastern and Central Europe. It fact, high
esteem of lexicographic practice has always been characteristic for this part of
the world. Suffice it to say that a famous linguist, J. Baudouin de Courteney,
once acted as the editor of the third edition of one of the largest Russian mono-
lingual dictionaries originally compiled by V.I. Dal' (1903–1909) and imple-
mented his theoretical beliefs in editorial practice.

Secondly, Apresjan himself made a point of introducing his theoretical considerations if not immediately into lexicographic description, then at least into lexicographic theory. He writes the following in the Introduction to the volume under consideration (Apresjan 2000: XI):

> Throughout the book I am concerned with two main issues — linguistic theory and lexicography. Until quite recently there has been a gap between the two. Moreover, in some quarters lexicography was looked down upon as a purely practical enterprise unworthy of scholarly interest. The present author is convinced, however, that sound lexicography can only be based on sound linguistic theory and that recent theoretical developments are of paramount importance for the practical skills of compiling a dictionary.

It should be noted in this connection that "sound linguistic theory" only is not sufficient for the production of an adequate reference work. In any lexicographic project, metalexicographic considerations, i.e. sound specifically lexicographic theory, are of no less importance. As I will try to show later, the latter, among other things, implies taking into account such lexicographic parameters of the reference work to be designed as its intended user, aims and objectives of the dictionary, a particular purpose the reference work is intended to serve, principles of lexicographic data selection, and a number of others.

To keep this article within manageable proportions, in the following we shall primarily occupy ourselves with a brief account of the issues of paramount importance that have always been the focus of attention in Apresjan's writings, namely: (a) the notion of "naïve picture of the world"; (b) metalanguage of semantic and/or lexicographic description; and (c) properties of synonyms and their representation within the framework of systematic lexicography.

Before addressing these matters we shall have to present a very brief, and by no means comprehensive, account of the theoretical background that has affected the interpretation of the aforementioned issues in the author's publications both translated for this volume and presented elsewhere.

2.   Theoretical Prerequisites

Apresjan's writings can be properly understood and evaluated not only within the framework of developments in mainstream theoretical linguistics, but more specifically, against the background of the linguistic school that originated in the former Soviet Union on the basis of Russian linguistic tradition. It is obvious enough that we shall discuss only the most general tendencies and cannot possibly account for particular theoretical constructs, which, unavoidably, presupposes a certain degree of simplification.

Moreover, that linguistic tradition is not a sole property or a characteristic feature of Russian linguistic thought, but also developed in other Slavonic countries, for instance in Poland and Bulgaria, and a number of successor states

to the former USSR.

Why Poland? For well-known historical reasons, cultural development of both countries has been related. Moreover, in connection with the matters under consideration, it should be specifically mentioned that Apresjan's beliefs were considerably influenced by what he referred to as "the Polish School of Semantics", particularly works of A. Bugusławski and A. Wierzbicka on meta-language of semantic description (see bibliography in Apresjan 2000).[2]

It seems questionable that semantic investigation conducted in Poland is homogeneous enough to be regarded as one school, I would, personally, vote for at least several trends in semantic research, including quite a significant cognitively-oriented one (e.g., Kalisz 1981; Lewandowska-Tomaszczyk 1996; Krzeszowski 1997; to mention just a few). Nevertheless, it should be empha-sized that a large number of linguists in both Russia and Poland, irrespective of their theoretical preferences, dedicated considerable attention to the study of culturally-induced elements of meaning and, even more specifically, cultural background encoded in, or underlying, the meanings and usage of lexical units (lexemes and phraseological units), particularly those belonging to the general vocabulary.

It is worthy of noting that thus acknowledged significance of widely-understood culture of a linguistic community, its shared assumptions about the world and society, ancient religious beliefs, even environmental conditions have motivated a number of lexicographic projects of a scholarly nature devel-oped within the framework of academic lexicography. Their primary goal was to account for culturally-significant elements of word meanings. By way of example one may cite a short dictionary of traditional Russian poetic symbols (Ivanov et al. 1977) or a dictionary of folk stereotypes and symbols *Słownik stereotypów i symboli ludowych* (1999) developed under the general editorship of J. Bartmiński. The latter is, in fact, a large-scale on-going project whose major aim is formulated as "an attempt to reconstruct the traditional view of the world and man using methodology of ethnolinguistic and folklore research" (Bartmiński 1996: 9).

A point of crucial importance for the present discussion is the way the cor-relation between semantics and grammar was interpreted within the scope of linguistic exploration that was conducted in the former USSR. In this respect, the latter drastically differed from, for instance, the mainstream American lin-guistics that predominantly developed within the framework of form-oriented research as specified by L. Bloomfield and his followers, and may be said to have been, until recently, domineered by generative models of language.

A characteristic feature of generative models in various modifications is the assumption that syntax and semantics form two autonomous subsystems of language. Within this scientific paradigm the concept of linguistic semantics was virtually reduced to lexical semantics. Obviously enough, in this situation, the very term "lexical semantics" was tautological, since only lexical units were supposed to have meanings.

Soviet linguistics rather adhered to more "traditional" postulates. One of these assumptions presupposed the existence of at least three major symbolic levels of the language structure, i.e. lexicon, morphology and syntax. In this respect, the trend of linguistic research under consideration is remindful of the cognitive linguistic approach to the study of language that originated fairly recently as a form of dissatisfaction with the drawbacks of generativism (cf. Langacker 1987: 2-3):

> Grammar (or syntax) does not constitute an autonomous level of representation. Instead, grammar is symbolic in nature, consisting in the conventional symbolization of semantic structure. … There is no meaningful distinction between grammar and lexicon. Lexicon, morphology and syntax form a continuum of symbolic structures, which differ along various parameters but can be divided into separate components only arbitrarily.

Thus, the trend of linguistic exploration that developed in the former USSR was based on the assumption that meaning is not the sole property of lexical items. Grammatical units (morphemes and syntactic structures) were regarded as kinds of linguistic signs, i.e. meaningful units. An immediate consequence of this standpoint is the implication that semantics as a linguistic discipline incorporates three subbranches: lexical semantics, morphological semantics and syntactic semantics.

For this very reason Apresjan's book titled *Leksicheskaia semantika* 'Lexical Semantics' came out in 1974 and was translated into Polish within six years of publication (Apresjan 1980), whereas its English counterpart, i.e. a monographic study with the same title, came out 12 years after (see Cruse 1986).

It is obvious enough that this theoretical-linguistic background justified Apresjan's conception of "integrated linguistic description" which is traceable to Shcherba's postulate that a dictionary and a grammar should produce a unified account covering the whole of language with no gaps (Shcherba 1940). This totally natural requirement has been repeated many times in its country of origin (e.g. see Seliverstova 1975). In fact, it could be reformulated elsewhere. Nevertheless, Apresjan's outstanding contribution to applied-linguistic investigation, among other things, encompasses an outline of such an integrated theory developed for the unified description of lexicon and grammatical system of a language, including: (a) rules of systematic representation of lexicographic information of this kind; (b) specification of corresponding formal devices of systematic lexicography; and (c) introduction of appropriate descriptive notions of crucial importance such as "lexicographic type" and "lexicographic portrait".

One of the principal categories in Apresjan's approach to the study of lexical (and grammatical) semantics is the notion of "naïve picture of the world", which appears indispensable in view of the aforementioned preoccupation with culture-specific elements of meaning. Simultaneously, this concept is assumed to be indispensable in lexicographic description, since it is stated that

the primary task of systematic lexicography is to reflect the worldview in-built in a particular language (Apresjan 2000: 104). This descriptive concept will be our major concern in the next section.

3.   Naïve Picture of the World

The cornerstone of Apresjan's approach to the study of meaning is based on Shcherba's seminal idea of the distinction between commonsense (everyday, folk) concepts and scientific notions (Shcherba 1940). The issue of framing of concepts and the structure of language, undoubtedly belongs to the key problems of linguistic semantics. A particular standpoint in this matter determines peculiarities of each trend in the study of meaning. Various interpretations of the correlation between commonsense knowledge and scientific knowledge structures as forms of categorization have been presented in the history of semantic description. For instance, Bloomfield maintained that commonsense notions in the definitions of lexical meanings should be substituted with corresponding scientific concepts. Thus, the lexeme *salt* should be defined as 'sodium chloride' or, alternatively, equated to the chemical formula of the substance designated by the lexical item in question (Bloomfield 1933).

Conversely, Shcherba's notion of commonplace concept as fundamentally different from the scientific concept may be said to be universally accepted in Russian linguistic tradition. By way of example he discussed the linguistic expression *priamaia* 'straight/direct (line)'. In geometry, a direct line may be specified as "the shortest distance between two points" (which is a scientific concept), whereas in everyday life we call "straight" or "direct" a line that deviates neither to the left, nor to the right (also neither upwards, nor downwards).

It is important to note that Shcherba introduced this distinction, which is, obviously, relevant for the theoretical study of meaning in linguistic semantics and lexicology, in a paper dedicated to general lexicographic theory. Since that time the notion of commonsense concept has been rediscovered and elaborated on by a number of semanticists and lexicographers. For instance, Hallig and Von Wartburg noted in the Introduction to their onomasiological (ideographic) dictionary that they aimed to account for "the worldview which is characteristic for an average intelligent speaker and is based on pre-scientific general concepts of which the language avails itself" (Hallig and Von Wartburg 1952: XIV). Thus, the lexicographers intended to reproduce the system of commonsense concepts in the classificatory scheme of their reference work. To what extent this goal was achieved in their dictionary and other lexicographic works of this kind is a different matter (for detailed analysis see Burkhanov 1999).

The concept of "naive picture of the world" was introduced into semantic and lexicographic description by Apresjan (1974 and other publications) to denote the aggregate of commonsense concepts underlying meanings of lexical items that form the general vocabulary. He has also noted that the idea of conceptual system forming a definite worldview is traceable to W. von Humboldt's

writings and Sapir–Whorf hypothesis of linguistic relativity. Apresjan's major postulate is that the naïve picture of the world is, at least to some extent, language-specific, hence unique, whereas the scientific picture of the world was said to be universal, well-organized, logical, and language-independent.

According to Apresjan, the naïve picture of the world encoded in a particular language has been formed over centuries. It represents cultural and spiritual experience of a given linguistic community and includes naïve physics, naïve geometry, naïve psychology, naïve ethics, naïve physiology, etc. A point of primary importance is the assumption that these conceptual configurations are of a systematic nature (cf. Apresjan 2000: 104):

> The naïve pictures of each of these areas are not chaotic but form definite systems and should therefore receive a homogeneous description in a dictionary. For this purpose, generally speaking, we should first reconstruct the corresponding fragment of the naïve picture of the world on the basis of lexical and grammatical meanings. In practice, however, in this case and others like it, the reconstruction and the (lexicographical) description go hand in hand and constantly provide each other with correctives.

The quotation above contains at least two statements that seem to be open to controversy. First, since each of the aforementioned semantic domains is assumed to be systematically organized, the naïve worldview should also form a homogeneous system opposed to the scientific worldview which is, by definition, even more systematic.

In connection with scientific categorization, it should be mentioned that there are various types of specialized knowledge which cannot possibly be referred to as scientific concepts or any system of these, for instance: categories of various branches of technology, sportive events, business, and other kinds of human activity which require expertise, professional training and specialist knowledge. These specific concepts are designated by lexical items of a terminological and semi-terminological character, but are beyond the scope of both commonsense knowledge and academic disciplines. The term "expert concept" seems to be a good candidate for the status of an overall category which, in addition to scientific concepts, includes the notions of the aforementioned heterogeneous human occupations. Thus, every scientific concept is an expert concept, but not all expert notions are scientific (see also appropriate entries in Burkhanov 1998).

Expert knowledge is not always universal; at least part of it may be culture-bound, i.e. induced by the peculiarities of cultural and/or historical development of the linguistic community. Every professional translator and/or expert in bilingual specialized lexicography has learned it by painful experience. Law, economics and politics may be cited as examples of typical culture-dependent fields (e.g., see Reed 1993; Riggs 1993).

Bergenholtz and Tarp (1995: 60-61) claim that the distinction between culture-dependent and culture-independent subject fields proves usable, though

sometimes it is not easy to draw a demarcation line between them and there are borderline cases that are difficult to be assigned to one class or the other.

It should be noted that the statement above includes a good deal of over-generalization. In fact, we cannot specify, with any degree of certainty, which subject area can be unquestionably assigned to one class or the other. We cannot even say that all academic disciplines are culture-independent. My own experience of compiling a dictionary of semantic terminology in Russian (Burkhanov 1995) and a dictionary of lexicographic terminology in English (Burkhanov 1998) amply demonstrated frequent discrepancies between translationally-equivalent terminological units both in connotations and, more importantly, in the scope of conceptual content within different national specialist terminologies (for a detailed analysis see Burkhanov 2003: 126-135).

Even in the case of well-established academic disciplines the organizing assumption of uniformity of scientific description seems to be arguable. An overall picture of the world presupposes integration of worldviews presented in individual branches of science, such as physics, astronomy, chemistry, biology, history, sociology, and so on; not to mention interdisciplinary fields. It should be mentioned that the attempts at integration of several disciplines within the framework of a joint approach that may be considered successful, such as cognitive science, are few and far between. It seems to be extremely difficult, if not impossible, to design an overall scientific picture of the world at the present-day stage in the development of disciplinary and interdisciplinary studies.

In fact, the assumption that commonsense notions form a homogeneous system, i.e. a unified whole, also arouses suspicions. In principle, Shcherba's sharp opposition of commonsense (everyday) notions vs. scientific concepts appears to be too rigorist. It should be emphasized in this connection that types of categorization are too heterogeneous to be described in terms of a dichotomous division "commonsense" vs. "scientific" (cf. Lehrer 1992: 244). It seems more appealing to assume that numerous worldviews can, and do, co-exist in the semantics of a natural language and make their impact on the structure of semantic relations in the lexicon. The semantics of a particular language encompasses the representation of specific features of natural environment and ways of life, commonsense considerations, elements of conceptual systems of the past including those knowledge structures which originated as a by-product of mythological contemplation of the world, moral and customary beliefs of old times, etc. In view of these considerations it is difficult to expect that all those heterogeneous knowledge structures are amalgamated into a unified whole.

Another issue that seems to be controversial is the postulate that a theoretical-semantic model of the naïve picture of the world is directly applicable in lexicographic description, i.e. could be transferred wholesale from theoretical semantics into lexicographic practice. It seems more profitable to consider such a model as raw material of lexicographic investigation that has to be selected,

prepared, and only afterwards presented depending on the dictionary parameters that have been chosen for a given lexicographic project.

In addition, a model of an overall naïve picture of the world presented by a semanticist may be found inconsistent or not comprehensive enough from the viewpoint of lexicographic description. By way of example, let us consider Apresjan's interpretation of the notion of man in the naïve picture of the world as manifested in the Russian language. In particular, Apresjan (2000: 105) noted that man performs three basic types of action: physical actions, intellectual actions, and speech acts. These types of action both in theoretical considerations and in actual semantic descriptions are considered to be sufficient for an adequate account of the variety of activities and acts performed by humans.

It seems reasonable to add to the aforementioned three types of action at least one more, namely: social actions. It will be fair to say that it is not only Apresjan's fault, other experts in theoretical semantics also failed to specify this type of action. The actions under consideration are designated by both appropriate action verbs as well as adjectival and nominal lexemes. The necessity to introduce this descriptive category becomes evident in the course of semantic analysis of the following lexical items, for instance: *disinherit*, i.e. 'to undertake certain social actions so that somebody, especially one's son or daughter, will not receive one's money and property after one's death' and *influential* as "having great social power", where *power* means "ability to perform actions of certain type" and *social power* stands for "possession of control, authority or influence over others", cf.: *John was disinherited by his father* or *He is a very influential politician*. These lexical items seem to warrant semantic description of the following kind:

influential = HAVE + ABILITY [perform social actions] + > NORM (where the sign > stands for "more than");

disinherit = CAUSE + STOP + HAVE [right (BEGIN + HAVE {money or property}] + TEMPORAL SPECIFICATION [after one's death].

Social actions often incorporate, or are manifested by, physical and/or verbal actions. Nevertheless, basic conceptual elements of the former stand for primarily socially and/or culturally significant acts or activities, cf.: *marry* (transitive) "to establish a social relationship [marriage] between (at least prototypically nowadays) a man and a woman". In view of Apresjan's principle of uniform systematic representation of relevant semantic information, social actions seem to require specification as a lexicographic type.

It can be argued that in some cases it is very difficult, if not impossible, to draw a sharp demarcation line between physical actions on the one hand and social actions on the other. Nevertheless, the same can be said about other postulated types of action, cf.: *discuss* which is usually interpreted as a speech act verb, though its meaning also encompasses mental activity.

Probably the failure to pay sufficient attention to such cases can be attrib-

uted to the attempt to specify lexical meanings in terms of definitions written in a formal metalanguage. As will be shown in the next section, the latter implies the minimum definition requirement as well as simplification of semantic relations.

4.   Metalanguage of Semantic and / or Lexicographic Description

For years Apresjan was an ardent adherent of the idea that dictionary definitions should be formulated in a special formal metalanguage of semantic description developed in the publications of the Moscow Semantic School (e.g., see Zholkowsky 1964; Mel'čuk 1974; Apresjan 1974; Mel'čuk 1982). Summarizing the achievements of theoretical semantics and the first attempts at lexicographic descriptions in terms of metalanguage, he suggested the following requirements for its basic components: vocabulary and syntax (Apresjan 1974: 70-106).

As for syntax, the major requirement as formulated by the adherents of the Moscow Semantic School is to ensure nonambiguous notation of meaning configurations. The syntax of semantic graphs and dependency trees was declared the ideal solution, whereas the use of a simplified version of the syntax of a natural language was regarded as no more than a palliative (e.g. see Mel'čuk 1974: 53; Apresjan 1974: 77-79; Apresjan 2000: 216).

The core of the vocabulary of the metalanguage was presumed to be composed of semantic primitives. Each vocabulary unit should designate only one, preferably elementary, meaning, whereas each elementary meaning should be designated by one and only one word. Hence, the major requirement presupposes one-to-one correspondence between meaning and form of the word. Thus, vocabulary units of the metalanguage, unlike lexical items of natural languages, should be devoid of polysemy and homonymy. Moreover, the number of elementary meanings should be limited (minimum definition requirement), though sufficient to account for the linguistic facts under consideration (exhaustive definition requirement).

The theoretical metalanguage is designed using "artificial words" or "word-constructs", i.e. lexemes of the natural language with stipulated meanings or specifically coined words with stipulated meanings. Some of them were borrowed from natural sciences, mathematics and logic, cf.: *mnozhestvo* 'set' (in the mathematical sense), *sila* 'force, power', *funktsia* 'function'; others were invented, cf.: *kauzirovat'* 'to cause', *potok faktov* 'current of facts'; and others, though they are general vocabulary units, are used in restricted sense, cf.: *veshch* 'thing', *kolichestvo* 'quantity', *norma* 'norm'.

The choice of pre-existing lexical items and coinage of new ones to be included in the vocabulary of the semantic metalanguage was motivated by the properties of the Russian language. The terminological unit *mnozhestvo* was formed on the basis of the literal meaning of the word 'a very large number of', cf.: *mnozhestvo faktov* 'a large number of facts', *mnozhestvo sluchaev* 'a large num-

ber of cases', etc. In the terminological sense it stands for a collection of objects or elements classed together on the basis of a common property. It should be mentioned that *mnozhestvo* is a better solution as compared, for instance, to its English translation equivalent for a two-fold reason. First, the meaning of the former is motivated and, hence, easier to understand, since it is related to a high-frequency general vocabulary lexical item *mnogo* 'many, much'. Simultaneously, the noun *mnozhestvo*, unlike *mnogo*, is nonambiguous, since it refers to collections of countable objects. The English word *set* is too polysemous; its appropriate interpretation in the sense under consideration is context-bound. Moreover, in everyday language *set* often designates a collection of artefacts that are specifically designed to be used together or for a particular purpose, cf.: *a set of china* or *a set of carving knives*. The latter sense could have unduly limited the scope of meaning representation.

Thus, the choice of a lexical item to be included in the metalanguage was very apt in the case of *mnozhestvo*. As for the other word, *sila*, certain difficulties may arise, since it designates both the ability to perform actions, particularly physical strength, and a property of an action, i.e. the energy exerted. It is not by chance that the lexical unit in question is often translated as *power* or *force* into English. This kind of polysemy characteristic for a natural language may lead to misinterpretation either on the part of the researcher who uses this word in a stipulated meaning in the metalanguage or on the part of the dictionary user. Of course, it may be argued that both the researcher and the dictionary user know that the word *sila* is supposed to be used in semantic descriptions in the sense that is closer to the scientific usage in physics. Nevertheless, a possibility of unwanted misinterpretation still exists.

Explications of lexical meanings in a formalized metalanguage of description were presented in a number of lexicographic projects (Mel'čuk and Zholkovsky 1984; Mel'čuk et al. 1984, 1988, 1992; Wierzbicka 1987). Nevertheless, all of them were of a rather experimental nature and developed within the framework of academic lexicography. Irrespective of the fact whether it was acknowledged by the lexicographers or not, these works of reference displaying outstanding scholarly values are primarily meant for a professional linguist, not a lay dictionary user. So far they have been no competition for monolingual dictionaries providing definitions in a natural language, particularly with respect to user-friendliness.

It is obvious enough that semantic descriptions in a formal metalanguage are quite applicable in the case of computational lexicography, information retrieval systems and machine translation. Suffice it to say that Mel'čuk continues to investigate the applicability of lexical functions for the purposes of text analysis and text generation with a group of other experts in the field (e.g. see Wanner 1996).

In 1994 Apresjan published an article in which he overtly specified a different conception of metalanguage of lexicographic description. First, by way of justification a distinction of semantic representation into surface and deep

sublevels, which was proposed in an earlier article, was featured. Among other things, he claimed that the idioethnic semantics of a particular language should be handled by the surface-semantic component of the model. In order to produce a comprehensive account of semantic facts it was proposed "to use not an artificial language, but a somewhat simplified and unified sublanguage of the object language", i.e. lexical items and syntactic structures of the language under consideration (Apresjan 2000: 217).

Further on the requirements imposed on thus understood metalanguage of lexicographic description were formulated: (a) the lexical stock should be limited; (b) the vocabulary of the metalanguage should be unified to eliminate synonymy and homonymy; etc. It is evident enough that these requirements are not too much different from those specified for defining vocabulary in what Apresjan refers to as "traditional lexicography" (cf. Svensén 1993: 135-138).

It is important to emphasize that the statement above meant an actual resignation from the principle of obligatory definitions based on metalanguage of semantic description. In fact, it was the major issue of controversy between the Moscow Semantic School (Mel'čuk, Zholkovsky, Apresjan and others) and the leading figures of contemporary linguistics in the USSR. The former insisted that adequate lexicographic works of reference could be produced only on the basis of a formal metalanguage, whereas the latter tried to prove the advantages of conventional lexicographic representation. This conflict resulted in publications intended to undermine the academic achievements of "the rebellious semanticists" and sometimes went far beyond the scope of academic discussions (e.g., see Kotelova 1975).

The evolution of Apresjan's views concerning the subject is the telltale evidence that the choice of a particular type of definition for a given lexicographic project depends on a number of considerations, some of them of a practical nature. In fact, one may say that the question of selection of defining techniques cannot be reduced to a straightforward, simplistic answer: "Which one is better?" Probably it should be reformulated as follows: "Which one is better for a particular type of dictionary to account for a particular class of linguistic facts?"

The answer to this query always implies satisfactory specification of a dictionary type in terms of a number of parameters of the reference work to be designed: (a) the intended dictionary user (age group, prior knowledge of the subject in the case of LSP reference works, level of proficiency in the foreign or second language in the case of pedagogical lexicography, etc.); (b) aims and objectives of the reference work (reference only, instruction, professional translation, etc.); purpose (production or text comprehension); (c) lexicographic material to be presented in it; etc.[4]

Instead of a rigid requirement to use metalanguage in dictionary definitions it seems more profitable to assume a more flexible approach which allows for the selection of not only defining techniques, but also the most efficient technique of lexicographic presentation (in reality their combinations) inclusive

of graphic representation, exemplification, usage and grammatical labels, cross-referencing, etc. In fact, the adequate description of these tools of lexicographic representation and their applicability remains a priority of metalexicography. Such a description should specify advantages and disadvantages of each and every technique so that practicing lexicographers could select appropriate combinations of these techniques to design the most efficient form of lexicographic presentation for a given reference work.

5.   Synonymy and its Representation in Systematic Lexicography

Apresjan's beliefs concerning lexicographic description have been put forward under the umbrella term "systematic lexicography". He formulated four major principles of systematic lexicography as follows: (a) the dictionary should be designed to accommodate the needs of active production, i.e. it should be oriented towards use in speech; (b) the dictionary should be integrated, i.e. lexicographic description should, in particular, take account of the differing grammatical conditions of use; (c) the dictionary should be systematic in the sense of taking account of the different lexicographic types to which the lexeme under consideration belongs; and (d) the dictionary should reflect the naïve picture of the world.

Thus, a basic tenet of systematic lexicography is that the lexicon should be described in terms of "lexicographic types", i.e. groups of lexemes with a shared property or properties (not necessarily semantic), subject to the same linguistic rules, which warrants a uniform description in the dictionary. Apresjan (2000: 102) provided the following definition of the notion under consideration:

> The central concept of systematic lexicography is that of the lexicographic type, a group of lexemes with at least one common property (semantic, pragmatic, communicative, syntactic, combinatorial, morphological, prosodic, etc.), to which the same rules of linguistic description (of 'grammar' in the broad sense) refer and which therefore requires a homogenous description in a dictionary.

Lexicography traditionally dealt with lexical-semantic (and partly grammatical) properties of lexical items. In the case of some languages which, like English, are characterized by a noticeable discrepancy between the orthographic and phonological properties of lexemes, transcriptions are provided intended to account for segmental phonemes composing appropriate phonological words and word stress. In some dictionaries combinatorial features of lexical items (particularly lexical-semantic valence) are described. Nevertheless, it will be fair to say that lexicographic description has been centred on the word and accounted for its properties only.

In view of the considerations above, Apresjan's intent to account for prosody appears innovative and extremely promising, particularly in the case of intonation which is usually neglected in both dictionaries and grammars. For instance, initial position is normal for so-called "sentence adverbs" in English:

naturally, *geographically*, etc. They are usually characterized by the fall-rise into-
nation pattern and separated from the rest of the utterance by a pause signify-
ing the syntagm boundary in the spoken mode of communication, whereas in
the written mode this boundary is represented by a comma.

This group of adverbs qualifies for the status of a lexicographic type by
virtue of at least three features in common: their syntactic function as well as
phonological and orthographic properties. It is only reasonable to assume that
these aspects of their usage should be accounted for in reference works, at least
in some dictionaries, usage guides, and reference grammars. Again, the ques-
tion of intended user immediately arises in this connection. Providing supra-
segmental features of the aforementioned type seems to be obligatory in
learner-oriented reference materials. From this viewpoint, it is natural to repre-
sent lexicographic information of the kind in the nonnative speaker-learner's
production dictionary (and/or grammar book), whereas this matter should be
reconsidered in the case of a general-purpose dictionary designed for the native
speaker-user.

It should be noted that Apresjan's description of verbs of perception in
Russian in terms of two main participants, the one who perceives and the sec-
ond the thing perceived, may be considered as a masterpiece of semantic ana-
lysis. No less imposing are "lexicographic portraits" of the verbs *shchitat'* 'con-
sider', *znat'* 'know', *vyiti* 'come out' and a number of other lexical items, which
are also represented in the book under consideration. Before we proceed, it is
necessary to discuss the notion of lexicographic portrayal.

The term "portrait" was introduced into descriptive semantic research by
A. Zholkovsky (1964). For him, a portrait should incorporate an exhaustive
description of word meaning attained by using it in the broadest possible range
of contexts and experimenting with its applicability for the description of the
broadest range of situations. Thus, the notion of portrayal was primarily lim-
ited to the specification of conceptual content and reference of a lexical item
obtained, and/or verified, by linguistic experiment. Apresjan suggested a
broader interpretation of the scope of this concept within the framework of
integrated description of a language (cf. Apresjan 2000: xvi):

> A lexicographic portrait is an exhaustive account of all the linguistically relevant
> properties of a lexeme, with particular emphasis on the semantic motivation of
> its formal properties. A certain property is considered to be linguistically rele-
> vant if there is a rule of grammar or some other sufficiently general rule (seman-
> tic rules included) that access this property.

This term based on metaphorical extension was intended to emphasize that a
dictionary entry should store lexicographic data hitherto not included in refer-
ence works and should considerably expand the amount of lexicographic
information that dictionaries traditionally contain. In addition to the afore-
mentioned prosodic features, one may note particular interest to lexical co-oc-
currence, i.e. combinability of lexical items. Apresjan (2000: 232) mentions that

the latter encompasses, in addition to lexical-semantic constraints, "prosodic, morphological, pragmatic, communicative and syntactic co-occurrence".

It should be mentioned in this connection that Apresjan developed his own interpretation of the pragmatics of a natural language. It has to be featured here because otherwise it is difficult to understand how to distinguish pragmatic properties of a lexeme from its communicative properties on the one hand, and its semantic ones on the other, particularly in view of the requirement that each of these types of lexicographic information at least was supposed to be represented in a different part of the dictionary entry.

In an article specifically dedicated to the issues of representing pragmatic information in a dictionary (Apresjan 1988) defines pragmatics as the expression of the speaker's attitude to reality, the message and/or the interlocutor, which is encoded in a linguistic sign, i.e. a lexical item, a morpheme or a syntactic structure. Moreover, the notion of pragmatic information in his interpretation sometimes includes lexical connotations and associations as opposed to components of meaning in the strict sense, e.g. connotation of monotony in the verb *pilit'* 'saw' which is characterized by a metaphorical extension 'nag', of abruptness in *rubit'* 'chop', and of speed in *streliat'* 'shoot'. Simultaneously, he also uses the terms "pragmatic information" or "pragmatic features" in the sense which does not deviate much from their generally-accepted usage; cf.: capacity for performative use or inadmissibility in reported speech and nontrivial illocutionary functions of a lexeme (e.g., see Apresjan 2000: 264). Nevertheless, in all these cases and similar ones, pragmatic information is regarded as a property of linguistic signs as units of a language-system; hence it requires uniform lexicographic presentation.

Thus, the notion of lexicographic type was introduced to emphasize the tendency to standardization of lexicographic presentation of shared properties, i.e. unification, whereas that of lexicographic portrait accounts for lexicographic presentation of what distinguishes lexical items, i.e. individualization.

The foregoing considerations enabled Apresjan to highlight basic principles of lexicographic description of synonymy in systematic lexicography, particularly requirements for a dictionary of synonyms. It will be fair to say that Apresjan's greatest contribution to metalexicography is the specification of basic principles of compiling a bilingual and a monolingual dictionary of synonyms.

In connection with bilingual lexicography, it is important to note that Shcherba in the aforementioned article made a tentative suggestion that a pedagogical dictionary should furnish not so much the translation equivalents of foreign language lexical items, but also and more importantly, detailed definitions in the learner's native language. From this it follows that a bilingual dictionary may be specifically designed either to serve the purposes of professional translation or to accommodate the needs of a nonnative language learner.

These two different purposes warrant different types of definitions (for a detailed analysis see Burkhanov 1999: 235-254; and Burkhanov 2003: 193-200).

Obviously enough, this approach opens up new perspectives for the development of bilingual lexicography which, for many experts in the field of lexicographic theory, still remains a synonym of translation lexicography (cf. Zgusta 1984). Moreover, with the further development of pedagogical lexicography it is reasonable to expect that more and more bilingual reference works will be intended to explicate meaning and usage of lexical items of the object language in the learner's native language. In this case, the latter will be used as a metalanguage of lexicographic description.

A successful attempt to implement Shcherba's ideas with respect to bilingual lexicography has been undertaken by a group of lexicographers under the leadership of Apresjan in the English–Russian dictionary of synonyms (Apresjan et al. 1979). Each entry (or rather dictionary article) of the reference work at issue, contains a detailed specification of the synonym series, i.e. the ordered list of synonyms, and consists of the following zones:

> (1) entry head comprising a list of synonyms starting with the lexical item with the most general meaning; (2) explanatory definition in Russian intended to specify the elements of meaning that are shared by all the synonyms, primarily the appropriate naïve concept and the evaluative component of meaning; (3) translation equivalents intended to simplify understanding explications and, at least partly, to compare the corresponding synonym series in English and Russian; (4) detailed explications of the synonyms in a given series with a particular emphasis on meaning discrimination, i.e. specification of their similarities and differences; (5) notes that account for those meanings of words making up the series which are close to the meanings under consideration as well as comments concerning figurative senses of those lexical units; (6) syntactic information; (7) co-occurrence constraints; and (8) illustrative examples.

As any other lexicographic publication, the dictionary under consideration is not free from drawbacks. Firstly, only one-word lexical items are included into synonym series; synonymous multiple-word lexical items, i.e. phraseological units, are beyond the scope of the dictionary. Secondly, synonyms are presented solely in a graphic form, whereas pronunciations of the corresponding phonological words are not provided, not to mention prosodic properties. Thirdly, the dictionary furnishes detailed specifications of synonyms on the scale that is unprecedented in bilingual lexicography, but only 350 synonym series are featured in it. It is not by chance that Apresjan himself called the dictionary "experimental". Nevertheless, this lexicographic project is, undoubtedly, an important milestone in the development of contemporary lexicography.

As for a monolingual dictionary of synonyms, Apresjan (2000: 55-100) suggested the following zonal structure of dictionary articles: (1) the zone of heading which comprises the dominant, i.e. a synonym with the most general

meaning, stylistic labels and grammatical notes, semantic groups within the series and their explication; (2) the zone of meaning specification that includes a synopsis (a short guide to the entry), similarities and differences in meaning between synonyms, etc.; (3) the form zone that accounts for the formal properties of usage; (4) the syntax zone; (5) the co-occurrence zone specifying combinatorial properties of the synonyms belonging to a series; (6) the illustration zone; and (7) the auxiliary zones. It is interesting to note that Apresjan decided to include phraseological synonyms into lexicographic material collected and analyzed for *Novyi ob'iasnitel'nyi slovar' sinonimov russkogo iazyka* 'New Explanatory Dictionary of Russian Synonyms' (Apresjan et al. 1997).

Hopefully the foregoing considerations have persuaded the reader that Juri Apresjan's book *Systematic Lexicography* is a very interesting and highly recommendable publication for both a practicing lexicographer and an expert in lexicographic theory. Not the least profit is to be gained by a theoretical semanticist who has never been exposed to Apresjan's stimulating and innovative ideas.

It should be specifically mentioned that a considerable part of the impact of this publication arises due to Kevin Windle's excellent translation. Paraphrasing a popular definition of translation, it can be said that he successfully coped with the extremely difficult task of "negotiating meaning in a different academic background".

Notes

1. For me, a linguist whose academic career began in the former USSR, Apresjan's book *Leksicheskaia semantika* originally published in 1974 was the first introduction to the works of A. Wierzbicka, Ch. Fillmore, G. Lakoff, and other internationally renowned experts in the study of meaning, not to mention that it was rightly regarded as a leading monographic study in semantic analysis. I am indebted to the Editor of *Lexikos* for giving me an opportunity to reconsider Apresjan's impact on the theoretical and applied study of language.
2. To keep the bibliography of the present article to the minimum and to make it easier for the prospective reader to locate the required information in the English translation, references to Apresjan's articles included in the volume under consideration will be made not to the original, but to the appropriate pages of the publication in question.
3. It is interesting to note that cognitive semanticists interpret the distinction between these two senses in terms of the distinction between primary and secondary domains. *Salt* in everyday sense is primarily associated with the domain of food, whereas the terminological unit *sodium chloride* which has the same reference, is understood against the domain of chemical composition (Langacker 1987: 165; Taylor 1989: 85-86). Thus, the standpoint of cognitive semantics and that of Russian linguistic thought are similar in this respect, though they are formulated within two different scientific paradigms.
4. More information concerning dictionary typology on the basis of distinctive features or lexicographic parameters can be found in Hartmann and James 1998: 147-148; Burkhanov 1998: 68-72; and other sources.

Bibliography

Apresjan, J.D. 1974. *Leksicheskaia semantika: sinonimicheskie sredstva iazyka*. Moscow: Nauka.

Apresjan, J.D. 1980. *Semantyka leksykalna. Synonimiczne środki języka*. Translated by Z. Kozlołowska and A. Markowski. Wrocław: Ossolineum.

Apresjan, J.D. 1988. Pragmaticheskaia informatsiia dlia tolkovogo slovaria. *Pragmatika i problemy intensional'nosti: Sbornik nauchnykh trudov*: 7-44. Moscow.

Apresjan, J. 2000. *Systematic Lexicography*. Translated by Kevin Windle. Oxford/New York: Oxford University Press.

Bartmiński, J. 1996. O "Słowniku stereotypów I symboli ludowych". Bartmiński, J. (Ed.). 1996. *Słownik srereotypów i symboli ludowych. Volume 1. Kosmos: ziemia, woda, podziemie*: 9-34. Lublin: UMCS Press.

Bergenholz, H. and S. Tarp (Eds.). 1995. *Manual of Specialised Lexicography*. Amsterdam/Philadelphia: John Benjamins.

Bloomfield, L. 1933. *Language*. New York: Holt and Co.

Burkhanov, I. 1995. *Uchebnyi slovar' sistemy poniatii lingvisticheskoi semantiki*. Rzeszów: University Press.

Burkhanov, I. 1998. *Lexicography: A Dictionary of Basic Terminology*. Rzeszów: University Press.

Burkhanov, I. 1999. *Linguistic Foundations of Ideography: Semantic Analysis and Ideographic Dictionaries*. Rzeszów: University Press.

Burkhanov, I. 2003. *Translation: Theoretical Prerequisites*. Rzeszów: University Press.

Cruse, D.A. 1986. *Lexical Semantics*. Cambridge: University Press.

Hartmann, R.R.K. and G. James. 1998. *Dictionary of Lexicography*. London/New York: Routledge.

Kalisz, R. 1981. *The Pragmatics, Semantics and Syntax of the English Sentences with **that** Complements and the Polish Sentences with **że** Complements*. Gdańsk: University Press.

Kotelova, N.Z. 1975. *Znachenie slova i iego sochetaemost' (k formalizacii v iazykoznanii)*. Leningrad: Nauka.

Krzeszowski, T.P. 1997. *Angels and Devils in Hell*. Warsaw: Energeia.

Langacker, R. 1987. *Foundations of Cognitive Grammar. Volume 1. Theoretical Prerequisites*. Stanford: University Press.

Lehrer, A. 1992. A Theory of Vocabulary Structure: Retrospectives and Prospectives. Pütz, Martin (Ed.). *Thirty Years of Linguistic Evolution*: 243-256. Amsterdam/Philadelphia: John Benjamins.

Lewandowska-Tomaszczyk, B. 1996. *Depth of Negation: A Cognitive Semantic Study*. Łódź: University Press.

Mel'čuk, I.A. 1974. *Opyt teorii lingvisticheskikh modelei "Smysl ⇔ Tekst"*. Moscow: Nauka.

Mel'čuk, I.A. 1982. Lexical Functions in Lexicographic Description. *Proceedings of the VIIIth Annual Meeting of the Berkeley Linguistic Society*: 427-444. Berkeley: University Press.

Reed, D. 1993. Some Terminological Problems of Translating Common Law Concepts from English to French. Sonneveld, H.B. and K.L. Loening (Eds.): 79-86.

Riggs, F. 1993. Social Science Terminology: Basic Problems and Proposed Solutions. Sonneveld, H.B. and K.L. Loening (Eds.): 195-222.

Seliverstova, O.N. 1975. *Komponentnyi analiz mnogoznachnykh slov*. Moscow: Nauka.

Shcherba, L.V. 1940. Opyt obshchei teorii leksikografii. *Izvestiia Akademii Nauk SSSP (otdelenie literatury i iazyka)* 3: 89-117.

Sonneveld, H.B. and K.L. Loening (Eds.). 1993. *Terminology: Applications in Interdisciplinary Communication*. Amsterdam/Philadelphia: John Benjamins.

Svensén, B. 1993. *Practical Lexicography: Principles and Methods of Dictionary-Making*. Transl. by J. Sykes and K. Schofield. Oxford/New York: Oxford University Press.

Taylor, J. 1989. *Linguistic Categorization: Prototypes in Linguistic Theory*. Oxford: Clarendon Press.

Wanner, L. (Ed.). 1996. *Lexical Functions in Lexicography and Natural Language Processing*. Amsterdam/Philadelphia: John Benjamins.

Wierzbicka, A. 1980. *Lingua Mentalis: The Semantics of Natural Language*. Sydney/New York: Academic Press.

Zgusta, L. 1984. Translational Equivalence and the Bilingual Dictionary. Hartmann, R.R.K. (Ed.). *LEXeter '83 Proceedings*: 147-154. Tübingen: Niemeyer.

Zholkovsky, A.K. 1964. Predislovie. *Mashinnyi perevod i prikladnaia lingvistika* 8: 3-16.

Lexicographic Publications

Apresjan, J.D., O.Iu. Boguslavskaia, I.B. Levontina, E.V. Urynson, M.Ia. Glovinskaia and T.V. Krylova. 1997. *Novyi ob'iasnitel'nyi slovar' sinonimov russkogo iazyka. Pervyi vypousk.* Moscow.

Apresjan, J.D., V.V. Botiakova, T.E. Latysheva, M.A. Mosiagina, I.V. Polik, V.I. Rakitina, A.I. Rozenman and E.E. Sretenskaia. 1979. *Anglo–russkii sinonimicheskii slovar'.* Moscow: Russkii iazyk.

Bartmiński, J. (Ed.). 1996; 1999. *Słownik srereotypów i symboli ludowych.* Vol. 1. *Kosmos: ziemia, woda, podziemie.* Lublin: UMCS Press.

Dal', V.I. 1863–1866. *Tolkovyii slovar' zhivogo velikorusskogo iazyka.* Third Edition. Edited by I.A. Baudouin de Courteney. 4 volumes. Saint-Petersburg, 1903–1909.

Hallig, R. and W. von Wartburg. 1952. *Begriffssystem als Grundlage für Lexikografie. Versuch eines Ordnungsschemas.* Berlin: Akademie Verlag.

Ivanov, V.V., V.M. Pan'kin, A.V. Filippov and H.M. Shansky. 1977. Kratkii slovar' traditsionnykh simvolov russkoi poezii. *Russkii iazyk.* 4-5.

Mel'čuk, I.A. and A.K. Zholkovsky. 1984. *Tolkovo–kombinatornyi slovar' sovremennogo russkogo iazyka.* Sonderband 14. Vienna: Wiener Slawistischer Almanach.

Mel'čuk, I.A., N. Arbatchewsky-Jumarie, L. Elnitsky, L. Iordanskaja and A. Lessard. 1984. *Dictionnaire explicatif et combinatoire du français contemporain. Volume I.* Montréal: Presses de l'Université de Montréal.

Mel'čuk, I.A., N. Arbatchewsky-Jumarie, L. Dagenais, L. Elnitsky, L. Iordanskaja, M.-N. Lefebvre and S. Mantha. 1988. *Dictionnaire explicatif et combinatoire du français contemporain. Volume II.* Montréal: Presses de l'Université de Montréal.

Mel'čuk, I.A., N. Arbatchewsky-Jumarie, L. Iordanskaja and S. Mantha. 1992. *Dictionnaire explikatif et combinatoire du français contemporain. Volume III.* Montréal: Presses de l'Université de Montréal.

Wierzbicka, A. 1987. *English Speech Acts Verbs: A Semantic Dictionary.* London/Port Worth/Philadelphia: Harcourt Brace Jovanovich.

Ou wyn in nuwe sakke: 'n Kritiese waardering van *Bijbels lexicon*

Anna Nel Otto, *Departement Afrikaans, Universiteit Vista, Port Elizabeth, Republiek van Suid-Afrika (otto-an@pelican.vista.ac.za)*

Opsomming: *Bijbels lexicon* is 'n Nederlandse leksikon wat woorde en uitdrukkings bevat wat aan die Bybel ontleen is en wat ná 1945 nog bekend was. Die teikengebruikers van hierdie leksikon is waarskynlik persone wat graag die Bybel lees en diegene wat belangstel in die oorsprong en agtergrond van die Nederlandse taal. Die inleiding tot die leksikon bevat baie inligting, onder andere oor hoe die leksikon saamgestel is, 'n uiteensetting van tipes leksikale items en motivering vir lemmakeuses en tydvak wat gedek word, hoe om leksikale items na te slaan, ens. Selfs niegelowiges sal die leksikon kan gebruik, aangesien die inleiding ook agtergrondsinligting in blokvorm bevat, nl. 'n uiteensetting van die inhoud van die Bybel, en inligting oor Nederlandse Bybelvertalings, spreekwoordversamelings en Bybelse name.

Die voordele van hierdie leksikon is dat dit slegs huidige gebruiksvoorbeelde bevat en daarom dus neologismes en veranderde betekenisse akkommodeer. Hoewel die outeurs in enkele gevalle afwyk van die terminologie vir verskillende tipes leksikale items, soos deur hulle self omskryf, is die hantering van woordeboekartikels oor die algemeen baie konsekwent en inligting is maklik toeganklik. Die waarde van die leksikon sou moontlik verder verhoog kon word deur die gebruik van meer kruisverwysings.

Sleutelwoorde: BETEKENIS, BYBEL, EIENAAM, HEDENDAAGSE GEBRUIK, HERKOMS, LEKSIKOGRAFIE, LEKSIKON, NEDERLANDS, NEOLOGISME, SITAAT, SPREEKWOORD, VASTE UITDRUKKING

Abstract: **Old Wine in New Bottles: A Critical Appreciation of *Bijbels lexicon*.** *Bijbels lexicon* is a Dutch lexicon which includes words and fixed expressions which have been borrowed from the Bible and which were still known after 1945. The target users of this lexicon are probably people who are fond of reading the Bible and those who are interested in the origin and background of the Dutch language. The introduction to the dictionary contains much information, *inter alia* on how the lexicon is structured, an exposition of different types of lexical items and motivation for the choice of lemmata and period covered, on how to look up lexical items, etc. Even non-believers could use the lexicon, as the introduction also contains background information in block format, i.e. an exposition of the content of the Bible, and information on Dutch Bible translations, proverb collections and Bible names.

The benefits of this lexicon are that it includes only examples of current usage and as such accommodates neologisms and changed meanings. Although the authors sometimes divert from the terminology for different types of lexical items, as defined by themselves, the treatment of dictionary articles is generally very consistent and information is easily accessible. The value of the lexicon could possibly be enhanced by using more cross-references.

Keywords: BIBLE, CITATION, CURRENT USAGE, DUTCH, FIXED EXPRESSION, LEXI-COGRAPHY, LEXICON, MEANING, NEOLOGISM, ORIGIN, PROPER NAME, PROVERB

1.   Inleiding

Bijbels lexicon is 'n Nederlandse leksikon waarin 'n bepaalde seleksie van woor-de, uitdrukkings en spreekwoorde opgeneem is wat aan die Bybel ontleen is. Dié seleksie is gebaseer op daardie woorde en uitdrukkings wat ná 1945 nog bekend was.

Die teikengebruikers word nie in die teks self geïdentifiseer nie, maar op die agterblad word wel gesê dat *Bijbels lexicon* 'n unieke boek is vir elke By-belliefhebber, gelowig of niegelowig, en dat dit onontbeerlik is vir elkeen wat belangstel in die oorsprong en die agtergrond van die Nederlandse taal.

In verskeie publikasies word die belangrikheid van 'n goed geïdentifiseer-de teikengroep en die vasstel van hulle behoeftes beklemtoon. De Schryver en Prinsloo (2000: 1-31) bepleit byvoorbeeld die konsep van "gelyktydige terug-voering", dit is dat terugvoering vanaf die teikengebruikers verkry moet word terwyl die samestelling van die woordeboek nog aan die gang is. In *Bijbels lexi-con* word daar slegs in die voorwoord aangedui dat die outeurs die versoek gekry het om 'n woordeboek saam te stel wat Bybelse woorde en uitdrukkings bevat wat steeds gebruik word. Dit is dus nie duidelik of formele en/of infor-mele marknavorsing gedoen is en wie presies die versoek aan die outeurs gerig het nie.

In 'n navorsingsprojek oor die behoeftes van Bybelgebruikers het die outeur van hierdie artikel bevind dat uit 'n totaal van 29 persone met "religieu-se beroepe" slegs 5 persone aangedui het dat hierdie inligting essensieel is in 'n Bybelwoordeboek, terwyl uit 'n totaal van 45 persone met "nie-religieuse" beroepe niemand aangedui het dat hulle hierdie inligting as essensieel beskou nie, hoewel 15 persone aangedui het dat hierdie inligting wenslik is. Die vraag wat in 'n vraelys aan respondente gestel is, lui soos volg: Watter van die vol-gende inligtingstipes beskou u as essensieel, wenslik of oortollig in 'n Bybel-woordeboek: aanhalings wat die huidige gebruik (ná 1945) van woorde en uit-drukkings uit die Bybel in ander media illustreer, bv. in tydskrifte, boeke, oor die radio, televisie, ens.? Dit wil dus voorkom asof die meeste respondente wat hoofsaaklik Suid-Afrikaners was, nie juis 'n behoefte aan hierdie tipe woorde-boek het nie.

2.   Werkwyse

By die beoordeling van *Bijbels lexicon* sal daar kommentaar gelewer word op die verskillende dele waaruit die woordeboek bestaan, deur eers aan te dui hoe die woordeboek saamgestel is en deur die kommentaar in dieselfde volgorde hierby aan te bied.

3.   Die niealfabetiese deel van die leksikon

3.1   Die toeligtingsgedeelte

Dit is verblydend dat die toeligtingsgedeelte so volledig is, afgesien daarvan dat 'n teikengroep of -groepe nie hier eksplisiet genoem word nie. Hierdie deel bestaan uit die volgende afdelings: 'n inleiding, 'n geskiedkundige oorsig waar-in beskryf word hoe Bybeltekste Nederlands sedert die vroegste tye beïnvloed het, 'n motivering vir die fokus op hedendaagse gebruik, 'n toeligting oor die verskillende tipes ontlenings, kriteria vir opname, verantwoording van bronne, 'n afdeling wat uiteensit hoe die artikels opgebou is en aanwysings oor hoe om woorde en uitdrukkings na te slaan.

3.1.1   Die inleiding

Die inleiding dui die samestelling van die leksikon volledig aan. Daar word eerstens gewys op die feit dat Nederlands baie woorde, uitdrukkings en spreekwoorde het wat aan die Bybel ontleen is. Baie van hierdie woorde en uitdrukkings is egter nie meer bekend nie. Dan word 'n motivering gegee vir die opname van hierdie woorde en uitdrukkings in die leksikon: dit is nie net interessant om hulle agtergrond te ken nie, maar vir die begrip van literêre en ander tekste uit die laaste eeu is dit ook nodig om hulle betekenis en herkoms te ken. Boonop gee hierdie woorde en uitdrukkings 'n beeld van die inhoud en sfeer van die Bybelverhale en daarmee ook van een van die belangrikste tradi-sionele elemente van die Nederlandse kultuur, aldus die outeurs.

Die inligtingstipes wat in die alfabetiese gedeelte opgeneem is, word uit-eengesit en die seleksie van woordeskatitems word aangedui, nl. woorde en uitdrukkings wat sedert 1945 bekend was/steeds is. Dié woordeskatitems is in alfabetiese volgorde opgeneem, voorsien van 'n betekenisverklaring, 'n toelig-ting oor die herkoms, 'n Bybeluittreksel uit 'n ouer vertaling en laastens enkele sitate ter illustrasie van die huidige gebruik uit verskillende soorte moderne bronne.

Daar word ook uiteengesit watter onderwerpe in die niealfabetiese gedeel-te behandel word.

3.1.2   Die geskiedkundige oorsig oor hoe Bybeltekste Nederlands beïnvloed het

Interessante feite word in hierdie afdeling vermeld. Slegs enkele feite wat die outeurs noem, word hier uitgelig. Daar word aanvaar dat Christelike leen-woorde al vanaf die agtste en negende eeu ontstaan het. Bestaande Germaanse woorde het nuwe, Christelike betekenisse bygekry: vergelyk woorde soos *hel, herder* en *ziel.* Hierdie nuwe Christelike woordeskat het betrekking gehad op kerktaal, soos *absolutie, hostie* en *bisschop*, en Bybeltaal, bv. *kruis, engel* en *priester.*

Woorde uit Grieks en Hebreeus het meestal via Latyn in Nederlands

gekom. Voorbeelde is *amen* en *psalm* (Hebreeus) en *evangelie* en *apostel* (Grieks). Uit onvertaalde Latyn het Nederlands ook enkele name van Bybelboeke (bv. *Genesis* en *Exodus*) oorgeneem, en Bybelse verbindings soos die eerste versreëls van gesange of gebede, wat later soms selfs as afsonderlike woorde 'n eie betekenis gekry het (bv. *lavabo*, 'ik zal wassen', later 'wasbekken').

3.1.3   Die huidige gebruik

'n Besliste voordeel van hierdie leksikon is dat die gebruiker verseker word dat slegs huidige gebruiksvoorbeelde opgeneem is. Die outeurs wys dan ook daarop dat daar twee belangrike verskille is tussen dít wat woordeboeke en spreekwoordversamelings ons ten opsigte van idiome, Bybels al dan nie, aanbied en daardie idiome wat werklik gebruik word. 'n Uitsondering is *Van Dale Idioom woordenboek* van 1999 wat ook net op werklike gebruik gebaseer is.

Eerstens bevat naslaanwerke dikwels verouderde materiaal. As voorbeeld noem die outeurs die herdruk van Huizinga (1998) wat twee uitdrukkings bevat, wat na hulle mening nie meer gebruik word nie, nl. *maaien waar men niet gezaaid heeft*, 'vruchten plukken van de inspanningen van een ander' en *gehoorzamen is beter dan offerande* 'het is beter eerst naar een ander te luisteren dan om meteen te handelen'. Die teikengebruikers se behoeftes, in samehang met dié tipe woordeboek, is deurslaggewend vir die opname van verouderde materiaal al dan nie. In 'n omvattende of handwoordeboek verwag gebruikers volledigheid, met dien verstande dat verouderde materiaal as sodanig geëtiketteer behoort te word. Indien daar 'n behoefte bestaan aan 'n woordeboek wat slegs huidige gebruik verteenwoordig, behoort verouderde materiaal uit die aard van die saak nie in so 'n woordeboek opgeneem te word nie.

Die outeurs merk tereg op dat daar 'n besliste afname is in mense se kennis van die Bybel. Bybeltekste word nie maklik meer spontaan aangehaal nie. Ook het die spontane vernoeming van voorwerpe na Bybelse persone of plekke sterk afgeneem. Hulle wys ook daarop dat veroudering saamhang met taalkringe. Wat uit die algemene Nederlands verdwyn het, kan in bepaalde Christelike kringe nog algemeen wees.

'n Tweede groot verskil tussen dít wat opgeneem is in woordeboeke en spreekwoordversamelings en die werklike gebruik daarvan, is die baie variasie ten opsigte van vorm by werklike gebruik. As daar in die woordeboeke *in goede aarde vallen* staan, dan tref 'n mens in die praktyk ook *vruchtbare aarde* aan, en vir die teenoorgestelde betekenis *verkeerde* of *dorre aarde.* Dit gebeur ook soms in die praktyk dat elemente van uitdrukkings weggelaat word, bv. *te licht bevonden* naas *gewogen en te licht bevonden.* Volgens die outeurs ontstaan analogieë ook soms in die praktyk: *de ware Jakob* lei tot *de ware Jakoba* en *Sara zien* kom naas *Abraham zien* voor.

Daar is nie net 'n groot mate van variasie ten opsigte van vorm nie, maar ook die reikwydte van toepassing val op. Baie woorde en uitdrukkings wat uit 'n ernstige Bybelgedeelte kom, kan gemaklik in 'n wêreldse konteks gebruik

word, bv. *de geest is wel gewillig, maar het vlees is zwak.* 'n Ironiese gebruik van hierdie uitdrukking kry 'n mens in Jan Wolkers se roman *Een roos van vlees* (p. 11) waar die hoofkarakter daarop wys dat hy wel 'n gesonde liggaam en 'n gesonde gees behoort te hê vanweë die gimnastiek wat hy beoefen, maar besef dat hy eintlik 'n ongesonde gees in 'n gesonde liggaam het: "Het vlees is wel gewillig, maar de geest is zwak." Dat die gebruik van die naam van *Jezus* in die informele vergelyking *zo lui als het paard van Jezus* redelik resent is, sou volgens die outeurs dalk verklaar kon word op grond van die gebrekkige optekening van die informele of plat spreektaal.

In literêre tekste vind 'n mens sowel ingeburgerde uitdrukkings, soms met kreatiewe variante, as meer regstreekse verwysings na 'n Bybelteks of aanhaling daaruit, waarby dit om die Bybelse inhoud gaan. Vergelyk weer eens enkele voorbeelde uit bogenoemde roman van Jan Wolkers:

Eet, drink, dit is mijn bloed hetwelk voor u vergoten is. (p. 10)
Daar ligt vlees van mijn vlees. (p.14)
Dood waar is uw prikkel, denkt hij. (p. 28)

Wat die politici betref, neem Bybelse taal ook by hulle af. Volgens die outeurs is dit jammer, want hulle hou uitdrukkings soos *iemands nieren proeven* en *een tocht door de woestijn* lewend en het gebruike soos die volgende in die taal gebring: *bergrede* vir 'rede over sociale politiek'.

3.1.4   Toeligting by die verskillende tipes ontlenings

Dit is belangrik dat leksikograwe 'n duidelike omskrywing gee van die verskillende tipes woordeskatitems wat in 'n woordeboek opgeneem word en dan konsekwent by hierdie omskrywings bly. In hierdie afdeling gee *Bijbels lexicon* 'n omskrywing van die terminologie wat gebruik word. Die term *uitdrukking* word gedefinieer as "een vaste verbinding van twee of meer woorden met een eigen betekenis". Onder die term *uitdrukking* moet ook spreekwoorde verstaan word. Hoewel *Bijbels lexicon* dus nie formeel 'n onderskeid tref tussen dít wat hulle "uitdrukkings" noem en spreekwoorde nie, omskryf hulle wel die term *spreekwoord*, soos hulle dit verstaan: "een levenswijsheid in de vorm van een volzin, die ook zonder context betekenis heeft."

Daar bestaan 'n wye verskeidenheid terme vir die begrip *vaste uitdrukking* of *vaste verbinding*. Gouws (1989: 98) beskryf vaste uitdrukkings as geykte komplekse leksikale items. Hulle is geleksikaliseerde eenhede met 'n betekenis wat nie sonder meer afleibaar is van dié van die samestellende dele nie. Vaste uitdrukkings en spreekwoorde verskil dus in dié opsig dat spreekwoorde se betekenisse wel deursigtig is. Verstraten (1992: 12) definieer die term *vaste verbinding* soos volg: "een vaste verbinding is een combinatie van minimaal twee lexicale elementen (woorden), die als gefixeerd geldt." Vastheid, al dan nie, word bepaal deur die terme komposisionaliteit, deursigtigheid, gemotiveerdheid en idiomatisiteit.

Die begrip **komposisionaliteit** het betrekking op die wyse waarop die betekenis van die vaste verbinding opgebou is: indien die betekenis beskou kan word as 'n optelsom van die betekenisse van die samestellende dele, dan het 'n mens te make met 'n komposisionele betekenis.

Die **deursigtigheid** van die verbinding is afhanklik van die mate waarin die taalgebruiker in staat is om raak te sien op watter wyse die betekenis van die verbinding uit dié van die samestellende dele tot stand gekom het. Die komposisionaliteit van 'n verbinding kan 'n invloed hê op die deursigtigheid daarvan. Deursigtigheid is egter 'n subjektiewe begrip, want dit is sterk afhanklik van die individuele kennis van die taalgebruiker.

Die begrip **gemotiveerdheid** word soos volg deur Lakoff (1987: 448) omskryf: "The relationship between A and B is *motivated* just in case there is an independently existing link L, such that A-L-B 'fit together'. L *makes sense* of the relationship between A and B."

Verstraten (1992: 14) beskou **idiomatisiteit** en **ondeursigtigheid** in beginsel as dieselfde.

Daar is egter nie altyd sprake van net twee kategorieë, nl. woorde of uitdrukkings, wat in *Bijbels lexicon* opgeneem is nie. By die lemma **aanstoot** staan die volgende:

> **Aanstoot**, iets waaraan men zich stoot; (fig.) persoon of zaak die ergernis of morele verontwaardiging wekt. Steeds in de volgende uitdrukkingen.
> **Aanstoot geven,** morele ergernis opwekken.
> **Aanstootgevend**, ergerniswekkend, kwetsend.
> **Aanstoot nemen aan,** zich ergeren aan; als kwetsend beschouwen.
> **Steen des aanstoots,** iets waaraan men zich ergert, oorzaak van ergernis.

Die sogenaamde uitdrukkings **aanstoot geven** en **aanstoot nemen aan** is eerstens kollokasies (aangesien die dele komposisioneel en deursigtig is) en nie uitdrukkings nie. Tweedens is die kollokasies nie konsekwent gehanteer nie. Die setsel **aan** sou eweneens, net soos by **aanstoot nemen (aan)**, by **aanstoot geven (aan)** kon optree.

Dit is belangrik dat uitdrukkings korrek en volledig met betrekking tot variante vorme opgeneem word. By die lemma **Abraham** word onder andere die volgende uitdrukkings aangegee: **aan Abrahams borst** of **in Abrahams schoot rusten**. Die korrekte vorme is egter: **aan Abrahams borst rusten** of **in Abrahams schoot zitten**, soos aangedui in *Bijbels lexicon* se voorbeeldsinne en bevestig deur *Grote Van Dale* en *Spreekwoorden verklaard*.

Die volgende tipes ontlenings word in die niealfabetiese deel van *Bijbels lexicon* onderskei:

(a) Woorde en uitdrukkings wat afkomstig is van 'n Bybelvertaling

Die outeurs verduidelik hierdie kategorie soos volg: dit is afkomstig van 'n Bybelvertaling en kon 'n gewysigde, byvoorbeeld figuurlike betekenis gekry het of het die reeds bestaande figuurlike betekenis behou, soos *aanfluiting*, *jubeljaar*,

manna en *talent.* Afleidings en samestellings word ook aangedui, bv. *profeteren* en *profetie* asook *onheilsprofeet* en *weerprofeet* onder *profeet.*

Nuutskeppings op basis van woorde wat uit 'n Bybelvertaling afkomstig is, sluit die volgende kategorieë (met voorbeelde daarnaas) in: afleidings: *evangeliseren, paradijselijk*; samestellings: *aanstootgevend, vredesapostel*; uitdrukkings: *apostel van x, geen jota van iets begrijpen.* Voorbeelde van uitdrukkings wat in die geheel op 'n Bybelse formulering teruggaan, is talryk: *naar de letter, met bezemen keren, zich niet onbetuigd laten, het is volbracht.* 'n Spesiale vorm het beginwoorde van gebede of liedere soos *Onzevader, Weesgegroet* en, uit die Latyn, *lavabo.* Spreekwoorde, bv. *aan de vruchten kent men de boom* en *een goede buur is beter dan een verre vriend,* kom blykbaar meestal in mondelinge taalgebruik voor.

Kommentaar op grond van 'n eie steekproef

Dit is interessant dat die woord *Pasen* as lemma opgeneem is en dat daar in die verduideliking na *paasfeest* verwys word, maar dat *paasfeest* nie as lemma met 'n kruisverwysing na *Pasen* aangedui word nie. Vergelyk die volgende voorbeeld:

> Na de begrensde en behaaglijke sfeer van de gevangeniscel, … waar wij ons om de honger te stillen overgaven aan fantasieën rond het paasfeest, bevond ik me in een wereld van ruime afmetingen. (NWT 16(3): 34, 1999)

Dit is eweneens nie duidelik waarom *schepping* as lemma opgeneem is, maar nie *Schepper* nie. By *schepping* word bloot terloops verwys na *Schepper*. Vergelyk enkele bewyse vir die hedendaagse gebruik van *Schepper*:

> Losgerukt van het continent waar witte ladders omhoogrezen tot de troon van de Schepper, vond ik mezelf terug op een eiland. (NWT 16(3): 40, 1999)

> God (heeft) ons naar Zijn beeld en gelijkenis geschapen, en zoals de Schepper zich verhoudt tot zijn schepping, zo verhoudt de scheppende mens, de kunstenaar, zich tot zijn kunstwerk. (NWT 16(3): 51, 1999)

(b) Nuutgevormde woorde en uitdrukkings vir begrippe uit die Bybel

Volgens die outeurs bestaan hierdie kategorie uit taalvorme wat nie letterlik so in 'n Bybelvertaling voorkom nie, maar wel Bybelse sake benoem. Die herkoms is nie altyd duidelik nie. Hulle kon moontlik in die spreektaal ontstaan het en hulle word reëlmatig deur uitgewers van Bybelvertalings gebruik in inleidende tekste of samevattings by die eintlike Bybelteks of as hoofstuktitel. Woorde uit hierdie groep is: *bergrede, beschermengel, kruiswoord* en *zondeval*; uitdrukkings: *de tien geboden, verloren paradijs, verboden vrucht* en *de verloren zoon.*

Kommentaar op grond van 'n eie steekproef

Die woord *kruisweg* is as samestelling onder die lemma *kruis* opgeneem. Hoewel *kruistog* nie direk van Bybelse herkoms is nie, sou 'n mens by *kruisweg* 'n

verwysing na *kruistog* waardeer ten einde die verskille tussen die twee woorde se betekenisse en herkoms aan te dui. Hier volg enkele voorbeelde ter illustrasie van *kruistog*:

> In 1984 lanceerde dit instituut *Losing Ground,* het werk van Charles Murray dat de bijbel zou worden van Reagens kruistocht tegen de verzorgingsstaat. (NWT 16(8): 60, 1999)

> In die omstandigheden is het moeilijk om een kruistocht voor radikale vernieuwing te voeren, zelfs als het grootste deel van het kiezerscorps het leven permanent somber lijkt in te zien. (NWT 16 (1):12, 1999)

Op 'n min of meer soortgelyke wyse sou 'n mens by *Kerst* 'n sublemma *kerstdag* wou hê, soos geïllustreer in die onderstaande voorbeeld:

> Ik begon over "A Child's Christmas in Wales". Het verhaal van Dylan Thomas, waarop het liedje is gebaseerd, gaat over een idyllische kerstdag uit Thomas' kindertijd. (NWT 16(1): 25, 1999)

(c) Eiename

Hierdie kategorie betrek name van persone en plekke, genoem in gestandaardiseerde of niegestandaardiseerde uitdrukkings (bv. *zo arm als Job, naakt als Eva*), as simbool of as metafoor (bv. *Benjamin, Babylon)*. Soms is daar ook persone in stereotiepe voorstellings, bv. *Petrus aan de hemelpoort*, wat nie 'n vaste uitdrukking is nie, maar dikwels as sodanig saamgebruik word. Van Bybelname word afleidings gevorm, wat dan wel opgeneem word, bv. *jeremiade* (van *Jeremia*), *onaneren* (van *Onan*), verdere samestellings soos *adamsappel, salomonsoordeel* en uitdrukkings soos *Job op de mestvaalt*, *ongelovige Tomas*. Laastens is daar ook kombinasies van name, soos *Adam en Eva*, *David en Goliat*, wat nie altyd so in kombinasie in die Bybel self genoem word nie.

Dit wil voorkom asof persoonsname redelik volledig opgeneem is. So is byvoorbeeld die volgende bewys gevind dat *Lazarus* steeds gebruik word:

> Ik haal mijn handen over mijn christelike gezicht om de bleekheid te verdrijven die me overviel bij het zien van deze Lazarus die nog rook naar de joodse dood. (NWT 16(3): 31, 1999)

In die volgende voorbeeld word twee persoonsname, nl. *Salomon* en *Hiram*, genoem. *Salomon* is tereg opgeneem, maar *Hiram* nie. Dit kan toegeskryf word aan die feit dat *Hiram* nie in 'n uitdrukking gebruik word nie en nie 'n metaforiese of simboliese betekenis het nie, terwyl daar algemeen na die wysheid van Salomo verwys word in die omgangstaal:

> In mijn sarmatische … boot roei ik naas hun cederhouten vlot, van het hout dat Hiram aan Salomon stuurde om hem te behagen. (NWT 16(3): 31, 1999)

Al verbetering wat 'n mens sou kon aanbring, is om die wisselvorm *Salomon*, naas *Salomo*, reeds by die lemma te gee, waar slegs die vorm *Salomo* opgeneem

is. Eers in die derde paragraaf word aangedui dat *Salomon* 'n ouer vorm van *Salomo* is.

Dit is opvallend dat die persoonsnaam *Delila* nie opgeneem is nie. In *Grote Van Dale* (Lys van Bybelse name in Aanhangsel III) staan daar by *Delila*: "Filistijnse vrouw die als minnares van Simson hem in handen speelt van de Filistijnen, na hem het geheim van zijn kracht ontlokt te hebben (Richt. 16: 4)." In *Verklarend Handwoordenboek der Nederlandse Taal* is *Delila* opgeneem en soos volg verklaar: "1 vrouw van Simson, een Filistijnse (*bijb. Richt. 16*); 2 (*fig.*) verleidelijke, trouweloze vrouw." Na my beste wete was *Delila* slegs Simson se minnares en het die woord vandag nog die figuurlike betekenis "verleidelike, listige vrou". In *Bijbels lexicon* by die artikel *Simson* word *Delila* Simson se vrou genoem.

Wat die spelling van eiename betref, sou 'n mens soms meer wisselvorme verwag of is die spelling waarskynlik foutief, aangesien dit nie ooreenstem met die spelling in die voorbeelde nie. Vergelyk *Thomas* wat in *Bijbels lexicon Tomas* gespel word, maar waar al die voorbeeldsinne die spelling *Thomas* het.

Die naam *Jesaja* het by die Katolieke die vorm *Isaias*, gevolglik behoort hierdie vorm ook opgeneem te word.

Objektiwiteit word meestal gehandhaaf deurdat sowel woorde en uitdrukkings uit die kanonieke as apokriewe boeke opgeneem word. Die woord *Susanna* en die uitdrukking *een kuise Susanna* uit 'n apokriewe boek is byvoorbeeld wel opgeneem, maar die spelling verskil van dié van *Grote Van Dale*, wat *Suzanna* en *een kuise Suzanna* met 'n z-spelling vermeld. Dit kan gebruikers verwar.

(d) Sitate

Daar is ook woorde en uitdrukkings wat, net soos die eiename hierbo beskryf, tot die taalgemeengoed behoort. Toevallig gebruikte Bybelse frases kom soms voor. Hulle word aangehaal as die woorde van iemand anders. Soms gaan hierdie aanhalings gepaard met 'n globale bronvermelding, bv. *zoals de bijbel zegt*, soms is dit presieser, soos in die volgende voorbeeld: *Niet alles wat toegestaan is, is wenselijk, om maar eens met Paulus te spreken*. As so 'n uitspraak uit meer as een bron bekend is, en aangehaal word in 'n algemene konteks, dus van toepassing is op 'n nie-Bybelse situasie, dan is hierdie uitspraak in die *Bijbels lexicon* opgeneem. Die rede hiervoor is dat hierdie uitsprake onderweg na 'n volwaardige uitdrukking kan wees. Regstreekse Bybelsitate met presiese bronvermelding is uit die aard van die saak nie opgeneem nie. Ook die sitate sonder bronvermelding vorm 'n problematiese kategorie, veral aangesien bepaalde literêre outeurs na hartelus in Bybelse woorde praat.

Kommentaar op grond van 'n eie steekproef

Die bogenoemde kategorie kan geldig wees, soos bewys deur die volgende voorbeeld:

> Waarom zouden we niet samen met koning Salomo zeggen: voor alles is een tijd. (NWT 16(8): 15, 1999)

En 'n mens wonder dan waarom slegs die gestandaardiseerde uitdrukkings *Er is een tijd van (een bepaald bezigheid) en van (de tegenovergestelde bezigheid)* en *Er is een tijd van komen en er is een tijd van gaan* opgeneem is in *Bijbels lexicon*. Minstens in Afrikaans hoor 'n mens gereeld die uitdrukking *vir alles is daar 'n tyd (en 'n plek)* en die feit dat dit gebruik word, is ook bevestig in 'n redakteursbrief in die tydskrif *rooi rose* van Februarie 2003, p. 8.

Die problematiek rondom die opname, al dan nie, van bogenoemde kategorie, blyk uit 'n voorbeeld soos die volgende, wat myns insiens waarskynlik 'n geïsoleerde geval verteenwoordig:

> Dank u, Heer, dat ik niet ben als zij. (NWT 16(8): 54, 1999)

Dit laat 'n mens onmiddellik Lukas 18: 11 in herinnering roep, wat soos volg in 'n Afrikaanse vertaling lui:

> Die Fariseër het gaan staan en by homself so gebid: "O God, ek dank U dat ek nie soos ander mense is nie: diewe, bedrieërs, egbrekers, en ook nie soos hierdie tollenaar nie."

'n Ander problematiese geval is *als 't God belieft*, wat nie in *Bijbels lexicon* opgeneem is nie en voorkom in die volgende voorbeeld:

> Op de tweede dag van de tweede maand van het jaar tweeduizend zal een klein meisje, als 't God belieft, haar tweede verjaarsdag vieren. (NWT 16(8): 54, 1999)

Die outeurs het waarskynlik nie hierdie uitdrukking opgeneem nie, aangesien hulle dalk nie bewyse van Bybelse herkoms kon vind nie. Die naaste uitdrukking hieraan waaroor inligting opgespoor kon word wat op 'n moontlike Bybelse herkoms dui, is *Deo volente* waaroor in die woordeboekartikel van **Deus** (in *Grote Van Dale*, p. 637) die volgende inligting verskaf word: "*Deo volente* (D.V.), zo het God behaagt, zo God wil (slechts éénmaal in de bijbel: Acta Apostolorum 18: 21)." *Deo Volente* is wel in *Bijbels lexicon* opgeneem.

Om dieselfde rede is die woord *bekeerd* waarskynlik nie opgeneem nie. Vergelyk die onderstaande voorbeeld:

> Ik ben niet bekeerd, of juist wel, net hoe je het bekijkt. Ik geloof, maar ik zwijg daar het liefst over. (NWT 16(8): 31, 1999)

Stam (2002: 4) haal Karina van Dalen-Oskam aan in hierdie verband: "Als wij niet zeker wisten of een uitdrukking aan de Bijbel was ontleend," en "omdat je die ook al in een heel vroeg stadium in wereldlijke teksten tegenkwam, namen wij het woord niet op. Het is dus verklaarbaar dat u een woord als 'bekering' mist. Wij wilden puur registreren." Samewerking tussen leksikograwe, etimo-

loë en teoloë sou waarskynlik tot groter sekerheid oor problematiese gevalle kon lei. In hierdie verband beklemtoon Tarp (2000: 201) samewerking tussen leksikograwe en vakspesialiste.

Ten opsigte van al hierdie kategorieë woorde en uitdrukkings sou 'n mens meer kruisverwysings wou hê ten einde verbande tussen lemmata makliker raak te sien. By *Abraham* is daar wel 'n verwysing na *Sara* en omgekeerd. Kruisverwysings tussen die volgende lemmata sou verhelderend optree: *antichrist* en *beest*, *apostel* en *discipel*, *Beëlzebul* en *duivel*, *Geest, God* en *Christus*, ens.

4.   Kriteria vir opname

Die leksikon bevat slegs woorde en uitdrukkings vir sover hulle ná 1945 nog bekend was. Die outeurs wys daarop dat daar in hierdie periode belangrike maatskaplike veranderinge plaasgevind het, soos vernuwing in die kuns en literatuur en toenemende ontkerkliking. Dit is ook belangrik dat daar aan die begin van hierdie periode 'n nuwe Protestantse Bybelvertaling verskyn het, die NBG-vertaling van 1951, soos die verantwoording dit stel.

5.   Verantwoording van bronne

Stam (2000: 2) wys tereg daarop dat *Bijbels lexicon* 'n boeiende boek geword het omdat dit uit so 'n breë spektrum bronne put. Benewens die vele Bybelvertalings en gesaghebbende handboeke, is daar ook 'n groot versameling literatuur, wat wissel van Baantjer tot Van Zomeren en van *Het grote kamerplantenboek* tot die *Tuinfluiter-Trilogie*.

Die sitate uit koerante en uit die politiek kom byna almal uit die 27 miljoen koerantkorpus en die 38 miljoen woordkorpus van die Instituut voor Nederlandse Lexicologie in Leiden, waaraan albei outeurs verbonde is. Vanweë die beperkings van die korpus word daar dikwels aangehaal uit *NRC Handelsblad*, die *Meppeler Courant* en *De Standaard* en selfs een keer uit *Playboy*, maar slegs selde uit *Trouw, Nederlands Dagblad* of *Reformatorisch Dagblad*. "Zodoende konden wij overigens duidelijker maken dat aan de Bijbel ontleende uitdrukkingen ook echt algemeen en in meer wereldlijke lectuur gebruikt worden," aldus Karina van Dalen-Oskam (Stam 2000: 2-3).

Stronks (2000: 1) lewer kritiek op die feit dat "een weinig representatiewe bron" soos die *Meppeler Courant* heel dikwels aangehaal word, daar verwys word na bronne soos geselsies en telefoongesprekke wat nie nagegaan kan word nie, en daar nie erkenning gegee word aan die verskille in die Noord- en Suid-Nederlandse taalsituasie nie.

Aangesien die outeurs streng kriteria het vir opname, kan daar verwag word dat sekere bronne waarskynlik prominenter as ander sal figureer. En dit is verblydend dat die spreektaal ook in ag geneem is. By baie woordeboeke ontbreek neologismes soms juis omdat die spreektaal nie voldoende in aanmerking geneem word nie.

6.   Slot

Bijbels lexicon is inderdaad 'n boeiende boek wat in terme van die kriteria vir opname en hantering van artikels so konsekwent en volledig moontlik saamge-stel is. Slegs enkele inkonsekwenthede kom voor. Dit bied baie interessante inligting en 'n mens kan verrassende ontdekkings maak, byvoorbeeld deur vas te stel dat 'n uitdrukking soos *op die hande dra* van Bybelse herkoms is. Uit *Bij-bels lexicon* kan 'n mens ook aflei hoe betekenisse met verloop van tyd verander het, bv. *vele zijn geroepen, maar weinigen uitverkoren* het deesdae in Nederlands die betekenis dat baie mense in bepaalde beroepe belangstel, maar dat min van hulle hiervoor geskik is. By gebrek aan 'n soortgelyke woordeboek in Afri-kaans, kan Afrikaanse lesers beslis baat vind by die lees van hierdie leksikon, aangesien die betekenisse van woorde en uitdrukkings in Nederlands dikwels ooreenstem met hulle Afrikaanse ekwivalente.

Bibliografie

Algemene bronne

De Schryver, Gilles-Maurice en D.J. Prinsloo. 2000. The Concept of "Simultaneous Feedback": To-wards a New Methodology for Compiling Dictionaries. *Lexikos* 10: 1-31.

Gouws, R.H. 1989. *Leksikografie*. Pretoria/Kaapstad: Academica.

Lakoff, G. 1987. *Women, Fire and Dangerous Things: What Categories Reveal about the Mind*. Chicago: Chicago University Press.

Stam, A. 2000. *De frisheid van eeuwenoude folianten* [aanlyn]. Beskikbaar: <http://oud.refdag.nl/boek/boekgs/001227boekgs02.html>. 27 Desember 2000.

Stronks, Els. 2001. Resensie van *Bijbels lexicon* [aanlyn]. Beskikbaar: <http://www.leidenuniv.nl/host/mnl/tntl/117/117-3/stronks.htm>.

Tarp, Sven. 2000. Theoretical Challenges to Practical Specialised Lexicography. *Lexikos* 10: 189-208.

Van Dalen-Oskam, Karina en Marijke Mooijaart. 2000. *Bijbels lexicon: Woorden en uitdrukkingen uit de bijbel in het Nederlands van nu*. Amsterdam: Prometheus.

Verstraten, Linda P. 1992. *Vaste verbindingen: Een lexicologische studie vanuit cognitief-semantisch perspectief naar fraseologismen in het Nederlands*. Amsterdam: L.P. Verstraten.

Woordeboeke en spreekwoordversamelings

Apeldoorn, C.G.L. en R. van Riet. 1999. *Spreekwoorden verklaard.* Utrecht: Het Spectrum B.V.

De Groot, H. (Hoofred.). 1999. *Van Dale Idioom woordenboek*. Utrecht/Amsterdam: Van Dale Lexi-cografie/The Readers Digest.

Geerts, G. en H. Heestermans in samewerking met C.A. den Boon en E.C.M. Vos. 1992[12]. *Van Dale Groot Woordenboek der Nederlandse Taal*. Etimologie deur P.A.F. van Veen in samewerking met N. van der Sijs. Utrecht/Antwerpen: Van Dale Lexicografie.

Koenen, M.J. en J.B. Drewes. 1977. *Verklarend Handwoordenboek der Nederlandse Taal*. Pretoria/

Kaapstad: Academica.

Kruijssen, A. e.a. (Reds.). 1998[4]. *Huizinga's spreekwoorden en gezegden: Herkomst, verklaring en verge-lijking met Frans, Duits en Engels*. Baarn: Trion.

Odendal, F.F. en R.H. Gouws (Reds.). 2000. *Verklarende Handwoordeboek van die Afrikaanse Taal*. Midrand: Perskor.

Bybelvertalings

Die Bybel: Nuwe vertaling. 1986. Goodwood: Bybelgenootskap van Suid-Afrika.

NBG-vertaling. 1951. *Bijbel met deuterocanonieke of apocriefe boeken*. Vertaling in opdracht van het Nederlands Bijbelgenootschap bewerkt door de daartoe benoemde commissies. Haarlem/Brussel: Nederlands Bijbelgenootschap/Belgisch Bijbelgenootschap, 1989 (eerste druk 1951).

Ander bronne

Boeke

Wolkers, Jan. 1981[28]. *Een roos van vlees*. Amsterdam: Meulenhoff.

Tydskrifte

De Schaepdrijver, **Sophie.** 1999. Stapvoets en rugwaarts de toekomst in. *Nieuw Wereldtijdschrift* 16(8): 52-56, Oktober 1999.

De Swardt, Liezl. 2003. Hoogskop met die blink skoene (redakteursartikel). *rooi rose*, Februarie 2003: 8.

McElvoy, Anne. 1999. De macht van het midden. *Nieuw Wereldtijdschrift* 16(1): 10-15, Februarie 1999.

Meeuse, Piet. 1999. Reproduction interdite. *Nieuw Wereldtijdschrift* 16(3): 48-61, April 1999.

Otten, Christine. 1999. Terug naar de altviool. *Nieuw Wereldtijdschrift* 16(1): 22-31, Februarie 1999.

Pankowski, Marian. 1999. De planeet Auschwitz. *Nieuw Wereldtijdschrift* 16(3): 30-40, April 1999.

Stout, Gerard. 1999. Wrak vee. *Nieuw Wereldtijdschrift* 16(8): 28-37, Oktober 1999.

Van Istendael, Geert. 1999. Een dag in het paradijs. *Nieuw Wereldtijdschrift* 16(8): 10-25, Oktober 1999.

Wacquant, Loic. 1999. De ideologie van de onveiligheid. *Nieuw Wereldtijdschrift* 16(8): 58-65, Oktober 1999.

Henri Béjoint. *Modern Lexicography: An Introduction*. 2000, xii + 276 pp. ISBN 0 19 829951 6. Oxford: Oxford University Press.

This book is a paperback reprint of *Tradition and Innovation in Modern English Dictionaries*, first published 1994, under a new title, *Modern Lexicography: An Introduction*.

It contains an update of the 1994 references, but no new material was added to the main text of this paperback edition. The almost ten years that have elapsed since the first edition, is very briefly covered in the "Preface to the Paperback Edition". A paragraph each is devoted to the following topics: the continuation of traditional practices in the design of new general-purpose dictionaries which were published within this period, the increasingly central role of the computer in dictionary work, and the negative effects computerization is having on lexicography as "industry".

Why the change in title? The author hints at an explanation in the "Preface to the Paperback Edition": Although the old title (*Tradition and Innovation in Modern English Dictionaries)* aptly captures some aspects of the content of the book (a sociolexicography of dictionaries in English-speaking societies), the new title captures in a more suitable way what the book has been used for these last ten years: as a general introduction to lexicography for students and young lexicographers.

The switch in title, I would suggest, could also be seen merely as a switch in focus on different aspects of the existing content of the book: from the language-specific parts (a sociolexicography of dictionaries in English-speaking societies) to those parts focusing on general metalexicographical issues.

One of the strong points of this book has precisely been the fact that the whole, including the presentation of the language-specific data, has a very strong metalexicographically guided conceptualization and structure. For example, various aspects of the general-purpose English dictionary as well as data on dictionaries are set against or analysed in terms of general definitions (e.g. definitions of *dictionary*; cf. Chapter 1), a general dictionary typology (cf. Chapter 1), the (general) functions of the general-purpose dictionary (cf. Chapter 4), and the study of the (general) reference needs and skills of the dictionary user. Each of these topics is well covered from a theoretical perspective, relevant research is discussed, and, where applicable, illustrated.

The outsider (French) perspective, user-friendly writing style, and presentation have made it a very usable textbook.

I do, however, have serious reservations about the way in which "innovations" in the field of (meta)lexicography have been dealt with in this paperback edition, namely by way of a meagre list of about 25 references to cover the period 1994–2000. The list itself reflects very little of the "evolutions" since 1994 which are referred to in the "Preface to the Paperback Edition".

For each of the innovations listed (e.g. corpus lexicography, design of machine-readable dictionaries on CD-Rom, on-line dictionaries, meaning-text type dictionaries, etc.), one can in fact come up with a host of relevant publications from which the author could at least have chosen three or four of the most

important ones and, given that this was indeed the way the publishers decided to update the book (i.e. by means of a reference list), provided the reader with an annotated bibliography.

The question, I suppose, you should ask yourself is: What kind of introduction to (meta)lexicography would you be providing to young students and lexicographers in the field if they are not introduced to these innovative aspects of lexicography? I think you would be providing them with only a partial (and in some respects a somewhat outmoded) overview of the field, and, equally important, you will be leaving out most of what makes (meta)lexicography an exciting field: the fact that there still is (in some fields, at least) even major innovation.

The fact that the publication has not been properly updated, sometimes leads to confusion. A case in point is Chapter 7. With the heading "Whither Lexicography?", it promises the reader a look into the future, and specifically into what influence research on idiomaticity and word meaning might in future have on practical lexicography. What the chapter in fact covers, is research on these two topics up to the time of the publication of the 1994-manuscript. Idiomaticity and word meaning itself, however, have been major areas of research in lexicography during the past number of years, but only a few of the publications listed in the updated references actually reflect these developments. Neither Chapter 7 nor the references themselves therefore give the reader any notion of the developments that have taken place since 1994.

As the text itself has not been revised, the reader will also sometimes come across sentences that are simply untrue. In the discussion of word meaning, for example, the author makes the following statement: "Lakoff's ideas have received little attention in metalexicography so far" (p. 233)! The fact is that, as one of *the* proponents of cognitive semantics, he has had a major influence on a large part of all metalexicographic research on lexical semantics since 1994.

Of course, for whatever purpose it has been used in the past, the text remains useful as handbook. However, the burden of updating the book, and of updating your course materials (should you use the text as handbook for a course in lexicography) is shifted to you as lecturer, and this makes it a less attractive option as handbook.

The decision whether to update a book (and to what extent it should be updated) in a field in which there has indeed been much innovation nearly ten years after its first publication, is undoubtedly not an easy one, given the costs involved. However, I would have been much happier with the end-product as an introductory text if the author had at least added an additional chapter covering the innovations in the field.

Piet H. Swanepoel
Department of Afrikaans and Theory of Literature
University of South Africa (UNISA)
Pretoria
Republic of South Africa
(swaneph@unisa.ac.za)

Monika Elisabeth Bründl. *Lexikalische Dynamik: Kognitiv-linguistische Untersuchungen am englischen Computerwortschatz.* Linguistische Arbeiten 443. 2001, XII + 293 pp. ISBN 3 484 30443 X. Tübingen: Max Niemeyer. Price: €72.00.

Nowhere is the dynamic character of the English lexicon more apparent than in the case of computer vocabulary. Elisabeth Bründl's study investigates this field with two aims in mind (as pointed out in chapter one): firstly, the author wants to find out which principles underlie the creation of words in this field; secondly, her study aims at the integration of cognitive insights into the traditional field of lexicology.

Research in the field of cognitive linguistics as well as its (lack of) application in lexicology and lexicography is reviewed in chapter two. Key concepts are the mental lexicon, metaphor and metonymy as central processes in the creation of new lexical units (as defined by Cruse 1986: 77), and iconicity as the cognitive answer to the structural tenet of the arbitrariness of the linguistic sign. Bründl argues that the cognitive approach so far has been mainly applied to the syntactic-semantic level but rarely to the level of word-formation (p. 8). Yet, the study of word-formation can profit from the integration of cognitive insights. For example, there is arguably a connection between the position of a concept in a conceptual hierarchy and its morphological realisation: the morphological realisation of basic level terms tends to be simple, whereas concepts above or below the basic level (in Rosch et al.'s (1976) sense) tend to be morphologically complex (p. 36). (One might add in parenthesis that this insight was utilized in reverse by Berlin and Kay as early as 1969 to distinguish basic from non-basic colour terms.) The cognitive approach also allows for submorphemic chunks of language to play a role in word-formation (p. 41f.) as long as they can be said to be cognitively salient. It therefore broadens the possibilities of word-formation.

However, the fact that Bründl's call for the integration of cognitive insights into lexicological studies is generally convincing does not mean that one always has to agree with her in detail. For example, her claim that the partial realization of morphemes in blends is iconic because it mirrors the fact that the fused concepts are only present in part as well, may work for a blend such as *Germlish* (p. 52), because this is neither fully German nor English, but it is much less convincing with other blends such as *televangelist* (who is not a hybrid of a television and an evangelist but rather an evangelist who (typically) appears on TV) or *sexploitation*.

Chapter three starts out by discussing lexical processes (word-formation and semantic shift) on the basis of the insights gained in chapter two. Bründl here distinguishes between rule-governed productivity and non-rule-governed creativity, emphasizing that from a cognitive point of view processes which are traditionally seen as creative (e.g. clippings, metaphor) can be seen as productive, i.e. rule-governed (p. 60, 65). The force behind productivity and creativity

alike is analogy, which can work on the formal level (word-formation, i.e. the creation of new lexemes) or the semantic level (semantic transfer, i.e. the creation of new lexical units). Once new lexemes have been created, they can be institutionalised and lexicalised (p. 67f.). Bründl then returns to the issue of the mental lexicon with the aim of integrating the dynamic aspects of the lexicon into that model, discussing suggestions by Meijs and Tournier. One central aspect here is the differentiation of existing and potential lexical items with the latter being governed by lexical processes, which need to be an integral part of a model of the mental lexicon.

In a somewhat surprising turn, the author proceeds by discussing methodological issues such as the character of the lexical material, which constitutes the empirical basis of her study. The designation *corpus*, which Bründl uses for her collection of more than 700 lexemes, is slightly misleading, since she gets her data from secondary sources — a range of general and specialized dictionaries — whereas the term corpus in linguistics usually designates a collection of primary sources, i.e. authentic text (cf. McArthur 1992 s.v. *corpus*). Bründl usefully subdivides the subject field of computing terms into those which belong to the common vocabulary of English and those which belong either to the official technical jargon of computing, *computerese*, or to the less offical *hackers' slang*. The focus of her study is on the first-mentioned class, i.e. those computer terms which have entered the common vocabulary of English. Unfortunately, this focus on predominantly one class of computer terms prevents a comparison of these classes with respect to the lexical processes employed in the creation of new lexical items in each class.

The last section of chapter three returns to more theoretical issues, discussing each lexical process which will feature in her analysis in turn, as well as discussing the "building blocks" which enter into the formation of complex lexemes. In both respects, Bründl goes beyond the limitations of traditional morpheme-based word-formation. For example, she not only considers morphemes as building blocks, but also submorphemic *splinters* such at *emot-* in *emoticon*. Also, following her cognitive orientation she supplements traditional word-formation by semantic and phonetic processes. The main categories of semantic processes are metaphor, metonymy, metaphtonymy (containing both elements of metaphor and metonymy) and shifts in application (the use of an established meaning in a new context without thereby creating a new meaning) (p. 106ff.). Phonetic processes are instrumental in explaining e.g. the formation of *screenager* in analogy to *teenager* (p. 116ff.). Again, while one can easily agree with Bründl's general outline, one need not necessarily agree with the details. For example, she gives the word *careware* as an example of phonetic-stylistic motivation by internal rhyme (*Binnenreim*; p. 116); while this is undeniably true, the more important fact about *careware* is the analogy to *shareware*, which is an example of external rhyme (*Echo-Reim*; p. 117).

The detailed analysis and discussion of Bründl's data with respect to the role of the different lexical processes discussed in the previous chapter makes

up the substance of chapter four. Her statistics show that semantic processes play a role in more than 50% of her data, either alone — as in *mouse* "animal" → "input device" — or in combination with morphological processes — as in *motherboard*. A statistical comparison of her data of computer terms with the frequencies given by Algeo (1991, which, unfortunately, is missing in her bibliography) for English neologisms in general shows that although the ranking of lexical process types is the same (with combining, i.e. compounding and affixation, as the most frequent type followed by semantic shifting), the computer vocabulary contains a much higher number of neologisms based on semantic shift. However, her contention that English computer vocabulary contains approximately 50% of semantically shifted neologisms whereas semantic shift accounts for only 11% of English neologisms in general (p. 145) must be questioned, since she counts neologisms of the combined morphological-semantic type (*motherboard*) as examples of semantic shift whereas it is not quite clear where these are put in Algeo's classification.

Since metaphor is the most important type of semantic process in her data, Bründl proceeds to discuss the role of metaphor in the creation of computer neologisms. She argues in favour of a cognitive approach to metaphor as opposed to a semantic feature approach, taking *spam* as one of her examples. The metaphorical relationship between *spam* meaning 'a type of canned meat' and 'mass advertising by email' is too complex, she argues, to be reduced to a single semantic feature. Instead she sees associations such as 'spam mails "taste" just a bad as the meat' and 'the flooding of mail boxes by spam mail is just as annoying as the spilling of the meat when opening the can' (p. 156). Regardless of the plausibility of these associations, it is clear that it is almost impossible to substantiate them, i.e. to show that it is these associations which lead to the semantic shift and which are foremost in average speakers' minds when they encounter the word *spam*. After all, the fact that specialists in the field try to give a more or less fanciful motivation for *spam* does not automatically imply that the word is motivated for the average computer user (cf. the similarly fanciful motivations for *cookie*, p. 199ff.). Furthermore, an (admittedly non-atomic) semantic feature such as UNPLEASANT JUNK might have served the same purpose; Bründl herself gives *junk e-mail* as a synonym for the computer meaning of *spam* without, however, drawing the obvious parallel to *junk food*.

Following Lakoff and Johnson's (1980) notion of the conceptual metaphor, Bründl identifies 13 cognitive concepts under which she subsumes the majority of metaphors occurring in her data. A central role is played by the concept COMPUTERS ARE HUMAN BEINGS, which results in metaphors such as *guest – host, master – slave* or *peer-to-peer*; also computers partake in human activities such as *read* and *write* and show human characteristics, being *ill-* or *well-behaved*. More unexpected and therefore also more interesting are metaphors of the type COMPUTING IS HANDLING FOOD, which can be seen not only in items such as *byte* (as a homophone of *bite*; "an informational unit consisting of 8 bits") and *cookie* ("a file stored on a computer by a website"), but also in verbs such as *feed (data into*

a computer) or *crunch (data)*. Metaphors in the vocabulary of computing thus turn out not to be an arbitrary and idiosyncratic set; rather, they are governed in the main by generalisable tendencies. Chapter four is completed by a discussion of other semantic processes (metonymy, shifts in application) found in her material.

Bründl herself readily admits that there is a certain amount of subjectivity in her method of analysis (p. 162). Consequently, a number of her analyses admit of a different interpretation. One problem concerned is the differentiation between metaphor and shift in application. For example, she analyses *browse* as a shift in application, by which a general term 'to look through or glance at casually' becomes specialised to mean 'read or survey data files' (p. 194). However, common collocators of *browse* in its non-computing meaning are *book, magazine* or *catalogue*, so *browse* could have been analysed as a metaphor of the A COMPUTER/A COMPUTER-NETWORK IS A BOOK type along with items such a *home page* and *bookmark*. On the other hand, she considers the names of some standard Windows programs such as *calculator* and *clock* as cases of metaphorical shift from the concrete to the virtual domain (p. 187; for some reason she only mentions the German equivalents *Rechner* and *Uhr*); here the analysis as shifts in application would have been more convincing, since the Windows *calculator* is just that, a calculator, and the *clock* is a clock although a virtual one.

In a similar vein, some of her metonymies would have been better analysed as metaphors. Thus, *crack* and *crash* are analysed as metonymies of the type SOUND FOR ACTIVITY (p. 190). However, in their application to the domain of computing, these items could have been analysed perhaps more convincingly as metaphors; *crack* meaning 'break into a computer system' as metaphorical extension of *cracking a safe* and *crash* meaning 'shutdown because of hard- or software malfunctioning' as extension of a *plane crash* (cf. the German equivalent *Absturz*). Surely, metonymy has played a role in the semantic shifts these items underwent from signifying a sound to signifying a process that involves this sound, but this happened in all likelihood at some earlier stage in the semantic history of these lexemes before they were applied to the domain of computing.

The analysis of *crack* and *crash* thus points towards the more general problem of which meaning of a polysemous item is to be taken as the source of a metaphor. Bründl discusses this question using *bomb* as an example (p. 163): starting out from the general meaning 'explosive device' as the source, the meaning in computing 'spectacular hard- or software failure' would be seen as a metaphor. However, there is also a slang meaning of bomb 'absolute failure; fiasco'; taking this as point of departure, the meaning in computing would be a mere shift in application. Bründl handles problems like these by always starting out from what she calls the "dominante, allgemeinsprachliche Primärbedeutung" (p. 163). However, a look at her data shows that finding the one dominant primary meaning, which would then serve as the source of the metaphor, is by no means a straightforward task. For example, Bründl sees a human

service provider (*Dienstleister*) as the source for the metaphorical extension of *server* meaning 'computer which provides access to data, etc. to other computers' (p. 169); as such it fits together nicely with *client*, which signifies the computer making use of the server's services. However, the meaning 'human service provider' is not mentioned at all in a synchronic dictionary such as the *Oxford Advanced Learner's Dictionary* (1995[5]). A common meaning mentioned, however, is 'implement used for putting a portion of food onto sb.'s plate'. Given the fact that Bründl herself assumes a general concept COMPUTING IS HANDLING FOOD, this meaning could easily be seen as the source of the *server* metaphor.

Similarly, Bründl sees the computing meaning of *shell* 'graphical user interface of a program or operating system' as a metaphorical extension of the type COMPUTER PARTS ARE OBJECTS/ARTEFACTS (p. 175); however, especially together with *kernel* 'the inmost layer of an operating system', which Bündl also mentions in this context, *shell* would also fit nicely into the category of FOOD-metaphors; after all, nuts consist of a shell which is around a kernel.

It is thus clear that in the absence of direct historical evidence, which is not always available, the analysis of a metaphor with respect to its source domain can only amount to informed guesswork.

Chapter five evaluates the relevance of cognitive concepts for lexicology on the basis of the analyses described in the previous chapter. Not surprisingly, Bründl argues in favour of the usefulness of these concepts. The computer lexicon is seen as a highly structured web of items, which are connected and motivated by such cognitive factors as metaphor, iconicity, underlying cultural knowledge and cognitive economy. Cognitive aspects provide a deeper understanding of lexicon processes not only of the semantic type (metaphor, metonymy) but also of the morphological type such as compounding, derivation or reduction (clipping, acronymy). While Bründl's conclusions are generally convincing, it has to be added that not all concepts, which she labels as "cognitive", are proprietary to cognitive linguistics. Thus, her "conceptual relations" have been known to structural semanticists in terms of antonymy, synonymy and hyponymy for a long time and Bründl openly admits this (p. 208). What is new here is not so much the insight, but rather the perspective from which it is approached.

In sum, Bründl's study shows the rich cognitive structuring of computer vocabulary and through that the value of the application of cognitive linguistic concepts to lexicology. At the same time, the considerable level of subjectivity inherent in her analyses also points at the price linguistics has to pay by embracing the cognitive approach. On the gradient between the sciences and the arts, the cognitive approach certainly moves linguistics closer to the arts.

References

Berlin, B. and P. Kay. 1969. *Basic Color Terms: Their Universality and Evolotion*. Berkeley: University

of California Press.

Crowther, J. (Ed.). 1995⁵. *Oxford Advanced Learner's Dictionary of Current English*. Oxford: Oxford University Press.

Cruse, D.A. 1986. *Lexical Semantics*. Cambridge: Cambridge University Press.

Lakoff, G. and M. Johnson. 1980. *Metaphors We Live By*. Chicago: University of Chicago.

McArthur, T. 1992. *The Oxford Companion to the English Language*. Oxford: Oxford University Press.

Rosch, E.H., W.D. Gray, D.M. Johnson and P. Boyes-Braem. 1976. Basic Objects in Natural Categories. *Cognitive Psychology* 8: 382-439.

Michael Klotz
Institut für Anglistik und Amerikanistik
Universität Erlangen-Nürnberg
Erlangen
Germany
(mklotz@phil.uni-erlangen.de)

The Dictionary Unit for South African English. *South African Concise Oxford Dictionary.* 2002, xx + 1368 pp. Cape Town: Oxford University Press. Price: R275.00.

The *South African Concise Oxford Dictionary* (henceforth SACOD) is a South African version of the *Concise Oxford Dictionary*, the first time that this particular hybrid has been prepared. It is testimony to the enduring success of the work of the Dictionary Unit for South African English at Rhodes University, headed by teams that included Jean and William Branford in the 1970s, Penny Silva in the 1990s and now, Kathryn Kavanagh. The lexicographical work from the unit saw the publication of four editions of the *Dictionary of Southern African English* (1978, 1980, 1987, 1991), a *South African Pocket Oxford Dictionary* (SAPOD) and the *Dictionary of South African English on Historical Principles* (DOSAEHP) (1995). SACOD differs from the rest in several ways. It is larger in scope than SAPOD, smaller than DOSAEHP, and unlike DOSAE and DOSAEHP, does not deal with South African words alone. Based on the 10th edition of the *Concise Oxford Dictionary* SACOD has excised some words from the parent, whilst adding many new words of general English as well as of South Africa.

In its layout and content SACOD resembles the Oxford Clarendon prototype. Using my copy of the 7th edition of the latter (1991) as a point of comparison, SACOD is thicker by 110 pages, and each page is longer and broader by a centimetre each way. It also shows some technical advances from the 1990s that are made possible mostly by the improved computer technology. SACOD uses the IPA system for the pronunciation of headwords, rather than the outdated (if less daunting to the novice) system of Roman orthography with modifications. There is even a table for the click sounds and their symbols. The computer age enables the use of a 'footer' containing the IPA symbols and their values on every page, with consonants on two facing pages alternating with vowels on the next two. SACOD also has little panels, covering points of usage (e.g. a box on *few* vs *less*) and word formation (e.g. a box on *ptero-*). These are quite useful and, I would argue, could have been used more frequently. The computer age makes its mark in another way; the way it dominates the new words recorded. As can be expected, words like *flame, spam, stiffy* (a South Africanism), *firewall, download, URL,* and *HTML* are very much part of the word-hoard recorded in SACOD. On a point of etymology, we learn from the entry on *spam* that it probably comes from a Monty Python sketch set in a café in which every item on the menu includes sp(iced h)am. It is also good to see topical words included like the use of *sledging* in cricket, acronyms like *FAQ* and informal truncations like *ex-con,* and even *ex-* as a noun meaning 'a former husband or wife'. I remember the effectivenesss of a recent women's magazine's headline debating the merits of 'Sex with your ex'.

It is quite clear then that SACOD has all the benefits one expects from the *Concise Oxford Dictionary.* Are these matched by the South African interpolations, one might wish to ask? In reviews of previous publications from the Dic-

tionary Unit for South African English, I have suggested that among their few weaknesses were a disproportionate amount of vocabulary from Afrikaans, especially a failure to exclude code-switching from established borrowings, and a tendency to accept folk etymologies for items of slang and informal words, especially those drawing on languages other than Afrikaans. These remarks are not generally applicable to SACOD, which discusses the code-switching/borrowing dilemma early on in its preface, and which is not faced with a plethora of etymologies for obscure words, since its selection of South Africanisms is small in comparison to DOSAE and especially DOSAEHP.

South Africanisms have become accepted in international English from the days of exploration and colonisation — the *trek – veld – laager* semantic field is well known. Later it was the spectacular attempt at social engineering that brought lexicographical fame: *apartheid, dompas, exit permit*, etc. The post-apartheid era makes its impact in a smaller way in SACOD with terms like *imbizo, lekgotla* and *African Renaissance*. Every now and then our dictionaries still give occasion for surprise in announcing the South Africanness of terms like *slasto* (a brand name), *slave bell* and *dam*. The entry for *dam* works rather well for SACOD, since its main sense is 'downloaded' from the *Concise Oxford Dictionary* (the barrier constructed to hold back water), while a second sense (the body of water itself) is marked as 'South African'. A complex case is the lack of a label specifying the status of *stop street,* which is not generally familiar to overseas visitors I encounter. Moreover, it is not listed in the *Concise Oxford Dictionary* (7th edition), nor in the *Encarta World English Dictionary.* Kathryn Kavanagh (personal communication) informs me that the term is, however, attested in the U.S.A. and Canada, thanks to a Google search. The question remains whether such terms localised to South Africa and one other territory should somehow be flagged.

Inevitably with a work of this scope some problems arise. One that may not have been foreseen by the editors is a clash of semantics of the many words arising ultimately from the Indian subcontinent. Whose pronunciations and meanings should the dictionary record: that of the British in the days of Raj, those of Indian/Pakistani/Bangla Deshi/Sri Lankan immigrants to 20th century Britain, or (more crucially for SACOD) those of Indians in South Africa? The latter now mostly speak English as a mother tongue, but have retained a rich vocabulary from India forged from personal and group circumstances in South Africa. Thus *bhaji* is given in SACOD as an equivalent to *bhajia* glossed as 'a small flat cake or ball of vegetables, fried in batter'. If correct, this must be a Britishism, as in South Africa the two terms are distinct. That is, whilst *bhaji* is a term for 'leaves of plants, herbs, cresses' that are cooked, *bhajia* is the small flat cake referred to, also anglicised in South Africa as *chilli bite*. Similarly the meaning of *dosa* 'a pancake made from rice flour' is the usual one in India and Britain, where the unsweetened pancake is stuffed with a curried vegetable filling. In South Africa however, *dosa* (pronounced [dosɛ]) is usually a sweet pancake, with no filling. Finally, SACOD gives the Indian and British form *dhal* or

dal for 'split lentils, split pulses, a soup made from these'. In South Africa the usual spelling is *dol* (pronounced [dɔ:l]), possibly on account of the raising and rounding of the [a:] vowel by white South Africans. Perhaps the same tactic as for *dam* would have been appropriate here, with the more widely spread forms given first, and the different South African sense or pronunciation given second.

In respect of etymologies and meanings, SACOD is of the highest quality, though I have a few quibbles. Fanakalo was not developed on the mines, as the entry for this headword suggests. It existed in a clearcut form decades before the mines were founded in Natal and to a lesser extent, the eastern Cape (Mesthrie 1989). The etymology of *larney* is given as 'from Tsotsitaal *lani(e)* 'white man', perhaps related to Malay *rani* 'rich'. A more plausible etymology, as I argued in Mesthrie (1997), was that the Tsotsitaal form was ultimately based on Cape vernacular Afrikaans *oolana* 'Hollander', which spread up north in the era of the gold and diamond rush. The entry for *bhuti* ('brother') correctly identifies it as taken from Xhosa, but for reasons unknown the fact that the term is based on Afrikaans *boetie* (as mentioned in DOSAEHP) is not stated. *Peri-peri* is given as a variant spelling of *piripiri*, which is traced to the Ronga language of Mozambique. In the South African context, I think the usual form is *peri-peri* (the headword in DOSAEHP), and while the link with Mozambique is correct, it may be optimistic to tie it to a particular language of the area. Is it definitely Ronga, but not the closely related Tonga and Tsonga? (In South Africa the term occurs as *pelepele* 'pepper' in Zulu). The word must be linked to the spice trading along the East African coast (*piripiri* or *pilipili* in Swahili), and prior to that the Arabic word *phil-phil*, ultimately based on Sanskrit *pipar,* the word for 'pepper'. Mozambiquan Portuguese is the most likely link between all of these and the South African term.

To conclude: SACOD will prove a most valuable resource for students and the general public in South Africa. The editorial team in Grahamstown and the publishers in Cape Town are to be congratulated on another splendid piece of lexicography from their joint stable.

References

Mesthrie, R. 1989. The origins of Fanagalo. *Journal of Pidgin and Creole Languages* 4, 2: 211-40.
Mesthrie, R. 1997. Lexicography from below. *Studia Anglica Posnaniensa* 31: 153-62.

Rajend Mesthrie
Linguistics Section
Department of English
University of Cape Town
Cape Town
Republic of South Africa
(raj@humanities.uct.ac.za)

Wilfrid H.G. Haacke and Eliphas Eiseb. *A Khoekhoegowab Dictionary with an English–Khoekhoegowab Index.* 2002, xiv + 740 pp. ISBN 99916-0-401-4. Windhoek: Gamsberg Macmillan. Price: N\$175.00 (Hard-cover), N\$150.00 (Soft-cover).

Having dedicated more than 20 years to this single major lexicographic project, the authors, Professor Wilfrid H.G. Haacke and Pastor Eliphas Eiseb, have produced a dictionary of amazing quantity and quality. The newly published *Khoekhoegowab Dictionary* comprises some 25 000 Khoekhoegowab entries and the English renderings in the English–Khoekhoegowab index amount to over 26 000 entries. While these figures are quite impressive in themselves, the reader will also find a thorough analysis of the Khoekhoegowab entries as well as refined English translations. The dictionary attempts to address the interests of linguists as well as the needs of the speech community. Unavoidably, the divergent interests of these two groups force the authors to accept a number of compromises that result in some inconvenience for both the target audiences. Despite these limitations, linguistic research on African languages will find the dictionary a source of eminent importance for both historical linguistics as well as typological studies. And mother-tongue speakers will greatly benefit from this publication, which, thanks to financial support from the German technical assistance agency GTZ, is available at a price that is affordable for the Namibian public.

The language documented in the dictionary is referred to as Khoekhoegowab. While quite a number of Namibians identify themselves and their language as Nama, Damara or Haiǁom, it seems, very few, if any, refer to themselves as Khoekhoen or speakers of Khoekhoegowab. The dictionary does not explain why "Khoekhoegowab" is used for the language of these three groups of people; it simply adds "revived authentic name" to the rendering under the entry *Khoekhoegowab* (Haacke and Eiseb 2002: 72). The reason for using "Khoekhoegowab", however, was provided in a lecture, which Professor Haacke delivered at the University of Namibia in the year 2000 (printed in 2002). There he states (Haacke 2002: 12):

> The word Khoekhoegowab … means Khoekhoe-language. This glossonym has, after Namibia's independence, been reintroduced as formal reference to the so-called Nama/Damara language at the initiative of my long-time colleague, Pastor Eliphas Eiseb, so as to avoid the dichotomy of Nama/Damara. … this term (Khoekhoegowab) is not an artificial creation but was the original name for the language, as already recorded by Jan van Riebeeck in January 1653.

With this new comprehensive dictionary, Khoekhoegowab is now by far the best-documented Khoesaan language. Only time will tell whether this more correct, but rather bulky language name, Khoekhoegowab, will enter into everyday usage, replacing Nama, Damara and Haiǁom.

The target users of the *Khoekhoegowab Dictionary* will be readers with wide-

ly different backgrounds and interests: either native speakers from Namibia, Botswana and the Republic of South Africa, or professional linguists of diverse nationalities. The only potential reader combining both characteristics is an exceptional case: Levi Namaseb, a Khoekhoe-speaking linguist, who teaches his language at the University of Namibia. Many of the Khoekhoegowab linguistic terms in the dictionary go directly back to him and their usage in his Khoekhoegowab courses. As a general principle, only notions from mother-tongue speakers have been considered, as the authors themselves "did not attempt to coin Khoekhoegowab words for English concepts, where such words are not already in existence" (Haacke and Eiseb 2002: iii).

In order to serve the needs of the Khoekhoegowab speech community, the dictionary has to be accessible to non-professional dictionary users. For this reason the authors decided to use the officially recognized standardised orthography of Khoekhoegowab.

Unfortunately, the conventions of the "sole official orthography for Khoekhoegowab" (Curriculum Committee for Khoekhoegowab 2002: xiv) do not always comply with international practice. The most problematic deviation in this respect is the use of b, d and g for lower tone melodies, in contrast to p, t and k for the higher melodies. Another confusion arises as a result of the representation of the clicks. In contrast to international conventions, a simple click symbol does not represent a simple click, but instead a click with glottal accompaniment. While a click followed by g otherwise refers to a voiced click, in the orthography of the *Khoekhoegowab Dictionary* it stands for a voiceless click.

Also to meet the requirements of linguistic researchers as well as of persons wanting to learn the language, the authors considered some modifications of the official orthography necessary. Tone marks were added and the "long" vowels are no longer written with a macron, but as double vowels instead. In order to mark nasalization, the standard orthography tops the vowel with a circumflex. In the modified form, the tiny circumflex is inserted between the vowels to allow for separate tone marks of the vowels. Being professionals in the field of linguistics, these scholars can however be expected to be able to adapt to the official orthography. In using the Khoekhoegowab lexical entries from the dictionary in comparative research, one has therefore to adjust them to international phonetic standards in order to avoid confusion.

The decision in favour of the official Khoekhoegowab orthography used in formal education from the elementary school up to university level was made for the benefit of mother-tongue users. The use of the dictionary by Khoekhoegowab speakers requires the coverage of semantic areas essential to everyday life. The dictionary provides Khoekhoegowab entries in areas such as transport (e.g. *car, car-wash, fuel, traffic circle* and *zebra crossing*), communication (e.g. *fax* and *telephone*), disease and its treatment (e.g. *AIDS, malaria, TB, vaccination* and *x-ray*). Dictionaries, like all books, also reflect the personal interests of the authors. Three semantic fields seem to have received special attention. Numerous entries from the religious context, such as *baptism, Bible, faith,* etc. may possibly

go back to Pastor Eiseb, as well as the large number of names for plants. Furthermore, Professor Haacke appears not to hide his passion for highly abstract Khoekhoegowab terms for phonetic and tonological concepts, for example, the entries of *vowel*, *toneme*, *tone melody*, etc. where the different kinds are listed. All in all, the wide coverage, also of modern aspects of life, makes the dictionary a useful source of reference for native speakers, be they interested laypersons or teachers of Khoekhoegowab.

Scholars will admire the richness of phonetic, lexical and structural information. Impressive is the number of compounds, which follow the main entries, along with other variations. The meanings have been analysed in great detail and even the usage of a given term is mentioned.

In 1999, the same authors published a *Khoekhoegowab–English/English–Khoekhoegowab Glossary* (for a review see Snyman 2000). This glossary, with approximately 10 000 basic entries and exclusive application of the standard orthography, aims to respond to the needs of the Khoekhoegowab community. With this former publication being available, and given the minor orthographic changes discussed above, one might simply omit the repeated entries in a revised edition of the dictionary. The forms in the standard spelling follow each Khoekhoegowab entry in parentheses { }, the main difference in the two forms being the omission of the tone in the standard orthography. Even non-linguists can easily derive the standard form from the main entries. The user of the dictionary is severely hindered by these redundant 25 000 entries.

A statement in the introduction, indicating that the authors consider Khoekhoegowab an "endangered language", at first comes as a surprise. The phrase "this last surviving language of the Central Khoesaan family" (Haacke and Eiseb 2002: iii) merits correction, as several other Central Khoesaan languages, such as Khwe and Naro are still spoken today. In fact, after Oshiwambo, with 713 919 speakers, Khoekhoegowab is the second largest speech community of Namibia, with 175 554 speakers (cf. census figures of 1991 by language spoken, Haacke 1996: 32). The results of several surveys reveal, however, that Nama and Damara seem not positively disposed to the formal use of their language. Khoekhoegowab speakers are reluctant to send their children to schools in which their language is employed as medium of instruction (Laurentius Davids, personal communication). Numerous instances from all parts of the world demonstrate that even languages spoken by large communities may indeed disappear, if abandoned by their speakers.

Up to now most research on African languages has been carried out by scholars from the USA, Europe and Japan. The authors of the most important previously available dictionaries of Khoekhoegowab were also foreigners, more precisely German missionaries: Johann Georg Kroenlein (1889) and Friedrich Rust (1969), who based his work on that of the former. With the present dictionary, the situation has obviously changed, as both scholars are born Namibians.

With the completion of this dictionary, which is a milestone in the field of

Khoesaan linguistics, Professor Haacke and Pastor Eiseb's work will occupy a prominent position in the study and teaching of African languages abroad. This new Khoekhoegowab dictionary may help to preserve the wealth of the language and contribute to efforts of the speakers of Khoekhoegowab to find a meaningful role for their language in modern life and in the future.

References

Curriculum Committee for Khoekhoegowab. 2003. *Khoekhoegowab, Orthography 3.* Windhoek: Gamsberg-Macmillian.

Haacke, Wilfrid H.G. and Eliphas Eiseb. 1999. *Khoekhoegowab–English/English–Khoekhoegowab Glossary/Mîdi Saogub*. Windhoek: Gamsberg-Macmillian.

Haacke, Wilfrid H.G. and Eliphas Eiseb. 2002. *A Khoekhoegowab Dictionary with an English–Khoekhoegowab Index*. Windhoek: Gamsberg-Macmillian.

Haacke, Wilfrid H.G. 1996. *Language and Literature in Africa.* Study guide for Core Module HSL 0111, part 1, Department of African Languages, University of Namibia.

Haacke, Wilfrid H.G. 2002. *Linguistic Evidence in the Study of Origins: The Case of the Namibian Khoekhoe-speakers.* Windhoek: University of Namibia.

Kroenlein, Johann Georg. 1889. *Wortschatz der Khoi-khoin (Namaqua-Hottentotten).* Berlin: Deutsche Kolonialgesellschaft.

Rust, Friedrich. 1969. *Nama Wörterbuch.* Pietermaritzburg: University of Natal Press.

Snyman, J.W. 2000. Review: Haacke, Wilfrid H.G. and Eliphas Eiseb. 1999. *Khoekhoegowab–English/English–Khoekhoegowab Glossary/Mîdi Saogub*. Windhoek: Gamsberg-Macmillian. *Lexikos* 10: 331-333.

Matthias Brenzinger
Institut für Afrikanistik
University of Cologne
Cologne
Germany
(Matthias.Brenzinger@uni-koeln.de)

Louis Molamu. *Tsotsitaal: A Dictionary of the Language of Sophiatown.* 2003, xxix + 130 pp. ISBN 1 86888 187 3. Pretoria: University of South Africa.

The compiler of this dictionary is perhaps the first and only person to have attempted to produce a dictionary of this 'language' or register, and he should be complemented on this. This is a dictionary in a class of its own, i.e. a specialised dictionary written to capture the vocabulary of Tsotsitaal, a language which may be regarded as 'obsolete' (or at least 'endangered') due to its nature and the versatility associated with it. It is gradually deteriorating and disappearing while changing from the original language of Sophiatown into various regional versions corrupted by the various other languages of South Africa. The compiler does not seem to have been motivated by knowledge of or experience in dictionary writing, but by a desire to preserve the vocabulary of the language he loves, and to share his knowledge with future generations who may have been unfortunate not to have encountered the lifestyle of Sophiatown.

It will be unfair of one to expect a dictionary of a language such as this to conform to all or most of the requirements for dictionary writing. And therefore, even though the dictionary has important features such as front and back matter, and micro- and macrostructural elements which are basic to dictionary writing, these are presented in a form suitable for a language:

— which has no proper and recognised vocabulary,

— which has no standard vocabulary and status,

— which is very much informal,

— which has no recorded literature,

— which is mostly restricted to the youth, especially gangsters,

— which is still little known and insufficiently researched, and

— which most of its users wanted to remain unknown to the general public.

The front matter includes most relevant data with even a detailed list of acknowledgements especially to those people who grew up with the author in Sophiatown and Johannesburg, and who contributed to the compilation of this dictionary specifically by supplying oral and written material. In the Introduction the author clearly outlines how the dictionary was compiled and what micro- and macrostructural elements the user can expect in it.

The paragraph entitled 'Sources of Information' gives a detailed description of the method the author used to gather information for the dictionary. This practice is usually not common in dictionary compilation, but in literary and linguistic research. This is due to the fact that the vocabulary of Tsotsitaal

is still very much a subject under research since it has so far only scantily been documented. The only documents available are those dealing with its usage and origin, and the influence it has had on the standard languages of South Africa. None of these contain a detailed vocabulary like the one that can be found in this dictionary. Another difficulty the compiler possibly had to face is the lack of any formal or informal orthography for this language. Researchers writing about this language utilise any spelling to try to render the pronunciation they hear the Tsotsitaal-speaking community use in a specific area. For the purpose of this dictionary, the author seems to have gathered data by visiting shebeens where most Tsotsitaal speakers usually gather, freely and informally conversing while drinking. For instance, the author (Molamu 2003: xv) says:

> The plethora of alcohol-related terms, phrases and expressions — including the often flowery names of the local drinking establishments — form an essential part of the lingua franca.

Besides shebeens, material for the dictionary was seemingly recorded at gatherings such as weddings, parties, funerals and other ceremonies where people meet and chat freely and informally.

What complicated the compilation of a dictionary of this type is the fact that this 'language' is changing continuously, and some of the terminology used refer to different, and sometimes conflicting concepts, depending on the environment, and the speech community in which it is used. For this reason, there are many terms in this dictionary which one may regard as having incomplete definitions.

Tsotsitaal is often ambiguous because in most cases the proper word adopted from one language is given an additional meaning, or its meaning is corrupted to refer to a related action or concept. Therefore, not every person may define the same word in the same way. For instance, definitions of words such as the following seem to be incomplete:

— **bind** (page 10). It is defined as follows: 'The term refers to the act of parking taxis at taxi-ranks.' To my knowledge, the term also refers to 'the incident when the taxis do not have enough passengers to transport, and, since the queue is not moving, most of them just remain parked'.

— **borg** (page 12). The definition reads: 'The term refers to money required as security against the temporary release of an offender pending trial; bail; surety.' As far as I know, the term also has the Tsotsitaal meaning of 'favour'.

One of the characteristics of Tsotsitaal is the many variants and synonyms, especially for those objects and concepts which were intended to be kept secret. This results in much cross referencing by the compiler to avoid reduplication of definitions in the dictionary, e.g. *boeta* (page 11), *bok* (page 12), *bombela* (page 12) and *boomba* (page 12) are referred to *boeda* (page 11), *skwiel* (page 96), *mbombela*

(page 65) and *fatty boom-boom* (page 31) respectively, where the definitions are given. The compiler does not, however, explain in the front matter of the dictionary where the use of cross-references is discussed, why certain words are preferred as lemmata for carrying the definitions. Should it be because they are the most commonly used variants and synonyms? This important information is not communicated to the user. If, however, there is no conclusive reason why the definitions should be given at a certain variant or synonym, the term which alphabetically comes first may be defined fully, and those that follow may be referred to the first where the required meaning can be found.

A few times lemmata in the dictionary appear in the wrong alphabetical order, e.g. on page 16, the lemmata *Cashbah* should be preceded by *Casbah Kids* and followed by *cashbar*, and on pages 65 and 108 respectively the lemma *matopana* should precede *matshingi*, and the lemma *Tshangi* should follow *tshama hansi*. These, however, are small inconsistencies compared to those that occur when terms are qualified and defined. While each and every page has more or less the same number of terms with the same inconsistencies, the following examples are all drawn from page 16.

The first irregularity concerns the indication of the part of speech to which each term belongs. Of the 21 terms on page 16, only eight are qualified, viz. the nouns or noun phrases *Cabin in the Sky*, *can't-gets*, *chandies*, *chanster* and *chara*, the verb *caza* and the adjective *chalkstripe*. Those left unqualified are: *cable*, *cancer stick*, *cards-up*, *Casanova*, *Casbah Kids*, *Cashbah*, *cashbar*, *casino*, *Cavalla*, *cazi*, *chacharag*, *chaile*, *chaisa* and *change marobalo*. One of these, viz. *cazi*, is cross-referred to *caza* which is qualified.

The second irregularity can be found in the form of the definitions. Of the 21 terms on page 16, thirteen are nouns, viz. *Cabin in the Sky*, *cancer stick*, *can't-gets*, *Casanova*, *Casbah Kids*, *Cashbah*, *cashbar*, *casino*, *Cavalla*, *chaile*, *chandies*, *chanster* and *chara*, six verbs, viz. *cable*, *cards-up*, *caza*, *cazi*, *chaisa* and *change marobalo* and the two adjectives, viz. *chacharag* and *chalkstripe*. For nine of these terms traditionally formulated definitions have been used, e.g.

caza/v. To impress.

chanster/n. A bully.

The definitions of the others when nouns, often start with statements such as 'The name of ..', 'An expression used to refer to …', 'The term referred to …', 'A term used for …', 'A pejorative term for …', and when adjectives, sometimes with a formulation such as 'The term describes …'. These two different ways of defining can sometimes be found in the same lemma, e.g.

chaisa/ To finish work. The term also means 'to collide'.

When the parts of speech to which terms belong are missing, and the wording of the definitions do not indicate the parts of speech, the way terms should be used can be obscured, especially in the absence of elucidatory example senten-

ces.

Because of the inconsistent spelling used in the dictionary, it is not always clear how some of the terms are pronounced. Phonetic transcriptions to clarify the pronunciation would have been a great help to the user who is unfamiliar with Tsotsitaal, or with some of the languages from which the vocabulary of Tsotsitaal has been derived.

In spite of the criticisms raised in the previous paragraphs, the dictionary should be recommended to anybody interested in the characteristics and origins of Tsotsitaal. It is not a simple task to compile a work of this nature of a language of which the orthography is often uncertain and the vocabulary little documented. This is an important attempt to outline the concept Tsotsitaal.

V.M. Mojela
Sesotho sa Leboa National Lexicographic Unit
University of the North
Polokwane
Republic of South Africa
(mojelav@unorth.ac.za)

Dalene Müller. *Skryf Afrikaans van A tot Z.* 2003, 718 pp. ISBN 1 86890 037 1. Kaapstad: Pharos Woordeboeke.

Vanuit die ivoortorings van Afrikaans word in suiwer, akademiese taal oor taalkwessies en die voortbestaan van Afrikaans besin, maar intussen is daar 'n geslag leerders en studente wie se taalvermoëns met rasse skrede aan die kwyn is. Oor die afgelope jare word daar voortdurend in die kolomme van koerante hieroor geskryf. In die lig hiervan is die verskyning van *Skryf Afrikaans van A tot Z* (*SAAZ*), 'n praktiese skryfgids oor bykans elke denkbare aspek van taalgebruik en -versorging, nie alleen aktueel nie, maar noodsaaklik.

Dalene Müller, gesoute taalpraktisyn van *Die Burger*, het dié gids oor 'n periode van tien jaar saamgestel met die hulp van 'n groot groep medewerkers uit veral die mediawêreld. Dit is aanvanklik bedoel as handleiding vir taalgebruikers in die mediabedryf (joernaliste, sub- of teksredakteurs, vertalers, proeflesers, rubriekskrywers, korrespondente en ander medewerkers), maar danksy haar breë ervaring op ander terreine kon sy die behoeftes van algemene taalgebruikers (uitgewersredakteurs, reklamekopieskrywers, setters, dosente, studente, onderwysers en leerders) ook identifiseer en in ag neem.

SAAZ is in ooreenstemming met en aanvullend by die 2002-uitgawe van die *Afrikaanse Woordelys en Spelreëls* (*AWS*), maar omdat dit in die eerste instansie as huisgids vir joernaliste moes dien en 'n eenvormige styl onder Afrikaanse koerante en tydskrifte in die Media24-groep van Naspers moes bevorder, is daar dikwels 'n keuse gedoen uit wisselvorme in die *AWS.* Hiermee word riglyne, eerder as voorskrifte (aldus die voorwoord) aan skrywers gegee.

Die waarde van *SAAZ* lê in die wye verskeidenheid gegewens wat in toeganklike, nieakademiese taal in twee afdelings vervat is. Deel I bestaan uit sowat 5 000 lemmas wat alfabeties gerangskik is om die vindbaarheid daarvan te vergemaklik. In die artikels word uitsluitsel gegee oor taalkwessies waaroor die meeste taalgebruikers twyfel. 'n Magdom inligtingstipes kom aan bod, bv. betekenis (sien **andersins x andersyds**), konnotasie (sien **toedig**), gebruik (sien **epidemie x pandemie**), kollokasies (sien **aandag**), woordsoorte (sien **verkeerd**), morfologie (sien **roman**), geslag (sien **doyen(ne)**), sinonieme (sien **almanak x kalender**), verwarrende woorde (sien **bygaande x meegaande** en **kwes x kwets**), leestekens (sien **ensovoorts**), skryftekens (sien **elite**), skryfwyse (sien **agterna**), spelling (sien **Xhosa**), afkortings (sien **CD**), register (sien **destyds x toenmalig x toentertyd**), vertaalekwivalente (sien **alledaags x daagliks**), idiome (sien **haak x hak**), clichés (sien **snel**), styl (sien **skool**) en nog vele meer.

Alhoewel *SAAZ* merendeels goeie, bruikbare taalleiding gee, is daar tog riglyne ten opsigte van die juistheid of wenslikheid van woorde en uitdrukkings wat nie ooreenstem met dit wat gesaghebbende woordeboeke erken nie. Volgens *SAAZ* is die gebruik van die voorsetsel *by* saam met *aandoen* verkeerd, terwyl die *Handwoordeboek van die Afrikaanse Taal* (*HAT*) die voorsetsel as opsioneel aandui. Terwyl *SAAZ* sê dat *pamperlang* nie in Afrikaans "vertroetel, verwen of bederf" kan beteken nie, word hierdie betekenisonderskeiding wel in

HAT opgeneem. By **paddastoel** in *SAAZ* word die volgende riglyn gegee: "Ge-boue het soos paddastoele verrys [NIE: opgeskiet]". Volgens *HAT* is sowel *ver-rys* as *opskiet* aanvaarbaar. *SAAZ* beskou die gebruik van *robot* vir "verkeerslig" as foutief, maar *HAT* bied by die lemma **robot** die volgende as primêre beteke-nisonderskeiding aan: "Toestel wat deur wisselende ligte die verkeer reël" en gee *verkeerslig* as sinoniem. Die uitdrukking *jou eie beuel blaas* is volgens *SAAZ* taboe, terwyl *HAT* dit ongekwalifiseerd opneem.

SAAZ gee nie oral ewe ondubbelsinnig uitsluitsel oor register nie en mislei daardeur die taalgebruiker. By **poog** verskyn tereg die volgende riglyn: "Ver-vang liefs dié deftige woord deur *probeer*." Hierteenoor is die inskrywing by **jok, lieg** onvoldoende. Dit lui soos volg: "Dié wwe. is soms verkieslik bo die lang uitdrukkings *'n leuen vertel*, *leuens verkondig/versprei* of *stories verkoop*." Daar word nagelaat om te sê dat *jok* en *lieg* minder formeel as die langer uit-drukkings is en dat hulle ook nie gelykwaardig is wat register betref nie. *Jok* kan versagtend vir die krasser *lieg* gebruik word. By **mal** word heel korrek aan-gedui dat dit as sinoniem vir *waansinnig* of *versteurd* 'n kwetsende woord is, maar by **gay** is geen kwalifikasie wat waarsku dat dié benaming as seksisties ervaar kan word nie.

Dit is nie duidelik wat *SAAZ* se beleid ten opsigte van die hantering van homonimiese lemmas in Deel I is nie. Die homonieme *geen* (s.nw.) en *geen* (vnw. of telw.) word afsonderlik gelemmatiseer en met verhewe syfers van mekaar geskei. In die artikel van **ry**, daarenteen, kom inligting voor wat na regte onder twee aparte homonimiese inskrywings tuishoort. Die gegewens wat op *ry* (ww.) slaan, hoort nie in dieselfde artikel as dié wat op *ry* (s.nw.) betrekking het nie.

Kruisverwysing tussen lemmas in Deel I van *SAAZ* geskied deur middel van die instruksie "Sien (ook)" of 'n pyl met daarnaas die woord waarna ver-wys word. Ongelukkig is SAAZ nie altyd konsekwent wat die gebruik van kruisverwysings betref nie en daardeur gaan waardevolle inligting vir die gebruiker verlore. In die artikel van **angstig x gretig** word die betekenisverskil tussen *angstig* en *gretig* verduidelik. **Gretig** is op sy alfabetiese plek gelemmati-seer met 'n kruisverwysing na **angstig x gretig** waar die taalgebruiker die nodige inligting kan bekom. Dit is goeie praktyk. By **weelderig x welig** word die onderskeid tussen *weelderig* en *welig* verduidelik. *Welig* is egter nie gelem-matiseer nie. Dit skep 'n probleem vir die taalgebruiker wat oor die betekenis of gebruik van *welig* twyfel en nie weet dat hy dit onder **weelderig x welig** moet soek nie. Vir die gebruiker wat besonderhede oor *welig* soek, gaan die verband met *weelderig* boonop verlore. Dieselfde probleem geld ook die inskry-wing **dis x gereg**. By **dis x gereg** word die sinonimiese verband tussen *dis* en *gereg* verduidelik, maar *gereg* is nie gelemmatiseer met 'n kruisverwysing na **dis x gereg** nie. By **dis x gereg** word daar verder van 'n dis en 'n gereg die vol-gende gesê : "dit word … voorgesit, opgedien of opgeskep [EERDER AS: bedien]". Dit is 'n herhaling van gegewens wat in die artikel van die lemma **be-dien x dien** voorkom. 'n Meer ekonomiese oplossing sou wees om by **dis x**

gereg 'n kruisverwysing na **bedien x dien** te maak. By **berig** word die gebrui-
ker verwys na **berig x artikel** wat nie as lemma verskyn nie, in plaas van na
artikel x berig wat wel ingesluit is. By **sulks** word 'n kruisverwysing na **so**
gegee, terwyl die inligting waarna die gebruiker soek, in die artikel van **soda-
nig** voorkom.

Skakeling tussen Deel I en Deel II word bewerkstellig deurdat daar by
lemmas in Deel I ook kruisverwysings na artikels in Deel II voorkom. Dit stel
die gebruiker in staat om meer inligting vinniger te ontsluit.

Die ensiklopediese Deel II bestaan uit 26 afdelings waarin, danksy die
loopkoppe en duimindeks, maklik nageslaan kan word. Ná die opskrif van
elke afdeling verskyn boonop 'n volledige lys van die artikels en die nommers
daarvan in die betrokke afdeling. Hiermee word die toeganklikheid van elke
afdeling verder verhoog.

In geen ander soortgelyke Afrikaanse bron is al so divers en uitgebreid oor
aspekte van Afrikaanssprekendes se taal- en leefwêreld verslag gedoen nie. Die
gegewens wat in Deel II saamgevat is, sluit onder meer in: die juiste afbreking
van woorde in die Nguni- en Sotho-tale, brei- en hekelafkortings, web- en e-
posadresse, die name van die heilige dae van Christene, Jode, Hindoes én Mos-
lems, persoonsname vir die bewoners van lande of plekke, die sewe wonders
van die Ou Wêreld, naamsveranderinge van damme, kanale, lughawes, ens.,
persoonseiename waarvan die spelling probleme oplewer, proefleestekens, die
Afrikaanse vertaling van die Volkslied, die name van bekende produkte en
maatskappye wat dikwels verkeerd gespel word, die skryfwyse van sporttel-
lings, die korrekte weergawe van telefoonnommers, die transliterasie van bui-
telandse eiename, die korrekte Afrikaanse benamings vir die verskillende voer-
tuigsoorte, bronverwysings, algemene riglyne by die gebruik van leenwoorde
in Afrikaans, die name van wynkultivars en spesiale name vir wynbottels van
spesifieke groottes.

Die diagramme, tabelle, kaarte en ander grafika in Deel II is oor die alge-
meen funksioneel en ter sake. Die diagramme wat byvoorbeeld in die afdeling
oor sport gebruik word om die indeling van die baan of veld en die speelposi-
sies te illustreer, dra daartoe by dat 'n groot hoeveelheid inligting ekonomies en
ondubbelsinnig oorgedra word. Vir die sportskrywer, afrigter of speler wat oor
'n wedstryd verslag moet doen, word die korrekte Afrikaanse terminologie
daarmee binne die breër konteks van die spel geplaas.

In Deel II word plek-plek swart raampies gebruik om waardevolle skryf-
inligting uit te lig. Dit help die gebruiker om riglyne raak te lees wat andersins
in die aaneenlopende teks sou verdwyn het. Die Afrikaanse ekvivalent vir
"Greater Durban", naamlik "Durban en omgewing" of die "Durbanse metro-
pool" (in die afdeling oor lande, streke en plekke) word op dié wyse onder-
vang.

Omdat *SAAZ* primêr 'n skryfgids is, behoort die klem deurgaans op inlig-
ting te val wat vir die korrekte skryf van Afrikaans relevant is. In Deel II word
daar ongelukkig meermale met saakbesonderhede oorboord gegaan. In die

afdeling oor reg en orde word daar byvoorbeeld onder die opskrif "Vuurwapens" nie alleen verduidelik wat die verskil tussen 'n koeël, patroon en patroondoppie is nie ('n onderskeid wat reeds in die artikel van die lemma **koeël x patroon** in Deel I uitgeklaar is!), maar ook beskryf hoe 'n koeël by die loop van 'n geweer uitgedryf word. Die laasgenoemde inligting is nie van belang wanneer 'n korrekte woordkeuse in 'n bepaalde skryfstuk gemaak moet word nie. Dieselfde probleem duik op waar die verskil tussen 'n rewolwer en 'n pistool bespreek word, behalwe dat **rewolwer x pistool** nie ook as lemma in Deel I verskyn nie. Die gebruiker wat twyfel oor die korrekte gebruik van *rewolwer* en *pistool* sal nie sonder meer by dié inligting uitkom nie, veral ook omdat nóg *rewolwer* nóg *pistool* in die register opgeneem is.

Die uitgebreide register agterin het nommerverwysings na artikels in Deel II. Daar is nie, anders as in die toeligting vermeld, bladsyverwysings na taalkwessies in Deel I nie, maar wel "Sien"-verwysings na lemmas in die alfabetiese gedeelte. Hiermee word soektogte vir die gebruiker vergemaklik. Ongelukkig is daar te veel nuttige inligting in Deel II wat nie in die register gereflekteer word nie en waarby die gebruiker alleen kan uitkom as hy dit toevallig raaklees. 'n Voorbeeld hiervan is die volgende: In die lys van persoonseiename waarvan die spelling soms las gee, is naas die naam **Aaliyah**, die volgende inskrywing : "R&B-sangeres (in 2001 oorlede)". Die gebruiker wat nie weet waarvoor *R&B* staan nie, sal bes moontlik in die afdeling oor afkortings en simbole daarna gaan soek. 'n Tweede opsie sou wees om in die afdeling oor lees- en skryftekens onder die opskrif "Ampersand" te kyk of *R&B* daar as voorbeeld gegee word. Raak hy nêrens geholpe nie, is 'n laaste moontlikheid om in die register na óf *R&B* óf *&* te soek. *R&B* is nie gelys nie, maar *ampersand (&)* wel. Geeneen van die bladsyverwysings wat in die register by *ampersand (&)* gegee word, werp lig op sy probleem nie. Dit is alleen wanneer hy toevallig op die lys van Engelse leenwoorde in die afdeling oor vreemde woorde afkom, dat hy sal agterkom dat *R&B* na "rhythm and/& blues" verwys.

SAAZ is nietemin 'n ambisieuse werk wat ongetwyfeld daartoe kan bydra om die taalvaardigheidspeil van die gebruikers van Afrikaans te verhoog.

Bronne

Gouws, R.H. 1989. *Leksikografie*. Kaapstad: Academica.

Odendal, F.F. en R.H. Gouws. 2000. *Verklarende Handwoordeboek van die Afrikaanse Taal*. Johannesburg: Perskor.

Suid-Afrikaanse Akademie. Taalkommissie. 2002. *Afrikaanse woordelys en spelreëls*. Kaapstad: Pharos.

A.E. Cloete
(cobusclo@iafrica.com)

Reino Ottermann en / and Maria Smit, Hoofredakteurs / Chief editors, bygestaan deur / assisted by Izak Grové, Winfried Lüdemann, Heinrich van der Mescht en / and Caroline van Niekerk. *Suid-Afrikaanse Musiekwoordeboek / South African Music Dictionary*. Tweede, hersiene en vermeerderde uitgawe/Second, revised and enlarged edition. 2000, 296 pp. ISBN 1-86890-019-3. Kaapstad/Cape Town: Pharos. Prys/Price: R69.95.

In the nearly thirty years since the publication of the Vaktaalburo's pioneering bilingual *Musiekwoordeboek/Dictionary of Music* (1973) there has been a marked transformation both in terms of musical ontologies generally and the identity of the South African nation-state in particular. On the one hand Philip Bohlman (1999: 17) has neatly pointed out in his contribution to the anthology of essays entitled *Rethinking Music* that 'music may be what we think it is: it may not be', while on the other hand Sabine Marschall (2002: 117) reminds us that 'nations are not naturally in place but are invented and in need of constant maintenance'.

So when one opens the covers of the *Suid-Afrikaanse Musiekwoordeboek/ South African Music Dictionary* published in 2000, one's expectations are — not unreasonably — rather different from those that may have been aroused in 1973. What indeed might one legitimately expect a South African music dictionary for a new millennium to be? A dictionary of South African music(s)? A music dictionary for South Africa? With the stunning wealth of diverse musics that are endemic to our country, a dictionary *of* South African music(s) would indeed be welcome, meeting an urgent but as yet unfulfilled need, while a well-planned music dictionary *for* South African use would be equally welcome, provided it was able to meet the diverse needs of its multicultural users. So whatever specific use might be intended for a South African music dictionary, it would surely need to take into account the sheer fecundity and plurality of our musics, celebrated with such enthusiasm and acclaim within the global community.

In the event this dictionary turns out to be neither of the above. As 'second, revised and enlarged edition' of the Vaktaalburo's *Musiekwoordeboek/Dictionary of Music* of 1973 (p. iv; the front cover declares it to be 'revised and greatly enlarged'), its 'main aim is to promote the standardisation of Afrikaans and English terminology within the music education system of South Africa' (p. 5). What this presumably really means is that the dictionary's intention is to provide Afrikaans-speakers with a standardized vocabulary for speaking and writing about music — in itself a perfectly unobjectionable end.

This is, then, not really a 'South African music dictionary' at all. According to the editors (Introduction, p. 6), the intended target group is

> mainly music pupils, students and teachers, as well as music lovers. Authors of textbooks, musical reviews and articles, newspaper and other magazine articles, especially in Afrikaans, as well as people in public media such as broadcasting, were also included in the potential users' group.

Fair enough — perhaps. But as a result the title of the *Suid-Afrikaanse Musiek-woordeboek/South African Music Dictionary* is entirely off the mark and highly misleading. It is also, we believe, insulting to a majority of South Africans of all colours and languages (both within the academy and outside it) whose musical interests and activities extend far beyond the boundaries of what Lydia Goehr (1992) has called the 'imaginary museum of musical works'. The words 'bilingual', 'Afrikaans' and 'English' in the title would certainly help define the dictionary's legitimate purpose more accurately.

Although this is emphatically *not* a 'South African music dictionary' in any satisfactory sense of the term, let us go along for the moment with the premises on which the dictionary is predicated. How well does it deal with terms from 'Western music culture'? On the whole, rather well. The editors have managed, within their limited scope, to get many things right: hundreds of terms have been added to those included in the original version of the dictionary. *Plainchant* is thankfully no longer translated as *gelyksang* but more sensibly as *cantus planus*. There are also appendices giving the names of notes in six (European) languages, the Afrikaans transliteration of Russian composers names (rightly including names of famous Russian singers and instrumentalists as well), names of orchestral instruments, and so on.

But there are also some surprising omissions and lapses. The Introduction specifically calls the reader's attention to what it calls the 'problematical' case of the term *canon*: 'There are so many different names for the different types of canon, particularly in Latin and the European languages, that one has difficulty in distinguishing the various types. An effort has then been made to simplify and only note the most common names' (p. 7). Yet despite having given attention to several rather obscure meanings for the term *canon* (including this definition: 'a portion following the Sanctus in the Roman Catholic mass'), the editors have chosen to ignore one of the most significant usages of the term in late twentieth-century musicology. This is *canon* in the sense of the repertory of Western musical works that over the past century and a half has been deemed to form the conceptual framework of Western musicology and which has increasingly been called into question during the past twenty years or so. One is left to question what kind of (Western) musical education some South African students are receiving in our institutions.

A random search for terms currently used in the analysis of twentieth-century music that one might expect to find revealed no entries for *set theory*, *integral serialism*, *pitch class* and *pitch-class set*. And several entries under late Renaissance and Baroque terms are suspect. It is curious that the meaning of the term *a cappella* is given as 'unaccompanied' rather than 'in the style of the chapel' (and hence by extension, though often this was not historically the case, 'without instrumental accompaniment'). *Hoog-Barok* (there is no entry for *High Baroque*) is defined as 'middle Baroque' ('Middel-Barok'). There are two entries for *ripieno* but none for *ripienista* or *ripienist*. *Coro spezzato* finds an entry but not *cori spezzati*, the more acceptable usage of the term; the definition given is 'di-

vided choir' rather than the correct translation 'broken choirs' (for *cori spezzati*). The entry for *Bach trumpet* (*Bach-trompet*), with a parenthetical definition 'a term used for Baroque trumpet or clarino', gives no suggestion to the unwary reader that this term is a total misnomer for a high trumpet used in some modern performances of Baroque music and that it is a nineteenth-century invention, completely unknown in Bach's day; furthermore, during the seventeenth and eighteenth centuries the term *clarino* referred to a register rather than an instrument. Another curious omission is *maestro di cappella*, found in any number of historical contexts from the sixteenth to the eighteenth centuries. Yet the less frequent term *maître de chapelle* is included.

And although the editors explain that the dictionary's function is 'mainly to **translate**, rather than to explain' (p. 5), it is surely not too much to expect a degree of explication in the interests of redressing common misconceptions. Take the case of the term *adagio*. Here it is translated simply as 'slow', without any qualification, despite the fact that its meaning has shifted considerably over the years and was highly ambiguous until the nineteenth century. During the eighteenth century the term usually implied first and foremost that unnotated embellishments were called for on the part of the soloist, hence a relaxation in the tempo. In fact, the flautist J.J. Quantz devoted an entire chapter to the subject of playing an Adagio in his famous *Versuch einer Anweisung die Flöte traversière zu spielen* of 1752. Should a student confronted with the term in a piece of seventeenth- or eighteenth-century music simply interpret it to mean singing or playing slowly, thereby missing the point of the marking? The least one might expect is that the term be defined along the lines of 'at ease, i.e. slow'. (And one might wonder how many South African music students receive appropriate instruction, on the basis of this inadequate definition, in singing or playing an eighteenth-century Adagio.)

Another example is the supposedly well-known term *forte*, which really means 'strong' and only by implication 'loud'. During the seventeenth and eighteenth centuries it frequently referred to a 'normal' dynamic in contradistinction to *piano* which indicated an echo effect or was an instruction to players — as often found in Bach and Handel's scores, for instance — that the passage in question was an accompaniment to the soloist. Once again one might argue that it is too much to expect a dictionary of this nature to indicate these niceties of meaning, but surely one of the aims of a dictionary is to enlighten.

In the Introduction (p. 6) we are told that the *Suid-Afrikaanse Musiekwoordeboek/South African Music Dictionary*

> concentrates on terms from Western music culture. A small number of terms used in the indigenous African musics in South Africa, which pupils and students will commonly encounter, have nevertheless been added.

This statement reveals a patronizing (if not downright offensive) attitude. For a start, it assumes that twenty-first-century South African music students, writers of newspaper and magazine articles, media representatives and so on, are

deemed to need only to know about 'Western music culture' plus a token smattering of indigenous African musical terms thrown in for good measure — this in a hard-won pluralistic and democratic society! And the dictionary's basic premise finds itself on even shakier ground, since there is no reason to assume that a majority of Afrikaans-speakers in the country as a whole is interested primarily in 'Western music culture', especially within a rapidly-changing educational system.

In short, it is difficult to understand the altogether dated rationale behind this dictionary, even given its unobjectionable intention (and largely satisfactory realization) of providing an authorised list of Afrikaans musical terminology. How is it possible in our present stage of historical, political and cultural development that the editors of a South African music dictionary could have called upon a 'circle of collaborators' (!), numbering well over thirty South African academics, not one of whom is black, is an ethnomusicologist, is a specialist in popular music, or is a recognised authority on indigenous South African musics? What kind of message does this send to South African music students and authors of music articles in the media? That the music of their own country may be regarded as some kind of afterthought? And how can it be assumed that Afrikaans-speaking students are unlikely to come across terms in their musical studies that lie outside the imagined orbit of 'Western music culture'?

So it is entirely in keeping with the exclusivist nature of the dictionary that one finds an entry for the medieval *istampita* but none whatever for *isicathamiya*, that highly acclaimed home-grown South African genre within which Ladysmith Black Mambazo has carved out an astonishingly successful international career. Some other uniquely South African forms of music-making that receive no attention in the dictionary — not an exhaustive list by any means — include *bubblegum music*, *kwaito*, *makwaya*, *marabi*, *mokorotla*, *mqashiyo* (all mentioned in the second edition of *The New Grove Dictionary of Music and Musicians*) yet *maqam*, *maquam*, *mascherata*, *matraca* and *mattacines*, obscure terms the South African student is unlikely to come across, are all there.

The publishers' blurb on the back cover tells us that this is 'an authoritative new dictionary *for all* music lovers, learners, teachers and lecturers' (emphasis added). Even on its own terms the *Suid-Afrikaanse Musiekwoordeboek/ South African Music Dictionary* is no more than a qualified success. That there is a need for such a dictionary is undeniable, but it could, and ought, to meet that need far better than it does. Its limited, outdated concept represents a lost opportunity to have created a truly *South African* music dictionary.

References

Bohlman, Philip. 1999. Ontologies of Music. Cook, Nicholas and Mark Everist (Eds.). 1999. *Rethinking Music*. Oxford/New York: Oxford University Press.

Goehr, Lydia. 1992. *The Imaginary Museum of Musical Works: An Essay in the Musical Philosophy of Music*. Oxford/New York: Oxford University Press.

Marschall, Sabine. 2002. *Community Mural Art in South Africa*. Pretoria: Unisa Press.

Sadie, S. and J. Tyrrell (Eds.). 2001². *The New Grove Dictionary of Music and Musicians*. London: Macmillan.

Vaktaalburo of the Suid-Afrikaanse Adademie vir Wetenskap en Kuns. 1973. *Musiekwoorde-boek/Dictionary of Music*. Cape Town: Tafelberg.

George King and Carol Steyn
Department of Art History, Visual Arts and Musicology
University of South Africa (UNISA)
Pretoria
Republic of South Africa
(kinggt@unisa.ac.za)
(steynfc@unisa.ac.za)

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. Lexicographica. Series Maior 108. ISBN 3 484 39108 1. Tübingen: Max Niemeyer. Price: €68.00.

The main title of this work might be a bit misleading, because it creates the impression that a very wide and general topic will be discussed. The subtitle, however, is much more to the point, since the author specifically investigates English and French monolingual dictionaries in view of certain aspects. The author makes it clear (2001: 210) that she does not want to describe general "influences" of a socio-cultural nature on the contents of dictionaries, but rather how socio-cultural trends and traditions had an impact on the methodological aspects of lexicography in Britain and France. This means that generalisation for other languages will only be possible after further in-depth studies.

The author sets out to demonstrate systematically and empirically how, in the two particular societies, differences exist in the conception of dictionary writing because of differences in the two language communities, and poses as a sub-question (2001: 2) whether one could say that dictionaries are "culturally-determined". She investigates five contemporary general monolingual dictionaries in French (*Le Nouveau Petit Robert* (1993) (NPR); *Le Petit Larousse Illustré* (1995) (PL); *Dictionnaire de la Langue Française*; *LEXIS* (1994) (LEXIS); *Le Robert Micro-Poche* (1988) (MR); and *Dictionnaire du Français au Collège* (1995) (DFColl)), and four contemporary general monolingual dictionaries in English (*Concise Oxford Dictionary of Current English* (1995[9]) (COD); *Oxford Advanced Learner's Dictionary of Current English* (1995[5]) (AOLD); *Longman Dictionary of Contemporary English* (1995[3]) (LCODE); and *Collins COBUILD English Dictionary* (1995[2]) (COBUILD)). For testing the results of her analyses, several other French and English dictionaries are also occasionally involved in her investigation. Rothe (2001: 2) states that her study is not intended as a contribution to the field of dictionary criticism, and therefore her main aim is not to suggest ways of improving the dictionaries discussed.

Rothe lists the criteria for the selection of dictionaries (2001: 14), which include that the chosen dictionaries should already have been on the market for a while, and should have undergone at least one new edition. They should more or less address the same target groups, and have more or less the same range with regard to their lemma collection and their microstructure. Even though her own study is mainly synchronic, Rothe places the dictionaries under discussion in their historical context.

The exposition starts, in Chapter 1, with contemplations on the notion of "culture" and how Rothe interprets this notion in her study. Reference is made to Hausmann's (1983, 1985) expression "dictionary landscape" ("*paysage dictionnairique*") (2001: 3-4) and Rothe (2001: 213) formulates her basic premise on Hausmann's (1997: 184) statement that "dictionaries, as cultural products, follow the cultural guidelines of the society". She also draws on Rey's publications

(1987a and 1987b) and Rey and Delasalle (1979) for her definition of "culture". Rey (1987a: 20, 34) contends that the influence of culture on dictionaries can be discerned most clearly in items for labelling, definitions, lexicographical examples, and the structure of dictionary articles. Other aspects identified by Rey are also taken into account by Rothe, such as preference for diastratic versus diaphasic criteria, preference for citations versus examples constructed by the lexicographer, the question whether the sources of citations are given, and the market situation of the dictionary in the country of origin. Ideological issues, such as the degree of normativity, should also, Rothe (2001: 8) explains, be taken into account.

Therefore, central issues investigated in this publication are: (a) metalinguistical attitudes as exemplified by traditions of linguistic politics, research results on language attitudes, and the presence of special-purpose dictionaries dealing with particular aspects of language (for example, new words or slang); (b) attitudes towards the dictionary as an institution and the social tasks of lexicography (inferred from statements from the mass media, from reports on user surveys, and from the front matter of the dictionaries); (c) linguistic theories and popular points of view on language that were developed or which prevail in a given society; and (d) the situation of the dictionary market and the dictionary as a commodity (Rothe 2001: 11, 277). In addition, consideration is given to structural differences between the two languages concerned, which may also have an influence on the methods used in dictionaries.

Chapter 2 gives a survey of recent developments in the dictionary market in Britain and France. It is interesting that English dictionaries are more focused on advanced foreign learners, whereas French dictionaries are more targeted towards a general group of educated French-speaking users. This means that in English, learner's dictionaries emphasise the encoding function of the dictionary (to help foreign learners), and general monolingual dictionaries are meant for decoding (for first-language speakers). In French dictionaries, this distinction between the two functions is not important, because dictionaries are targeted towards French-speaking users who would not have problems with encoding. In addition, "user-friendliness" is a concept which is of great importance in English dictionaries. This includes several characteristics of dictionaries, such as the clarity of the metalanguage in items used for the construction of articles and in definitions, as well as the ease with which information can be found.

Rothe carefully and precisely describes the actual empirical work in Chapters 3 to 7. Chapter 3 deals with the selection of lemmata, usage labels and usage notes. Even though labelling practices are arbitrary in many instances, and dictionary compilers tend to copy from other dictionaries and in the process may violate the linguistic realities of the languages described, Rothe detects certain interesting trends with regard to usage labels and usage notes.

Normativity is very important in this respect. Reference is made to Ripfel's (1989) distinction between "normative", "descriptive" and "covertly normative" dictionaries. The last-mentioned category is found when dictionaries exclude

certain lemmata because of e.g. vulgarity (as sometimes happens in school dictionaries). Normativity therefore has an impact on the selection of material to be used as lemmata. It also affects the choice of lexicographical examples and the formulation of definitions. To determine the theoretical points of departure of the dictionary editors, Rothe (2001: 35) recommends that one should compare the scrutinised material with the explanations given in the front matter of the dictionary. For example, the number of markers given is not necessarily, according to Rothe (2001: 58), an indicator of the normativity in the dictionary. It is quite possible that a particular dictionary sets its standards of "standard language" very high, and that certain lemmata were excluded before compilation started. Then, of course, such examples of "low" or "offensive" usage would not be included in the dictionary, and also not marked as such. This method would constitute a policy of covert normativity.

Labelling practices of stylistically and diachronically marked vocabulary are also investigated in Chapter 3 because they shed light on the issue of normativity. Interesting differences between the two countries come to the fore, as well as some common trends. French dictionaries tend to work more with diastratic criteria indicating levels of social stratification (such as *argot*, *populaire*), whereas English dictionaries prefer to use diaphasic criteria indicating register (such as *slang*, *spoken*, *written*). Rothe (2001: 46) concludes that English dictionaries use pragmatic and non-evaluative considerations for the grouping of labels, whereas French dictionaries tend to make use of a continuum which distributes labels according to a hierarchical and evaluative system. It is even possible to ascribe these differences in the dictionaries to parallels with recent theoretical linguistic trends which developed in both countries. Pragmatics has been very popular in English linguistics, but in France this has not been the case.

Another interesting fact given by Rothe (2001: 50) is that in English, class stratification is usually determined by pronunciation and not so much by lexical differences. In French, however, the opposite is true. Rothe (2001: 51-52) ascribes this to the fact that in France, the Académie Française was very influential in establishing the *bon usage*, while there has never been such an important, unifying body in English-speaking countries. In England, Rothe (2001: 72) claims, the "spirit of English liberty" did not tolerate too much state intervention in language matters. The *bon usage* in France was, from the beginning, defined diastratically. Literary usage by *les bons auteurs* (the great authors) was set as a good example of *bon usage*, and the label *littéraire* can still be seen abundantly in French dictionaries as a way of encouraging "correct" usage. When looking at current corpora, it is clear that in England actual spoken language and written "non-literary" language (such as texts taken from the mass media) play an important role.

According to Rothe's analyses (2001: 60), English dictionaries are also more prepared to include material with "lower" diastratic marking. French dictionaries, by contrast, generally tend to filter their inclusion of everyday lan-

guage. Rothe's conclusion is that French dictionaries are on the whole more normative than their English counterparts, and that in the case of French learner's dictionaries, there are certain tendencies towards covert normativity. In addition, it seems that French general monolingual dictionaries and French learner's dictionaries are much more homogeneous than the English ones, where general monolingual dictionaries differ in many respects from learner's dictionaries (2001: 81). This might once again be traced back to the fact that English learner's dictionaries are usually compiled for non-English speakers.

With regard to diachronic marking, Rothe finds that French dictionaries tend to take language use of past eras into account, continuing the tradition of *bon usage*, whereas English dictionaries lean towards contemporary usage. For example, COBUILD explicitly states in its front matter that it aims to be non-historic and that it has included material exclusively from the 1990s. This is also true of the inclusion of neologisms, where French dictionaries are generally more conservative than the English ones.

In Chapter 4, Rothe investigates differences in the formulation of lexicographical definitions. She looks at defining techniques and defining styles, and for this purpose she analyses 300 definitions for noun lemmata, and 200 definitions for adjectival lemmata in each English and French dictionary. Rothe (2001: 87) draws on the definition typology by Rey-Debove, which include defining by means of (a) the typical Aristotelian definition where interchangeability plays a role (Rey calls this type *inclusion*); (b) analysis; (c) synonymy; (d) opposition; and (e) a metalinguistic definition. In this typology, Rothe only takes the first section of a definition into account, and uses Hanks' (1987: 120) term *multiple-bite strategy* to refer to definitions which consist of more than one section. It seems that in the case of nouns, the Aristotelian (inclusion) type of definition is used most frequently in both languages, even though French dictionaries use it even more than English dictionaries. In addition to this, English dictionaries use definition by means of synonym more often than French dictionaries. The multiple-bite strategy is used far more frequently in English dictionaries than in French ones. In defining adjectives, English dictionaries clearly prefer giving chains of synonyms (as in COD *babyish* 'childish, simple') and participial constructions (e.g. *factional* 'belonging to a faction'), whereas the French prefer relative clauses (e.g. *babillard* 'qui aime à babiller').

In interpreting the data on definition types she collected from the dictionaries in question, Rothe engages in an interesting discussion about the underlying theoretical and semantic points of departure in the two countries. In French dictionaries, one can often detect connections with structural semantics, and the quite prominent use of componential analysis features, resulting in the frequent application of the Aristotelian definition with the *genus proximum* and *differentiae specificae*. For example, LEXIS states in its front matter that it uses componential analysis and rejects the use of definitions by means of chains of synonyms. Rothe, however, warns that one should not see this preference in French dictionaries as a direct influence of structural semantic theory (2001: 92-

93). It would be more appropriate to speak of a certain relationship between these theories and the preference for certain definition styles. English dictionaries, on the other hand, show preference for prototype semantics, as has been developed in the 1980s in Anglo-American circles in opposition to structural semantics. The focus is on "typification" (Hanks 1979: 33) instead of "complete analysis". Of course, COBUILD with its complete sentences which serve as "explanation" rather than "definition", presents a class of its own.

Rothe looks even deeper into linguistic theory when she distinguishes between the so-called mentalistic approach to meaning and the operational theories on meaning. The mentalistic approach proceeds from the assumption that words have a conceptual "core" meaning, and that the relationship between words and the categories of objects to which they refer exists on a mental level. Meaning is an idea, a concept, which is established independently from the use of a word. French dictionaries tend to prefer this approach in determining meaning. Rothe also points out that French lexicographers tend to take diachronic matters more seriously than is the case in English lexicography. In English dictionaries, historical information is usually limited to etymological information.

The operational theories of meaning which developed in English linguistic circles, on the other hand, are "theories of usage" and draw on Wittgenstein's thesis that "the meaning of words lies in their use". Together with this view, Firth (1957a: 180, 1957b: 194) also proposed that the collocation of a word is very important in determining its meaning. Corpus linguistics, in which word frequencies and collocations play a dominant role, also, Rothe (2001: 128, 144) asserts, influenced English lexicography in this regard.

Chapter 5 deals with article structures and the ordering of senses in articles. Rothe selected strongly polysemous verbs and adjectives for this analysis, because they would be more context-dependent than nouns. Her conclusions are that English dictionaries generally tend to use more primary meanings and less sub-meanings and glosses, whereas French dictionaries have considerably less primary meaning discriminations. Contextual information is often moved to a separate "idioms section" in English dictionaries, or treated in a separate entry altogether. English articles typically consist of many relatively short paragraphs, whereas French articles tend to be longer and more continuous.

Even the ordering of the different aspects of meaning in dictionary articles shows different trends in the two languages under consideration. Rothe (2001: 130) distinguishes between (a) the historical principle, where the oldest meaning is presented first; (b) the "logical" principle, where the "core" meaning is presented first; (c) the frequency-oriented principle, where the most frequent meaning is presented first, and (d) the distributional principle, where the syntactical distribution or the different semantic contexts in which a word can occur, determine the order of presentation. On the whole, contemporary English dictionaries seem to prefer the frequency principle. French dictionaries are more heterogeneous in this regard, but the logical principle seems to be the

most popular among French lexicographers.

Preferences for linear versus hierarchical structures, and macro-structural ordering — and the relation to language structures (i.e. the derivational systems of the two languages under consideration) — also receive attention. Here, French dictionaries tend to prefer formal methods to use hierarchical structures on the primary levels of meaning discrimination, linking to a well-established tradition of treating polysemy inclusively in one article by means of the logical principle. English dictionaries generally prefer linear structures in the definition, which is, for instance, constituted by the presentation of chains of synonyms. They normally only use hierarchical structures on the secondary levels of meaning discrimination, which are divided by paragraph headings. This also links to the focus on "user-friendliness" in English lexicography, where it is argued that secondary meanings can be found more easily when the typography provides for headings.

When looking at the macrostructure of the dictionaries, Rothe detects that English dictionaries in most cases show the tendency to group formal and semantically motivated suffix derivations together in one entry, even when it means that the alphabetical order is violated. Most of the dictionaries use so-called "text blocks" where derivations are grouped together by means of nesting. French dictionaries, on the other hand, generally do not deviate from the alphabetical order. They mostly use a strictly initial-alphabetical, or straight-alphabetical ordering of lemmata. The question whether this state of affairs has anything to do with the structural characteristics of the two languages is investigated by Rothe. English derivations are often much further from the original Latin-Roman forms in the word family than is the case in French, as, for example, in *father* – *paternal*. Presenting these semantically motivated derivational families in English poses a special problem which is less crucial in French.

Even in the presentation of lexicographical examples, Rothe finds differences between the two lexicographical traditions. French dictionaries generally tend to use literary citations, also preferring to give references to sources. Even if English dictionaries use citations, they normally do not give references to the sources of citations. But English dictionaries generally rely rather on examples constructed by the lexicographer, or draw on examples from computer-based corpora. These corpus-based examples are not from literary works, but usually from written texts from mass media such as newspapers and magazines. Rothe links this situation to the fact that normativity plays an important role in French lexicography, and that the statistically proven language usage of the majority is important in English lexicography.

Rothe ends her investigation of lexicographical methods and cultural backgrounds by a specific study of collocations — their lexicographical treatment have proved to be "culturally significant" in the previous chapters.

It seems that English lexicography on the whole shows more innovative tendencies than French lexicography. Rothe ascribes this to the big market for English language dictionaries, and to the fact that France has a longer lexico-

graphical tradition. French dictionaries are more homogeneous in their characteristics, and in the lexicographical methods used. French dictionaries are more "linguistically" oriented because they tend to take linguistic theories into account, and English dictionaries are generally more user-oriented because they are pragmatically oriented towards the market.

The author warns (2001: 216) against an over-interpretation of certain clues if the relationship between the dictionary structure and aspects of culture is not explicitly stated in the front matter of the dictionary. Lexicographical texts should be seen as complex products of several factors: the socio-cultural setting prevailing at the time of compilation, established lexicographical traditions, and the market which puts pressure on the lexicographer to create original dictionaries without violating certain cultural norms. Rothe concludes that an element of a given culture has the greatest chance of becoming lexicographically relevant if the methodological options which it implies are already rooted in a given lexicographical tradition, and if sales are likely to be increased.

This publication is important for scholars who study the impact of cultural and linguistic traditions on general monolingual dictionaries. Although the author focuses on general monolingual dictionaries and learners' dictionaries in English and French, the issues she raises and the procedures she uses in her investigation may inspire new thinking and new insights for other languages, and for other types of dictionaries.

References

Dictionaries

COBUILD 1995[2] = Sinclair, John (Ed.). 1995[2]. *Collins COBUILD English Dictionary*. London: HarperCollins.

COD 1995[9] = Thompson, Della (Ed.). 1995[9]. *Concise Oxford Dictionary of Current English*. Oxford: Clarendon Press.

DFColl 1995 = Dubois, Jean (Ed.). 1995. *Dictionnaire du Français au Collège*. Paris: Larousse.

LEXIS 1994 = Dubois, Jean (Ed.). 1994. *Dictionnaire de la Langue Française*. Paris: Larousse.

LCODE 1995[3] = Summers, Della (Ed.). 1995[3]. *Longman Dictionary of Contemporary English*. Harlow: Longman Dictionaries.

MR 1988 = Rey, A. (Ed.). 1993. *Le Robert Micro-Poche*. Paris: Dictionnaire Le Robert.

NPR 1993 = Rey-Debove, Josette and Alain Rey. 1995. *Le Nouveau Petit Robert*. Paris: Dictionnaires Le Robert.

OALD 1995[5] = Hornby, A.S. (Ed.) and Jonathan Crowther. 1995[5]. *Oxford Advanced Learner's Dictionary of Current English*. Oxford: Oxford University Press.

PL 1995 = Maubourget, P. (Ed.). 1996. *Le Petit Larousse Illustré*. Paris: Larousse.

Other literature

Firth, J.R. 1957a. A Synopsis of Linguistic Theory, 1930-55. Palmer, F.R. (Ed.). *Selected Papers of J.R. Firth*: 168-205. London: Longmans.

Firth, J.R. 1957b. Modes of Meaning. Firth, J.R. *Papers in Linguistics, 1934-1951*: 190-215. London: Oxford University Press.

Hanks, P. 1979. To What Extent Does a Dictionary Definition Define? Hartmann, R.R.K. (Ed.). *Dictionaries and their Users. Papers from the 1978 BAAL Seminar on Lexicography*: 32-38. Exeter Linguistic Studies 4. Exeter: University of Exeter.

Hanks, P. 1987. Definitions and Explanations. Sinclair, J. (Ed.). *Looking Up. An Account of the COBUILD Project in Lexical Computing and the Development of the Collins COBUILD English Language Dictionary*: 116-136. London: Collins ELT.

Hausmann, F.J. 1983. Wörterbücher in Frankreich und Deutschland. Ein Vergleich. Wiegand, H.E. (Ed.). *Studien zur neuhochdeutschen Lexikographie III. Germanistische Linguistik* 1-4/82: 119-155. Hildesheim: Olms.

Hausmann, F.J. 1985. Trois paysages dictionnairiques. La Grand Bretagne, la France et l'Allemagne. Comparaisons et connexions. *Lexicographica* 1: 24-50.

Hausmann, F.J. 1997. Das Zitat in der französischen Lexikographie. Geschichte und Gegenwart. Lieber, M. and W. Hirdt (Eds.). *Kunst und Kommunikation. Betrachtungen zum Medium Sprache in der Romania. Festschrift zum 60. Geburtstag von Richard Baum*: 173-189. Tübingen: Stauffenburg.

Rey, A. 1987a. Le dictionnaire culturel. *Lexicographica* 3: 3-50.

Rey, A. 1987b. La notion du dictionnaire culturel et ses applications. *Cahiers de Lexicologie* 51(2): 243-256.

Rey, A. and S. Delasalle. 1979. Problèmes et conflits lexicographiques. *Langue française* 43: 4-26.

Ripfel, M. 1989. Die normative Wirkung deskriptiver Wörterbücher. Hausmann, F.J., O. Reichmann and H.E. Wiegand (Eds.). *Wörterbücher. Ein internationales Handbuch zur Lexikographie. Band 1./Dictionaries. An International Encyclopedia of Lexicography. Volume 1./Dictionnaires. Encyclopédie internationale de lexicographie. Tome 1*: 189-207. Handbücher zur Sprach- und Kommunikationswissenschaft 5.1. Berlin/New York: De Gruyter.

Maria Smit
Department of Music
University of Stellenbosch
Stellenbosch
Republic of South Africa
(msmit@sun.ac.za)

Juan C. Sager (Selector and Editor). *Essays on Definition.* Introduction by Alain Rey. 2000, viii + 254 pp. ISBN 90 272 2327 0 (Eur.), 1 55619 773 X (US). Amsterdam/Philadelphia: John Benjamins.

This publication is volume 4 in John Benjamins' series *Terminology and Lexicography Research and Practice.* It contains a "Preface" and an introductory essay "Defining Definition" (pp. 1-4) by Alain Rey. This is followed by a collection of texts, selected and edited by Juan Sager, which examine various aspects of definition from the point of view of philosophy.

The essays span the period from Plato and Aristotle to the 19th century, covering the major Western philosophical traditions. Besides essays of Plato and Aristotle, essays of Isidoro of Seville, Blaise Pascal, Benedict de Spinoza, John Locke, Gottfried Wilhelm Leibniz, George Berkeley, Immanuel Kant, John Stuart Mill, and Heinrich Rickert have also been included.

Although all included texts are classified as "essays", there is some variety in the form in which the authors present their views on definition: besides the essay, there are also genres such as the Platonic dialogue, letters (cf. the extracts from the correspondence of Leibniz to colleagues), and even a full monograph by Heinrich Riekert. There is also considerable variety in the length of the essays and the amount of space allocated to the different authors. Some of the texts constitute no more than a few lines, others are full-length essays. Besides the well-known essays of Aristotle on defining and definitions, most space has been allocated to a text in fact unknown to me, namely, a monograph of Heinrich Riekert "The Theory of Definitions" (pp. 191-254).

The main aim of this series of John Benjamins is to provide in-depth studies and background information pertaining to lexicography and terminology. It follows that this volume is aimed at specialists or advanced students in the field, and more specifically at those with the necessary background in philosophy to be able to access the texts and link and integrate the philosophical issues under discussion with current theoretical approaches to definition in lexicography and terminology.

This approach by the editor and publisher is corroborated by the fact that the selected texts are not provided with any introductory remarks or annotations giving data on the author's general views on defining and definitions, the topics of the essays, and/or how these topics under discussion link up with "modern" lexicographic or terminological definition theory.

The selected texts indeed make for very interesting reading, precisely because they remind us once again of exactly how reductionistic current theoretical views of definitions are in nature once you isolate them from the the philosophic traditions from which they have emerged . This reductionism is, of course, brought about by the practical goals of lexicography and terminology (compiling dictionaries and terminologies). These pragmatic goals have, however, lead us through a process in which we have inevitably delimited our theoretical constructs and framed them in such a way that they reflect little of

the critical issues underlying them — issues which have plagued philosophers for ages (e.g. the nature of knowledge and knowledge acquisition). In a number of these essays we are actually reminded of the fact that we have reduced some of these complex issues to simple dichotomies (e.g. encyclopaedic vs. linguistic definitions) or complex phenomena to neat taxonomies (e.g. a neat list of definition types).

Besides the more well-known texts, as, for example, the work of Aristotle (cf. pp. 25-89), John Stuart Mill (pp. 174-190), and Kant (pp. 163-171), there are also a number of lesser known texts, which really need the attention of lexicographers and terminologists. I would like to focus on three of these which have been translated by Sager.

The first is Isidoro of Seville's typology of definition types (pp. 91-94) which has been extracted from a book by Marius Victori. The author lists fifteen definition types, gives their Greek and Latin names, illustrates each with an example and succinctly explains what the characteristic defining technique of the definition type is. Most of the well-known definition types found in handbooks of lexicography are in the list, but a number of lesser known and/or used ones also appear. For example, the definition *eodem et de alterero* ("from one and the other") is explained as follows: "when one asks what difference there is between a king and a tyrant, and by means of the difference one is defined as well as the other, saying: 'The king is moderate and good-tempered the tyrant is godless and cruel' " (p. 92). The poetic "ring" of this definition type is characteristic of some of the other lesser known definition types listed, but characteristic of defining techniques in genres other than the dictionary.

The second is Heinrich Rickert's "The Theory of Definitions", which has been translated from the German, and of which the third edition of the monograph from 1929 has been included (pp. 199-249). This essay (including the extracts from the three prefaces) is really a remarkable piece of writing in which a number of philosophical issues concerning definitions — of which some feature in the other essays — come under critical scrutiny. Some of the topics discussed include the origin and original meaning of definition, word-explanation and definition, the purpose of definition, the inadequacies of existing theories; definitions in law, the natural sciences and mathematics; the inadequacy of the existing theories of the concept; the concepts of genus and essence in the empirical sciences; and nominal and real definitions — essential reading matter for terminologists, and, for its broader implications, also for lexicographers. A summary of the author's points of view will, however, not do justice to the eloquence of the phrasing, formulation, argumentation and structure of this study.

The third essay is the thought-provoking one by Blaise Pascal, "The Art of Persuasion" (pp. 108-117). This essay is particularly interesting as it is one of the few in which definitions are linked to the rhetorical tradition. The rules for logical persuasion and definition are presented in this essay within a general theory of persuasion. Pascal describes the latter as follows: "The art of persua-

sion is necessarily related to the way in which we agree to what is presented to us and to the nature of the things we are expected to believe" (p. 108). Before presenting his way to logical reason, Pascal goes to some length in explaining for what nature of things it is appropriate to use the logical method. These exclude, for example, divine truths. This is made clear in a paragraph in which he explains how man accepts divine truths in contrast to the acceptance of profane things.

The author argues that divine truths stand above nature and that we come to know them through our hearts, not our minds. That is why the saints exhort us to love divine things in order to come to know them. For natural things, however, God has reversed this order: to know (secular) things we must know them with our minds first. However, people have corrupted this order and have only come to believe things that they have come to love: "we believe almost nothing except that which pleases us" (p. 109). Although Pascal argues that little passes through our minds, he nevertheless accepts that the mind and the heart are like two doors by which these thoughts can enter our soul, and that each of these two doors have their own principles and stimuli for actions (p. 109).

Pascal has all kinds of ideas about persuasion through the heart (Aristotele' s *pathos)*, but the major aim of the essay is to show how definitions are to be used in logical arguments (Aristotele's *logos)* to persuade people (i.e. through the door of the mind). In essence this boils down to what Pascal outlines as "methodologically perfect proofs", which consist of three essential parts: define the terms you use with clear definitions, propose evident principles or axioms for proving the matter at hand, and, in demonstrations, always substitute the definitions for the defined concept (cf. p. 111) — simple rules that still lie at the heart of rational argumentation. This finally leads to a set of rules that will secure complete definitions, namely, (a) do not define perfectly well-known terms; (b) do not use somewhat obscure or ambiguous terms without defining them; and (c) when defining terms, use only well-known words or already explained words (cf. p. 112).

This essay illustrates precisely how complex issues of defining and definitions — despite such simple rules — can in fact become. They take on added complexity once they are seen against their larger functions (persuasion) within a certain discourse context (rational argumentation), and if they are linked to specific assumptions about knowledge acquisition and persuasion (the role of cognition versus emotion). These issues are, furthermore, tied in with assumptions about a moralistic (corrupt human nature) and a religious universe.

The introductory essay by Alain Rey, "Defining Definition", does not, as one would have initially suspected, provide the reader with an overarching framework within which the ensuing philosophical discussions can be systematically and coherently integrated and used as basis to link the topics under discussion in the essays to current theoretical approaches to definition in lexicography and terminology.

Rey's main argument in his essay is that it would be impossible to do so, given the enormous variety of definitions of the term *definition*, the various definition types, and the variety of functions for which they are utilized in the different disciplines. In some cases, it would in fact lead to confusion and bad descriptive practice (cf. his references to Diderot and d'Alembert's *Encylopédie*, p. 7) if the different, often incompatible, viewpoints and traditions that have evolved in the various sciences (as diverse as, for example, lexicography, terminology, logic, philosophy, law and religion) are not respected.

Rey, therefore, aptly remarks that a "summarising study of definition, like Richard Robinson's well-known manual (1950) (*Definitions,* PHS), is an impossible undertaking because it can only list and try to relate incompatible points of view" (p. 8). The only cogent kind of description one can actually come up with, is, as Rey argues, a discipline-specific theory of definition.

Although the essays presented in this volume, present a philosophical view, they do by no means present a cogent view. Exactly this makes them challenging reading.

Piet H. Swanepoel
Department of Afrikaans and Theory of Literature
University of South Africa (UNISA)
Pretoria
Republic of South Africa
(swaneph@unisa.ac.za)

Publikasieaankondigings / Publication Announcements

Juri Apresjan. *Systematic Lexicography*. Translated by Kevin Windle. 2000, xviii + 304 pp. ISBN 0-19-823780-4. Oxford/New York: Oxford University Press. (Review article in this issue.)

Henri Béjoint. *Modern Lexicography: An Introduction*. 2000, xii + 276 pp. ISBN 0 19 829951 6. Oxford: Oxford University Press. (Review in this issue.)

Monika Elisabeth Bründl. *Lexikalische Dynamik: Kognitiv-linguistische Untersuchungen am englischen Computerwortschatz.* Linguistische Arbeiten 443. 2001, XII + 293 pp. ISBN 3 484 30443 X. Tübingen: Max Niemeyer. Price: €72.00. (Review in this issue.)

Buro van die WAT. *Elektroniese WAT: Woordeboek van die Afrikaanse Taal (A–O).* 2003. ISBN 0-9584401-1-5. Buro van die WAT/Van Schaik Electronic. Prys: R450.00.

Committee for Political and Related Terminology/Komitee vir Staatkundige en Verwante Terminologie. *Modern Political Dictionary: Terms, Concepts and Usages in Politics and the Political Sciences/Nuwerwetse Politieke Woordeboek: Terme, begrippe en gebruike in die politiek en politieke wetenskappe.* 2002, 526 pp. ISBN 0-86970-561-X. Johannesburg: Centre for Political and Related Terminology in Southern African Languages (CEPTSA), Rand Afrikaans University/Sentrum vir Politieke en Verwante Terminologie in Suider-Afrikaanse Tale (SEPTSA), Randse Afrikaanse Universiteit. (Resensieartikel in hierdie uitgawe.)

A.P. Cowie. *English Dictionaries for Foreign Learners: A History*. 2002, xiii + 232 pp. ISBN 0-19-925084-7. Oxford: Oxford University Press.

The Dictionary Unit for South African English. *South African Concise Oxford Dictionary.* 2002, xx + 1368 pp. Cape Town: Oxford University Press. Price: R275.00. (Review in this issue.)

Henrik Gottlieb, Jens Erik Mogensen and Arne Zettersten (Editors). *Symposium on Lexicography X: Proceedings of the Tenth International Symposium on Lexicography May 4–6, 2000 at the University of Copenhagen*. Lexicographica.

Series Maior 109. 2002, XVII + 330 pp. ISBN 3-484-39109-X. Tübingen: Max Niemeyer. Price: €92.00.

Wilfrid H.G. Haacke and Eliphas Eiseb. *A Khoekhoegowab Dictionary with an English–Khoekhoegowab Index.* 2002, xiv + 740 pp. ISBN 99916-0-401-4. Windhoek: Gamsberg Macmillan. Price: N$175.00 (Hard-cover), N$150.00 (Soft-cover). (Review in this issue.)

Samukele Hadebe (Chief Editor). *Isichazamazwi SesiNdebele*. 2001, xlvii + 556 pp. ISBN 1-77900-404-4. Harare: College Press.

Werner Hüllen. *English Dictionaries 800–1700: The Topical Tradition*. 1999, xvii + 525 pp. ISBN 0-19-823796-0. Oxford: Clarendon Press.

Philippe Humblé. *Dictionaries and Language Learners*. 2001, 211 pp. ISBN 3-89846-110-6. Frankfurt am Main: Haag und Herchen. Price: €23.00.

Hans-Dieter Kreuder. *Metasprachliche Lexikographie: Untersuchungen zur Kodifizierung der linguistischen Terminologie*. Lexicographica. Series Maior 114. 2003, X + 271 pp. ISBN 3-484-39114-6. Tübingen: Max Niemeyer. Price: €78.00.

Louis Molamu. *Tsotsitaal: A Dictionary of the Language of Sophiatown.* 2003, xxix + 130 pp. ISBN 1 86888 187 3. Pretoria: University of South Africa. (Review in this issue.)

Dalene Müller. *Skryf Afrikaans van A tot Z.* 2003, 718 pp. ISBN 1 86890 037 1. Kaapstad: Pharos Woordeboeke. (Resensie in hierdie uitgawe.)

Reino Ottermann en/and Maria Smit, Hoofredakteurs/Chief editors, bygestaan deur/assisted by Izak Grové, Winfried Lüdemann, Heinrich van der Mescht en/and Caroline van Niekerk. *Suid-Afrikaanse Musiekwoordeboek/South African Music Dictionary*. Tweede, hersiene en vermeerderde uitgawe/Second, revised and enlarged edition. 2000, 296 pp. ISBN 1-86890-019-3. Kaapstad/Cape Town: Pharos. Prys/Price: R69.95. (Review 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. Lexicographica. Series Maior 108. ISBN 3 484 39108 1. Tübingen: Max Niemeyer. Price: €68.00. (Review in this issue.)

Juan C. Sager (Selector and Editor). *Essays on Definition.* Introduction by Alain Rey. 2000, viii + 254 pp. ISBN 90 272 2327 0 (Eur.), 1 55619 773 X (US). Amsterdam/Philadelphia: John Benjamins. (Review in this issue.)

Ewoud Sanders. *Voor een dubbeltje op de eerste rang: 1001 spreekwoorden en zegs-wijzen over Nederlands geld*. 2001, 242 pp. ISBN 90 446 0106 7. Amsterdam/Rotterdam: Prometheus/NRC Handelsblad.

Ewoud Sanders. *Van Nergenshuizen tot Absurdistan: Verzonnen plaatsnamen in het Nederlands*. 2003, 240 pp. ISBN 90 446 0195 4. Amsterdam/Rotterdam: Prometheus/NRC Handelsblad.

Gabriele Stein. *Better Words: Evaluating EFL Dictionaries*. 2002, x + 246 pp. ISBN 0 85989 719 2. Exeter: University of Exeter Press. Price: £39.50 (Hardback).

Karina van Dalen-Oskam en Marijke Mooijaart. *Bijbels lexicon: Woorden en uit-drukkingen uit de bijbel in het Nederlands van nu.* 2000, 443 pp. ISBN 90 5333 923 X. Amsterdam: Prometheus. (Resensieartikel in hierdie uitgawe.)

G.J. van Wyk (Tegniese Redakteur), A.E. Cloete, A. Jordaan, H.C. Liebenberg en H.J. Lubbe. *Etimologiewoordeboek van Afrikaans*. 2003, xiii + 596 pp. ISBN 0 9584401 3 1. Stellenbosch: Buro van die WAT. Prys: R195.00.

Ursula Wingate. *The Effectiveness of Different Learner Dictionaries: An Investigation into the Use of Dictionaries for Reading Comprehension by Intermediate Learners of German*. Lexicographica. Series Maior 112. 2002, X + 301 pp. ISBN 3-484-39112-X. Tübingen: Max Niemeyer. Price: €88.00.

362

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 evaluerend aangebied word.

(3) **Resensieartikels:** Navorsingsartikels wat in die vorm van 'n kritiese resensie van een of meer gepubliseerde wetenskaplike bronne aangebied word.

Bydraes in kategorieë (1)-(3) word aan streng anonieme keuring deur onafhanklike akademiese vakgenote onderwerp ten einde die internasionale navorsingsgehalte daarvan te verseker.

(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 nuusvrystellings van leksikografiese verenigings wat veral vir die praktiserende leksikograaf van waarde sal wees.

(8) **Verslae:** Verslae van konferensies en werksessies.

Bydraes in kategorieë (4)-(8) moet almal aan die eise van akademiese geskrifte voldoen en word met die oog hierop deur die redaksie gekeur.

2. Wetenskaplike standaard en keuringsprosedure
Lexikos is deur die Departement van Onderwys van die Suid-Afrikaanse Regering as 'n gesubsidieerde d.w.s. inkomstegenererende navorsingstydskrif goedgekeur.

Artikels sal op grond van die volgende aspekte beoordeel word: taal en styl; saaklikheid en verstaanbaarheid; probleemstelling, beredenering en gevolgtrekking; verwysing na die belangrikste en jongste literatuur; wesenlike bydrae tot die spesifieke vakgebied.

3. Taal van bydraes
Afrikaans, Duits, Engels, Frans of Nederlands.

4. Kopiereg
Nóg die Buro van die WAT nóg die African Association for Lexicography (AFRILEX) aanvaar enige aanspreeklikheid vir eise wat uit meewerkende skrywers se gebruik van materiaal uit ander bronne mag spruit.

Outeursreg op alle materiaal wat in *Lexikos* gepubliseer is, berus by die Beheerraad van die Woordeboek van die Afrikaanse Taal. Dit staan skrywers egter vry om hulle materiaal elders te gebruik mits *Lexikos* (AFRILEX-reeks) erken word as die oorspronklike publikasiebron.

5. Oorspronklikheid
Slegs oorspronklike werk sal vir opname oorweeg word. Skrywers dra die volle verantwoordelikheid vir die oorspronklikheid en feitelike inhoud van hulle publikasies.

6. Gratis oordrukke en eksemplare
Skrywers ontvang vyf gratis oordrukke van elke navorsings-, beskouende of resensieartikel van hulle wat gepubliseer is asook een gratis eksemplaar van die uitgawe waarin sodanige artikel(s) verskyn het. Skrywers van suiwer evaluerende resensies en van bydraes tot die rubrieke Leksikovaria, Projekte en Verslae ontvang vyf gratis oordrukke van hulle bydraes. In laasgenoemde vier kategorieë kan die redaksie egter, afhangend van die aard en omvang van die bydraes, besluit om ook 'n eksemplaar van die betrokke uitgawe aan 'n skrywer toe te ken.

7. Uitnodiging en redaksionele adres
Alle belangstellende skrywers is welkom om bydraes vir opname in *Lexikos* te lewer en aan die volgende adres te stuur:

Die Redakteur: LEXIKOS
Buro van die WAT
Posbus 245
7599 STELLENBOSCH
Republiek van Suid-Afrika

B. VOORBEREIDING VAN MANUSKRIP

Die manuskrip van artikels moet aan die volgende redaksionele vereistes voldoen:

1. Lengte en formaat van artikels
Bydraes moet verkieslik nie 20 getikte A4-bladsye met teks in dubbelspasiëring en ruim kantlyne (ongeveer 2,5 cm) oorskry nie. Manuskrip moet verkieslik in elektroniese formaat as ASCII-teks, as volledig geformateerde Microsoft Word (DOS of Windows) lêers of as WordPerfect (DOS of Windows) lêers op rekenaarskyf (360 KB tot 1.44 MB) voorgelê word. 'n Rekenaardrukstuk van die artikel moet die skyf vergesel. Elke artikel moet voorsien wees van 'n Engelse opsomming van tussen 150 en 250 woorde, sowel as tussen 10 en 30 Engelse sleutelwoorde.

2. Grafika
Een stel duidelike oorspronklike illustrasies, tabelle, grafieke, diagramme, of kwaliteitsafdrukke daarvan, moet voorgelê word. Die plasing van grafika binne die teks moet duidelik aangedui word.

3. Bibliografiese gegewens en verwysings binne die teks
Kyk na onlangse nommers van *Lexikos* vir meer inligting.

INSTRUCTIONS TO AUTHORS

*(For a more detailed version of these instructions, please contact the Bureau of the WAT (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 usere 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 rezenzieren.

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 Freiexemplare
Autoren erhalten fünf Sonderdrucke ihrer veröffentlichten Forschungsartikel, kontemplativen Artikel oder Rezensionsartikel gratis sowie ein Freiexemplar 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 Freiexemplar 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 au

teurs sont libres d'utiliser leurs textes dans d'autres publications, à condition d'indiquer *Lexikos* (collection AFRILEX) comme source de la publication originale.

5. Originalité
Seules les contributions originales seront acceptées pour la publication. Les auteurs conservent l'entière responsabilité de l'originalité et du contenu factuel de leur texte.

6. Tirés-à-part et exemplaires gratuits
Les auteurs recevront gratuitement cinq (5) tirés-à-part de chaque article 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*.