Lexikos 25

Lexikos 25

Redakteur

Editor

Elsabé Taljard

Resensieredakteur

Review Editor

T. Harteveld

African Association for Lexicography

AFRILEX-REEKS 25:2015
AFRILEX SERIES 25:2015

BURO VAN DIE WAT

STELLENBOSCH

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

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

Uitgewer Publisher

BURO VAN DIE WAT
Posbus 245
7599 STELLENBOSCH

Eerste uitgawe 2015

Tipografie en uitleg deur Tanja Harteveld en Hermien van der Westhuizen
Bandontwerp deur Piet Grobler
Geset in 10 op 12 pt Palatino

Gedruk en gebind deur SUN MeDIA Stellenbosch
Office 5, The Woodmill, Vredenburg Road, Stellenbosch

ISBN 978-0-9946528-1-2
ISSN 1684-4904

Lexikos is elektronies beskikbaar by http://lexikos.journals.ac.za/ (in kleur)
Lexikos is available online at http://lexikos.journals.ac.za/ (in colour)

Lexikos is elektronies beskikbaar by Sabinet, AJOL, Ebsco en Proquest
Lexikos is available online from Sabinet, AJOL, Ebsco and Proquest

Indekse Indexes
Arts and Humanities Citation Index®, Current Contents®/Arts & Humanities, Current Contents®/Social and Behavioral Sciences, ERIH Plus, Index Copernicus Journals Master List, Journal Citation Reports/Social Sciences Edition, Social Sciences Citation Index®, and Social Scisearch®; Linguistic Bibliography Online; Linguistics Abstracts Online; Linguistics and Language Behavior Abstracts; MLA Inter-national Bibliography; R.R.K. Hartmann's Bibliography of Lexicography; Scopus

Herbert Ernst Wiegand
zum 80sten Geburtstag gewidmet

Dedicated to
Herbert Ernst Wiegand
on his 80th birthday

Opgedra aan
Herbert Ernst Wiegand
by geleentheid van sy 80ste verjaarsdag

Adviesraad / Advisory Board

Redaksiekomitee / Editorial Committee

Inhoud / Contents

Voorwoord

Sedert 2011 is die redakteurskap van *Lexikos* gesamentlik deur proff. Rufus Gouws, Danie Prinsloo en Elsabé Taljard behartig, met een van die drie as finale redakteur vir elke uitgawe. Tydens die afgelope Afrilexkongres het prof. Gouws aangedui dat hy die redakteurstuig gaan neerlê. Dr. Steve Ndinga-Koumba-Binza het ingewillig om sy plek in te neem en die oorblywende twee redaksielede wil hom graag hartlik welkom heet. Die bydrae wat prof. Gouws tot die uitbou van *Lexikos* as 'n tydskrif van keuse vir beide plaaslike en internasionale leksikograwe gelewer het, kan kwalik oorskat word. Sy goeie oordeel en grondige kennis van beide die praktiese en teoretiese leksikografie het hom 'n uitmuntende redakteur gemaak.

Die aantrekkingskrag wat *Lexikos* ook vir internasionale leksikograwe het, word weerspieël in die verskeidenheid tale waarin daar in hierdie uitgawe gepubliseer is — benewens artikels in Engels, is daar ook bydraes in Frans, Duits en Afrikaans. Die bydraes in *Lexikos* 25 verteenwoordig verder 'n balans tussen artikels wat die metaleksikografie as tema het, en ander wat die toepassing van die nuutste teorieë op praktiese wyse illustreer. Die leksikografiese gesprek tussen Afrika en Europa word in hierdie uitgawe voortgesit, met stemme uit China wat tot die gesprek toetree.

Elke uitgawe van *Lexikos* is die resultaat van 'n spanpoging. My opregte dank aan al die outeurs — dit is hulle bydraes wat die voortbestaan van *Lexikos* as gesaghebbende internasionale vaktydskrif verseker. Ewe belangrik is die keurders, wat onselfsugtig hul tyd en kundigheid beskikbaar stel om van elke *Lexikos* 'n gehaltetydskrif te maak. Die nougesetheid waarmee outeurs gewoonlik die kommentaar van keurders opvolg en implimenteer, weerspieël die waarde wat aan die insette van keurders geheg word. Deeglike eweknie-evaluering dra ongetwyfeld by tot die akademiese kwaliteit van die tydskrif.

Wat die tegniese en administratiewe aspekte van die publikasie van *Lexikos* betref, rig ek graag 'n besondere woord van dank aan me. Tanja Harteveld van die kantoor van die Buro van die WAT. Saam met me. Hermien van der Westhuizen, was sy verantwoordelik vir die setwerk, en haar oog vir detail en nougesette werkswyse het veel bygedra tot die tegniese kwaliteit van die publikasie.

Voornemende outeurs word uitgenooi om manuskripte vir moontlike publikasie in 'n volgende *Lexikos* deur middel van die aanlynsisteem wat op die webblad beskikbaar is aan ons te stuur.

Elsabé Taljard
Redakteur

Foreword

Since 2011 the editorship of *Lexikos* has been jointly run by Proff. Rufus Gouws, Danie Prinsloo and Elsabé Taljard, with one of these being the final editor for each volume. During the past Afrilex conference, Prof. Gouws indicated that he was relinquishing his part of the editorship. Dr Steve Ndinga-Koumba-Binza has agreed to replace him and the remaining two editors would like to extend a warm welcome to him in this position. The contribution that Prof. Gouws has made in establishing *Lexikos* as a journal of choice for both national and international lexicographers can hardly be overestimated. His sound judgement and knowledge of both practical and theoretical lexicography have made him an outstanding editor.

The appeal that the *Lexikos* journal has for international lexicographers is reflected in the variety of languages in which articles appear in this issue — in addition to articles in English, there are also contributions in Afrikaans, French and German. The contributions in *Lexikos* 25 further represent a balance between articles with a metalexicographical theme, and others that illustrate the application of the latest theories in a practical way. The lexicographical dialogue between Africa and Europe continues in this issue, with voices from China entering the conversation.

Each volume of *Lexikos* is the result of a team effort. My sincere thanks to all the authors — it is their contributions that have ensured the continued existence of *Lexikos* as an authoritative international journal. Equally important are the adjudicators, who have generously made their time and expertise available to render each *Lexikos* volume a journal of high academic standing. The rigour with which the authors usually follow and implement the adjudicators' comments, reflects the high regard for the adjudicators' efforts. Thorough peer reviewing undoubtedly contributes to the academic quality of the journal.

Regarding the technical and administrative aspects of the publication of the journal, I would like to say a special word of thanks to Ms Tanja Harteveld of the office of the Bureau of the WAT. Together with Ms Hermien van der Westhuizen she was responsible for the final typesetting, and her eye for detail and methodical approach have contributed vastly to the technical quality of the journal.

Prospective authors are invited to submit manuscripts for possible publication in a future volume of *Lexikos* through the online system which is available on the website.

Elsabé Taljard
Editor

'n Woord van AFRILEX

AFRILEX wil prof. Elsabé Taljard, ons AFRILEX-tesourier, bedank dat sy die belangrike rol van redakteur vir vanjaar se nommer, *Lexikos* 25, vervul het. Prof. Taljard is 'n bekende deskundige en akademikus binne die vakgebied van die taalkunde en die leksikografie, en 'n talentvolle vakkundige in navorsing oor inheemse Afrikatale. Soos in die geval van die vorige redakteur, prof. Rufus Gouws, is *Lexikos* met prof. Taljard in die hande van 'n deskundige. Namens AFRILEX wil ons prof. Taljard hartlik bedank en wens ons haar krag en leiding toe om hierdie verantwoordelike taak voort te sit. Ons wil ook by voorbaat vir dr. H.S. (Steve) Ndinga-Koumba-Binza bedank vir en gelukwens met sy besluit om deel te word van die *Lexikos*- redaksionele span in die plek van prof. Rufus Gouws wat onlangs vrywillig uitgetree het as redakteur. Dr. Ndinga-Koumba-Binza sluit by die span aan as die derde redakteur, saam met prof. Danie Prinsloo en prof. Elsabé Taljard. Met hierdie drie aan die stuur, is die toekoms van hierdie internasionale tydskrif rooskleurig deur hul diensbaarstelling aan die ontwikkeling van die leksikografie, veral op die Afrika-kontinent.

As mondstuk vir die leksikografie het *Lexikos* sedert sy ontstaan 'n uitstekende platform vir leksikograwe, taalvakkundiges en AFRILEX-lede geskep om hul leksikografiese navorsingsuitsette met kollegas, sowel nasionaal as internasionaal, te deel. Hierdie tydskrif het veel bygedra tot die ontwikkeling van die leksikografie, veral op die Afrika-kontinent. Vanjaar se nommer handhaaf steeds 'n hoë standaard met betrekking tot die kwaliteit van die gepubliseerde artikels. Binne AFRILEX is ons verheug om te sien hoe hierdie tydskrif, wat beide kwantiteit en kwaliteit betref, van krag tot krag gaan met die verskyning van elke nommer sodat dit gunstig kan kompeteer met verskeie tydskrifte wat elders in die wêreld gepubliseer word. Al hierdie suksesse kan alleenlik toegeskryf word aan die werk van die redakteurs en die eweknie-beoordelaars wat baie ure deurbring om te verseker dat die *Lexikos*-manuskripte op 'n toepaslike vlak is vir insluiting. Ons wil ook die outeurs van die artikels bedank, veral die AFRILEX-lede, wat voortgaan om *Lexikos* steeds van kwaliteitsartikels te voorsien wat jaar na jaar die status van hierdie tydskrif verhoog en handhaaf.

Weereens wil ons alle deelnemers wat referate by AFRILEX-konferensies gelewer het, aanmoedig om hul artikels vir publikasie in hierdie geakkrediteerde tydskrif voor te berei en om hul navorsingskundigheid met ander vakkundiges te deel, insluitende die oorweldigende meerderheid wat nie die konferensies kan bywoon nie.

Maropeng Victor Mojela
President: AFRILEX

A Few Words from AFRILEX

AFRILEX would like to thank Prof. Elsabé Taljard, our AFRILEX treasurer, for playing the important role of editor for this year's issue, *Lexikos* 25. Prof. Taljard is a well-known expert and academic in the field of linguistics and lexicography and a gifted scholar in indigenous African language research. Just as in the case of the previous editor, Prof. Rufus Gouws, with Prof. Taljard *Lexikos* is in the hands of an expert. On behalf of AFRILEX, we want to thank Prof. Taljard whole-heartedly and trust that she will receive guidance and strength to continue to carry out this responsible task. We also wish to thank and congratulate in advance Dr H.S. (Steve) Ndinga-Koumba-Binza for accepting to join the *Lexikos* editorial team, substituting Prof. Gouws, who recently took a voluntary retirement from the position of editor. Dr Ndinga-Koumba-Binza joins the team as the third editor of *Lexikos* together with Prof. Danie Prinsloo and Prof. Elsabé Taljard. With these three at the helm, the future of this international journal is bright for catering for the development of lexicography especially on the African continent.

As mouthpiece for lexicography, *Lexikos* has since its inception, created an excellent forum for lexicographers, linguistic scholars and AFRILEX members to share their lexicographical research output with colleagues, both nationally and internationally. This journal has contributed much to the development of lexicography, especially on the African continent. This year's issue still holds the standard high with regard to the quality of the published articles. In AFRILEX, we are pleased to see this journal going from strength to strength in both quantity and quality with the advent of every issue, competing effectively with various journals published elsewhere in the world. All these successes are solely ascribed to the work of its editors and the peer reviewers who spend many hours making sure that the *Lexikos* manuscripts are at a suitable level for inclusion. We also want to thank the authors of the articles, especially the AFRILEX members, who keep the lexicographical flag flying by continuously supplying *Lexikos* with quality articles to elevate and sustain the status of this journal year by year.

Once more we want to encourage all participants who read papers at AFRILEX conferences to prepare their papers for publication in this accredited journal and to share their research expertise with other scholars, including the overwhelming majority who are unable to attend the conferences.

Maropeng Victor Mojela
President: AFRILEX

Redaksionele doelstellings

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

Die strewe van die AFRILEX-reeks is:

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

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

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

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

Editorial Objectives

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

The objectives of the AFRILEX Series are:

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

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

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

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

http://lexikos.journals.ac.za

Phases and Steps in the Access to Data in Information Tools

Henning Bergenholtz, *Centre for Lexicography, Aarhus University, Aarhus, Denmark (hb@bcom.au.dk) and Department of Information Science, University of Pretoria, Pretoria, South Africa and Department of Afrikaans and Dutch, Stellenbosch University, Stellenbosch, South Africa*

Theo J.D. Bothma, *Department of Information Science, University of Pretoria, Pretoria, South Africa (theo.bothma@up.ac.za)*

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

Abstract: In the information society, it is much easier for someone to find relevant data if s/he has an information need because of the availability of databases and electronic information tools. In information science this topic is usually treated under the topic *information behaviour*. In lexicography the term *access process* is used (Bergenholtz and Gouws 2010). It can be shown that this process beginning with the "origin of the problem" leading to an "information source usage situation" contains different parts, and that each part contains different phases, with the pre-consultation phase and the intra-consultant phase containing different steps. The most important concepts here are the *access route* and the *access time*. In this paper some experiments in two case studies are described to show how different access processes in different user situations take place.

Keywords: ACCESS TO DATA, COGNITIVE, COMMUNICATIVE, INFORMATION SOURCE SELECTION, INFORMATION TOOL, INTERPRETATIVE, INTRA-CONSULTATION PHASE, META-REFLECTIONS, OPERATIVE, POST-CONSULTATION PHASE, PRE-CONSULTATION PHASE

Opsomming: Fases en stappe in die toegang tot data in inligtingswerktuie. In die inligtingsgemeenskap is dit baie makliker vir iemand om relevante data te kry as hy/sy 'n inligtingsbehoefte het omdat databasisse en elektroniese inligtingswerktuie beskikbaar is. In die inligtingswetenskap word hierdie onderwerp gewoonlik behandel onder die onderwerp *inligtingsgedrag*. In die leksikografie word die term *toegangsproses* gebruik (Bergenholtz en Gouws 2010). Dit kan aangetoon word dat hierdie proses wat begin met die "oorsprong van die probleem" en lei na 'n inligtingsbrongebruiksituasie uit verskillende dele bestaan en dat elke deel verskillende fases bevat met die prekonsultasiefase en die intrakonsultasiefase wat verskillende stappe bevat. Die belangrikste begrippe hier is die *toegangsroete* en die *toegangstyd*. In hierdie artikel word enkele eksperimente in twee gevallestudies beskryf om te wys hoe verskillende toegangsprosesse in verskillende gebruiksituasies plaasvind.

Sleutelwoorde: INLIGTINGBRONKEUSE, INLIGTINGSWERKTUIG, INTERPRETATIEWE, INTRAKONSULTASIEFASE, KOGNITIEWE, KOMMUNIKATIEWE, METAREFLEKSIES, OPERATIEWE, PREKONSULTASIEFASE, TOEGANG TOT DATA

1. Data in the information society

We are living in the so-called information society. This does not mean that we only now need information that we acquire in the form of knowledge and experience. We need knowledge and skills to be able to communicate successfully with one another (communicative skills and communicative knowledge). We need knowledge to understand the world and all relationships in the world (cognitive knowledge). We need knowledge and skills to be able to execute certain actions (operative knowledge and operative skills). We need knowledge to be able to understand certain non-linguistic signs and to act accordingly (interpretative knowledge), cf. Tarp (2008). This has always been the case. What is new is the fact that the collection of data and the complexity of data relations in the last 50–70 years have increased to a previously unknown extent. As part of this more complex world the amount of data needed and of accessible data have increased in such a way that we can refer to a data explosion. Where we collected the data before 1450 in handwritten manuscripts and handed down experiences and after the introduction of the art of printing could preserve it in books, magazines etc. we can collect such enormous amounts of data and access it today like never before in the history of mankind.

The problem of 100 years ago no longer exists. Today one does not have to travel far in order to be able to question learned people who have the necessary knowledge or to find the lonely library with many old and new books. We have, on the contrary, the problem of such an enormous amount of data at our disposal that it can lead to information overload and eventually even to information death. When executing a Google search one often finds millions of hits.

There is, however, a solution to the problem. There is in fact an information tool that is easy to use and that can be used without a user manual

— it is small and handy
— it is always available (outside, inside, on holiday, etc.)
— it provides answers to all questions
— it provides answers in a way that the user understands the answer
— it gives a correct answer, but never a more detailed answer than is necessary to solve the problem.

That this information tool is also quite small and convenient is self-evident. In a concrete way it is described how this information tool offers assistance, e.g.

— when looking for help to find the right way when not knowing the direction
— to know whether a particular plant is toxic or not

— to help understand what a gorilla tells you and to show you how to talk to a gorilla
— to help understand an inscription in an extinct language
— to help understand a word in a language used by only a few people living on a small island

We are talking about the *Junior Woodchucks Guidebook* used only by Huey, Dewey, and Louie. However, there is a Donald Duck magazine which indicates that a first edition of the *Junior Woodchucks Guidebook* already existed before the library in Alexandria burned down. Since then, all data from all the books from all the libraries are part of the *Junior Woodchucks Guidebook*. This information tool is in fact an intelligent computer, which is able to know what Huey, Dewey, and Louie want to know and that gives them the answers in a way that they can understand. It gives the answer on a small hand-held computer, which has a book cover as protection (usually red, sometimes yellow). The answer comes quickly, it is true and it is complete in the sense that all the relevant data are given — and nothing more.

This is an optimal solution: An information tool as a personal information tool that is able to know what you want to know, and is able to provide an answer in a way that you get access to exactly the required data — and nothing more. However, theoretical debates still fail to explain access to data even in these simple information tools in a convincing way.

2. Access to data in lexicography and information science

In recent publications, lexicographers have discussed the relationship between lexicography and information science, e.g. Bergenholtz (2011), Fuertes-Olivera and Nielsen (2012). There is no doubt that lexicography should not be regarded as a sub-discipline of linguistics, although linguistic knowledge is important for some kinds of dictionaries. Lexicography often has much more in common with information science, or at least with some parts of information science.

Lexicography deals with information tools containing data for users in communicative situations (reading, writing, translating and proof reading), in cognitive situations giving information to archive general or special knowledge about a certain topic, as well as in interpretative and operative situations; see Bergenholtz and Tarp (2002, 2003), Fuertes-Olivera and Tarp (2014), Tarp (2008, 2008a). However, there is a great difference in the way the two disciplines describe what lexicography calls "access to data" and what information science calls "information behaviour".

Lexicographers generally distinguish between the pre-lexicographic and post-lexicographic phases (or extra-lexicographic phases) and the search for and in the information tool, viz. the intra-lexicographic phase:

- An extra-lexicographical pre-consultation phase where a user with specific characteristics finding him- or herself in a specific context or situation and:
  - experiences an information need,
  - becomes aware of the information need,
  - determines whether this need is lexicographically relevant,
  - and decides to start a lexicographical consultation.
- An intra-lexicographical consultation phase where the user:
  - selects the relevant lexicographical information tool,
  - accesses the relevant data,
  - verifies that s/he has found the right data, i.e. relevant to the information need in question,
  - and retrieves the needed information from the data.
- An extra-lexicographical post-consultation phase where the user:
  - makes use of the retrieved information in order to solve a communicative or cognitive problem, to store it as knowledge, to perform a task or to interpret a sign, signal, symbol etc. (Bothma and Tarp 2012: 92-93).

In information science there are more differentiations and therefore a number of terms not commonly used or known in lexicography.

This article takes as its point of departure the terms for access to data in lexicography as defined by Bergenholtz (2009) and Bergenholtz and Gouws (2010), taking into account insights from information science. As in information science, lexicography has different uses for the terms data and information. Many lexicographers use *information* as a synonym for *data*. In this paper we refer to *data* when referring to the data selected for the lexicographical database and shown in the articles in dictionaries; we use *information* as a cognitive term for the individual user's interpretation of the data. The interpretation of the same data, i.e. the information retrieved from that data, may differ from one user to another though identical data were found in the dictionary.

Somehow lexicography has always dealt with access. Lexicographers did not always use this specific term but described the problems to be solved by the dictionary, the selection of a dictionary, and the search in the dictionary. More often they spoke very generally about dictionary use. In the 90s the term *access structure* was used regularly. The use of this term was very general, somehow pretending that all users would choose the same type of access according to the macro- and microstructure of a given dictionary. The access structure was in this sense described in Nielsen (1995: 219-223) who referred to the sequence of single access steps and the relationship between those steps using the differentiation between outer and inner access structure:

> Access structure may be explained as the structure of the lexicographical indicators directing the user to the information required. A distinction is made between outer access structure, which is the structure of the indicators directing the user to the lemma, and inner access structure, which is the structure of the indicators directing the user to the required information inside the dictionary article. The access structure may be realised in a number of ways, all of which are intended to direct the user to the information required in the shortest possible consultation time (Nielsen 1995: 219).

The outer access structure contains all the steps before the user comes to the dictionary article; the inner access structure contains the steps within the article. It is very imprecise and also unclear because the reader does not learn anything concrete about the different steps in the access process. Recently there has been some progress — with Wiegand (2007) one of the first to continue the discussion. For the Wiegand-school, the paper by Wiegand and Beer (2013) is the latest and the clearest one. First of all — as the title of the paper says — this contribution only deals with access to printed dictionaries, not to e-dictionaries and not to any other kind of information tools. In reality it is a contribution to access to dictionaries with specific access structures for each single dictionary. These access structures are closely related to the macro- and microstructures in certain printed dictionaries (see Wiegand and Gouws 2013, Wiegand and Smit 2013).

The macro- and also the microstructures in some printed dictionaries are often complicated and successful access to data can only be achieved by a "well-informed user" (Wiegand and Beer 2013: 111). If the user has not read and understood the user guide, presented in the front matter of the dictionary, s/he can only use the dictionary in a haphazard way:

> The knowledge of non-alphabetical inner access structures which is particularly related to partially and completely condensed dictionary articles has to be gained by means of an intensive reading of the guidelines for use. Only after doing so, the user is able to execute internal access procedures without looking for something in longer dictionary articles. In the guidelines for use, non-alphabetical inner access structures are not explained to the user explicitly, but implicitly by means of the introductions to the so-called "composition of article". (Wiegand and Beer 2013: 140)

The well-informed user is not only someone who has read the user guide, but is also a very experienced user, according to Wiegand and Beer (2013). Both assumptions are not realistic when looking at the average dictionary user. Most users do not read the user guide, and most users are only using a given dictionary infrequently. It is true that familiarity with the structures of a dictionary makes it easier to use that dictionary and that this knowledge about the structures makes it easier for the user to find the data s/he is searching for. But to presuppose this knowledge from the average dictionary user is not realistic. In a two-year period, only a few users had read the user instructions, both the

short and the more detailed user instructions, of THE DANISH DICTIONARY OF FIXED EXPRESSIONS and THE DANISH MUSIC DICTIONARY. Bergenholtz and Johnsen (2013: 563) provide data from the log files of these two dictionaries. Over a longer period with 3 038 932 searches in total in THE DANISH DICTIONARY OF FIXED EXPRESSIONS, only a fraction of these (0.19%) were directed at the user instructions:

short user instructions	3 089
detailed user instructions	2 717
	5 806

In THE DANISH MUSIC DICTIONARY, there were 160 157 searches in total, with only 3.18% directed at the user instructions:

short user instructions	1 822
detailed user instructions	3 267
	5 089

This is very discouraging for those lexicographers who emphasise the importance of users' familiarity with and understanding of the structure of the specific dictionary — unless the presentation is so uncomplicated that user experience and knowledge about the dictionary structures is not a prerequisite for successful dictionary consultation. The users' lack of knowledge about and interest in user instructions can partially be seen from the fact that within a three-year period, which included 18 million searches, 80% of all users of the Danish Internet Dictionary (or, to be more precise, each specific IP address) had only used the dictionary three times, two times or one time (Bergenholtz and Norddahl 2012: 209).

There is one more unrealistic aspect of the description of access structures by looking at only one dictionary. In some cases one dictionary can be sufficient. In most cases the user has to consult more than one dictionary and not only dictionaries, but every kind of information tool to which the user has access.

The same objections also apply to Bergenholtz (2009) and Bergenholtz and Gouws (2008, 2010). In those papers the concrete search processes are related to a specific dictionary and not — what would have been more realistic — only to the concrete information needs. The common term in the whole process is described under the same keyword: **searching**. In Bergenholtz (2009) there are 18 different terms describing the search process, in Bergenholtz and Gouws (2010) there are 19 terms, as listed in Table 1.

The origin of the problem	Combined search strings
Information source usage situations	Access by means of an alphabetical macrostructure
Choice of information source	Access by means of a systematic macrostructure
Choice of the component of an information source	Index access

Consultation of an information source	Search in a part of a component
Search string	Search route
Search option	Search step
Situation-oriented access	Search speed
User type-oriented access	Search time
Accuracy of the access and the data presentation	

Table 1: Terms in Bergenholtz and Gouws (2010) for describing the access process

The different steps were originally identified outgoing from tests where test persons were asked to solve text reception and knowledge (cognitive) problems. The single steps were described as different steps but they were not grouped into clusters for different phases in the access process. The number and use of steps were not very different for the two different types of situations. No tests for operative and interpretative problems were conducted. Therefore it is unclear whether additional steps will be found.

The main objection to the metalexicographical proposals and the long list of access terms in Bergenholtz and Gouws (2010) is the notion of "step" for each of those terms. Should e.g. a term like *search time* rather not be called a step, but rather be regarded as a part of the users' time experience and for the metalexicographer a part of the meta-reflections? But also the "real" steps in the access process are not separated at all in the different phases. Without doing that the description of the access process steps remains unclear.

In information science, one mostly deals with cognitive situations, but sometimes with operative situations, too, when looking for help and advice. As long as information science has existed, the access to data has always been a main issue. But where lexicography has dealt only with the single steps, information science has focused on the main phases in the access process. In information science the information seeking process is embedded in the broader term *information behaviour*. This process is in essence the same as in lexicography, but there are smaller or larger terminological differences. Some of the terms are the same, for example *information need*, *information source*, *search string*. The term *search* in lexicography is, however, used differently in information science and it includes the information science concepts of *searching*, *browsing* and *navigating*. *Searching* in information science implies that the user has to define a search string (which could consist of a single word, a phrase, or a number of words/phrases combined by means of Boolean operators), i.e. it requires a specific intellectual input from the user to identify the specific words/phrases and to combine them in the correct way by means of Boolean operators (either in a command line interface, or through drop-down menus). *Navigating* and *browsing* are also clearly distinguished; both imply following predefined links provided to the user (either by the system or by the original author), but with the difference of intentionally following a path to find specific information, in contrast to serendipitously following links without necessarily

knowing where this will lead. Finding the required relevant information often requires, in information science terminology, a combination of searching and navigating, but each aspect of the process is clearly defined; see also Bergenholtz and Bothma (2011) and Bothma (2011).

3. Case studies

It is well known that most information needed could be found somewhere in the Internet, but it is also well known that information overload and too long and difficult access routes often make it difficult or impossible to find the required data. In this section, we will discuss the protocols of some experiments from two case studies, each with one single test person. The two test persons are trying to solve different information needs in different user situations. The aim was to determine whether different access procedures apply to different user situations. For this purpose a single test person is sufficient.

Such case studies should not be regarded as either a real survey or a set of questions put to an incidental passer-by. Two controlled case studies have been performed albeit that only one test person was used in each. The value and status of a case study in scientific research, even with only one test person, should not be underestimated, cf. Flyvbjerg (2006) in this regard. Kuhn (1987) argues convincingly that a discipline without a large number of thoroughly executed case studies lacks a systematic production of exemplars and this leads to it becoming ineffective, cf. Flyvbjerg (2006: 242). The two test persons used in these case studies were asked to answer specific questions. There is no doubt that different test persons would not have used the same access route to find a solution and would also not have used the same consultation time. Neither is it a question that varying experience in using information tools would have rendered different results, with the more experienced test persons having a quicker and the less experienced user a slower access. The reading abilities and experience of different respondents would most likely also have led to different results. But it was not the aim of the case studies to lead to general remarks regarding reading abilities and experience in using different kinds of information tools.

In these case studies insights from the modern functional theory (Bergenholtz and Tarp 2003, Tarp 2008, 2008a) have been used. A distinction is made between four main types of situations when using information tools: communicative situations, operative situations, interpretative situations and cognitive situations. Each one of these main types has different subtypes, e.g. text reception problems, text production problems, translation problems and text correction problems are four subtypes of communicative situations. A communicative situation is characterised by a starting point within a text — oral or written. After using an information source the user can use the information found in the text and can then proceed with the text. If s/he did not find an answer, s/he can proceed anyway or stop reading, writing etc. this specific text. In an operative situation, the user does not know how to act in a certain situation. After

using an information tool the user can (if s/he gets the needed help) act in the world as s/he has learned from the information tool. In an interpretative situation the user doesn't understand a sign or a signal. After using an information tool, the user can act or not act in the way the symbol indicates. In a cognitive situation the user has a knowledge need or interest. After using an information tool, s/he has learned or not learned what s/he wanted and in the case of a successful consultation can store the acquired information in the brain — forever or for a short period. This knowledge can also be used in another situation, but this would be a new communicative situation. In all four main user situations we can describe the process and the results and then reflect on the process and results, which can be summarized as follows:

— communicative situations: text → information source → act in the text → meta-reflections;
— operative situations: knowledge problem how to act → information source → act in the world after having got the needed information → meta-reflections;
— interpretative situations: acting problem reacting to a sign or a symbol → information source → act or not act in the world after having obtained the needed information → meta-reflections;
— cognitive situations: knowledge interest → information source → no acting but storing the found information in the brain (forever or for a short period) → meta-reflections.

3.1 Access experiments about driving in South Africa

Driving in South Africa: experiment 1 deals with a communicative situation, a text reception situation, where the respondent is a person with German as mother tongue, with a good knowledge of English and some knowledge of Afrikaans. She reads the following test and does not understand the words *e-tolling* and *gantries*:

> Whether we like it or not, e-tolling has arrived in Gauteng. Up until now the gantries erected across the province's highways have been a new addition to the urban landscape and the object of much anger, but they have not yet played a role in picking our pockets. But that will all changed at midnight on December 2, when the gantries began to start functioning according to their designed purpose. To date the gantries have reportedly only been gathering traffic information and undergoing testing but no deductions from fitted e-tags have yet been made.

With regard to the word *e-tolling* the test person

1. opens Google;
2. searches for *e-tolling* in Google;
3. gets the result as a long list of hits;
4. chooses a website which would not open;

5. chooses instead another website about electronic tolling in South Africa: http://en.wikipedia.org/wiki/E-toll_(South_Africa);
6. opens this site and finds a satisfactory explanation.
(The total access time is 1 minute 3 seconds.)

Driving in South Africa: experiment 2 deals with a communicative situation, a translation situation. The respondent is asked to find the Afrikaans equivalent for *e-tolling*:

The test person does not have a printed English–Afrikaans dictionary, but

1. opens Google;
2. searches for *e-tolling* in Google;
3. the first page of links provides no help;
4. opens instead Google translate;
5. writes *e-tolling*;
6. does not get any result (steps 1-6 take 2 minutes 40 seconds);
7. looks for an English–Afrikaans dictionary and searches for *dictionary English–Afrikaans*;
8. follows the link: *English–Afrikaans Translation*;
9. writes *e-tolling* and gets a result (steps 7–9 take 3 minutes 56 seconds).
(The total access time is 6 minutes 36 seconds.)

Driving in South Africa: experiment 3 deals with a communicative situation, a text production situation. The respondent is asked: What is the correct orthographical version, if any, of the following possibilities: *driver licence, drivers licence, driver license, drivers license?* The respondent

1. thinks about the four variants and is not sure whether anyone of them is correct;
2. opens Google;
3. does a search for *driver license*;
4. in the first link to Wikipedia sees that *driver's license* is the term used in the USA and *driving licence* in GB;
5. does a new search in Google for *driving license + South Africa*;
6. concludes that all four variants in the question are wrong, *driving licence* is the right form.
(The total access time is 3 minutes 33 seconds.)

It is interesting that the test person did not find the official form for SA which is *driving licence* (with two c's), but only concluded that this is the correct form.

Driving in South Africa: experiment 4 deals with an operative situation: While planning a holiday in SA, you find out that there are new toll roads in Gauteng. If you have a rental car, how do you pay for the toll fees?

The test person

1. opens Google;
2. searches for *Toll road Gauteng* in Google;
3. the first page of links provides no help;
4. does a new Google-search for *Toll road Gauteng new*;
5. chooses the first link for an e-toll calculator;
6. sees that this page gives no answer (steps 1–6 take 2 minutes 59 seconds.);
7. does a new Google-search for *toll road Gauteng rental car*;
8. chooses link number two *E-tolling Gauteng*;
9. gets a very long text without any content overview;
10. comes to *e-toll cost and toll gate location*;
11. has to choose dates for departure and return and does it;
12. gets no result (for steps 7–12 she uses 8 minutes 36 seconds.);
13. does a new Google-search and writes a question: *How to pay e-toll renting car Gauteng?*;
14. uses the first link *How e-tolling will impact on your car rental*;
15. reads that Europcar puts the cost for the e-toll on the bill;
16. knows now the answer for Europcar, but not necessarily for other car rental companies, but stops the search (steps 13–16 use 3 minutes 30 seconds.).
    (The total access time is 15 minutes 5 seconds.)

Driving in South Africa: experiment 5 deals with an interpretative situation: How should I act when I see the following traffic sign?

The respondent

1. opens Google:
2. writes *traffic signs South Africa* in the Google search field;
3. uses the first link with the title *traffic signs*;
4. reads the text;
5. click on the link *signs*;
6. has to select between *control*, *command* and other links;
7. selects the *command sign*;
8. this link does not help, she goes back to the *sign* site;

9. opens the link *control sign*;
10. finds the sign, but no explanation, but learns that the sign is called a *four way stop sign*;
11. goes back to Google and does a search for *four way stop sign*;
12. clicks on the link *who goes first*;
13. clicks on the first site with the Google results called *Yahoo answers*;
14. the first answer gives the rule.
    (The total access time is 6 minutes 6 seconds.)

Driving in South Africa: experiment 6 deals with a cognitive situation where the respondent gets three related questions:

1. Is a German driver's license valid in South Africa or do I need an international driver's license?
2. Is a German driver's license valid in South Africa if I am going to live in South Africa for many years?
3. Is there a time limit until you need a South African driver's license? If yes, what is the time limit?

The respondent

1. opens Google;
2. searches for *German driver's license in South Africa*;
3. Google recommends an alternative search for *Drivers license in South Africa for foreigners*;
4. does this search for the recommended text;
5. opens the site *Using a foreign driving licence/South Africa*;
6. reads the text and gets the answer for question 1;
7. continues reading the same text for the next two question and gets the answers.
   (The total access time is 6 minutes 40 seconds.)

An analysis of this experiment can be done in the following way, dividing the access into three phases.

3.1.1 Pre-consultation phase

The pre-consultation phase for all six experiments is not a real life situation but part of a case study. We therefore requested the test person to act even though she had the option to decline to do the individual experiments. The process did therefore not terminate but continued with the intra-consultation phase in each single case.

3.1.2 Intra-consultation phase

In the information source selection during the intra-consultation phase the test

person was requested not to make use of a human source and only to use electronic sources. In all cases the test person decided to start with a Google search, even in cases where she eventually decided to use an e-dictionary, e.g. in experiment 1 and 2, where there was a text reception problem and a translation problem. In experiment 3 where the test person had a text production problem and the user could have used a dictionary as well, she, however, did not use a traditional dictionary but rather opted for Wikipedia, followed by a Google-search. In all other cases the test person relied on standard Google searches. In terms of the content of the questions the test person was a layperson (as the test questions were not designed for a semi-expert or an expert). Since the test person did not use any specialised resources a distinction could not be made between monofunctional and polyfunctional sources.

When using the selected information sources, the test person never consulted the full document. However, in the case of the dictionary articles she did read the full dictionary article. In all other cases she consulted one or more components of the information found in the specific source. The test person read larger or smaller parts of the text in the components and used one or more of the links but never all the given links. In the case of experiment 5, she opened the page and this first web site could be regarded as a table of contents. Here she could select further pages to see if the required information appeared on one of the linked pages. In experiment 3 where the test person used Wikipedia she could have used the typical Wikipedia table of contents but the required information was given in the paragraph preceding the table of contents. In the case of experiment 1 and 2 where the test person used a traditional dictionary, markers in the dictionary article were used to indicate whether the subsequent item was a definition, an item giving the origin, etc. The test person could therefore limit the amount of information accessed by means of only reading the relevant part of the article. Although an index is a very common access component in printed information tools it appeared in some sources but not in the sources used by the test person.

In all cases the user combined searching with following the given links. She did not serendipitously follow links (therefore she did not browse) but purposely followed selected links in the hope of finding relevant data.

3.1.2.1 Information retrieved

In all cases the test person retrieved information. However, she was not always satisfied with the information retrieved in the first effort and then had to repeat the search or follow different links. This happened in most of the experiments, resulting in an iterative process. In such cases she had the option not to carry on but since she had been requested to find a (fairly) acceptable answer in each case she persevered. The nature of the iterative processes is clear from the description of the experiments. In all cases, therefore, the test person arrived at a satisfactory or partly satisfactory answer and her information needs were

satisfied or partly satisfied.

3.1.3 Post-consultation phase

In the post-consultation phase the information which has been retrieved is used either to act in the real world (in communicative, operative and interpretative situations) or it is stored as knowledge. The four main functions of the modern function theory have been illustrated by means of the set of very simple experiments.

In addition to the concrete access phases there are different kinds of meta-reflections.

3.1.4 Routes and steps

Since only one test person was used only one route with a single set of steps was observed in the experiment. However, the authors of this article did the experiments beforehand to find out whether feasible answers could be obtained for the different questions. They used completely different routes and therefore totally different steps to arrive at answers. It is therefore evident that access routes and the different steps followed by an individual will be unique and that there is not only one possible route or set of steps. From teaching information retrieval we know that not all students get the same answers, some students get incorrect answers and some students tend to terminate the process at different stages.

3.1.5 Speed and time

Speed is the time it takes to complete the individual steps, and time is the total time of all the steps. This was calculated in the experiments above (however, in some cases not for all individual steps but for some connected steps). For the user those four variables (route, step, speed and time) are important because they can influence whether the user carries on with the retrieval process until s/he arrives at a satisfactory answer or decides to abandon and terminate the process. Contextual variables such as time available and importance to satisfy the information need, etc. will influence the user in this respect. For the lexicographer and the information specialist those four variables are to be taken into consideration when designing an information tool, writing the content and/or evaluating the effectiveness or efficiency of the information tool.

3.1.6 Success

In our "Driving in South Africa" experiments it was clear that the test person in

some cases arrived at an answer with which she was very satisfied. And based on her evaluation of the information sources she consulted, she decided that the information was correct. In a couple of cases, however, she indicated that she was not absolutely sure whether the answers that she found were correct. To solve this problem she should either have gone into a further iterative process or consulted a human source such as an expert. It could, of course, also be that the answer was incorrect but that the test person did not realise this at all and used the information she found as correct data without querying the correctness. In such cases in a work task situation, only a review by an expert will reveal the error.

3.1.7 Multiple information needs in a work task

In a typical work task a user may have many individual/interrelated information needs. For each of these needs the user will typically go through the whole process and then reflect not only on the single information need but also on the totality of information needs. This was obviously not part of the current set of experiments but we foresee that reflection on multiple information needs in a work task may provide interesting results. The characteristics of a work task and the characteristics of the user were discussed in Bergenholtz and Bothma (2011).

3.2 Access experiments about accounting problems

The access process for more complicated problems from a specific field shows the use of the same phases but different steps and especially very different meta-reflections.

Accounting experiment 1

The first experiment deals with a communicative situation where the test person has a translation problem. He does not know how to translate *deemed costs* in the following English text into Danish:

> "Operating costs include some deemed costs and actual costs. Accordingly, the total operating costs will be different to those relevant for income tax purposes."

The access process for solving this problem is quite short, a total of 35 seconds, with 12 seconds for step 1, 4 seconds for step 2, 14 seconds for step 3 and 5 seconds for step 4.

1. The user knows the Internet address of THE ENGLISH–DANISH ACCOUNTING DICTIONARY. He writes the address of this dictionary and gets the search page.
2. He writes *deemed cost* in the search box.

3. He gets the following result and reads the dictionary entry:
   "Deemed cost is an amount used instead of cost or depreciated cost at a specific date. Any following amortisation or depreciation is made on the assumption that the enterprise initially recognised the asset or liability at a cost equal to the deemed cost."
   The translation is found: *fastsat kostpris.*
4. He writes *fastsat kostpris* in his Danish translation.

Only one test person was used in the experiment. One can therefore, as in Bergenholtz and Gouws (2010), not compare the same experiment as executed by different test persons. However, the following conclusions can be reached, albeit not generalisations but only indications resulting from a case study:

— Both the intra-consultation phase and the post consultation phase are short and quick.
— Meta-reflections will show differences in route, steps, step time, consultation time and success by different users. The most important fact is that the post-consultation phase is short when using the result in the translated text.

Experiment 2

The second experiment deals with a communicative situation where the test person has a text reception problem. Reading the text, he does not understand the meaning of *commercial activities*:

> "Many companies nowadays consider schools to be an important setting for marketing to children. However, important concerns can be raised from a health promotion perspective about the potential negative impact of commercial activities on the health and well-being of pupils."

Does the term *commercial activities* refer to a special kind of business? Or is every kind of business a commercial activity?

Compared with the first experiment with a problem in a communication situation, this one is obviously more complicated and time consuming with many steps:

1. The user knows the Internet address of THE ENGLISH–DANISH ACCOUNTING DICTIONARY and writes the address of the dictionary. (8 seconds)
2. Here he sees the search box and writes *commercial activities* in the search box. (8 seconds)
3. He gets no result and instead writes *commercial activity.* (8 seconds)
4. He gets no result again and looks for another tool by using a Google search where he gets a list of finance dictionaries.
5. He uses the first proposal in the list, the information tool called INVESTOPEDIA (http://www.investopedia.com/dictionary/) and looks for *commercial activity.*

6. He gets many hits; the first one gives the definition: "activity undertaken by an individual or a group".
7. This result is not convincing, he goes back to the Google search. (step 4–7: 5 minutes 2 seconds.)
8. He uses the link to the FREE FINANCIAL DICTIONARY and looks for *commercial activity* and gets the following definition: "activity undertaken as part of a commercial enterprise". (3 minutes 28 seconds.)
9. He is still not quite convinced and tries once again in the Google links and looks under FINANCE GLOSSARY but receives the message: "no results found". (46 seconds.)
10. He tries again: the DICTIONARY OF FINANCE AND BANKING. He does not get an answer, but needs a password or has to buy a password. (56 seconds.)
11. He tries again: the MERRIAM WEBSTER. It gives no help but only the message: "The word you've entered isn't in the dictionary". (36 seconds.)
12. The user gives up and has to believe the definition from step 6 or 8. Those two answers give some hint for the understanding. (24 seconds)

Totally the whole process, without a good and convincing answer, took 11 minutes 36 seconds. We can see that:

— the intra-consultation phase is long and not satisfying; the post-consultation phase is short;
— meta-reflections will surely show differences in route, steps, step time, consultation time and success by different users. The most important fact is that the post-consultation phase is quite short, also in a case with a problematic answer. In other cases with more convincing information, the post-consultation phase could be even shorter.

Experiment 3

In the third experiment our test person is asked to act as a managing director of a minor Danish Company. He has to find answers for some related operative problems:

1. Through which medium do I have to send the financial statements to the authorities?
2. How often do I have to send it? (as annual or interim accounts?)
3. In which language should the report be sent?

The first question is not a trivial one. There has been a real case where a company was closed down by the authorities because the owner did not send the financial statements in the proper way. And this required way was the result of a new decision.

1. He finds THE DANISH ACCOUNTING DICTIONARY and does a search for *regnskab* (accounting). He finds no relevant information. (18 seconds)
2. He searches for *årsregnskab*. Here he finds the following example:

"Årsregnskaber skal indsendes til Erhvervsstyrelsen senest 5 måneder efter regnskabsårets afslutning." (The financial statements must be submitted to the Danish Commerce and Companies Agency within a time limit of 5 months after the end of the financial year.) The answer to question (2) seems to be once a year, but (1) and (3) are still open. (28 seconds)

3. He does a Google search for *regnskabsloven* (law on annual report). The first hit has a link to the law. Having found the text with the accounting law, it takes time to come to the relevant Chapter 19 "Indsendelse til Erhvervs- og Selskabsstyrelsen" (= Submission to the Danish Commerce and Companies Agency). The text is long and difficult to understand. In §3 he learns that the report can be sent in paper form or electronically. But there is a link to new guidance. Here he learns that almost everyone must send the report in an electronic form, with the exception of "regnskabsklasse D" (reporting class D). According to the used language, he sees in §3.1 (6) that the report must be written in Danish, but there are exceptions, e.g. for foreign companies. After 32 minutes and 16 seconds, he is still not absolutely sure about the answers to the 3 questions.

Both the intra-consultation and post-consultation phases are long and difficult.

In meta-reflection one can, for this single case with this single test person, show a complicated route with four main steps. But the reception of the text that was found, has several single steps requiring several re-readings of an earlier part. The four steps takes 25, 18 and 28 seconds and the last one 32 minutes and 16 seconds (= step time). The total consultation time is 33 minutes and 2 seconds. There is no clear success. The test person has to ask an expert to be sure.

Experiment 4

We will not conduct experiments for interpretative situations in this paper because they are not very relevant for problems in accounting. They could, for example, be very relevant for traffic and traffic planning problems. Here we will close the experiments with a cognitive problem:

> What is *deemed cost* really? The question came up by reading Kilgarriff (2012: 27), who criticizes THE ENGLISH–SPANISH ACCOUNTING DICTIONARY for providing a too long definition:
>
> "Deemed cost is an amount used instead of cost or depreciated cost at a specific date. Any following amortisation or depreciation is made on the assumption that the enterprise initially recognised the asset or liability at a cost equal to the deemed cost."

Kilgarriff (2012) has instead a much easier solution, one every lexicographer should use in his running work: "Because they could have found a shorter and better one by a Google search: Surrogate for cost at a given day."

Which version is correct? Or asked in another way: Is the short definition

that was quickly found in Google better than the one in an accounting dictionary — a tool explaining accounting terms for accounting experts and semi-experts? The test person tries in the first way to find the answer making a Google search:

1. He makes a Google search for *deemed cost*.
2. He gets 106 000 hits. The first one is one with a quotation from IFRS, the official international standard commission (http://annualreporting.info/definiciones/deemed-cost): "Deemed cost is an amount used as a surrogate for cost or depreciated cost at a given date. Subsequent depreciation or amortisation assumes that the entity had initially recognised the asset or liability at the given date and that its cost was equal to the deemed cost".
3. He learns from this that the short sentence is not enough. *Deemed costs* have to be recognised and noted in the company's books as such before the term can be used, e.g. by *amortisation*.

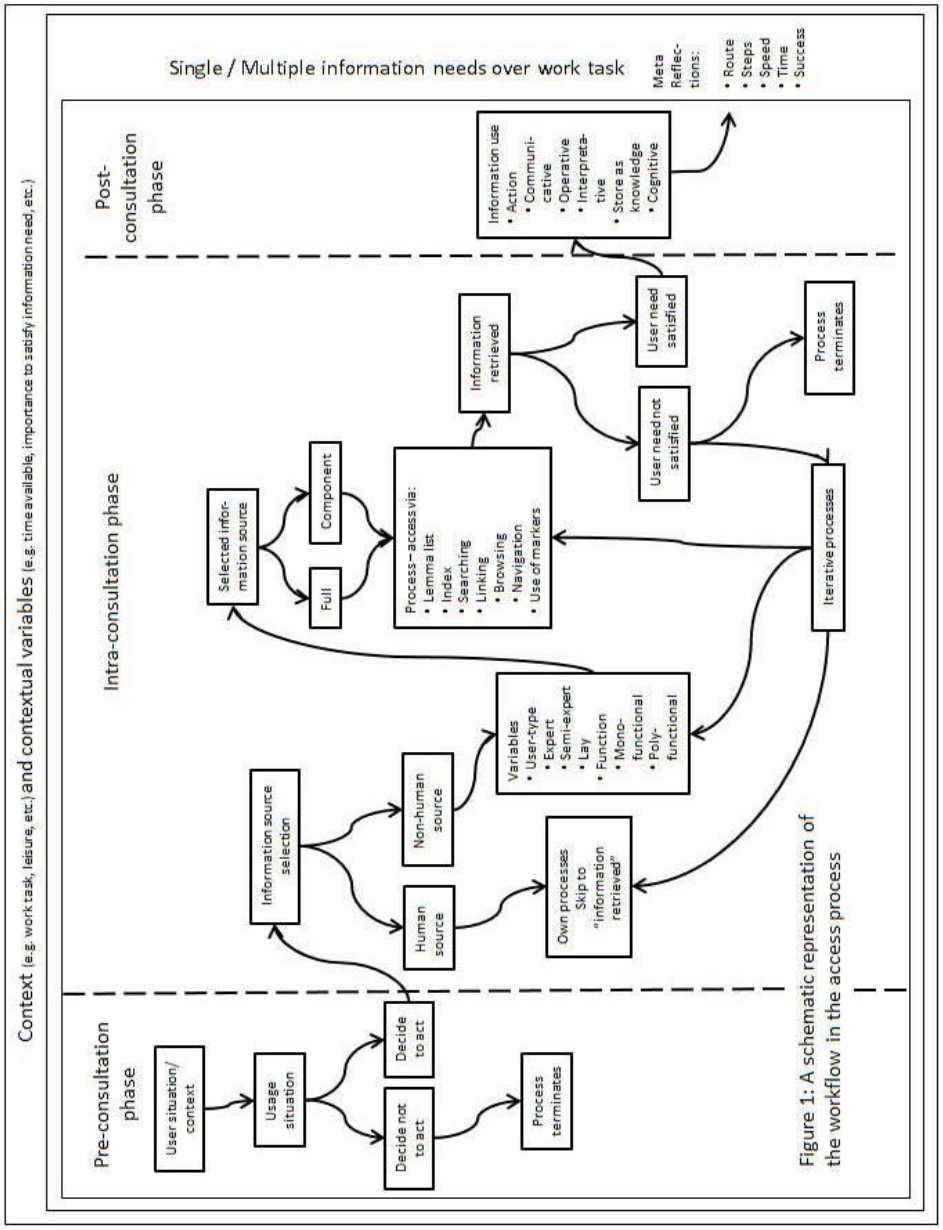
In the meta-reflections it is noted that the whole process took only 1 minute 12 seconds. The intra-consultation phase is quite quick and the post-consultation phase too. It does not mean that cognitive problems can be solved so quickly and so easily every time. In Bergenholtz and Gouws (2010) there are other cases with a very long and complicated process for cognitive problems — similar to the one described above with an operative accounting problem.

4. Proposals for a common theoretically founded set of terms for access to data

The biggest difference between the earlier proposals and the approach followed in this paper is the systematic description of the access process within the three main phases mentioned in Section 1 above, with subdivisions in the intra-consultation phase, viz. information source selection, consultation process and information retrieved.

It is, however, evident that there are a number of issues that are relevant to the full process and not only to a single phase. These issues are not "steps", but are reflections that both the end user of the information tool and the lexicographer should take into account — the end-user when using (or planning to use) the tool, and the lexicographer in designing the tool, writing the content and/or evaluating the efficiency of the tool. We call these issues "meta-reflections". In the diagrammatic representation they are listed outside the formal flow diagram to indicate that they are to be taken into account for the full process. Within a work task, users may have any number of information needs following one another during the work task (see Bergenholtz and Bothma (2011)), and we therefore indicate that the meta-reflections apply not only to a single information need, but also to multiple information needs in a work task. All of this occurs within a specific context, with specific contextual variables (see also Bothma (2011), Ingwersen (2007), and Ingwersen and Järvelin (2004, 2005)).

Figure 1: A schematic representation of the workflow in the access process



As can be seen Figure 1 the real access process can therefore have several iterative partial processes, and it is foreseen that during the same process different information tools can be used. Explained in words, but in a more systematical way as in Bergenholtz and Gouws (2010) we have three main access phases and a following meta-reflection phase:

4.1 Pre-consultation phase

4.1.1 The origin of the problem

There can be many reasons for the origin of an information need. They could be called extra-lexicographic and could be conceptually separated from the usage situations of the information source. These usage situations are divided into communicative, cognitive, operational and interpretative situations; cf. above in Section 4. There are various communicative situations, for example text production, text reception and translation in the mother tongue or from/to a foreign language. Cognitive situations refer to situations where the users for one or other reason need to add to their existing knowledge, e.g. where they have doubts or problems about a specific topic. The needs do not have to be of a concrete nature, but it could also be some form of inquisitiveness or even boredom. Especially in log files it can be seen that users often take the opportunity to look at recently completed dictionary articles when the lexicographer makes them available for the users in a special list in an outer text. Operative functions are known at best from manuals and user guides, but they also occur in text books and dictionaries when one is instructed to do something or to refrain from something, or when there is an explanation of how to do something.

In monofunctional information sources the situation type is predetermined. It could, for example, be a dictionary for the communication situation of text production or text reception or translation (as is the case with many bilingual dictionaries). Similarly, there are information sources for cognitive situations (e.g. many encyclopaedias and the majority of text books) or for operative situations (such as most user guides). In polyfunctional information sources, the user can try to access that part of a dictionary article or of a chapter in a text book that seems relevant to satisfy a specific need. In electronic databases one can formulate one's own search strings in the search box supplied by the specific information source to obtain different articles according to the given function.

Printed dictionaries can be categorized according to the intended user groups, e.g. for laypersons, semi-experts, experts or school learners in a specific age group. In electronic dictionaries, access to predefined subsets of data from the dictionary database is possible via a function button for each user group and/or type of user need.

4.1.2 Information source usage situations

An information source usage situation refers to a situation where there is at least one available information source, as well as the necessary time and energy to consult one or more of these sources. If, however, these conditions are not met, the information access process is terminated.

4.2 Intra-consultation phase

4.2.1 Selection of an information source

The situation and the type of user will determine whether someone looking for information will turn to a non-human source (such as a printed dictionary, an electronic dictionary, a text book, a Google search), or a human source (by asking a person). In this article we do not discuss humans as information sources, but only the process of consulting non-human sources.

4.2.1.1 Choice of the component of an information source

The most important information source will be dictionaries, lexicons and encyclopaedias. In accordance with the reference needs stemming from a communicative, cognitive, interpretative or operative situation one would consult bigger or smaller or even minute components of an information source. This could depend on the level of expertise of the user, i.e. whether the user is a layperson, a semi-expert or an expert. It can also be determined by whether the information source is a monofunctional or a polyfunctional source. In accordance with the choice of information source, one would first consult the table of contents, index or another outer text. In all information tools, including dictionaries, one would usually go directly to the lemma list which could be ordered alphabetically or systematically. However, in dictionaries with a dictionary grammar or a systematic introduction one might first look there.

4.2.2 Consultation of an information source

For each situation where a source is consulted, one can talk about a consultation. In some instances a single consultation could be linked to different search strings. There may be chains of consultations in instances where information may create the need for more information and follow-up consultations — therefore an iterative process.

4.2.2.1 Access by means of metadata

Metadata is data about data. We divide access through metadata into three

categories, viz. access through a table of contents, an index and markers.

— **Table of contents**

A table of contents is a standard method of access in most paper-based information sources, where one gets directions to move directly to a specific part of an information source, e.g. a reference to a page number. In e-resources, items in such a table of contents can be linked to the specific item, and the user can then navigate directly to the linked item by clicking on it.

— **Index**

Users can employ index access only in printed dictionaries with one or more indexes. In electronic information sources this is possible in a few works by means of the access mode developed by the author of the work, with the possibility of reaching the required place by means of a partial index or the complete index.

— **Markers**

Markers can be either a sign such as a graphic symbol or a word/phrase indicating subsections in articles, for example to distinguish between collocations, idioms, examples, style, etc. The user therefore has the option to skip those sections which are not relevant for his/her information needs.

4.2.3 Consultation in a part of a component

When a user reaches a specific part of a component, either a section in a text book or a dictionary article, the article structure (in dictionaries) or subheadings or headwords on the edge of the page can be helpful to achieve a more rapid access. It is important to note that the real person looking for information does not proceed in a linear way, i.e. not from the beginning of a chapter or a dictionary article, but sometimes from the back, repeating some partial processes because new ones did not give the necessary results.

4.2.3.1 System features

— **Searching**

Searching implies that the user has to define a search string which can consist of one or more words by means of which an information source is searched. In a very detailed information need a user could also combine various search strings by means of brackets or Boolean operators (and, or, not). The user, in addition, has the option to specify whether search terms should be truncated and can also specify so-called "wild cards", i.e. to make provision for spelling

variations such as *optimise/optimize* and *behavior/behaviour*, or distance between items (word X within Y words from word Z). Search interfaces could be "simple" (e.g. the standard Google interface) or "advanced" which may make use of drop down menus or fill-in forms (e.g. the advanced search option in Google Scholar) or a command line interface (as is found as one of the options in most bibliographic databases). Search therefore implies cognitive input from the user; s/he has to decide which words/phrases/strings to use, how to combine them, how to specify alternatives etc. After completing the input, the user then clicks on search and the system does retrieval based on the specified criteria. This applies only to electronic information tools.

In printed books one would rather talk about *search word* or *search expression*. It is the expression that the user has in mind when starting the access process. In a printed dictionary the user knows from experience that one usually looks for the basis form of an inflected lexeme which can be found in the guiding element of a dictionary article, i.e. the lemma. However, in, for example, printed specialised dictionaries, idiom dictionaries or dictionaries of proverbs, one can also look for multiword expressions. This also applies to all forms of text books and manuals. In all forms of printed reference works with a register of meanings or concepts one can execute a semantic search by employing an appropriate search word or expression. In printed works one can employ all types of search words and expressions to access both the expression and the contents by going directly to a semantic field or a field with meaning annotations.

In a similar way, access options are found in printed books too where one would accordingly search in the table of contents or in the chapter headings, and in dictionaries at the start of the lemma field but in dictionaries of fixed expressions often in the middle of the lemma field, e.g. in a dictionary of idioms where one can look for the expression *cut off one's* ***nose*** *to spite one's face* with the search string *nose.*

— **Linking**

Linking in the electronic environment implies clicking on links that are indicated in the information source. Links are typically indicated by means of some or other visual representation, for example a change in colour of the words/ phrases or underlining the text as is standard on the web. In some cases, such visual indications are absent and all words in the text could be clickable (e.g. linking all words to a glossary or cross references in dictionaries). In printed works, a cross-reference can start a new access process or continue an existing one, as is the case in electronic works, where there may be many links that could also be used extensively.

Linking could have a specific goal in mind, i.e. the user has a specific need and tries to solve this problem by following one or more links in sequence. The user could also serendipitously follow links based on a possible interest in what may be available, even leading to infotainment or lexicotainment. Typi-

cally these two types of linking are distinguished by means of different terms such as navigation and browsing; see also Bothma (2011).

— **Combining searching and linking**

In a typical consultation process, users combine searching and linking iteratively to obtain the required information. A user may conduct a search and the results of the search may be a list of items that are clickable; the user then has to click on one of these links to obtain the required information. This is obvious in any web search where the user has to enter a search string in a search box, and after executing the search the user obtains a list of results which could be one item or millions. The user then has to click on some of these results to acquire the required information. In an e-dictionary environment, exactly the same applies, for example in THE DANISH DICTIONARY OF FIXED EXPRESSIONS: If the search gives one result the dictionary article is displayed; however, in the case of multiple results, a brief summary is displayed and the user has to click on one of these to view the dictionary article.

4.2.4 Information retrieved

The retrieval process may result in the user retrieving information satisfying his/her information need, or the retrieval may yield no relevant information at all. If the user has retrieved no information or the information that has been retrieved has not satisfied the user's need, the user has two options. The user could start from scratch and through an iterative process decide on different information sources or different access strategies. Alternatively, the user can decide to abandon the process and this process therefore terminates. If the user has relevant information, s/he may decide to act on the information. However, at this stage it is not evident that the data are necessarily correct and/or to be trusted. The data are only perceived as relevant because the user has not yet evaluated the data for validity.

4.3 Post-consultation phase

4.3.1 Use of the found information

The use of the information depends on the type of user situation: In a communicative situation, one has a concrete problem in a text. After having found the needed information, s/he goes back to the text and uses this information. The same can be said in an operative and an interpretative situation. After having obtained the required information, the user knows how to act or what a sign means and can continue the acting process s/he was in. In a cognitive situation the user obtained an answer to his/her need for knowledge. Whether s/he uses this knowledge now, later or never is not the topic here. The information can

play a role for further acting but this is not required at this stage, as the information is only added to the user's knowledge base for possible future use.

4.3.2 Meta-issues

4.3.2.1 Consultation route

The German equivalent for this term is *Suchpfad*, and in English it is also referred to as *access route*. It encompasses the process from the selection of an information source up to reaching the destination or destinations in the information source and the eventual conclusion of the search in terms of the specific consultation as either successful or unsuccessful.

4.3.2.2 Consultation step

This refers to the clearly distinguishable steps from the first stage in the search to the next, for example from the selection of a dictionary up to the selection of a component of the dictionary or from the stage in the article stretch of the letter *S* up to the required article starting with the letter *S* has been found. This also applies to all other types of information sources: With every change in the search route, when the direction of the search or the place of the search is changed, a new search step is introduced. The use of this term is not unambiguous because it depends on the circumstances of how the search is observed. When a user is observed who talks to him-/herself aloud during the access process or has been instructed to do so, the investigation will be able to show more search steps compared to an investigation without such comments.

4.3.2.3 Step time

This term refers to the time from the start of one stage up to the next stage. In Bergenholtz and Gouws (2010) the step time was not recorded; in the current experiments, it was, however, done, as indicated in Section 4 above. It could be interesting in further investigations to obtain more empirical results using the same experiments with different test persons.

4.3.2.4 Consultation time

This refers to the total time of a given consultation, i.e. from identifying the problem up to the successful conclusion or a negative termination of the search. Consultation time can then be seen as the duration between the first search step and the last, or the sum of step time.

4.3.2.5 Success

Success implies that the user is satisfied with the result and in his/her perception the information fulfilled the information need s/he had. However, this does not necessarily mean that the information is correct in terms of factual correctness.

5. Conclusion, future research and future information tools

From the preceding discussions, it is clear that sequential phases consisting of different steps structure the access. In addition, it is clear that some of these terms are not *steps*, but issues on which the lexicographer and/or the end-user should reflect after completing a work task.

Our description of the complexity of the access process and the missing help for operative problems in some accounting dictionaries should lead to the conception of more advanced information tools — not like the current traditional paper dictionaries, but more like those from the 18th century. They, or at least some of them, have extensive operative parts of the dictionary articles; see Tarp and Bothma (2013). We have shown in one example for an operative problem the complexity of the information needs. No dictionary gave any kind of help, but in future dictionaries it could be done, perhaps not in same way as in the paper dictionaries of the 18th century, but with the same function in new information tools. And then it is not important whether we call such a new tool a dictionary or something else.

As it is now, the situation is very complex. The *Junior Woodchucks Guidebook* is up to now only the solution in the Donald Duck comics. Nowadays users have to consult multiple sources and therefore have either to

— know about these sources, or
— rely on Google or specific databases to find these sources.

There is no single integrated solution available. In these cases, additional information sources were needed to resolve the information need. We need more empirical evidence to illustrate this.

This empirical evidence could be found through extensive experiments, as we have shown for two case studies, with six and four experiments respectively, each, however, with only one respondent. It should be done with a greater number of test persons. We are here not speaking about a representative survey, but case studies with something between 20 and 50 test persons for different kinds of knowledge fields. In addition, it should not only be done for a single problem, but for a whole work task, e.g. testing the access process for reception problems reading a whole text, or translating problems translating a whole text, or operative problems connected to one and the same acting problem. Lexicographers can play an important role together with subject field specialists to provide such integrated sources.

6. References

Bergenholtz, H. 2009. Schnellerer und sicherer Datenzugriff in gedruckten und elektronischen Fachwörterbüchern und Lexika. *Revue française de linguistique appliquée, dossier: terminologie orientations actuelle* 14(2): 81-97.

Bergenholtz, H. 2011. Access to and Presentation of Needs-adapted Data in Monofunctional Internet Dictionaries. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011. *e-Lexicography: The Internet, Digital Initiatives and Lexicography*: 30-53. London/New York: Continuum.

Bergenholtz, H. and T.J.D. Bothma. 2011. Needs-adapted Data Presentation in e-Information Tools. *Lexikos* 21: 53-77.

Bergenholtz, H. and R.H. Gouws. 2008. The Access Process in Dictionaries for Fixed Expressions. *Lexikographica. International Annual for Lexicography* 23: 237-260.

Bergenholtz, H. and R.H. Gouws. 2010. A New Perspective on the Access Process. *Hermes. Journal of Language and Communication Studies* 44: 103-127.

Bergenholtz, H. and M. Johnsen. 2013. User Research in the Field of Electronic Dictionaries: Methods, First Results, Proposals. Gouws, R.H., U. Heid, W. Schweickard, H.E. Wiegand. 2013: 556-568.

Bergenholtz, H. and B. Norddahl. 2012. Ordbogsartikler, som ingen læser. *LexicoNordica* 19: 206-222.

Bergenholtz, H. and S. Tarp. 2002. Die moderne lexikographische Funktionslehre. Diskussionsbeitrag zu neuen und alten Paradigmen, die Wörterbücher als Gebrauchsgegenstände verstehen. *Lexicographica. International Annual for Lexicography* 18: 253-263.

Bergenholtz, H. and S. Tarp. 2003. Two Opposing Theories: On H.E. Wiegand's Recent Discovery of Lexicographic Functions. *Hermes. Journal of Linguistics* 31: 171-196.

Bothma, T.J.D. 2011. Filtering and Adapting Data and Information in the Online Environment in Response to User Needs. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011. *e- Lexicography: The Internet, Digital Initiatives and Lexicography*: 71-102. London/New York: Continuum.

Bothma, T.J.D. and S. Tarp. 2012. Lexicography and the Relevance Criterion. *Lexikos* 22: 86-108.

DICTIONARY OF FINANCE AND BANKING = A Dictionary of Finance and Banking. (4th revised edition) http://www.oxfordreference.com/view/10.1093/acref/9780199229741.001.0001/acref-9780199229741 (accessed December 2014).

FINANCE GLOSSARY = *Financeglossary. The Online Dictionary of Financial Terms*. http://www.finance-glossary.com/ (accessed December 2014).

Flyvbjerg, B. 2006. Five Misunderstandings About Case-Study Research. *Qualitative Inquiry* 12(2): 219-245.

FREE FINANCIAL DICTIONARY = The Free Dictionary by Farlex. http://financial-dictionary.thefreedictionary.com/ (accessed December 2014).

Fuertes-Olivera, Pedro A. et al. (Eds.). 2015. *Diccionario Inglés-Español de Contabilidad: Recepción*. Odense: Lemma.com.

Fuertes-Olivera, P.A. and S. Nielsen. 2012. Online Dictionaries for Assisting Translators of LSP Texts: The Accounting Dictionaries. *International Journal of Lexicography* 25(2): 191-215.

Fuertes-Olivera, P.A. and S. Tarp. 2014. *Theory and Practice of Specialised Online Dictionaries: Lexicography versus Terminography.* Berlin/New York: De Gruyter.

Gouws, R.H., U. Heid, W. Schweickard and H.E. Wiegand (Eds). 2013. *Dictionaries. An International Encyclopedia of Lexicography*. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography. Berlin/Boston: Walter de Gruyter.

Ingwersen, P. 2007. Context in Information Interaction — Revisited 2006. Bothma, T.J.D. and A. Kaniki (Eds.). 2007. *ProLISSA 2006: Proceedings of the Fourth Biennial DISSAnet Conference, Farm Inn, Pretoria, South Africa, 2–3 November*: 13-23. Pretoria: Infuse.

Ingwersen, P. and K. Järvelin. 2004. Context in Information Interaction. Bothma, T.J.D. and A. Kaniki (Eds). 2004. *Progress in Library and Information Science in Southern Africa (ProLISSA): Proceedings of the Third Biennial DISSAnet Conference, Farm Inn, Pretoria, South Africa, 28–29 October*: 301-310. Pretoria: Infuse.

Ingwersen, P. and K. Järvelin. 2005. *The Turn: Integration of Information Seeking and Retrieval in Context.* Dordrect: Springer.

INVESTOPEDIA = *(I)Investopedia. Dictionary.* http://www.investopedia.com/dictionary/ (accessed December 2014).

Kilgarriff, A. 2012. Review of Fuertes-Olivera, Pedro A. and Henning Bergenholtz (Eds.). 2011. *e-Lexicogaphy. The Internet, Digital Initiatives and Lexicography*. *Kernerman Dictionary News*, July 2012: 26-29.

Kuhn, T.S. 1987. What are Scientific Revolutions? Kruger, L., L.J. Daston and M. Heidelberger (Eds.). 1987. *The Probabilistic Revolution, Volume I: Ideas in History*: 7-22. Cambridge, MA: MIT Press.

MERRIAM WEBSTER = *Merriam Webster. Dictionary*. http://www.merriam-webster.com/dictionary/finance/ (accessed December 2014).

Nielsen, S. 1995. Access Structure. Bergenholtz, H. and S. Tarp (Eds.). 1995. *Manual of Specialised Lexicography: The Preparation of Specialised Dictionaries*: 219-223. Amsterdam/Philadelphia: John Benjamins.

Tarp, S. 2008. Revival of a Dusty Old Profession. *Hermes. Journal of Linguistics* 41: 175-188.

Tarp, S. 2008a. *Lexicography in the Borderland Between Knowledge and Non-knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography.* Tübingen: Max Niemeyer.

Tarp, S. and T.J.D. Bothma. 2013. An Alternative Approach to Enlightenment Age Lexicography: The Universal Dictionary of Trade and Commerce. *Lexicographica* 29: 222-284.

THE DANISH ACCOUNTING DICTIONARY = Nielsen, S., L. Mourier, H. Bergenholtz: *Den Danske Regnskabsordbog: Betydning.* Database: Almind, R. and J.S. Nielsen. Odense: Ordbogen.com 2014. http://www.ordbogen.com (accessed December 2014).

THE DANISH DICTIONARY OF FIXED EXPRESSIONS = Bergenholtz, H. in collaboration with E. Bjærge: *Viden om faste vendinger.* Database: Almind, R. Odense: Ordbogen.com 2014. http://www.ordbogen.com (accessed December 2014).

THE DANISH INTERNET DICTIONARY = Bergenholtz, H. in collaboration with H. Agerbo in cooperation with A. Bock Michelsen, A. Bodilsen, K. Brosbøl Eriksen, C. Brun, C. Kastorp, I. Koed, J. Nguyen, J. Nørgaard Poulsen, H. Thers and A. Toftgaard Madsen. Database: Almind, R. and M. Carlsen. *Den Danske Netordbog.* Odense: Ordbogen.com 2015. http://www.ord bogen.com (accessed August 2015).

THE DANISH MUSIC DICTIONARY = Bergenholtz, I. in collaboration with H. Bergenholtz. Database: Almind, R. and M. Gyde Poulsen: *Viden om musikudtryk.* Odense: Ordbogen.com 2014. http://www.ordbogen.com (accessed December 2014).

THE ENGLISH–DANISH ACCOUNTING DICTIONARY = Nielsen, S., L. Mourier and H. Bergenholtz: *Den Dansk–Engelske Regnskabsordbog: Oversættelse.* Database: R. Almind and J.S. Nielsen. Odense: Ordbogen.com 2014. http://www.ordbogen.com (accessed December 2014).

Wiegand, H.E. 2007. Über Zugriffspfade in Printwörterbüchern. Ein Beitrag zur Schnittstelle von Benutzungshandlungen und Wörterbuchform. *Lexikos* 17: 180-211.

Wiegand, H.E. and S. Beer. 2013. Access Structures in Printed Dictionaries. Gouws, R.H., U. Heid, W. Schweickard and H.E. Wiegand (Eds). 2013: 110-149.

Wiegand, H.E. and R.H. Gouws. 2013. Macrostructures in Printed Dictionaries. Gouws, R.H., U. Heid, W. Schweickard and H.E. Wiegand. 2013: 73-110.

Wiegand, H.E. and M. Smit. 2013. Microstructures in Printed Dictionaries. Gouws, R.H., U. Heid, W. Schweickard and H.E. Wiegand. 2013: 149-214.

'n Evaluasie van *Pharos Aanleerderwoordeboek vir skole/ Learner's Dictionary for Schools* en *Longman-HAT Afrikaans Dictionary and Grammar for English Speakers* in 'n Afrikaanse taalverwerwingskursus op universiteitsvlak*

Dané Claassen, *Departement Afrikaans en Nederlands, Universiteit Stellenbosch, Stellenbosch, Suid-Afrika (16018583@sun.ac.za)*

Opsomming: Navorsing toon dat 'n goed gekose woordeboek, naas die opvoeder, 'n belangrike hulpmiddel is om die taalverwerwingsproses te verryk. Daar bestaan tans nie 'n eentalige of tweetalige *aanleerderwoordeboek* wat spesifiek gerig is op *nultaalsprekers* wat *Afrikaans* op tersiêre vlak aanleer nie. Taalverwerwingsdosente en taalfasiliteerders moet dus die gaping ten beste probeer vul deur *aanleerderwoordeboeke* soos Pharos se *Aanleerderwoordeboek vir skole/Learner's Dictionary for Schools* en Longman-HAT se *Afrikaans Dictionary and Grammar for English Speakers* vir hierdie aanleerders voor te skryf. In hierdie artikel word dié twee *aanleerderwoordeboeke* empiries en kwalitatief geëvalueer ten einde vas te stel of daar wel ruimte is vir die samestelling van 'n *aanleerderwoordeboek* wat spesifiek gerig is op *nultaalsprekers* wat *Afrikaans* aanleer op 'n universiteitsvlak. Hierdie spesifieke twee woordeboeke word geëvalueer omdat dit onderskeidelik voorgeskryf word vir *nultaalsprekers* by die Universiteit van Stellenbosch en die Nelson Mandela Metropolitaanse Universiteit. Die mate waarin 'n spesifieke woordeboek voldoen aan die bevrediging van 'n sekere teikengroep se leksikografiese behoeftes word nie slegs deur die leksikograaf bepaal nie, maar ook deur die teikengebruiker. Die huidige *woordeboekkultuur*, soos geformuleer in die *Nasionale Kurrikulum- en Assesseringsbeleidsverklaring (NKAV)* wat tans in die Suid-Afrikaanse skoolsisteem gevolg word, word onder die vergrootglas geplaas. Op grond hiervan word voorstelle vir formele *woordeboekonderrig* op universiteitsvlak gemaak, wat beide algemene en spesifieke *woordeboekonderrig* insluit.

Sleutelwoorde: AANLEERDERWOORDEBOEK, SENTRALE LYS, NULTAALSPREKERS, AFRIKAANS, BUITETEKSTE, WOORDEBOEKVAARDIGHEDE, WOORDEBOEKONDERRIG,

* Hierdie artikel is gebaseer op 'n gedeelte van 'n M.A.-verhandeling getiteld *'n Aanleerderwoordeboek vir nultaalsprekers in 'n Afrikaanse taalverwerwingskursus op universiteitsvlak* wat eersdaags voltooi word. Dit is ook aangebied as lesing by die Twintigste Jaarlikse Internasionale Konferensie van die African Association for Lexicography (AFRILEX), Universiteit van KwaZulu-Natal, Durban, Suid-Afrika, 6–8 Julie 2015.

GEVALLESTUDIES, WOORDEBOEKKULTUUR, NASIONALE KURRIKULUM- EN ASSESSERINGSBELEIDSVERKLARING (NKAV), DEPARTEMENT VAN BASIESE ONDERWYS

Abstract: An evaluation of *Pharos Aanleerderwoordeboek vir skole/Learner's Dictionary for Schools* and *Longman-HAT Afrikaans Dictionary and Grammar for English Speakers* in an Afrikaans language learning course at university level. Research shows that a well-chosen dictionary is the next best resource after the facilitator to enhance the language learning process. There are currently no monolingual or bilingual *learner's dictionaries* specifically targeting *foreign learners* who are learning *Afrikaans* at a tertiary level. Educators and facilitators must thus compensate by prescribing *learner's dictionaries* such as Pharos's *Aanleerderwoordeboek vir skole/Learner's Dictionary for Schools* and Longman-HAT's *Afrikaans Dictionary and Grammar for English Speakers* for these learners. In this article these two dictionaries are being evaluated empirically and qualitatively to determine if there is a need for the composition of a *learner's dictionary* which specifically targets *foreign learners* learning *Afrikaans* at a university level. These specific two *learner's dictionaries* are being evaluated because they are prescribed for *foreign students* respectively at, the University of Stellenbosch and the Nelson Mandela Metropolitan University. The degree to which a certain dictionary succeeds in satisfying the lexicographical needs of a specific target group does not only depend on the lexicographer, but also on the target group itself. The current *dictionary culture* in South Africa is brought to light through reviewing the *Curriculum Assessment Policy Statements (CAPS)*, which is currently being followed in the South African school system. On this basis, suggestions are made for formal *dictionary education* at tertiary level which includes general and specific *dictionary education*.

Keywords: LEARNER'S DICTIONARY, CENTRAL LIST, FOREIGN LEARNERS, AFRIKAANS, OUTER TEXTS, DICTIONARY SKILLS, DICTIONARY EDUCATION, CASE STUDIES, DICTIONARY CULTURE, CURRICULUM ASSESSMENT POLICY STATEMENTS (CAPS), DEPARTMENT OF BASIC EDUCATION

1. Inleiding

'n Aanleerderwoordeboek is 'n woordeboek wat saamgestel word vir volwasse aanleerders van 'n vreemde taal (Tarp en Gouws 2004: 279). Só 'n woordeboek is gerig op die verbetering van niemoedertaalsprekers se kommunikatiewe vermoëns (Otto 1999: 136). Dit is dus 'n woordeboek wat 'n aanbieding en bewerking van leksikale items bied waarmee die teikengebruiker in die daaglikse taalgebruik gekonfronteer word (Tarp en Gouws 2004: 279). Die hooffunksies van hierdie tipe woordeboek is teksresepsie in die vreemde taal, teksproduksie in die vreemde taal, 'n bewusmaking van die leksikon van die vreemde taal en 'n bewusmaking van die grammatika van die vreemde taal (Tarp 2004: 223).

Tydens die afgelope dekades is daar baie navorsing oor aanleerderwoordeboeke gedoen. Ondanks al hierdie navorsing is die rol van aanleerderwoordeboeke in tersiêre onderrig met nultaalsprekers as teikengroep nog nie naastenby na behore ondersoek nie. Dit is die onderwerp van hierdie artikel.

Uit die navorsing wat oor aanleerderwoordeboeke gedoen is, blyk dit dat

hierdie woordeboeke 'n belangrike rol in enige taalverwerwingsproses speel. Vergelyk in hierdie verband Lew (2004), Shanshan (2008) en Augustyn (2013). Die opvoeder is nie altyd beskikbaar nie en dus nie altyd 'n opsie as hulpmiddel nie. Daarom is 'n goed gekose woordeboek waarskynlik een van die beste hulpmiddels, naas die opvoeder, om leksikale en algemene taalgebruiksprobleme aan te pak wat ondervind kan word gedurende die taalverwerwingsproses (Lew en Adamska-Sałaciak 2014: 1).[1]

Daar bestaan tans nie 'n eentalige of tweetalige aanleerderwoordeboek wat spesifiek gerig is op nultaalsprekers wat Afrikaans aanleer op tersiêre vlak nie. Daar moet dus gekompenseer word deur woordeboeke soos Pharos se *Aanleerderwoordeboek vir skole/Learner's Dictionary for Schools* en Longman-HAT se *Afrikaans Dictionary and Grammar for English speakers* vir nultaalsprekers voor te skryf.[2]

Woordeboekgebruik is 'n aktiewe proses tussen die gebruiker en die woordeboek; die gebruiker is nie 'n passiewe deelnemer nie. Hierdie posisie wat die gebruiker inneem as aktiewe deelnemer moet gevestig word in 'n woordeboekkultuur. 'n Woordeboekkultuur verwys na die mate waarin 'n gemeenskap vertroud is met woordeboeke, die tipologiese verskeidenheid, die inhoud, strukture en funksies van woordeboeke asook die gebruiksvaardighede. Die ideaal is dus 'n woordeboek wat 'n gebruikersvriendelike aanbieding en bewerking bevat, wat nultaalsprekers se leksikografiese behoeftes in 'n bepaalde konteks bevredig en wat toeganklik is vir nultaalsprekers sodat dit die taalverwerwingsproses kan vergemaklik (Hausmann 1989: 15). Instellings wat 'n woordeboekkultuur handhaaf om hierdie ideale woordeboek te ondersteun, is ook belangrik.

'n Nultaalspreker verwys, in die konteks van hierdie artikel, na 'n volwassene wat Afrikaans as derde of verdere taal (vreemde taal) op universiteitsvlak aanleer en wie se eerste taal nie Engels is nie, maar 'n ander taal.

In hierdie artikel word die gegewe aanleerderwoordeboeke geëvalueer met behulp van twintig studente wat ingeskryf is vir die module Afrikaanse Taalverwerwing 178 aan die Departement Afrikaans en Nederlands by die Universiteit van Stellenbosch. Hierdie evaluering geskied in die vorm van 'n gevallestudie en terugvoer van hierdie prototipe-teikengebruiker. Studente is gevra om as deel van gewone werk in hierdie module basiese woordeboekvaardighede te gebruik om oefeninge te voltooi sonder, asook met behulp van onderskeidelik Pharos se woordeboek en Longman-HAT se woordeboek. Hulle is ook verder gevra om skriftelike kommentaar te lewer op die gegewe woordeboek wat aan hulle toegeken is.

Voorts word die huidige woordeboekkultuur vasgestel deur die Nasionale Kurrikulum- en Assesseringsbeleidsverklaring (NKAV), wat tans in die skoolsisteem gevolg word vir die vak Afrikaans as Addisionele taal, onder die loep te neem. Woordeboeke word onder "kernmateriaal" in dié kurrikulum aangegee, maar daar is geen aanduiding gegee oor watter tipe woordeboek benodig word, of hoe om 'n woordeboek te gebruik nie (Departement van Basiese Onderwys 2011). Hierdie onvoldoende leiding rakende die gebruik van woor-

deboeke op skoolvlak lei tot 'n onvoldoende woordeboekkultuur op skoolvlak en sodoende ook op universiteitsvlak.

2. Gevallestudie

Gevallestudies word al reeds vir dekades lank in wetenskaplike navorsing gebruik. Die waarde van hierdie gebruik van gevallestudies in wetenskaplike navorsing is ook reeds deurentyd bewys. Vergelyk in hierdie verband Bergenholtz en Gouws (2010), Flyvbjerg (2006) en Kuhn (1987).[3]

Vir hierdie ondersoek het al die studente wat op 'n gegewe dag klas vir die module Afrikaanse Taalverwerwing 178 aan die Departement Afrikaans en Nederlands by die Universiteit van Stellenbosch bygewoon het, deelgeneem aan hierdie gevallestudie. Dit was twintig uit die vyf en twintig geregistreerde studente. Daar is dus gebruik gemaak van 80% van die studente wat hierdie module neem. Daar is 'n toets saamgestel uit werk wat deel vorm van die studiegids van die module. Hierdie "toets" is in die klas deur die studente gedoen as deel van hulle normale klaswerk. Dieselfde toets is as toets 1 (T1) en toets 2 (T2) gebruik. Die toets het uit vyftig punte bestaan en daar is vyftig minute (die tydperk van 'n lesing) vir elk van die toetse gegee.

Die toetse is aan die begin van die jaar gedoen voordat die studente werklik begin Afrikaans aanleer het. Die rede hiervoor was sodat die resultate van T2 grotendeels toegeskryf kon word aan 'n woordeboek as hulpmiddel. Albei toetse is gedoen voordat enige vorm van evaluering van die module deur die dosent met die studente gedoen is. Die dosent het ook verder bevestig dat die werk wat in die toets is, nog nie met die studente behandel is nie (Adendorff 2015).

T1 is gedoen na slegs drie weke van lesings. Studente is gevra om alle hulpmiddels weg te sit en as deel van hul klaswerk, die toets te doen en in te vul wat hulle kon. T2 is die volgende week gedoen. T2 is gedoen met behulp van óf Pharos se *Aanleerderwoordeboek vir skole/Learner's Dictionary for Schools* óf Longman-HAT se *Afrikaans Dictionary and Grammar for English Speakers*. Die betrokke woordeboeke is op 'n lukrake wyse aan die studente uitgedeel en hulle kon nie kies watter een hulle wou gebruik nie. Die woordeboeke is egter gelyk verdeel. Tien studente het Pharos se *Aanleerderwoordeboek vir skole/Learner's Dictionary for Schools* as hulpmiddel gebruik en tien studente het Longman-HAT se *Afrikaans Dictionary and Grammar for English Speakers* as hulpmiddel gebruik. Die rede vir die lukrake toekenning van die woordeboeke was om die intelligensievlak en woordeboekvaardighede van elke individuele student as veranderlikes te hanteer wat nie beheer kan word nie en om die invloed van elk te minimaliseer.

Die toets is saamgestel om die studente se begrip van Afrikaans te toets asook om die effek van 'n woordeboek as hulpmiddel, in geheel, te toets eerder as om die studente se woordeboekvaardighede te toets. Dit is dus nie as 'n woordeboektoets aan die studente bekend gestel nie. Die doel van die toets was wel om die effek van 'n woordeboek as hulpmiddel onder die vergrootglas te

bring, maar ook om op 'n implisiete wyse sekere strukture van die betrokke woordeboeke te evalueer. Die doel was dus ook om 'n woordeboek met 'n meer tradisionele struktuur soos Pharos se *Aanleerderwoordeboek vir skole/Learner's Dictionary for Schools*, wat slegs bestaan uit 'n sentrale lys, te vergelyk met 'n woordeboek wat 'n nuwer benadering het soos Longman-HAT se *Afrikaans Dictionary and Grammar for English Speakers*, wat ook naas die sentrale lys buitetekste bevat waarin 'n wesenlike deel van die relevante leksikografiese data aangebied word. Dit is op 'n implisiete wyse gedoen deur die volgende vrae wat daarop gerig is om die gebruikers se toegang tot die data en die onttrekking van inligting te evalueer.

Toetsvraag	**Pharos se woordeboek Waar kan antwoord gevind word:**	**Longman-HAT se woordeboek Waar kan antwoord gevind word:**
1. Gee die korrekte meervoudsvorme (plural forms) vir die volgende woorde	Woordeboekartikel	Woordeboekartikel
2. Gee die korrekte verkleiningsvorme (dimunitive) vir die volgende woorde	Woordeboekartikel	Woordeboekartikel
3. Gee die korrekte Afrikaanse vertalings (translations) vir die vier seisoene	Woordeboekartikel	Middeltekste
4. Gee die regte vorm (correct vorm) van die byvoeglike naamwoord (adjective) tussen hakies	Woordeboekartikel	Woordeboekartikel
5. Vul die korrekte intensiewe vorm (intensive form) in.	Woordeboekartikel	Agtertekste
6. Trappe van vergelyking (Degrees of comparison)	Woordeboekartikel	Woordeboekartikel
7. Bestudeer die onderstaande prentjies en voltooi die sinne deur die korrekte voorsetsel (preposition) in te vul	Woordeboekartikel	Middeltekste

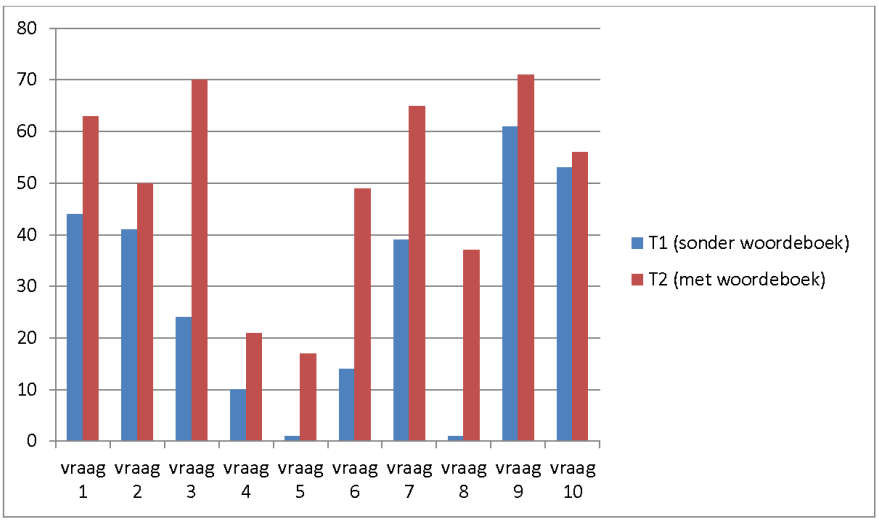
8. Maak 'n sin (make a sentence) met die volgende woorde	Woordeboekartikel	Middeltekste
9. Is die volgende stellings waar (true) of onwaar (false)?	Woordeboekartikel	Woordeboekartikel
10. Kies (choose) die korrekte woord tussen hakies (brackets)	Woordeboekartikel	Woordeboekartikel

Tabel 1: Toegangstrukuur van die twee woordeboeke

3. Resultate en bespreking

Toets 1 is gedoen sonder enige hulp van 'n woordeboek. Die gemiddelde punt van die twintig studente was 18 uit 50. Die gemiddelde persentasie van T1 was dus 36%. Toets 2 is gedoen met behulp van óf Pharos se woordeboek óf Longman-HAT se woordeboek. Die algehele gemiddeld (afgesien van watter woordeboek) van T2 was 28 uit 50. Die gemiddelde persentasie van T2 was dus 56%. Dit wil sê daar was 'n algehele verbetering van 20% met een van die betrokke woordeboeke as hulpmiddel. Dit dui op 'n beduidende verbetering in punte wat grotendeels aan die hulp van 'n woordeboek toegeskryf kan word. Alhoewel dieselfde vrae in beide T1 en T2 gevra is, is die korrekte antwoorde nie na T1 aan die studente gegee nie en omdat T2 in die daaropvolgende week gedoen is, kon daar nie veel van die werk wat in die toetse gevra is deur die dosent behandel word nie. Die studente moes ook die vrae inhandig na T1.

In figuur 1 volg 'n opsomming van die resultate van die twee toetse, gevolg deur 'n bespreking van die belangrikste tien toetsvrae waarin die toegangstrukture en dataverspreidingstrukture van albei woordeboeke implisiet geëvalueer is. In die toetsvrae word veral aandag geskenk aan die interne toegangstruktuur waarvan die sukses medebepaal word deur die dataverspreidingstruktuur, ordening van datatipes in die mikrostruktuur en die dataargitektuur (Steyn 2004: 288-290). Een van die grootste probleme wat teikengebruikers ondervind in hulle interne soektog is die onduidelike verdeling tussen verskillende datakategorieë (Steyn 2004: 290). Al hierdie strukture word implisiet geëvalueer omdat dit soms problematies kan wees en sodoende 'n groot impak kan hê op die sukses van die onttrekking van die korrekte inligting.



Figuur 1: Toetsresultate met en sonder woordeboeke

Vervolgens 'n bespreking van tien geselekteerde toetsvrae:

(1) Gee die korrekte meervoudsvorme (plural forms) vir die volgende woorde:

1. Blom _____

2. Oog _____

3. Broer _____

4. Taak _____

Die gemiddelde korrekte antwoorde vir hierdie vraag gedurende T1 was 1.75 uit 4. Die meervoudsvorme van *oog* en *taak* het die meeste probleme veroorsaak. Die gemiddelde korrekte antwoorde vir hierdie vraag gedurende T2, met behulp van een van die betrokke woordeboeke, was 2.5 uit 4. Die studente wat Longman-HAT se woordeboek as hulpmiddel gebruik het, het 'n gemiddelde punt van 1.8 uit 4 vir hierdie vraag gekry, terwyl die studente wat Pharos se woordeboek gebruik het 3.2 uit 4 vir hierdie vraag gekry het. Al vier die betrokke woorde is in die sentrale lys van Pharos se woordeboek as lemmas opgeneem. Daarteenoor is slegs *oog* en *taak* in Longman-HAT se sentrale lys te vinde. Die woord *broer* word nie opgeneem in Longman-HAT se woordeboek nie, maar *boetie* verskyn wel as lemma. Dit is 'n vreemde keuse deur die leksikograaf om eerder *boetie* as *broer* as lemma op te neem, aangesien *broer* meer gebruiksfrekwent is.

(2) Gee die korrekte verkleiningsvorme (dimunitive) vir die volgende woorde:

1. Berg _____
2. Hond _____
3. Seun _____
4. Klip _____

Die gemiddelde korrekte antwoorde vir hierdie vraag gedurende T1 was 1.65 uit 4. Die studente het baie met die verkleiningsvorme gesukkel en sommige studente het nie eers geweet wat 'n "dimunitive" is nie. Dit beklemtoon weereens dat 'n woordeboek vir hierdie teikengebruikers eenvoudige Engels moet bevat omdat hulle eerste taal nie Engels is nie. Die gemiddelde korrekte antwoorde vir hierdie vraag gedurende T2 was 2 uit 4. Die studente wat Longman-HAT se woordeboek gebruik het, het 'n gemiddelde punt van 2.3 uit 4 vir hierdie vraag gekry, terwyl die studente wat Pharos se woordeboek gebruik het slegs 1.7 uit 4 gekry het.

Verkleiningsvorme word in die woordeboekartikels van Longman-HAT verskaf terwyl dit nie dieselfde geval is in Pharos se woordeboek nie. Die woorde *hondjie* en *seuntjie* word wel in Pharos se woordeboek opgeneem as lemmas en as sodanig bewerk. Daar bestaan nie 'n vertalingsekwivalent vir die woord *puppy* in Afrikaans nie. Die opname van die lemma *hondjie* deur die leksikograaf kan dus ondersteun word as 'n hantering van zero-ekwivalensie. Die opname van die lemma *seuntjie* is egter 'n lukrake keuse waarvan die besluit van die leksikograaf nie duidelik is nie. Alhoewel *dogtertjie* ook as lemma opgeneem word, word daar nie veel ander verkleiningsvorme opgeneem in Pharos se woordeboek nie. Die wyse waarop Longman-HAT se leksikograaf verkleiningsvorme opneem as deel van die verpligte mikrostruktuur is meer konsekwent en gebruikersvriendelik, alhoewel die weglating van 'n verkleiningsvorm nie gemerk word nie.

(3) Gee die korrekte Afrikaanse vertalings (translations) vir die vier seisoene:

winter _____
spring _____
summer _____
autumn _____

Die gemiddelde korrekte antwoorde vir hierdie vraag gedurende T1 was 0.95 uit 4. Die meeste van die studente het *winter* korrek gehad weens die feit dat dit dieselfde spelling as Engels het en die studente bloot die meeste van die woorde oorgeskryf het. Die gemiddelde korrekte antwoorde vir hierdie vraag gedurende T2 was 2.8 uit 4. Die studente wat Longman-HAT se woordeboek gebruik het se gemiddeld was 2.1 en die gemiddeld van die studente wat Pharos se woordeboek gebruik het, was 3.5. Hierdie gemiddelde is interessant

omdat albei woordeboeke die vertalings van al vier seisoene bevat. Dit word slegs op verskillende maniere in die betrokke woordeboeke aangebied. In Pharos se woordeboek kan dit gevind word in die sentrale lys van die English-Afrikaans afdeling en dus deur middel van die hoofmakrostruktuur en in Longman-HAT se woordeboek kan dit gevind word in middeltekste onder die *Thematic Dictionary/Tematiese* woordeboek onder *Seasons/Seisoene*. Die aanbieding in Longman-HAT se woordeboek lyk soos volg:

Seasons	**Seisoene**
spring	*lente*
summer	*somer*
autumn	*herfs*
winter	*winter*

Aangesien albei woordeboeke al vier die vertalings van die vier seisoene verskaf, kan daar slegs beweer word dat die studente wat Longman-HAT se woordeboek gebruik het, bloot nie die seisoene in die middeltekste opgespoor het nie. Die moontlike rede hiervoor is omdat die studente nie gewoond is aan buitetekste nie.

(4) Gee die regte vorm (correct form) van die byvoeglike naamwoord (adjective) tussen hakies.

1. **Jy moet** (droog) **klere aantrek, anders gaan jy siek word.**
2. **Die kinders het vir die** (snaaks) **man gelag.**
3. **Pas op vir die** (glad) **klip.**
4. **Hy stel 'n** (sleg) **voorbeeld.**

Die gemiddelde korrekte antwoorde vir hierdie vraag gedurende T1 was 0.4 uit 4 en gedurende T2 was dit 0.85 uit 4. Die studente wat Longman-HAT se woordeboek as hulpmiddel gebruik het se gemiddeld was 1 uit 4 en dié wat Pharos se woordeboek gebruik het se gemiddeld was 0.7 uit 4. Verboë vorme bly 'n moeilike aspek in Afrikaans en die teikengebruiker benodig eksplisiete leiding rakende hierdie aspek. Albei woordeboeke bevat die meeste van die woorde in bogenoemde vraag en verskaf ook die verbuigings van die woorde wat benodig word in die gegewe sinne. Hierdie verboë vorme word verskillend aangebied in die twee woordeboeke. Pharos se woordeboek se aanbieding is meer gebruikersvriendelik omdat dit tussen hakies aangebied word en die leksikograaf eksplisiet aandui dat dit 'n attributief is, terwyl Longman-HAT se woordeboek dit in dieselfde hakie as die trappe van vergelyking aanbied. Nie een van hierdie aanbiedings is egter voldoende vir hierdie teikengebruikers nie. Die resultate van hierdie vraag dui heel waarskynlik slegs op 'n tekort aan kennis van algemene taalreëls weens die kort tydperk waarin die studente nog in kontak was met Afrikaans. Alhoewel 'n attributief in Engels ook 'n "attributive" is, is daar reeds bevind dat Engels nie hierdie teikengebruikers se eerste taal is

nie en hulle dalk nie kennis dra van wat 'n "attributive" is nie. Dit is myns insiens belangrik dat verboë vorme ook bewerk moet word en aangebied word onder die lemma waarop dit van toepassing is.

Let op die aanbieding van *snaaks* en *glad* in Longman-HAT se woordeboek:

snaaks *adjektief* (**snaakse, snaakser, die snaaksste**)
1 [WAT JOU LAAT LAG] [MAKING YOU LAUGH]
► ***funny*** *'n snaakse storie* a funny story
► ***humorous*** *snaakse opmerkings* humerous remarks
2 [VREEMD/ONGEWOON] [STRANGE/UNUSUAL]
► ***funny*** *'n snaakse smaak* a funny taste

glad[1] *adjektief* (**gladde, gladder, die gladste**)
1 [MET 'N GELYK/SAGTE OPPERVLAK] [WITH AN EVEN SURFACE]
► ***smooth*** *Sy het 'n pragtige, gladde vel*. She has a lovely smooth skin.
2 [GRASIEUS, SONDER ENIGE RUKKERIGE BEWEGINGS] [GRACEFUL, WITH NO SUDDEN MOVEMENTS]► ***smooth*** *'n gladde landing* a smooth landing
3 [SONDER PROBLEME] [WITHOUT PROBLEMS]
► ***smooth*** *'n gladde proses* a smooth process

In bogenoemde woordeboekartikels kan daarop gelet word dat die voorbeeldsinne wat verskaf word slegs gerig is op die verboë vorme van die adjektiewe *snaaks* en *glad* en dat daar geen voorbeeldsinne is van hierdie adjektiewe se onverboë optredes nie. Alhoewel hierdie voorbeeldsinne suksesvolle leiding verskaf van die verboë optredes van die gegewe lemmas, moet daar leiding rakende die onverboë optredes ook verskaf word veral omdat die onverboë vorm juis die lemma is wat opgeneem is. Daar kan dus nie geen leiding verskaf word rondom die lemma self nie. Die volgende hipotetiese artikels as voorbeelde van moontlike verbeterde/hersiene artikels van Longman-HAT:

snaaks *adjektief* (**snaakser, die snaaksste**)
Verboë vorm ***Inflected form*** **snaakse**
1 [WAT JOU LAAT LAG] [MAKING YOU LAUGH]
► ***funny*** *die storie is snaaks* the story is funny
'n **snaakse** *storie* a funny story
► ***humorous*** *die opmerkings is snaaks* the remarks are humorous
snaakse *opmerkings* humorous remarks
2 [VREEMD/ONGEWOON] [STRANGE/UNUSUAL]
► ***funny*** *dit smaak snaaks* it taste funny
'n **snaakse** *smaak* a funny taste

glad[1] *adjektief* (**gladder, die gladste**)
Verboë vorm ***Inflected form*** **gladde**

1 [MET 'N GELYK/SAGTE OPPERVLAK] [WITH AN EVEN SURFACE]
► ***smooth*** *Haar vel is pragtig en glad* Her skin is lovely and smooth
Sy het 'n pragtige, **gladde** *vel*. She has a lovely, smooth skin.
2 [GRASIEUS, SONDER ENIGE RUKKERIGE BEWEGINGS]
[GRACEFUL, WITH NO SUDDEN MOVEMENTS]► ***smooth***
Die landing was glad The landing was smooth
'n **gladde** *landing* a smooth landing
3 [SONDER PROBLEME] [WITHOUT PROBLEMS]
► ***smooth*** *Die proses verloop glad* The process is going smooth
'n **gladde** *proses* a smooth process

Teksverdigting in woordeboeke bly 'n struikelblok vir aanleerders en daarom is dit belangrik dat verboë vorme geëkspliseer moet word. Omdat die leksikograaf soveel as moontlik inligting aan die gebruiker wil verskaf met minimale ruimte gebruik, is die herorganisering van inligting en aanwending van bykomende struktuurmerkers heel waarskynlik die beste opsie om gebruikers te lei ten opsigte van adjektiewe se verboë vorme. Deur die verboë vorme te onderstreep en in vetdruk te plaas kan die teikengebruiker duidelik sien waar hierdie vorme in voorbeelde optree. Daarteenoor bly die onverboë vorme in kursief sonder enige bykomende struktuurmerkers. Omdat die verboë vorm van die gegewe lemma nie meer tussen hakies gegee word nie en apart aan die gebruiker verskaf word, sal daar ook minder verwarring wees ten opsigte van die trappe van vergelyking wat in dieselfde hakie verskaf word in Longman-HAT se woordeboek. Dit is egter belangrik dat hierdie aanwending van bykomende struktuurmerkers deeglik aan die gebruiker verduidelik word in die inligtingsgedeelte en konsekwent gehandhaaf word.

(5) Vul die korrekte intensiewe vorm (intensive form) in. Iets wat:
1. baie oud **is** _____
2. baie maer **is** _____
3. baie lelik **is** _____

Geen van die studente het een van die intensiewe vorme gedurende T1 korrek gehad nie. Gedurende T2 het die studente wat Longman-HAT se woordeboek gebruik het, 'n gemiddeld van 0.5 uit 3 vir hierdie vraag behaal terwyl die studente wat Pharos se woordeboek gebruik het steeds nie een van die intensiewe vorme korrek gehad het nie. Pharos se woordeboek bied geen leiding rakende intensiewe vorme vir die gebruiker nie. Longman-HAT se woordeboek bied wel leiding rakende intensiewe vorme in die agterteks, ook bekend as die *Grammar Guide*. Let op die gedeeltelike uittreksel uit Longman-HAT se woordeboek ter voorbeeld:

Intensive forms of the adjective (*Intensiewe vorme*)
When something is very expensive, we say that it is **peperduur**

and when it is very cheap, it is **spotgoedkoop**. These are examples of intensive form. Intensives are used to add feeling to the adjectives. Here are some more:

blitsvinnig	bloedjonk	bloedrooi
brandmaer	broodnodig	dolleeg
doodeenvoudig	doodmoeg	galbitter
glashelder	goudgeel	hemelsbreed

Soos wat uit bogenoemde uittreksel blyk, bied Longman-HAT die nodige leiding rondom intensiewe vorme. Daar kan dus weereens slegs beweer word dat die rede waarom die studente wat hierdie woordeboek as hulpmiddel gekry het nie beter in hierdie vraag gevaar het nie, was omdat hulle bloot nie hierdie leiding opgespoor het nie. Dit is hier waar 'n woordeboekkultuur en die vestiging en handhawing daarvan 'n groot rol speel. Gebruikers moet algemene woordeboekopleiding ontvang, maar ook vaardighede aanleer ten opsigte van die spesifieke woordeboek wat vir hulle voorgeskryf is en wat hulle dus raadpleeg. Die woordeboekkultuur in Suid-Afrika word later in die artikel bespreek.

(6) Trappe van vergelyking (Degrees of comparison):

Stellende trap	Vergrotende trap	Oortreffende trap
groot	_____	_____
oud	_____	_____
interessant	_____	_____

Die gemiddelde korrekte antwoorde vir hierdie vraag gedurende T1 was 0.85 uit 6 en die gemiddelde korrekte antwoorde gedurende T2 was 2.95 uit 6. Die korrekte antwoorde vir hierdie vraag gedurende T2 het dus meer as verdriedubbel met behulp van 'n woordeboek. Albei woordeboeke bied die trappe van vergelyking aan van die gegewe lemma waarop dit van toepassing is. Longman-HAT se woordeboek bied wel die korrekte vorm van die oortreffende trap deur die lidwoord "die" voor die oortreffende trap vir die teikengebruiker te gee, wat Pharos se woordeboek nie verskaf nie. Omdat slegs een van die studente wat Longman-HAT se woordeboek gebruik het die korrekte vorme van die oortreffende trappe gegee het, is dit nie in berekening gebring vir hierdie vraag nie.

(7) Bestudeer die onderstaande prentjies en voltooi die sinne deur die korrekte voorsetsel in te vul:

B. Die kat sit _____ die stoel.
C. Die kat staan _____ die rekenaar.
F. Die hond sit _____ die skildery.
K. Die boekrak staan _____ die katte.

Die gemiddelde korrekte antwoorde vir hierdie vraag tydens T1 was 1.55 uit 4 en tydens T2 was dit 2.6 uit 4. Alhoewel Pharos se woordeboek geen eksplisiete leiding bied ten opsigte van voorsetsels nie het die studente wat hierdie woordeboek gebruik het beter gevaar in die vraag as dié wat Longman-HAT se woordeboek gebruik het. Dit is hier waar veranderlikes soos intelligensievlak heel waarskynlik ter sprake gekom het en die resultate van die vraag in T2 is nie noodwendig 'n akkurate weerspieëling van die leiding wat die betrokke woordeboeke ten opsigte van voorsetsels bied nie. Longman-HAT se woordeboek bied eksplisiete, gebruikersvriendelike leiding rondom voorsetsels in beide agtertekste en middeltekste, wat Pharos se woordeboek nie bied nie. Let eerstens op lukrake uittreksels ter illustrasie van hierdie leiding in die agterteks (*Grammar Guide*):

Prepositions in fixed use (*Voorsetsels in vaste gebruik*)

agter	Ons probeer **agter** die waarheid kom. (get to the truth) Die misdadiger sit nou **agter** slot en grendel (behind bars)
per	Ons het **per** trein/vliegtuig/motor gereis. (by train, etc.) Sy het per ongeluk/abuis op die kat se stert getrap. (by accident)
vir	Een **vir** een het die kinders vorentoe gekom. (one by one) Jan is kwaad **vir** sy suster. (cross with)

Omdat daar nie werklike, vaste reëls vir die gebruik van voorsetsels in Afrikaans is nie, is bogenoemde eksplisiete leiding ten opsigte van voorsetsels in vaste gebruik iets waarby hierdie teikengebruikers baie sal kan baat. Woorde en uitdrukkings soos "per abuis" en "agter slot en grendel" mag egter problematies wees vir hierdie gebruikers en dit is belangrik dat eenvoudiger sinne as voorbeeldmateriaal gegee word. Woorde en uitdrukkings wat gebruik word in voorbeelde soos hierbo en nie op enige ander plek in die woordeboek self gevind kan word nie, sal heel waarskynlik verwarring veroorsaak.

Longman-HAT se woordeboek bied verdere leiding ten opsigte van voorsetsels in die middeltekste aan. Hierdie suksesvolle leiding word soos volg aangebied:

There's a boy walking **across** the street.
*Daar is 'n seun wat **oor** die straat stap.*

The car is parked **in front of** the house.
*Die motor is **voor** die huis geparkeer.*

She is **inside** the house.
*Sy is **binne-in** die huis.*

Voorsetsels word hoofsaaklik gebruik om ruimtelike verhoudings aan te dui, soos byvoorbeeld "op," "onder," "in," "bo," "voor," "oor" ensovoorts. Ruimtelike verhoudings kan beter deur die nie-verbale aanduider (illustrasies) aangedui word omdat dit dikwels ingewikkeld is om hierdie verhoudings in woorde om te sit. Illustrasies word effektief in bostaande voorbeelde gebruik omdat dit die ruimtelike verhouding of aksie wat beskryf word beter aan die teikengebruiker verduidelik deur middel van 'n visuele illustrasie. Dit is 'n baie geslaagde metode vir nultaalsprekers; nie slegs vir die aanduiding van voorsetsels nie, maar ook om hierdie gebruikers bewus te maak van die korrekte gebruik van Afrikaanse voorsetsels. Die gebruik van illustrasies dien dus as aanvulling en ondersteunende leiding vir die voorbeeldsinne wat verskaf word. Alhoewel Longman-HAT se woordeboek goeie leiding ten opsigte van voorsetsels bied, kan daar beweer word dat weens die tekort aan 'n woordeboekkultuur, wat woordeboekvaardighede insluit, die studente bloot nie hierdie leiding in die gegewe middeltekste ontdek het nie.

(8) Maak 'n sin (make a sentence) met die volgende woorde:

1. **Troeteldier**

   _____

2. **Swaer**

   _____

3. **Pantoffel**

   _____

Hierdie woorde is oor die algemeen moeilike Afrikaanse woorde en daarom het die meeste studente hierdie vraag heeltemal, tydens T1, uitgelos. Die gemiddelde korrekte antwoorde van die studente wat Longman-HAT se woordeboek gebruik het, was 0.3 uit 3 en dié wat Pharos se woordeboek gebruik het, was 1.9 uit 3. Slegs *swaer* is nie in die sentrale lys van Pharos opgeneem is nie. Dit is ook juis hierdie woord wat die meerderheid van die studente wat hierdie woordeboek gebruik het, uitgelos het. Daarteenoor word nie *troeteldier*, *swaer* of *pantoffel* in die sentrale lys van Longman-HAT se woordeboek opgeneem en as sodanig bewerk nie. Die woorde *swaer* en *pantoffel* word wel in die middeltekste onder die *Thematic Dictionary/Tematiese woordeboek* verskaf en vertaal. Longman-HAT se lemmaversameling word grotendeels uitgebrei deur hierdie woordelyste en fraseslyste wat in die middeltekste verskaf word. Die enigste kritiek is egter dat dit beter bewerk kon word om meer leiding aan die teikengebruiker te verskaf. Daar kan verder ook beweer word dat die studente wat Longman-HAT se woordeboek gebruik het nie vertroud is met die middeltekste nie en weens onkunde moed opgegee het toe hulle nie die gegewe woorde in die sentrale lys kon opspoor nie. Die ideaal is om alles wat in die buitetekste opgeneem word, ook in die sentrale lys op te neem en met behulp van kruisverwysings die teikengebruiker te lei na die buitetekste en die verdere leiding wat ten opsigte van daardie lemma gegee word. Die leksikograaf sal sodoende die gebruikersvriendelikheid van die woordeboek verhoog.

(9) Is die volgende stellings waar (true) of onwaar (false)?

1. **'n Tandarts is iemand wat mense se tande versorg** _____
2. **'n Kelner bedien mense in 'n bank** _____
3. **'n Onderwyser maak siek mense gesond** _____
4. **'n Predikant werk gewoonlik in 'n kerk** _____

Die gemiddelde korrekte antwoorde vir bogenoemde vraag tydens T1 was 2.45 uit 4 en tydens T2 was dit 2.85 uit 4. Die groep studente wat Longman-HAT se woordeboek gebruik het se gemiddelde korrekte antwoorde was 2.7 uit 4 en dié wat Pharos se woordeboek gebruik het, was 3 uit 4. Daar was dus nie 'n beduidende verskil tussen die verskillende woordeboeke as hulpmiddels nie. Al vier die beroepe in die bogenoemde vraag is opgeneem in Pharos se sentrale lys en slegs *onderwyser* en *predikant* is in Longman-HAT se sentrale lys opgeneem. Hierdie betrokke woorde word in Pharos se woordeboek op so wyse in die woordeboekartikels aangebied dat daar deur middel van die voorbeeldsinne afgelei kan word of die gegewe vraag waar of onwaar is. Daarteenoor word daar geen voorbeeldsinne verskaf in Longman-HAT se woordeboek vir die aanbieding van *onderwyser* óf *predikant* nie. Longman-HAT se woordeboekartikels verskaf bondige betekenisparafrases wat die funksie van teksresepsie verhoog en sodoende beter leiding aan die teikengebruikers verskaf ten opsigte van hierdie funksie. Hierdie bondige betekenisparafrase speel 'n rol in die vraag onder bespreking omdat dít en 'n vertaalekwivalent al leiding is wat in

die aanbieding van *onderwyser* en *predikant* vir die gebruiker gegee word. Daar sou dus verwag kon word dat studente wat Pharos se woordeboek gebruik het 'n beduidender verskil in punte sou toon.

Myns insiens kan/moet daar nóg beter leiding rondom beroepe in 'n aanleerderwoordeboek vir hierdie teikengebruikers verskaf word. Omdat hierdie teikengebruikers studente op universiteit is, wat juis studeer om 'n sekere beroep te volg, sal dit van toepassing wees om 'n *Careers list/Beroepslys* in 'n middelteks aan hierdie teikengebruikers te verskaf. Vergelyk byvoorbeeld:

dentist	tandarts
teacher	onderwyser
lawyer	prokureur
lecturer	dosent
waiter	kelner

Omdat hierdie beroepe nie noodwendig 'n groot deel van die kursusraamwerk hoef uit te maak nie, kan dit slegs in die vorm van 'n lys verskaf word. Só word die lemmaversameling verbreed en word daar aanspraak gemaak op die student as teikengebruiker.

(10) Kies (choose) die korrekte woord tussen hakies (brackets):
Gertina sal 'n **(kies/keuse)** moet maak — sy moet **(sluit/besluit/gesluit)** of sy wil bly of sy wil **(huistoe/ huis toe)** gaan. Haar **(hart/hard)** is seer omdat die fliek te **(duur/deur)** is en sy nie genoeg geld het nie.

Die gemiddelde korrekte antwoorde vir hierdie vraag tydens T1 was 2.65 uit 5 en tydens T2 was dit 2.8 uit 5. Byna al die woorde in die vraag is in albei woordeboeke se sentrale lys opgeneem en bewerk. Daar kan dus slegs beweer word dat daar nie 'n beduidende verskil in punte was nie weens die feit dat dit die laaste vraag was en die studente nie genoeg tyd gehad het om dit deeglik te voltooi nie óf dat hulle eerder verkies het om te raai wat die korrekte antwoord is. Die spelling van woorde bly 'n problematiese kwessie omdat aanleerders/ studente geneig is om eerder te raai wat die korrekte spelling van 'n gegewe woord is. Hierdie neiging van aanleerders is iets wat nie bloot verander sal kan word deur die gebruik van 'n woordeboek nie, maar 'n positiewe impak kan gemaak word op hierdie neiging deur die vestiging en handhawing van 'n woordeboekkultuur.

4. Kommentaar vanuit die gebruiker se perspektief

Al twintig studente wat Afrikaanse Taalverwerwing 178 neem en deel gevorm het van die gevallestudie, is ook gevra om kortliks skriftelike kommentaar te lewer op die woordeboek wat aan hulle toegeken is. Daar is vir hulle gesê dat

die kommentaar hulle algehele gevoel oor die betrokke woordeboek moet oordra.

Dit het die volgende kommentaar tot gevolg gehad (direk oorgedra, slegs redigering ten opsigte van hoofletters):

Pharos

— "The dictionary was useful. I was able to find most words I was looking for. It definitely assisted me much more than I expected. I also think that I would have performed much better in this test in comparison to my first attempt without the dictionary."
— "There were words I could not find because I was not aware that it has English words to Afrikaans and Afrikaans to English. If I had realised early maybe I would have finished because the dictionary was helpful."
— "The dictionary helped alot but it never had all the words I needed."
— "It had most of the words but it was missing different forms of the words."
— "The dictionary helped me a lot, although it did not have all the words that I needed to look up. I learnt a few new Afrikaans words too."
— "It definitely helped a lot with the test. Feels like I can actually pass now."
— "The dictionary was easy to use. Some translations were not available, and some Afrikaans words which added on to the translation were not available in English. However it was overall a good dictionary to use. It helped me a lot in comparison to the previous test."
— "Much better than doing the test without the dictionary. However, this wasn't the best dictionary to make use of. Did not contain all info needed."
— "A good dictionary. Helped with most of the questions."
— "The dictionary was helpful in some of the exercises, but it was very unhelpful in most."

Longman-HAT

— "It was a nice dictionary to use although some words that should be there are missing. Finding the translation of a word often took a while as there was a lot of unnessesary text before I could read the translation."
— "Dictionary did help with vocabulary."
— "The dictionary helped me a lot in understanding the words that I did not understand in the previous task."
— "This dictionary blows! I understand they could only put so many words in but they could have at least been consistent with which type of words they included and excluded. For instance the only season in here was Summer. Nevertheless, the dictionary was better than nothing."
— "This is the worst dictionary I have ever used. A lot of really basic words are missing in both languages. They took the time & ink to include unnecessary pictures but failed to include necessary words. The blue part at the back was kind of helpful."

— "This dictionary would be helpful if it show every word clearly and difinition because I couldnot find many word that I was looking for. It is compliceted!"
— "Is this a dictionary or a study guide? Could not find the words I was looking for but it did help with intensive form, preposition and correct form."
— "The dictionary I used has very limited vocabulary, may words that I looked for were not there. It would also help if the dictionary indicated verbs, nouns, adjectives and so on."
— "The dictionary helped a little bit I couldn't find quite a few words & even with the dictionary it was difficult to translate a whole sentence."
— "Bad that I didn't recognise or connect some Afrikaans words with English words. But good that I can go and look for them and learn more."

Uit bogenoemde kommentaar kan afgelei word dat daar positiewe en negatiewe gevoelens oor albei woordeboeke was, alhoewel die kommentaar grotendeels negatief was oor Longman-HAT se woordeboek. Negatiewe kommentaar soos "unnessesary text," "only season in here was Summer" en "would also help if the dictionary indicated verbs, nouns, adjectives and so on" dui op 'n blote afwesigheid aan 'n woordeboekkultuur (wat woordeboekvaardighede insluit). Só kan negatiewe kommentaar op Pharos se woordeboek soos "I was not aware that it has English words to Afrikaans and Afrikaans to English" en "it was missing different forms of the words" ook aan hierdie tekort toegeskryf word. Dit word toegeskryf aan 'n tekort aan 'n woordeboekkultuur omdat al hierdie aspekte juis in albei woordeboeke verskaf word. Alhoewel die aanbieding van verskillende leksikale items gebruikersvriendelik moet wees, is woordeboekgebruik 'n interaktiewe proses tussen die gebruiker en die woordeboek. Die dringende tekort aan 'n woordeboekkultuur het grotendeels geblyk uit die kommentaar van die gegewe studente.

5. Die woordeboekkultuur in die Suid-Afrikaanse skoolsisteem

Die Nasionale Kurrikulum- en Assesseringsbeleidsverklaring (NKAV) waarvolgens Afrikaans as addisionele taal in die Intermediêre fase (graad 4 tot 6) in Suid-Afrika onderrig word, is eerstens onder die loep geneem. Afrikaans as addisionele taal word dus ook aangeleer vir leerders wie se moedertaal nie Afrikaans is nie en dalk ook nie Engels nie. Al leer hierdie leerders dit heel waarskynlik as tweede taal aan en nie as derde taal nie, kan 'n woordeboek die taalverwerwingsproses verryk en ook hierdie leerders kommunikatief bemagtig. Die belangrikste is dat die insluiting van woordeboekonderrig in hierdie kurrikulum (en so ook alle taalkurrikulums) 'n woordeboekkultuur sal vestig en leerders van jongs af woordeboekvaardighede sal aanleer om hulle sodoende ook vir die res van hul lewens kommunikatief te bemagtig.

Die kurrikulum stipuleer dat daar "[i]n die Intermediêre Fase op die grondslagfase wat in graad R–3 vasgelê is, voortgebou word." Daar word verder voorgestel dat hierdie "oorgangsfase" suksesvol ingelei kan word deur middel van lees — groepsbegeleide lees, selfstandige lees en lees in pare (Departement van Basiese Onderwys 2011: 11).

Let op 'n uittreksel van hierdie leesproses wat in die kurrikulum verskaf word (my skuinsdruk):

> **Die leesproses**
> Die leesproses bestaan uit 'n pre-lees, gedurende lees en 'n post-leesstadium. Aktiwiteite wat leerders kan uitvoer, kan soos volg opgesom word:
>
> **Pre-lees (Voorbeelde)**
> — Aktiveer bestaande kennis.
> — Stel vas wie die bron, die skrywer en wat die publikasiedatum is.
> — Lees die eerste en laaste paragrawe van 'n afdeling.
> — Maak voorspellings oor onder andere die inhoud en tema.
>
> **Gedurende lees (Voorbeelde)**
> — Neem ruspouses om begrip te toets en idees te laat insink.
> — Vergelyk die inhoud met wat afgelei is.
> — Maak gebruik van die inhoud om waar moontlik die betekenis van onbekende woorde vas te stel, *anders kan 'n woordeboek gebruik word*.
> — Visualiseer wat gelees word.
> — Leerders moet aanhou lees al verstaan hulle nie sekere dele nie.
> — Leerders lees gedeeltes weer oor indien hulle dit glad nie verstaan nie. Lees dele hardop en teen 'n stadiger tempo indien inhoud verwarrend is.
> — Vra 'n maat om te help om moeilike gedeeltes te verstaan.
> — Maak leesmerke en maak notas van hoofpunte.
>
> (Departement van Basiese Onderwys 2011: 11-12)

Soos wat daar uit bogenoemde uittreksel afgelei kan word, word daar wel in die kurrikulum na "'n woordeboek" verwys. Daar word verwys na 'n woordeboek in die "gedurende lees." Die kurrikulum stel voor dat die inhoud van die leesstof gebruik kan word om onbekende woorde te identifiseer óf dat 'n woordeboek gebruik kan word om hierdie onbekende woorde te identifiseer. Die verwysing na "'n woordeboek" is nie slegs baie vaag nie, maar ook onvoldoende. Dit is nie vansselfsprekend dat opvoeders enige kennis dra van woordeboeke nie, dus moet daar beter leiding verskaf word ten opsigte van watter tipe woordeboek gebruik moet word. Daar moet 'n spesifieke woordeboek voorgeskryf word deur die Departement van Basiese Onderwys sodat die opvoeders sodoende ook vertroud kan wees met hierdie woordeboek en dus beter leiding aan sy/haar leerders kan verskaf.

Daar word verder soos volg na 'n woordeboek in die leermateriaal verwys (my skuinsdruk):

Leerders behoort toegang te hê tot die volgende leermateriaal in graad 4–6:

Graad 4 – 6		
Kernmateriaal		
Voorgeskrewe handboeke vir Eerste Addisionele Taal		√
'n Woordeboek		√
Leesboek/e wat die volgende tekstipes bevat		
Stories / vervolgverhale	√	
Tonele / kortverhale	√	
Poësie	√	
Inligtingstekste	√	
Sosiale tekste	√	
Media- / grafiese tekste	√	
Mediatekste		
Koerante	√	
Tydskrifte	√	
Televisieprogramme	√	
Radioprogramme	√	

(Departement van Basiese Onderwys 2011: 15)

Daar word weereens, onder kernmateriaal wat leerders behoort te hê, slegs verwys na "'n woordeboek." Dit is onvoldoende leiding deur die Departement van Basiese Onderwys. Daar behoort 'n spesifieke woordeboek, eentalig of tweetalig, voorgestel te word. Dit kan gedoen word deur leiding en samewerking van 'n leksikograaf wat die leksikografiese behoeftes van hierdie teikengebruikers kan vasstel. Sodoende kan hierdie leerders eers algemene woordeboekonderrig ontvang en daarna onderrig ten opsigte van die woordeboek wat vir hulle voorgeskryf word. Al sal 'n graad 6 leerder heel waarskynlik 'n "moeiliker" woordeboek kan gebruik, kan graad 4's vertroud gemaak word met hierdie woordeboek sodat hulle dit teen graad 6 volkome sal kan bemeester.

Woordeboekaktiwiteite word wel ingesluit in die amptelike "inhoud en onderrigplanne vir eerste addisionele taal" vir graad 4 tot graad 6. Hierdie woordeboekaktiwiteite het egter minimaal te make met die werklike gebruik van 'n woordeboek en eerder met die skep van 'n persoonlike woordeboek. Hier volg die instruksies wat in die beplanning aan die opvoeders gegee word:

Werk met woorde

— Skryf woorde en hul betekenis in 'n persoonlike woordeboek neer
— Gebruik sinne, tekeninge of verduidelikings om betekenis van woorde te verduidelik
— Spel bekende woorde korrek

— Vokale en konsonante
— Sinonieme en antonieme

(Departement van Basiese Onderwys 2011)

Bogenoemde aktiwiteit word deurentyd voorgestel in die beplanning van die lesse vir graad 4 tot graad 6 onder "taalstrukture en –konvensies" as vaardighede. Dit is 'n problematiese aktiwiteit omdat die leerders nie noodwendig weet hoe 'n woordeboekartikel lyk nie omdat hulle geen woordeboekonderrig ontvang het nie. Daar word vir leerders gesê om die uitleg van 'n woordeboekartikel deur te gee sonder enige leiding oor hierdie uitleg. 'n Leerder sal onmoontlik weet waar om sinonieme, antonieme of voorbeeldsinne te plaas. Al is dit heel waarskynlik nie die hoofdoel van die aktiwiteit om die uitleg korrek te kry nie, sou dit hier die ideale geleentheid wees om eers die leerders vertroud te maak met 'n woordeboekartikel in sy eenvoudigste vorm en daarna te laat oefen deur hul eie een te skep met 'n woordeboek byderhand as hulpmiddel. Daar kan wel geredeneer word dat die Departement van Basiese Onderwys hierdie aktiwiteite tot 'n mate sien as woordeboekonderrig. Alhoewel dit dalk die geval mag wees, is dit onvoldoende onderrig.

Naas die skepping van 'n persoonlike woordeboek is die enigste ander woordeboekaktiwiteit wat in die onderrigplan gegee word, om woordeboeke te gebruik vir die korrekte spelling van woorde. Dit is wel een van die mees frekwente redes waarom woordeboeke geraadpleeg word, maar dit is sekerlik ook nie die enigste rede nie. Leerders word sodoende van jongs af "grootgemaak" met die wanopvatting dat woordeboeke slegs gebruik word deur mense wat moeilikheid ervaar met spelling. Dit is juis in die skoolsisteem waar hierdie stereotipe gebreek moet word deurdat leerders van jongs af woordeboekonderrig moet ontvang. Sodoende sal leerders nie slegs die veelvuldige ander elemente van 'n woordeboek ontdek nie, maar 'n woordeboekkultuur sal ook gevestig word omdat leerders bewus gemaak sal word dat woordeboeke vir meer as slegs die spelling van woorde geraadpleeg kan word. Die persepsie van woordeboeke as "vervelig" sal dalk uit die weggeruim kan word en 'n woordeboekkultuur sal tot op universiteitsvlak gehandhaaf kan word.

In die Nasionale Kurrikulum- en Assesseringsbeleidsverklaring (NKAV) vir Afrikaans as Eerste Addisionele Taal vir die Senior fase (graad 7 tot 9) word "'n woordeboek" ook voorgeskryf onder kernmateriaal. Daar word egter geen verdere woordeboekaktiwiteite of enige vorm van woordeboekgebruik voorgestel in die lesbeplanning vir hierdie grade nie. Daar kan dus aangeneem word dat die kurrikulum vir hierdie vak vir graad 10 tot graad 12 nie beduidend sal verskil nie en dat woordeboeke ook nie 'n groot rol in dié beplanning speel nie.

Al is slegs die Nasionale Kurrikulum- en Assesseringsbeleidsverklaring (NKAV) vir Afrikaans as Eerste Addisionele Taal onder die loep geneem, blyk die dringende tekort aan 'n woordeboekkultuur duidelik daaruit. Die tekort aan 'n woordeboekkultuur op skoolvlak lei dus ook tot 'n tekort aan 'n woordeboekkultuur op universiteitsvlak. Die prototipe-teikengebruikers van hierdie

ondersoek is van verskillende lande afkomstig en beskik moontlik oor verskillende woordeboekkulture. Dit is dus noodsaaklik dat indien hulle in 'n mate 'n woordeboekkultuur het, daardie woordeboekkultuur gehandhaaf word op tersiêre vlak. Die ideaal is wel dat 'n woordeboekkultuur op laerskoolvlak gevestig sal word en regdeur gehandhaaf sal word tot op universiteitsvlak. Dit is egter 'n moeilike proses waarmee leksikograwe vir dekades lank sukkel. Myns insiens hoef universiteite nie te wag totdat hierdie woordeboekkultuur op skoolvlak geïnisieer word nie en moet woordeboekonderrig 'n formele deel van enige taalaanleerprogram word. Voorstelle, wat dien as voorbeelde, vir formele woordeboekonderrig as deel van die Afrikaanse Taalverwerwing 178 kursus by die Universiteit van Stellenbosch word in die volgende afdeling gemaak.

6. Voorstelle vir formele woordeboekonderrig

Myns insiens moet woordeboekonderrig aan hierdie studente gegee word as deel van die formele kursusraamwerk. Die studente moet eerstens algemene woordeboekonderrig ontvang. Beyer (s.a.) beweer dat daar sekere kernelemente is wat 'n woordeboekpedagogieprogram moet bevat. Alhoewel Beyer (s.a.) van mening is dat hierdie kernelemente reeds op skoolvlak moet deel vorm van formele onderrig, is dit myns insiens nodig dat dit solank op universiteitsvlak geïmplementeer moet word as deel van formele onderrig. Let op hierdie aangepaste kernelemente wat deur Beyer voorgestel word (my opsommings in hakies):

— Vertroudheid met 'n basiese woordeboektipologie (watter tipes woordeboeke is daar alles beskikbaar?)
— Aanvaarding vir die behoefte van inligting (jy is nie dom as jy 'n woordeboek raadpleeg nie)
— Kennis van watter inligtingsbehoeftes leksikografies relevant is (watter tipes inligting kan ek alles uit 'n woordeboek onttrek?)
— Die vermoë om die gepaste woordeboek te kies (benodig ek 'n eentalige of 'n tweetalige woordeboek om hierdie probleem op te los?)
— Die vermoë om die inskrywings wat die relevante inligting op te spoor (wat beteken daardie woord, sin of struktuurmerker in die woordeboekartikel?)
— Die vermoë om hierdie gevonde inligting, toe te pas (ek het die relevante inligting opgespoor, wat maak ek nou daarmee?)

Bogenoemde kernelemente is elemente wat van kardinale belang is en deel moet vorm van die algemene woordeboekonderrig wat hierdie prototipe-teikengebruikers eerstens moet ontvang. Naas hierdie algemene woordeboekonderrig, moet hierdie gebruikers spesifieke woordeboekonderrig ontvang. Hierdie

tipe woordeboekonderrig verwys na vaardighede wat aangeleer moet word met spesifieke verwysing na die woordeboek wat vir hulle voorgeskryf is; dit wil sê met byvoorbeeld spesifieke verwysing na óf Pharos se *Aanleerderwoordeboek vir skole/Learner's Dictionary for Schools* óf Longman-Hat se *Afrikaans Dictionary and Grammar for English Speakers.*

Hieronder volg 'n hipotetiese woordeboekaktiwiteit met die gebruik van die lemma *horlosie* in Pharos se woordeboek:

> **horlosie 1** watch *"What is the time?" — "My **watch** says it's five past ten."* "Hoe laat is dit?" **—** "My **horlosie** sê dis vyf oor tien." **2** clock *The **clock** on the wall is a few minutes fast.* Die **horlosie** teen die muur is 'n paar minute voor.
> ♦ **op 'n horlosie kyk** tell the time. *My little brother has just learnt to **tell the time**.* My boetie het pas geleer om **op 'n horlosie** te **kyk**.
> □ **hor·lo·sie** *selfstandige naamwoord* (**meervoud horlosies**)

As 'n **horlosie** om 11:00 wys dat dit 10:55 is, is hy vyf minute **agter**; as hy 11:05 wys, is hy vyf minute **voor**.

1. **What is the translation of *horlosie*?**
2. **How many meanings does *horlosie* have?**
3. **Write down the example sentences to indicate the meanings.**
4. **How many syllables is the word *horlosie* divided into?**
5. **Which syllable is being emphasised?**
6. **What is the plural (meervoud) of *horlosie*?**
7. **What does *selfstandige naamwoord* tell you about the word *horlosie*?**
8. **Why is the extra textblock given?**
9. **Make your own sentence with the word *horlosie*.**
10. **What time is it now on the class *horlosie*?**

Bogenoemde woordeboekaktiwiteit kan eerstens deur die dosent saam met die studente gedoen word en daarna kan 'n soortgelyke oefening met 'n ander lemma as voorbeeld aan die studente gegee word om op hul eie te doen. Die vrae is in Engels gestel omdat dit nie hier gaan om die aanleer van Afrikaans nie en ook omdat woordeboekonderrig aan die begin van die kursus aangebied moet word. Hierdie studente sal dus nog nie die nodige kennis van Afrikaans dra om die vrae te verstaan as dit in Afrikaans gestel is nie. Myns insiens is spesifieke woordeboekonderrig net so belangrik soos algemene woordeboekonderrig omdat die woordeboekartikel en ook die strukture wat aangewend word, verskil van woordeboek tot woordeboek. Ten opsigte van die gebruik van Longman-HAT se woordeboek sal die studente ook spesifiek bewus gemaak moet word van die middel- en agtertekste en daar sal vrae uit hierdie buitetekste vir die studente gegee moet word.

7. Gevolgtrekking

Dit is belangrik om kennis te neem van die feit dat nie Pharos se woordeboek óf Longman-HAT se woordeboek saamgestel is vir hierdie spesifieke prototipe-teikengebruiker binne 'n universiteitskonteks nie. Daar is dus 'n leemte in die aanleerderwoordeboekversameling as tipologiese onderafdeling wat gevul moet word.

Laastens blyk dit ook uit die resultate van hierdie ondersoek dat die beskikbaarheid van woordeboeke nie noodwendig die suksesvolle gebruik daarvan verseker nie. Die Nasionale Kurrikulum- en Assesseringsbeleidsverklaring (NKAV) vir Afrikaans as Addisionele Taal in die Suid-Afrikaanse skoolsisteem is onder die loep geneem om die dringende afwesigheid aan 'n woordeboekkultuur onder die vergrootglas te bring. Tersiêre instellings hoef egter nie te wag totdat 'n woordeboekkultuur op skoolvlak gevestig word nie en algemene en spesifieke woordeboekonderrig moet solank deel vorm van enige taalkursus op universiteitsvlak. Indien 'n moontlike aanleerderwoordeboek vir die teikengebruikers van hierdie gevallestudie saamgestel word, kan die vestiging van 'n woordeboekkultuur op tersiêre vlak in Suid-Afrika 'n belangrike rol in die sukses van hierdie woordeboek speel.

Endnote

1. Vergelyk ook in hierdie verband Klein (2007), Abecassis (2008), Van der Merwe (2009) en Chen (2012).
2. Die gekose woordeboeke wat die kern van hierdie studie vorm, is gekies omdat dit tans (2015) vir nul-taalsprekers voorgeskryf word. Dit word onderskeidelik voorgeskryf by die Universiteit van Stellenbosch en die Nelson Mandela Metropolitaanse Universiteit.
3. 'n Gevallestudie is gebruik omdat hierdie studie nie betrekking het op 'n breë spektrum van deelnemers wat gevolg word deur 'n veralgemening oor een van die twee woordeboeke nie.

Bronnelys

Woordeboeke

Du Plessis, M. 2005. *Pharos Aanleerderwoordeboek vir skole/Learner's Dictionary for Schools.* Kaapstad: Tafelberg-Uitgewers.

Luther, J., T. McLachlan en A. Barnby (Reds.). 2012. *Longman-HAT Afrikaans Dictionary and Grammar for English Speakers*. Kaapstad: Pearson.

Ander bronne

Abecassis, M. 2008. The Ideology of the Perfect Dictionary: How Efficient can a Dictionary Be? *Lexikos* 18: 1-14.

Adendorff, E.M. 2015. Persoonlike onderhoud. Stellenbosch, 3 Maart.

Adendorff, E.M. 2015a. Afrikaanse Taalverwerwing 178 Studiegids. Klasnotas. Stellenbosch: Universiteit van Stellenbosch.

Augustyn, P. 2013. No Dictionaries in the Classroom: Translation Equivalents and Vocabulary Acquisition. *International Journal of Lexicography* 26(3): 362-385.

Bergenholtz, H. en R.H. Gouws. 2010. A New Perspective on the Access Process. *Hermes* 44: 103-127.

Beyer, H.L. s.a. *Wat is 'n woordeboek*. Pearson: *Longman-HAT* [Intyds]. Beskikbaar: http://longman-hat.co.za/?r=17. [18 Maart 2015].

Chen, Y. 2012. Bilingualized Dictionaries with Special Reference to the Chinese EFL Context. *Lexikos* 22: 139-158.

Departement van Basiese Onderwys. 2011. *Nasionale Kurrikulumverklaring (NKV)*. Pretoria: Departement van Basiese Onderwys.

Flyvbjerg, B. 2006. Five Misunderstandings about Case-study Research. *Qualitative Inquiry* 12(2): 219-245.

Hausmann, Franz Josef. 1989. Die gesellschaftlichen Aufgaben der Lexikographie in Geschichte und Gegenwart. Hausmann, Franz Josef, Oskar Reichmann, Herbert Ernst Wiegand and Ladislav Zgusta (Eds.). 1989. *Wörterbücher. Dictionaries. Dictionnaires. An International Encyclopedia of Lexicography* I: 1-19. Berlyn/New York: Walter de Gruyter.

Klein, J. 2007. Implementing a Dictionary Culture in South Africa: An Attempt at a General Framework for African Languages. *Lexikos* 17: 416-423.

Kuhn, T.S. 1987. What are Scientific Revolutions? Krüger, L., Lorraine J. Daston and Michael Heidelberger (Eds.). 1987. *The Probabilistic Revolution* 1: 7-22. Cambridge, MA: MIT Press.

Lew, R. 2004. *Which Dictionary for Whom? Receptive Use of Bilingual, Monolingual and Semi-bilingual Dictionaries by Polish Learners of English*. Poznań: Motivex.

Lew, R. en A. Adamska-Sałaciak. 2014. A Case for Bilingual Learners' Dictionaries. *ELT Journal* [Intyds]. Beskikbaar: http://eltj.oxfordjournals.org/. [9 Januarie 2015].

Otto, A. 1999. Kriteria vir 'n Afrikaanse aanleerderwoordeboek. *Lexikos* 9: 135-151.

Shanshan, C. 2008. Monolingual, Bilingual Dictionaries and Language Study. *Cross-cultural Communication* 4(1): 65-73.

Steyn, M. 2004. The Access Structure in Learner's Dictionaries. *Lexikos* 14: 275-298.

Tarp, S. 2004. Basic Problems of Learner's Lexicography. *Lexikos* 14: 222-252.

Tarp, S. en R.H. Gouws. 2004. Wie leer wat uit Afrikaanse (aan)leerderwoordeboeke. *Tydskrif vir Geesteswetenskappe* 44(4): 276-296.

Van der Merwe, M.F. 2009. Die opvoedkundige waarde van woordeboeke: Voorstelle vir woordeboekonderrig in Suid-Afrika. *Lexikos* 19: 297-313.

On the Most Innovative Outer Access Structure of any Bantu Dictionary: The *Lexique kikongo–français* by Charles Polis (1938)

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

Abstract: In this article a little-known dictionary manuscript from the 1930s, the *Lexique kikongo-français* by the Jesuit missionary Charles Polis, is analysed in great detail. Section 1 expounds on the goal and raison d'être of the study, Section 2 introduces the manuscript, its author as well as the Kikongo variety dealt with, Section 3 presents the inner workings of the *Lexique* on macro-, micro- and mediostructural levels, Section 4 gives a lexicographical appreciation based on a large selection of the entries, Section 5 joins the international debate on the exact nature of a dictionary's macrostructure, access structure and access route, and Section 6 compares Polis's work with a dictionary from the same region and period. Conclusions are offered in Section 7, chief among them the fact that Polis designed the most innovative outer access structure of any Bantu dictionary.

Keywords: BANTU, KIKONGO, KINTANDU, FRENCH, DICTIONARY MANUSCRIPT, BILINGUAL LEXICOGRAPHY, MISSIONARY LEXICOGRAPHY, LEMMA SIGN, MACROSTRUCTURE, OUTER ACCESS STRUCTURE, OUTER ACCESS ROUTE, INNOVATION

Samenvatting: Over de meest innovatieve externe toegangstructuur in de Bantoelexicografie: het *Lexicon Kikongo–Frans* van Charles Polis (1938). In dit artikel wordt een weinig bekend woordenboekmanuscript uit de jaren 1930, het *Lexicon Kikongo–Frans* van de jezuïet-missionaris Charles Polis, in detail onderzocht. Deel 1 licht het doel en de bestaansreden van de studie toe, Deel 2 introduceert het manuscript, zijn auteur, alsook de behandelde variant, Deel 3 beschrijft minutieus de macro-, micro- en mediostructuur van het *Lexicon*, Deel 4 geeft een lexicografische appreciatie die is gebaseerd op een ruime selectie van materiaal uit het werk, Deel 5 draagt bij tot het internationale debat over de ware aard van de macrostructuur, toegangstructuur en toegangsroute van een woordenboek, en Deel 6 vergelijkt Polis' werk met een woordenboek uit dezelfde regio en tijd. Conclusies worden aangeboden in Deel 7, de belangrijkste het feit dat Polis de meest innovatieve externe toegangstructuur uit de Bantoelexicografie ontwierp.

Sleutelwoorden: BANTOE, KIKONGO, KINTANDU, FRANS, WOORDENBOEKMANUSCRIPT, BILINGUALE LEXICOGRAFIE, MISSIONARISLEXICOGRAFIE, LEMMATEKEN, MACROSTRUCTUUR, EXTERNE TOEGANGSTRUCTUUR, EXTERNE TOEGANGSROUTE, INNOVATIE

Lexikos 25 (AFRILEX-reeks/series 25: 2015): 56-135

1. Goal and raison d'être of the present study

The main goal of this research article is to present and analyse an unpublished dictionary manuscript for Kikongo: the *Lexique kikongo–français* by Charles Polis s.j. (1885–†1943). Apart from the fact that this manuscript, of which there are still a handful of extant copies left, deserves to be widely known, what makes this work especially intriguing is that it employs a highly original outer access structure to get at the data. After a cursory examination of the work one could be forgiven to conclude that this is the first example of a Bantu dictionary in which the lexicographer has finally succeeded to atomise a Bantu language's lexicon, breaking it down to its most basic components, to then reassemble the lexicon (and to compile a dictionary in the process) from those building blocks. If this were indeed the case, this would be the first example of a purely stem-based lemmatisation strategy in Bantu lexicography, rather than the traditional approach which results in a lemmatisation strategy that is neither purely stem-based nor purely word-based, but sits somewhere on the sliding continuum between these two extremes (see De Schryver 2008: 86-88). Questions that immediately arise upon studying Polis's original presentation include:

— Could this outer access structure be applied to all Bantu languages?

— Is this outer access structure perhaps universal?

— Could such an outer access structure perhaps be a solution for all non-corpus-based lexicographic endeavours?

The latter question is a direct result of Polis's claim, in his introduction, that his approach allows for the systematic identification of the missing forms in a dictionary (Polis 1938: Part I, i). Regardless of the answers to these questions, a presentation of this work is important in its own right, as it contains — as will be argued below — a wealth of unique language data on a less-resourced Kikongo variety with a considerable time-depth (over 75 years, down to a century), which is rare in Bantu lexicography.

2. The manuscript, its author and the Kikongo variety dealt with

The manuscript consists of 719 typed-up pages, each page slightly larger in size than an A4. Polis, whose name appears on the bottom-right of the first page, 'transcribed' the material from his notes during the course of 1938, while in Leuven.[1] His manuscript was stencil duplicated and distributed in nine fascicules (Van de Casteele 1968).[2] It is not known how many copies were made. The Ghent University library has had the copy which used to belong to the agricultural engineer Lode De Wilde (whose name appears on the first page of each of the nine fascicules) for some time, and in 2015 also acquired the copy which used to belong to the missionary-linguists Jan Daeleman s.j. (1922–†2014) and Gaston van Bulck s.j. (1903–†1966).

The manuscript has no title page, which has resulted in a proliferation of designations. The copy of De Wilde was catalogued as *Dictionnaire Kikongo* (http://lib.ugent.be/catalog/rug01:000184510), while the Daeleman–Van Bulck one was catalogued as *Dictionnaire KiKongo–Français, schikking van K. Polis* (http://lib.ugent.be/catalog/rug01:002189704). Other copies that are known to us include a couple at the Leuven University library, where it is the *Dictionnaire de la langue congolaise* (http://www.unicat.be/uniCat?query=sysid:11277820), and one at the Antwerp University library, where it is the *Dictionnaire congolais (en formation)* (http://anet.be/record/opacanet/c:lvd:355728/N). We have chosen to refer to the work with *Lexique kikongo–français*, as done by Van de Casteele (1968) in his obituary of Polis, mainly because Polis himself points out in the introduction to his work that it is not yet a dictionary, but rather contains material that may lead to the compilation of a dictionary (Polis 1938: Part I, i). The Ghent University fascicules have been bound into three volumes, those at Leuven University into two, while at Antwerp University the catalogue entry shows nine 'volumes' (in all likelihood the nine fascicules). The version consulted for the present study is the one of De Wilde: the stencilled pages are generally of poor quality, often with faint vertical sections down the middle, and/or (mostly) horizontal ink smears. Several letters are also typed on top of one another, or are otherwise unreadable. With a bit of effort, these issues can of course be overcome (but they give even trained OCR software a hard time). During the binding of the De Wilde copy, on some pages a (very limited) number of lines at the bottom got cut off.

Even though there is no title page, we are certain the language dealt with is indeed Kikongo. This is amongst others clear from the illustrative material, which includes the phrases (1) to (4) — phrases that also tell us something about the process by which the material was collected:[3]

(1) **FuM + O, VERBE** (Part II, p. 251): ... *tufuma ndiinga kikoongo*, nous avons une masse (de travail) (dans cette étude) du kikoongo; donc bcp. à combiner, à réfléchir [we have a lot (of work) (in this study) of Kikongo; thus a lot to combine, to work out] ...

(2) **FiK + OII, SUBST., mi** (Part II, p. 261): ... *ba balaanda beeto, sa bamona mmfiku, sa babaka kikoongo mu mmf.*, ceux qui viendront après nous, trouveront la difficulté supprimée; ils acquerront le kikoongo sans peine, à bon compte (réflexion d'un aide) [those who will come after us, will find that the difficulty is gone; they will acquire Kikongo without efforts, on the cheap (remark of a helper)] ...

(3) **KoZ & KoS + OI, VERBE** (Part II, p. 402): ... *muundele ukoonza Na Kosi kikoongo*, le Blanc est en train d'épuiser tout le kikoongo de Monsieur Kosi [the White man is busy exhausting all the Kikongo of Mr Kosi] ...

(4) **KoZ & KoS + OI, V.DER., uNA** (Part II, p. 403): ... *konzununa kikoongo*, rechercher tout le kikongo mot par mot [research all of Kikongo, word after word] ...

What is thus immediately apparent from these four example phrases, is that Polis worked with native-speaker language consultants — one of them even

called Mr Kosi — with whom he tried out all possible 'combinations' in an attempt to pinpoint *all* the words of Kikongo. Those informants felt that Polis was really getting the maximum out of them, and as a result they concluded that anyone wishing to learn Kikongo in the future would now have an easy time thanks to their efforts.

Knowing that one is dealing with Kikongo is not enough, however, as 'Kikongo' is actually "a disparate continuum of closely related Bantu languages" known as the Kikongo Language Cluster or 'KLC' (De Schryver et al. 2015). The KLC may be divided into four subgroups — North, West, East and South — together with a Central buffer zone, as shown in Figure 1.

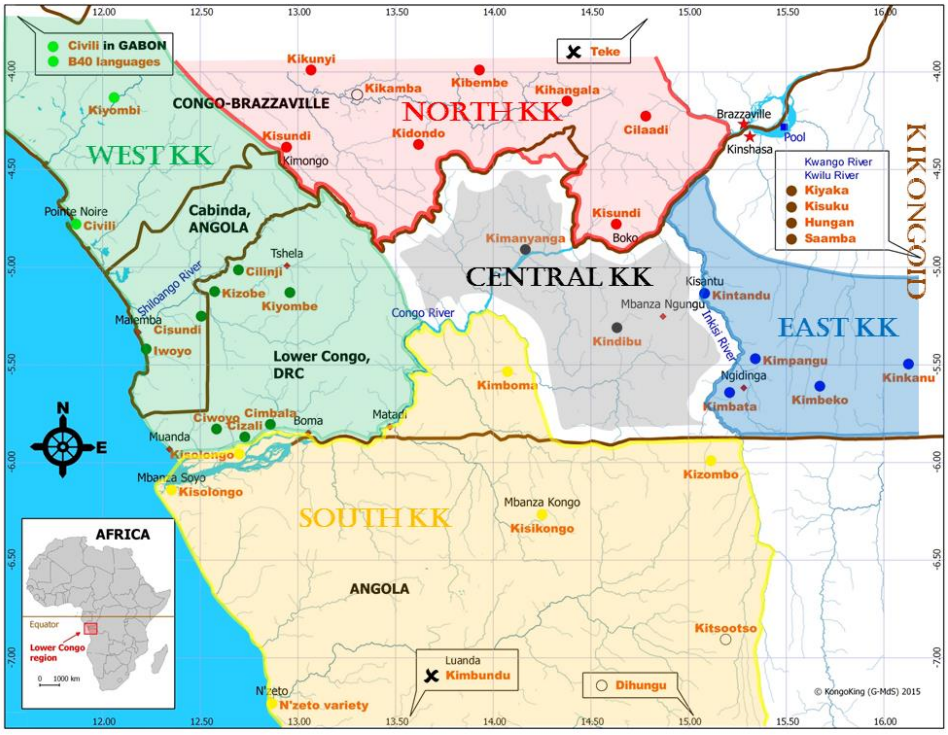


Figure 1: The Kikongo Language Cluster (KLC), with its sub-groups mapped. Colours indicate sub-group membership within the KLC. (Kimbundu and Teke are not part of the KLC, and are therefore preceded by black crosses.)

In order to determine to which subgroup the Kikongo variety described by Polis belongs, and perhaps even to pinpoint the variety itself, we can first turn to the literature. Over the past four years, the KongoKing research group has built up and digitised a documentation corpus which currently stands at over a

thousand sources on the Kongo kingdom and/or on the Kikongo language. Disappointingly, Polis is only mentioned in seven sources: in the obituary already referred to, in the documentation of one map, in the introduction of one dictionary, in two MAs, in one PhD, and in one scientific article.

From the obituary (Van de Casteele 1968) we learn that Charles Polis was born in Antwerp on November 1, 1885. As a Jesuit he arrived in the Kwango for missionary work on August 15, 1911, and basically stayed there for the rest of his life, being variously posted as a teacher/priest to Kimpako, Leverville, Lemfu, Mbanza-Mboma and Ngidinga. All of these locations are east of the Inkisi River, and all but one (Leverville, the present Lusanga) are close to that river, which thus strongly suggests that Polis is covering one (or several) East Kikongo varieties in his manuscript. Kimpako and Mbanza-Mboma are locations close to Kisantu, where Kintandu is spoken, while Lemfu is situated halfway between Kisantu and Ngidingi, the latter where Kimbata is spoken. See Figure 1 for the major locations, the Inkisi River, and the KLC varieties. Polis's stay in the Congo was interrupted four times for return visits to Belgium, the longest of these during the entire World War I. He died of health complications in a Léopoldville (today Kinshasa) hospital on June 27, 1943.

In the supporting documentation of Boone's (1973: 145) ethnographic map of the Congo, Polis is mentioned for a short contribution on the Mbata (Polis 1942). In contrast, in his dictionary Swartenbroeckx (1973: v) refers to "the excellent Kintandu works" of Polis which he was alas "not able to acquire in time" to help during compilation, while Makaya Lutumba (1999: 9) merely lists Polis together with René Butaye and Joseph Van Wing as contributors to Kintandu. In neither of these, references to actual works in Kintandu are however provided. Makolo Miaka (2000: 8), who clearly copied various pages verbatim from Makaya Lutumba (1999), repeats the exact same paragraph and thus statement regarding Kintandu. Lastly, and arguably the best evidence, Daeleman (1966) points out in his PhD on Kintandu:

> *Veel nut, inzonderheid bij de studie van de werkwoordafleidingen, hebben we getrokken van de rijke mijn van voorbeelden die amper aangeboord is in het als handschrift gestencilde 'Dictionnaire kikongo–français' van wijlen P. K. POLIS.* [A great help, especially when it comes to the study of verbal derivations, was the goldmine of examples found in the hardly perused stencilled manuscript *Dictionnaire kikongo–français* by the late Father C. POLIS.] (Daeleman 1966: 7)

In his article on the PB reflexes in Kintandu, Daeleman (1983) also refers to and uses data from Polis's manuscript. In the entire academic literature, then, there is but one scholar who explicitly refers to the work under study: Jan Daeleman, not surprisingly another Jesuit working from Leuven, who put the information contained within Polis's manuscript to good use in the course of his doctoral studies and subsequent research. Since then, Polis's work has regrettably remained untapped. The literature just reviewed further also suggests that the (main?) variety dealt with by Polis must be Kintandu.

Could the Kikongo variety be derived from Polis's manuscript itself? A good starting point is to look at the (frequencies of) place names used in the illustrative material,[4] which clearly favour Kisantu and neighbouring locations over Ngidinga: *Kisantu* (25 x), *Kimvulu* (12 x), *Kimpese* (6 x), *Lemfu* (4 x), *Ngidinga* (3 x), *Kimpako* (2 x), *Kivuunda* (2 x), or *Mbe* (an old village close to Lemfu, 1 x). Also compare this to locations much farther away: *Mpuumbu* and *Stanley Pool* (6 x) or *San-Salvador/S. Salvador* (3 x). Rivers are not mentioned often enough to be conclusive, but are indicative: *Inkisi* (1 x) and *Bongolo* (1 x) vs. *Congo* (0 x) and *Kwilu* (0 x). While the *Ntandu* (East), *Mpangu* (East), *Mbeko* (East), *Ndibu* (Central), *Zombo* (South), *Solongo* (South) and *Laari* (North) are each mentioned about once or twice, the *Mbata* (East) are mentioned numerous times. Further analysis reveals that this is actually done to contrast the speech of the Mbata, mostly labelled with the abbreviation *MB.*, with the main variety covered in the manuscript. There are well over 180 instances of the labels *MB.* and *Mbata* throughout the microstructure. Additionally, a note at the start of fascicule G/K- is very clear on the fact that Kimbata should be considered the dialectal form in this manuscript:

(5) **G/K-** (Part II, p. 293): Nota: le sigle (MB.g) indique que le g est maintenu dans le dialecte de Mbata, qui le remplace généralement par v bilabial. [Note: the label '*MB.g*' indicates that g is maintained in the dialect of Mbata, where it is generally replaced by the bilabial v.]

While the linguistic explanation is only approximate here, this note does point out an important difference between Kintandu and Kimbata regarding the phonetic outcome of the PB ***p** in intervocalic position, as seen in (6) vs. (7):[5]

(6) Kintandu

***p > ɣ/V_V[-closed] & *p > ɣ/Ṇ_V[-closed]**

***-páan-** 'give' (BLR 2345) > -**ɣáán**- (Daeleman 1983: 382)

(7) Kimbata

***p > v/V_V[-closed] & *p > v/Ṇ_V[-closed]**

***-páan-** 'give' (BLR 2345) > -**vana** (KongoKing Fieldwork 2012)

As shown in Bostoen et al. (2013: 63-66), given **β**, **v**, **h,** **ɣ̊** and **ɣ** are all found as unconditioned reflexes of PB ***p** within the KLC, these phonetic outcomes may be seriated most economically by postulating the intermediate proto-sound **°ɸ**, as shown in (8):

(8) ***p**-lenition within the KLC

			>	**β**	>	**v**		
***p**	>	**°ɸ**						
			>	**h**	>	**ɣ̊**	>	**ɣ**

Kintandu, then, is the endpoint of the bottom series of the split seriation shown in (8), while Kimbata is the endpoint of the top series. Polis covers the forms of the bottom series, thus Kintandu.

Other KLC varieties that are labelled in Polis's manuscript include the southern Kisikongo, as *Kis*. (63 x), and *Kizoombo* (2 x), as well as the western Kiyombe, as *May*. (22 x). There are even a few instances of the Kikongoid *Kiyaka* (10 x), and the non-related eastern *Kiluba/Luba* (2 x). In order to contrast 'his Kikongo' with other KLC varieties, Polis must have had access to other existing dictionaries, as he is not known to have travelled widely throughout the Lower Congo region, and thus to have had first-hand experience with other KLC varieties. That he indeed had access to other reference works is evidenced by the fact that he names his colleagues and thus indirectly his sources by name: *Laman/Lam*. (23 x), *Bittremieux* (1 x), *Butaye* (1 x), *Gillet* (1 x), *Georges de Gheel* (1 x), *Bentley* (1 x), and *Delplace* (1 x). These can be linked to dictionaries for the central Kimanyanga variety by Laman (1936), the western Kiyombe variety by Bittremieux (1922, 1927), the eastern Kintandu variety by Butaye (1909) and Gillet and Pâque (1910), the southern Kisikongo variety by Van Gheel (1652)[6] and Bentley (1887, 1895), and the concocted Kikongo by Delplace (1895). All of these reference works indeed appeared before 1938. The gleaned information was in such cases mostly literally lifted from those other sources, as is for instance the case for the excerpt shown in (9) with data copied from and cross-referenced to (10):

(9) **SuS + OII, SUBST., ma** (Part II, p. 144): ... <u>*dinsusu-nsusu*</u>, légume apprécié (*ocimum* sp., petit basilic à feuilles aromatiques servant aux assaisonnements; sec. Fr Gillet) [much-liked vegetable (*ocimum* sp., small basil with aromatic leaves used to season; according to Brother Gillet)] ...

(10) **Dinsusu**. — *Ocimum* sp. LABIÉES.

Nsusu = poule (generice).

Petit basilic, à feuilles aromatiques servant aux assaisonnements.

(Gillet and Pâque 1910: 14)

While it is a truism that one cannot prove a negative, the total absence of labels to mark Kintandu, in combination with the fact that other varieties *are* labelled and contrasted with what is being described lexicographically, should be sufficient additional proof that Polis indeed deals with Kintandu.

The only other known sizeable general-language dictionary for Kintandu is Butaye's (1909) *Dictionnaire kikongo–français, français–kikongo*. Even though that dictionary was published three decades before Polis transcribed his material (it was even published before Polis ever set foot in the Congo), we can be rather certain that Polis did not frequently consult Butaye's dictionary to compile his own work. This assertion is not only based on Polis's idiosyncratic approach to the macrostructure, but also on the actual coverage of the lexicon in his work, the original microstructural contents he presents, and his use of a

non-standard spelling for Kikongo/Kintandu. Certainly, a valid hypothesis could have been that Polis merely 'pretended' to be devising a new outer access structure, while he was in actual fact rearranging the data of an existing dictionary (including the possibility of the rearrangement of the draft of a dictionary of his own making). This hypothesis is inspired by the fact that it seems rather overwhelming, truth must be told, to imagine a situation whereby one departs from the most basic building blocks of a language — CV(C) clusters in this case — which are subjected to some systematic manipulations and the addition of other building blocks, the results of which are only kept when meaningful words are the outcome, to end up with a balanced and representative coverage of a language's lexicon. Can this type of introspection really lead to the full coverage of the lexicon — or at any rate, an acceptable one? As we will show in the subsequent sections, this is indeed possible, and the dictionary compilation procedure gleaned from the extracts shown in (1) to (4) must therefore reflect reality: Polis and his team of native speakers pulled off quite a feat.

3. Presentation of Polis's *Lexique*

3.1 The macrostructure of the *Lexique*

3.1.1 Metalexicographical context

It is well-known that modern dictionaries for the Bantu languages are corpus-based (De Schryver and Prinsloo 2000a, b), with the very best even aiming to be corpus-driven (De Schryver 2010). In the pre-corpus era the main strategies employed for the actual compilation of the macrostructure were either (i) random, (ii) rule-oriented, or (iii) enter-them-all approaches. In the random approach "words are simply added whenever they happen to cross the compiler's way", in the rule-oriented approach "a set of rules/guidelines presented in the dictionary's front matter must be followed whenever a word cannot be looked up directly" (so the assumption is that everything is covered 'in theory'), and in the enter-them-all approach "the compilers are obsessed to include all conceivable nominal and verbal derivations [working] through a modular paradigm in order to pursue such a comprehensiveness" (De Schryver and Prinsloo 2001: 219-225).

This is not to say that still other strategies have not been tried out. Prinsloo and De Schryver (1999: 264-267) discuss an interesting case whereby Vermeersch (1922) sought out and brought together the vocabulary of Cilubà according to morpho-lexical fields. This approach has been found wanting (Kalonji 1993: 134-136), and may be seen as a variant of the random approach in that there is no system to ensure a systematic coverage. A more recent alternative approach is that by Mbatha (2006) for Zulu, who argues that only content words belong in a dictionary and that just four word classes merit to be recog-

nised anyway: nouns, verbs, interjections, and ideophones. This could be seen as a variant of the rule-oriented approach, whereby all sorts of meanings now somehow have to be forced onto extremely low-frequent to non-existing verb and noun stems in order to fit in, say, what everyone else recognises to be adjectives, adverbs, etc. For a critique of Mbatha's method, with supporting examples, see De Schryver and Wilkes (2008: 829-830). Polis's approach, now, may be seen as a variant of the enter-them-all approach, also known as the paradigm approach, but unlike the blind generation of often highly infrequent, dubious and even non-existent forms that characterise other modular approaches (see Prinsloo 2014), Polis was only interested in what really exists (and in all openness even added question marks when he was uncertain).

3.1.2 The Vertical Base

So how is Polis's *Lexique* structured on the macrostructural level? To begin with, one needs to acquaint oneself with his alphabetical ordering for consonants (11) and vowels (12):

(11) Consonants

B/P, M, V/F, G/K, D~L/T, N, Z/S (, Y, W)

(12) Vowels

a, e, i, o, u (, diphthongs)

As may be seen from (11), Polis first works through the labials (bilabials B/P and M, and labiodentals V/F), then the velars (G/K), followed by the alveolars (D~L/T, N and Z/S), and concludes with the semi-vowels (Y and W). The consonants are thus grouped according to place of articulation (mostly from front to back), and within each group according to manner of articulation (from plosive over nasal, etc.). At each of those levels Polis considers the voiced ones before the unvoiced ones (where relevant). D and L are considered on a par, as they are in complementary distribution in Kintandu. The vowels, shown in (12), follow the standard ordering.

In a **first stretch** of his manuscript Polis combines each consonant (C) with either a following vowel (–) or a following vowel and itself, as shown in (13), going through all the consonants in his defined ordering (11), and associating each of the vowels in the standard ordering (12) to each of the consonants, one at a time:

(13) **C–**

C–C

An extract from the resulting ordering may be seen in (14):

(14) ... Pa, Pe, Pi, Po, Pu, PaP, PeP, PiP, PoP, PuP, Ma, Me, Mi, Mo, Mu, MaM, MeM, MiM, MoM, MuM, ... Sa, Se, Si, So, Su, Sau, Say, SaS, SeS, SiS, SoS, SuS. (The semi-vowels Y and W are not considered here.)

Voiced and unvoiced renderings are treated on a par, before moving to the next consonant. Furthermore, given that D and L are in complementary distribution, this means that to for example go from DaD to DeD to DiD etc. one needs quite some gymnastics, as illustrated in (15):

(15) ... Da, De, Di, Do, Du, La, Le, Li, Lo, Lu, Lau, **DaD**, DaL, LaL, LaD, **DeD**, DeL, LeL, LeD, **DiD**, DiL, LiL, LiD, **DoD**, DoL, LoL, LoD, **DuD**, DuL, LuL, LuD, Ta, Te, ...

In a **second stretch** Polis combines each consonant (C) with each other consonant (C'), as shown in (16), running through all the vowels before going to the next consonant:

(16) **C–C'**

This results in the ordering seen in (17):

(17) B-P (an empty category), PaB, PeB, PiB, PoB, PuB, BaM, BeM, BiM, BoM, BuM, PaM, PeM, ... SaT, SeT, SiT, SoT, SuT, ZaN, ZeN, ZiN, ZoN, ZuN, SaN, SeN, SiN, SoN, SuN, ZaS & SaZ, ZeS & SeZ, ZiS & SiZ, ZoS & SoZ, ZuS & SuZ. (The semi-vowels Y and W are not considered here.)

For D~L one now obtains sequences like (18):

(18) ... PaK, PeK, PiK, PoK, PuK, **BaD**, BaL, **BeD**, BeL, **BiD**, BiL, **BoD**, BoL, **BuD**, BuL, BaT, BeT, BiT, BoT, BuT, **PaD**, PaL, **PeD**, PeL, ...

While all of this is of an impeccable logic, this ordering is of course not user-friendly. Do note that this was not meant to be, as this work is only a temporary tool for Polis, so the usual dictionary criticism is not warranted. At the same time, Polis did get tangled up in his own system as he starts his dictionary with the sequence shown in (19)a, while it should have been as in (19)b:

(19) **a.** Ba, BaB, Be, BeB, Bi, BiB, Bo, BoB, Bu, BuB, ...
b. Ba, Be, Bi, Bo, Bu, BaB, BeB, BiB, BoB, BuB, ...

The difference is not insignificant, as the sequence shown in (19) runs over 25 pages. Overall, there are very few other errors in the actual ordering (the headings are sometimes wrong though). Also, some categories are lumped together, as may be seen from the ampersand '&' in (17). At times, this leads to slip ups: there is for instance an entry for GaD, and another for GaL & GaD.[7]

In a **third stretch** Polis specifically deals with the semi-vowels Y and W, as well as the diphthongs. For Y, formulas (20) and (21) are applicable, which results in the sequence shown in (22):

(20) **Y–**

Y–Y

(21) **Y–C,W**

C–Y

(22) Ya, Ye, Yi, Yo, Yu, YaY, YeY, YiY, YoY, YuY, YaB, YeB, YiB, ... YoZ & YoS, YuZ & YuS, YaW, BaY, BeY, BiY, BoY, PaY, PeY, PiY, PoY, PuY, MaY, MeY, MiY, ... ZaY, ZeY, ZiY, ZoY, ZuY, SaY, SeY, SoY, SuY.

Likewise for W, formulas (23) and (24) are applicable, which results in the sequence shown in (25):

(23) **W–**

W–W

(24) **W–C**

C–W

(25) Wa, We, Wo, Wu, WaW, WeW, WiW, WoW, WuW, WaB, WeB, WiB, ... WoZ & WoS, WuZ & WuS, BaW & PaW, BeW & PeW, ... DoW & LoW, DuW & LuW, TaW, ToW, Z-W, S-W.

The sequence for the diphthongs (26) is as shown in (27):

(26) **VV**

(27) ay, au (monosyllabic), au (polysyllabic), ey, eu, oy.

Clearly, not all combinations exist, and when one doesn't, Polis normally does not provide a heading for it, as may be deduced from the gaps in the sequences seen in (22) and (25). When treating each vowel in its own right is not worth the effort, Polis also lumps the material, as is the case for Z-W and S-W in (25). Again, very few errors were found in the sequence, except that an entry Y-W follows SuY, while that entry should actually have replaced (and the data merged with) the earlier YaW (which follows YuZ & YuS).

Finally, there is also a shorter **zeroth stretch**, which precedes the three stretches already mentioned, where Polis lists what he calls the 'pre-grammatical forms'. The ordering here mimics the ordering used in stretches one to three, thus first (13) and then (16), presented as (28) and (29). This is followed by a section dedicated to vowels and varia (30), sections on W, as (31) rather than (23) and (24), and Y, as (32) rather than (20) and (21), and concludes with a section on diphthongs (33):

(28) **C–**

C–c

(29) **C–c'**

(30) **V, varia**

(31) **W–**

W–c,w,y

C–w

(32) **Y–**

Y–c

C–y

Y–y

(33) **VV**

Following Polis, the second consonant in the pre-grammatical part has been written in lowercase (although he doesn't follow this convention consistently), which should ease in differentiating this part from the rest of the work. By and large, one notices that the pre-grammatical part thus follows the same sequence as the rest of the work, but not in the details. In addition to the swapping of the semi-vowel sections Y and W, the changes within those sections, and the addition of a vowel section, the pre-grammatical equivalent of (15), for instance, is also different, as shown in (34):

(34) ... Kik, Kok, Kuk, Da, De, Di, Do, Du, La, Le, Li, Lo, Lu, **Dad**, **Ded**, **Did**, **Dod**, **Dud**, Dal, Del, Dil, Dol, Dul, Lal, Lel, Lil, Lol, Lul, Lad, Led, Lid, Lod, Lud, Ta, Te, ...

There are also a few errors, such as the erroneous placement of the F– sequence before the V–v sequence, where it should be the reverse.

For all these reasons, and in order to make this first level of the macrostructure more accessible to future users, we felt that it was necessary to create an index of the full CV(C) sequence. We have termed that sequence the 'Vertical Base', where even the shorthand 'CV(C) sequence' is an approximation, as it refers to the sequence of formulas (28), (29), (30), (31), (32), (33), (13), (16), (20), (21), (23), (24), and (26). The result is shown in Addendum 1. In perusing Polis's manuscript, checking and searching (the electronic version of) Addendum 1 greatly speeds up the look-up process.

We are now in a position to make more sense of the different fascicules of the manuscript, as summarised in Table 1. The short introduction (numbered i-iii by Polis) together with the pre-grammatical forms (numbered 1-67, including one bis-page), constitutes one fascicule. All the subsequent sections (for which the page numbering restarts at 1, running up to 646, including two bis-

pages) each constitute a fascicule in their own right. Given the double numbering, we use Part I to refer to the pre-grammatical part, and Part II to refer to the bulk of the work. The dates in the manuscript indicate that Polis first transcribed the simple and double labials and linguals (fascicules 1 and 2 — corresponding to Stretch 1 of his CV(C) sequence), followed by the introduction and pre-grammatical forms (fascicule 3 — Stretch 0). He then proceeded with all the mixed consonants (fascicules 4 to 8 — Stretch 2), and concluded with the semivowels and diphthongs (fascicule 9 — Stretch 3).

Table 1: Details of the nine fascicules of Polis's *Lexique*, linked to the four different CV(C) stretches of the macrostructure's Vertical Base

No.	Title	Pp.	Fasc.	Stretch	Part	Signed
0	Introduction	i-iii	3	—	—	24 Apr. 1938
1	Pre-grammatical forms	1-67	"	0	Part I	21 May 1938
2	Simple & double labials	1-80	1	1	Part II	21 March 1938
3	Simple & double linguals	81-144	2	"	"	17 Apr. 1938
4	B/P with other consonants	145-225	4	2	"	26 Aug. 1938
5	M and V/F with other consonants	226-292	5	"	"	2 Sept. 1938
6	G/K with other consonants	293-406	6	"	"	10 Sept. 1938
7	D~L/T with other consonants	407-495	7	"	"	20 Sept. 1938
8	N and Z/S with other consonants	496-599	8	"	"	25 Sept. 1938
9	Words with Y, W and diphthongs	600-646	9	3	"	9 Oct. 1938

The information summarised in Table 1 and the CV(C) sequence shown in Addendum 1 constitute the first entry point to (and the first part of the macrostructure of) Polis's work. Addendum 1 reveals that there are 1 592 CV(C) clusters: these, then, are the basic building blocks around which Polis proceeds to construct the entire lexicon of Kintandu. Except for a few empty CV(C) clusters, for which the heading is listed but no data is provided, each of those clusters is now subjected to a number of manipulations in order to arrive at real words.

3.1.3 The Horizontal Base

Still on the macrostructural level, Polis follows up with three types of manipulations of each CV(C) cluster in Part II: Steps 1, 2 and 3 — to which we will collectively refer to with the term 'Horizontal Base'. To enable the presentation of Step 1, each CV(C) cluster may now be written as shown in (35):

(35) **C V**

C_1 V C_2

Step 1 is concerned with either the palatalization or labialization of the first consonant (C or C1) and/or the prenasalization of the second consonant (C2, or for the CV cluster simply the consonant C). Considering either the presence of palatalization (coded Y), the presence of labialization (W), or the absence of both (O), in combination with either the presence (I) or absence (II) of prenasalization, results in a 3 x 2 matrix, or thus six possibilities, for which Polis uses the codes shown in the column 'Code 2' of Table 2. The linguistic implication of those codes is shown in the last two columns of Table 2, where the superscript ^{j} stands for palatalization, the superscript ^{w} for labialization, and the superscript ^{N} for prenasalization (with N a non-syllabic homorganic nasal). Optional items are shown between round brackets; variation within the options is separated by a comma.

Table 2: Step 1 of the Horizontal Base in Polis's *Lexique*

Code 1	Code 2	Meaning	Linguistic formula CV	CVC
Y		palatalization of C or C_1	$^{(N)}\mathbf{C}^{j}\mathbf{V}$	$\mathbf{C_1}^{j}\mathbf{V}\,^{(N)}\mathbf{C_2}$
	YI	palatalization of C or C_1, and prenasalization of [C or] C_2	$[^{N}\mathbf{C}^{j}\mathbf{V}]$	$\mathbf{C_1}^{j}\mathbf{V}\,^{N}\mathbf{C_2}$
	YII	palatalization of C or C_1, but no prenasalization of [C or] C_2	$[\mathbf{C}^{j}\mathbf{V}]$	$\mathbf{C_1}^{j}\mathbf{V}\,\mathbf{C_2}$
W		labialization of C or C_1	$^{(N)}\mathbf{C}^{w}\mathbf{V}$	$\mathbf{C_1}^{w}\mathbf{V}\,^{(N)}\mathbf{C_2}$
	WI	labialization of C or C_1, and prenasalization of [C or] C_2	$[^{N}\mathbf{C}^{w}\mathbf{V}]$	$\mathbf{C_1}^{w}\mathbf{V}\,^{N}\mathbf{C_2}$
	WII	labialization of C or C_1, but no prenasalization of [C or] C_2	$[\mathbf{C}^{w}\mathbf{V}]$	$\mathbf{C_1}^{w}\mathbf{V}\,\mathbf{C_2}$
O		no palatalization nor labialization of C or C_1	$^{(N)}\mathbf{C}\,\mathbf{V}$	$\mathbf{C_1}\,\mathbf{V}\,^{(N)}\mathbf{C_2}$
	OI	no palatalization nor labialization of C or C_1, but with prenasalization of [C or] C_2	$[^{N}\mathbf{C}\,\mathbf{V}]$	$\mathbf{C_1}\,\mathbf{V}\,^{N}\mathbf{C_2}$
	OII	no palatalization nor labialization of C or C_1, and no prenasalization of [C or] C_2	$[\mathbf{C}\,\mathbf{V}]$	$\mathbf{C_1}\,\mathbf{V}\,\mathbf{C_2}$
I		prenasalization of [C or] C_2	$[^{N}\mathbf{C}^{(j,w)}\mathbf{V}]$	$\mathbf{C_1}^{(j,w)}\mathbf{V}\,^{N}\mathbf{C_2}$
II		no prenasalization of [C or] C_2	$[\mathbf{C}^{(j,w)}\mathbf{V}]$	$\mathbf{C_1}^{(j,w)}\mathbf{V}\,\mathbf{C_2}$

Analysis of the manuscript reveals that Polis also uses the codes shown in the column headed by 'Code 1', thus focusing on the palatalization (Y), labialization (W), or absence of both (O), as well as the prenasalization (I) or absence

thereof (II), without considering (or specifying) the other level. The use of the single codes Y, W and O is the favoured (but not the sole) approach for CV clusters, the use of the single codes I and II is the favoured (but not the sole) approach used in the semi-vowel-initial section of the third Stretch, while the use of the double codes (YI, YII, WI, WII, OI and OII) is the favoured (but not the sole) approach in the bulk of the work, thus for the CVC clusters. Not all Codes are used with each CV(C) cluster; only the applicable ones, and in the order seen in Table 2, from top to bottom.

The codes for the formulas shown between square brackets in the penultimate column of Table 2, namely those to indicate whether or not there is prenasalization of CV clusters, could have been used, but they haven't. Not only does Polis not codify the prenasalization of CV clusters, he does not do so for the first consonant of CVC clusters either. Rather, he silently lumps its potential presence at Step 3 of the Horizontal Base. Likewise, he does not codify the palatalization or labialization of the second consonant, again silently lumping its potential presence at Step 3 of the Horizontal Base. While this is not linguistically sound, his reasoning must have been that in Step 1 he was only interested in changes around the vowel of each CV(C) structure.

Step 2 could be said to be the signpost for the word class or part of speech; the two major categories being VERBE to introduce base verbs, and SUBST. to introduce base nouns. Verbs and nouns, as well as occasionally other word classes, derived from this are grouped in a section preceded by V.DER. Word forms that cannot be derived from the previous are brought together under VARIA. While pre-grammatical forms are to be found in their own separate part (i.e., Part I) of the dictionary, references to them are, when applicable, signnalled at the start of Step 2 with the signpost PREGR. Step 2 is repeated for each (relevant) Code from Step 1, in the order just described.

Step 3 is used with two of the signposts from Step 2: SUBST. and V.DER. At SUBST. the plural prefix of the nouns that follow is shown, as if to say that a collection (hence the plural) of nouns from that gender is announced.[8] Table 3 lists the labels used for the nouns at Step 3, together with their actual singular and plural prefixes as well as the gender numbers in use today.

Table 3: SUBST. 'noun prefixes' used at Step 3 of the Horizontal Base in Polis's *Lexique*

SUBST.	Singular / Plural prefixes	Modern gender numbers
ba	mu-, mw-, Ṇ- / ba-	1/2
bi	ki-, ø- / bi-	7/8
bu	bu-	14
ma	di-, ø- / ma-	5/6
mi	mu-, mw-, Ṇ- / mi-, mw-, Ṇ-	3/4
tu	lu- / tu-	11/13
zi	N- / N-	9/10

(ku)	ku- / ma-	15/6	(Part II, pp. 369, 381, 475, 478)
ga	ga-	16	(Part II, p. 314)
(ku)	ku-	17	(Part II, p. 318, 636)
gu	*gu-*	*17 ?*	*(Part II, p. 314)*
mu	*mu-*	*18*	*?*
fi	fi-	19	(Part II, p. 47)

The abbreviation N at gender 9/10 stands for the non-syllabic homorganic nasal, to contrast it with the syllabic homorganic nasal Ṇ found in classes 1, 3 and 4.[9] Orthographically, Polis differentiates the syllabic from the non-syllabic nasals by doubling the nasal when it is syllabic, for instance: *nnkeento/bakeento* 'woman/-en' (1/2) and *mmvu* 'year(s)' (3/4) vs. *nzo* 'house(s)' (9/10) and *mbwa* 'dog(s)' (9/10). With only the plain characters of a typewriter at hand, Polis came up with a wonderfully simple and straightforward way to draw attention to the difference between syllabic and non-syllabic nasals, which is most definitely an improvement over the earlier works by Butaye (1909, 1910), who doesn't mark the difference and only uses simple nasals.[10] Further note that Polis orders the nouns in Step 3 using his usual alphabetical ordering, as seen in (11) and (12): B, M, T, Z — followed by the addition of the relevant vowels for each of these.[11] Only those labels for which nouns are eventually presented are listed. The noun classes under the dashed line in Table 3 are infrequently used; the few instances we noted are referenced.

At V.DER. a large battery of labels for 'extensions' (real or perceived, and the eventual words themselves often with further extensions) is used. The main and more productive ones are shown in Table 4. This table lists the labels in the ordering used by Polis (from left to right, and within each column from top to bottom), which again follows the orderings (11) and (12). Of course no single V.DER. signpost has all these labels; what is not relevant is simply skipped.[12]

Table 4: V.DER. 'extensions' used at Step 3 of the Horizontal Base in Polis's *Lexique*

V.DER.							
1 ⇨	2 ⇨	3 ⇨	4 ⇨	5 ⇨	6 ⇨	7 ⇨	8
BA	MA	NGA	KA	LA	TA	NA	SA
aBA	aMA	aNGA	aKA	NDA	aTA	aNA	aSA
iBA	iMA	iNGA	iKA	aLA	eTA	iNA	iSA
uBA	uMA	oNGA	uKA	iLA	iTA	uNA	oSA
		uNGA		oLA	uTA		uSA
				uLA			

In the case of verbs with verbal extensions, Table 4 reveals that these are thus unfortunately not explicitly marked. Implicitly, though, a possible (first) verbal extension may be guessed from the label, such as the rollative (**-Vb-**) in column 1,

the stative (**-am-**) in column 2, the imperfective (**-ang-**) in column 3, the impositive (**-ik-**) and intransitive oppositive (**-uk-**) in column 4, the agitative (**-al-**), applicative (**-il-**) and oppositive (**-ul-**) in column 5, the contactive (**-Vt-**) in column 6, the reciprocal (**-an-**), expressive (**-in-**) and debilitative (**-un-**) in column 7, or finally the ergative (**-as-**) and causative (**-is-**) in column 8.[13]

The Horizontal Base in Part I simplifies that seen in Part II: Step 1 is kept, Step 2 is skipped, and for Step 3 only a set of numbers is used to represent extensions to the base clusters, namely 1 for -B-, 2 for -M-, 3 for -NG-, 4 for -K-, 5 for -L-, 6 for -T-, 7 for -N-, and 8 for -S-. This is analogous with the V.DER. extensions which are otherwise used at Step 3 (compare with Table 4), with this difference that combinations of extensions are marked explicitly here, resulting in for instance 15 for *-abala*, 38 for *-engese*, 77 for *-anana*, 546 for *-olokoto*, etc.

At last we are now in a position to combine the Vertical Base and the Horizontal Base, as illustrated with a number of selected examples in Table 5, to arrive at 'real words' in Polis's work.

Table 5: Combining the Vertical Base and the Horizontal Base in Polis's *Lexique*: Selected examples for each of the Codes used

Codes	With CV clusters	With CVC clusters
Y	**Da + Y, SUBST., bi** > *kindya*, *kidya* (Part II, p. 82)	**NeM + Y, V.DER., iTA** > *nyeemita* (Part II, p. 498)
YI	—	**ZuG + YI, V.DER., aLA** > *zyungalakana* (Part II, p. 547)
YII	—	**FoT + YII, VARIA** > *fyote* (Part II, p. 282)
W	**Fa + W, SUBST., zi** > *mfwa* (Part II, p. 45)	**SaM + W, V.DER., uNA** > *kiswaamunu* (Part II, p. 538)
WI	—	**TeG + WI, SUBST., bi** > *tweengi* (Part II, p. 463)
WII	—	**SeT + WII, V.DER., aMA** > *sweetama* (Part II, p. 591) **Set + WII, 4** > *sweetika* (Part I, p. 59)
O	**Ka + O, Varia** > *aka* (Part II, p. 63)	**SiN + O, SUBST., bi** > *kisina* (Part II, p. 595)
OI	—	**KoD & KoL + OI, V.DER., iSA** > *koondisa* (Part II, p. 363)
OII	—	**VuT + OII, SUBST., zi** > *mvutu* (Part II, p. 272)
I	—	**YaG + I, VERBE** > *yanga* (Part II, p. 609)
II	—	**YeK + II, V.DER., iKA** > *yekikila* (Part II, p. 614)

The words listed in Table 5 are however not the only words generated by each respective Vertical Base + Horizontal Base. For instance, **ZuG + YI, V.DER., aLA** also leads to *zyungalakasa* (i.o.w., the ergative replaced the reciprocal as the final verbal extension), or **KoD & KoL + OI, V.DER., iSA** also leads to *kondisanga* (where the imperfective was added), or **YaG + I, VERBE** also leads to *iyanga* (now with a preceding reflexive), or even **YeK + II, V.DER., iKA** which also leads to simply *yekika* (where the applicative was dropped). Therefore, where and how are these different 'words' (i.e., the ones in italics in Table 5) now listed in Polis's work? This is the topic of the next section, as we move to the microstructure.

3.2 The microstructure of the *Lexique*

3.2.1 The microstructure in Part II

Our discussion will be revolving around a single CVC cluster, namely the **ZaZ** cluster (Part II, pp. 127-128), reproduced in Addendum 2. This one CVC cluster runs over nearly two pages, and is obviously an example of one of the longer ones (recall there are 1 592 CV(C) clusters, on 716 pages). It was chosen as it is representative of a large number of the features of Polis's work. The full **ZaZ** cluster may be synthetized as shown in Table 6. Getting to the **ZaZ** cluster has been explained in Section 3.1.2 (on the Vertical Base), while the first three columns seen in Addendum 2 as well as Table 6 were covered in Section 3.1.3 (on the Horizontal Base). All of this may be considered as a multi-pronged approach that leads directly into the 48 'items' listed in column 4 of Table 6.

The items are often, but by no means always, the first 'word' of the microstructure. If not, in all but a few cases the formulaic macrostructure (i.e., the codified Vertical Base + Horizontal Base) enables one to pinpoint the item that is the focus of the lexicographical description. Items that are not at the start of their entry in the microstructure include the verb *zanzalaanga* at **ZaZ + OI, V.DER., LA** for which only an example sentence is provided: "*nkasa iye zanzalaanga mu nitu yakulu*, le poison se répand à travers le corps" [the poison is spreading through the body], or the noun *mazaaza* at **ZaZ + OII, SUBST., ma** which is used in the example: "*ndiinga ifwiidi mazaaza*, voix rauque (on perçoit des tremblements *za za za*)" [the voice is hoarse (one notices a trembling *za za za*)]. The item *nzanzumune* is even imbedded in a children's song. Also, some items are actually collocations: *nzanza madyaadi*, *'carton' di nzanza*, *kweenda nzanzuba*, *nzanzala koondi(i)*, *zaaza tiiya*, *zaza uzaza*, *zaaza ki bampaangi*, *nzazi nzazi*, and *mmbanda nzazi*. The items are moreover not always presented in their canonical form, given that some remain unlemmatised, as at **ZaZ + OI, V.DER., MA** where one of the entries is the following example sentence: "*ikizanzumuna*, je m'explique, m'excuse, me défends" [I explain/excuse/defend myself] rather than the reflexive verb *-kizanzumuna* as item.

Table 6: Synthesis (and translation) of the ZaZ cluster in Polis's *Lexique*

Code	POS	Affix	Item	Grammar ...	NTD	Meaning	COL	EG
YI	PREGR.		cf. previous	—	—	—	—	—
	VERBE		*zyaanza*	> *zyanzidi & zyeenze*	✓	to get out quickly	4	—
	V.DER.	MA		*-zyanzamene*	—	be lash; be straight	—	3
		KA	*zyanzika*	caus. of *zyanzama*	—	—	—	—
YII	PREGR.		cf. previous	—	—	—	—	—
	VERBE		…	—	—	—	—	—
	SUBST.	mi	*nnzyaasi*	—	—	a straight thing	—	3
			nnzyaasi	—	—	name of a bush liana	—	—
	V.DER.		*zyaazama*	—	—	be long, be stretched out (of snake, ...)	—	—
WI	…		…	—	—	—	—	—
WII	PREGR.		cf. previous	—	—	—	—	—
	VERBE		…	—	—	—	—	—
	SUBST.	bi	*zzuaazu*	—	—	fly whisk	—	—
		zi	*nzuaazu*	—	—	whipping; noise that sounds like *zua*	3	—
OI	PREGR.		cf. previous	—	—	—	—	—
	VERBE		*zaanza*	ant. *batalala*	—	to be up front, to be rising, to be in line	—	3
	SUBST.	bi	*kinzanza*	—	—	tin, tin can	2	—
		ma	*zanza*	cf. *nza baantu*	—	large quantity	4	—
			manzanza	—	—	fetish *nkita* which had its *kimpasi*	—	—
			dinzanza	cf. *nzaanza*	—	small boat	—	—
		mi	*nnzaanza*	—	—	bush between streams	—	—
			nnzaanza	syn. *ntaantu*	—	liana or pole serving as a bridge	—	—
		tu	*lunzanza*	—	—	kind of fly very eager for blood	—	—
			lunzaanza	—	—	wide-openness	1	—
		zi	*nzanza*	—	—	arrow	—	2
			nzanza madyaadi	—	—	edible rat, which roams the *madyaadi* grass	—	—
			nzaanza	—	—	whale boat, steel boat	—	—
			'carton' di nzanza	—	—	tin sheet used as the back of a cupboard	—	—
			nzaanza	—	—	line	—	1
	V.DER.	BA	*zaanzaba*	—	—	to crawl (of snakes, insects, …)	—	—
			nzanzaba(i)	—	—	large insect which climbs in trees	—	—

Code	POS	Affix	Item	Grammar ...	NTD	Meaning	COL	EG
			kinzanzabala	—	✓	a young boy, a small girl	—	—
			nnzanzuba(u)	—	—	winged insect, comes out of the ground	—	—
			kweenda nzanzuba	—	✓	to hurry	—	—
		MA	*zaanzama*	—	—	be in line, to make a line	—	3
			zanzumuna	—	—	to do a little and then move on	—	1
			-kizanzu-muna	—	—	to explain/excuse/ defend oneself	—	—
				-zaanzu-mukini	—	be having pins and needles	1	—
			zanzumuni ... ?	—	—	to oscillate, to go quickly	—	—
			nzanzumune	—	—	(maybe just used for the sound effect?)	—	—
		NGA	*nzanzaanga*	the 1st 'a' H, the 2nd L	—	forest tree whose wood is not very strong	—	—
		KA	*zaanzika*	caus. of *zaanzama*	—	—	—	—
		LA	*zaanzala*	—	—	to crawl; …?	—	—
			bizaanzala	—	—	that which crawls (i.e., insects, ants)	—	—
			nzanzala koondi(i)	—	—	spider web	—	—
			zanzalaanga	—	—	to spread	—	—
OII	PREGR.		cf. previous	—	—	—	—	—
	VERBE.		*zaaza*	> *zeeze*	✓	be very ripe	—	—
			zaaza tiiya	syn. *zaasa ?*	—	to light a line of fire (along the forest)	—	—
			zaza uzaza	—	—	be talkative	—	—
	SUBST.	bi	*zaaza ki bampaangi*	—	—	instinct to imitate, fear of other's judgement	—	—
		ma	*mazaaza*	+ trembling voice	—	hoarse	—	—
		mi	*nnzazi*	—	—	a strip of bush being farmed	—	—
			nzazi nzazi	—	—	along	—	—
		zi	*nzazi*	syn. *Ni Kongo*	—	lightning	—	—
			mmbanda nzazi	—	—	kind of plant	—	—
	V.DER.	LA	*zaazila*	—	—	to tremble	2	—
			nzaazila	—	—	trembling, vibration	—	1

Following the signpost VERBE the base verbs are shown, here *zyaanza* at YI, *zaanza* at OI, and *zaaza* at OII. Even though one could be led to assume that everything that follows is related semantically to each of these starting points, this is of course not the case. While base verbs are often accompanied by their perfect forms (as seen at *zyaanza* > *zyanzidi & zyeenze* and *zaaza* > *zeeze*), perfects of verbs with verbal extensions may be items in themselves (such as *-zyanzamene* and *-zaanzumukini*), and are thus included without a mention of their more canonical form.

Although Polis is not explicit in this regard, and certainly not systematic, he does include information on grammar (perfects, verbal extensions, etc.), pronunciation (especially on tone), lexical relations (synonyms, antonyms, etc.), various cross-references, and in other CVC clusters also labels and even usage notes. This type of information has been grouped in column 5 of Table 6.

At times, and in addition to a translation equivalent, Polis also provides a paraphrase in the language itself, information which typically precedes the translation equivalent proper. In Table 6 the presence of such information is indicated with a tick mark (✓) in the column headed by NTD (for Kintandu).

The most important information contained in any dictionary is of course the meaning; in a bilingual dictionary consequently the translation equivalents. Here Polis does rather well although one often needs to 'derive' the generic meaning from a series of collocations (COL) or example sentences (EG). The number of additional collocations and example sentences used to achieve this is indicated in the last two columns of Table 6: these come over and above the items which are collocations or examples in themselves.

3.2.2 The microstructure in Part I

By and large, the microstructure seen in Part I mimics that of Part II, albeit that it is on the whole simpler. This has to do with the nature of what Polis terms pre-grammatical forms. These turn out to be mostly ideophones (Doke 1935: 118), a decidedly elusive word category to characterise semantically as well as lexicographically (see De Schryver 2009). Various specific sounds and exclamations are also covered in part I, as well as the occasional infrequent word class. A sample page is shown in Addendum 3, with the synthesis presented in Table 7.

Ideophones, while an important word category of Bantu languages, are often overlooked in dictionaries. Polis's lists are a true treasure trove. Especially interesting are his and his team's attempts to derive other word classes from them, such as their suggestions that *nùù/núú* gave rise to the verb *nuuka*, or that *nyau* gave rise to the verb *nyaanga*. While the paraphrase and folk etymology seen at *nnniii(w)iiiii* vacillates between being intriguing and funny, the information does give extra insight into the compilation approach: even casual interlocutors (no doubt grabbed from their bicycles while on the way to the market) provided input on the words' uses, meanings and etymologies.

Table 7: Synthesis (and translation) of a pre-grammatical page (Part I, p. 24) from Polis's *Lexique*

Vertical Base	Horizontal Base	Item	Meaning	Grammar ...	NTD	COL	EG
Ne	Y	*nye*	of emptiness	—	—	—	2
		nnye	of crushing	—	—	—	1
		nyee	of piercing pain	—	—	—	2
		nyeee	cry of the cicada *mungyeenga*	—	—	—	—
		nye nye	the crying of small children	—	—	—	—
	W	*nwe*	of pinching	—	—	—	—
		nue nue	of small movements	—	—	—	1
		nue nue	sound of muffled fart	—	—	—	—
	O	*ne*	of solitude	—	—	—	—
Ni	W	*nwi*	of rapidness	—	—	—	—
		nwi	sound of muffled fart	—	—	—	—
		nwi	(*May*.) of sweetness	—	—	—	—
	O	*ni ni ni*	cry of small children when they see a parent arrive	—	—	—	—
		nni	buzzing (of flies, wasps, ...)	—	—	—	—
		nnniii(w)iiiii	sound of a bee, with (w) a small change in sound corresponding to the moment the bee stops its wings to sting	(w) has become *u* in the word *nyuki* 'bee' (according to a casual interlocutor)	—	—	—
No	YW	...	—	—	—	—	—
	O	*nno, no*	of dripping	—	—	—	1
Nu	Y	*nyu*	(*May*.) of piercing pain	—	✓	—	—
		nuu	of bad smell (with grimace)	L tone; > v. *nuuka* smell	—	—	—
		nuu	of good smell	H tone; > v. *nuuka* smell	—	—	—
Nau		*nyau*	mewing of a cat	> v. *nyaanga*	—	—	—
Nan	Y	*nyana*	of tiptoeing	—	—	—	1
		nya nya	exclamation of surprise, of disapproval	—	—	—	—
		nyanika	of twinkling	—	—	—	1
	W	*nuana*	of rapidness	—	—	—	—
		nuanuna	of rapidness	—	—	—	—
		nanika	of movement	—	—	1	1

3.3 The mediostructure of the *Lexique*

The mediostructure of a dictionary is the system of cross-referencing (Wiegand 1996, Gouws and Prinsloo 1998). A frequent type of cross-reference found in Polis's work is the one from Part II to Part I, linking the main dictionary material to the pre-grammatical forms. As could be seen from Addendum 2 and Table 6, the reference marker used for this is *cfr antea* [cf. previous]. The reference position is always the same, namely the microstructural slot following the signpost PREGR. in Step 2 of the macrostructure's Horizontal Base, while the reference address remains undefined. However, given the structure of the dictionary the reference address can be said to be implicit. For instance, the external reference addresses for the reference positions of this type in the **ZaZ** cluster (see Table 6) cross-refer to the corresponding entries in the **Zaz** cluster, reproduced in Addendum 4 and synthetized in Table 8.[14]

Table 8: Synthesis (and translation) of the pre-grammatical **Zaz** cluster (Part I, p. 26) in Polis's *Lexique*

Vertical Base	Horizontal Base	Item	Meaning
Zaz	YII	*zyazya*	of gushing forth (e.g. blood)
	WII	*zwazwa*	alternating hitting sound of a flexible stick used to cut grass left and right
	OI	*zanzumuni zazumuni*	of rhythm (dance, fast walk)
	OII	*zazaza*	of trembling (arrow, lightning, dance, voice)

Several other cross-reference markers are found, including *cfr* / *cfr.*; *cf* / *cf.*; *cfr supra*; *cfr Prov. et devinn.* / *cfr devinn.* [for references to proverbs and riddles]; and *syn.* These typically link items within a single CVC cluster or from different CVC clusters to one another. In addition to such explicit cross-references to items (i.e., full words) within the microstructure, explicit cross-references to specific Steps in a CVC cluster's macrostructure are also found, such as "*cfr NGA, KA*" at **DoK & LoK + OII, V.DER., uLA** (Part II, p. 458).

Lastly, the excerpts (36) to (38) show that the stacking of near-synonyms in Kintandu (seen at (36)) en lieu of a proper translation equivalent should also be considered to be (implicit) cross-references, for which alternatively *aussi* [also] (as seen in (37)) or simply *cfr* [cf.] (as seen in (38)) are used. These three items (*maanga*, *mpyaata* and *ngoombe*) cross-refer to one another, but in three different ways:

(36) **MaG + OI, SUBST., ma** (Part II, p. 228): *nkisi maanga, mpyaata, ngoombo; ta maanga ma nkele*, faire le jugement du fusil (tirer de près, s'il touche coupable, sinon libre) [do the gun judgement (shoot at close range, when he hits: guilty, if not: free)]; *twe teesa keti fimanga*, allons consulter (un brin de fétiche) [let's consult 'a bit of fetish'].

(37) **PaT + YII, SUBST., zi** (Part II, p. 209): *mpyaata*, fétiche pour découvrir la cause magique d'une maladie [fetish to discover the magical cause of an illness]. aussi [also] *ngoombo, maanga*. ...

(38) **GoB + OI, SUBST., zi** (Part II, p. 295): ... *ngoombo mu baabila ina*, la boîte-fétiche est en train de trembler (elle indique le coupable) [the fetish box is trembling (it indicates the guilty person)], cfr *maanga*, *mooko*, *mpyata*. ...

4. Lexicographical appreciation of Polis's *Lexique*

Before Polis's *Lexique* can be subjected to a true evaluation, a number of additional lexicographical elements must be pointed out and clarified further.

4.1 On the language and human resources

We already concluded that the main KLC variety covered is Kintandu, but is it the 'pure' Kintandu? In more than one place in the dictionary's microstructure, one finds example sentences such as the one seen in (39):

(39) **BiB + OI, SUBST., zi** (Part II, p. 16): *kikongo kizole, ki mbiimba*, le kikongo second (en opposition à l'ancien, celui des indigènes), celui des écoles, de fabrication étrangère, avec des obscurités [the second Kikongo (in opposition to the historical one, the one of the natives), the one of the schools, of foreign fabrication, with obscurities] ...

As so often with the laudable missionary efforts to codify a language and to reduce it to writing, they were at times either overzealous in their manipulation of the language, or the mother-tongue speakers felt that despite their efforts, the result still sounded stilted. One thus has to keep in mind that the data may not be as 'pure' as one would want them to be.

That said, one must also realise that Polis tried to give as accurate a picture as possible, being open about disagreements between his collaborators, as for instance seen in (40) and (41):

(40) **BuB + OII, V.DER., TA** (Part II, p. 25): ... *muuntu utabuubuta mu nzo*, qqun est en train de tâtonner dans la maison (contesté par d'autres qui disent, *baabita*) [someone is busy groping along in the house (contested by others who say *baabita* (rather than *buubuta*))] ...

(41) **KuD & K3uL + OI, V.DER., uBA** (Part II, p. 368): *nzau mu kwisa kunduba ina*, l'éléphant s'amène de son pas pesant, majestueux [the elephant arrives in his heavy, majestic step]; *mwaana mu k. ina*, l'enfant rampe à 4 pattes (sens contesté, il faudrait *kuluba*) [the child crawls on all fours (contested meaning, one would need *kuluba* (rather than *kunduba*))] ...

Polis also went to great length specifying where he recorded the material, whether or not in Mbata for instance, as in (42) vs. (43), or showing the differ-

ences between KLC varieties, especially the differences between Kintandu and Kimbata, as in (44) where the plural of *kutu* 'ear' is *makutu* in Kintandu but *matu* in Kimbata:

(42) **TaD & TaL + OII, V.DER., iSA** (Part II, p. 481): quelle direction prenez-vous? [which direction do you take?]; - *ntadisa mpuumbu*, la direction du Pool, le Nord [the direction of the Pool, the North]; - *nt. Koongo*, celle de S. Salvador, le Sud [the direction of S. Salvador, the South]; - *nt. Mbaamba*, celle de Mbaamba, l'Est [the direction of Mbamba, the East]; - *nt. Nsuundi*, l'Ouest (orientation des Bambata) [the direction of the Nsundi, the West (orientation of the Bambata)] ...

(43) **LuL + O, SUBST., tu** (Part II, p. 103): ... *Luula*, le pays des Baluula (*Luula Lumeene*); on connait ici (& à Mbata) *Luula lu Mpese*, et *Luula lu Ntari* (dans la région de Thysville) [the country of the Baluula (*Luula Lumeene*); here we know (& at Mbata) *Luula lu Mpese* and *Luula lu Ntari* (in the region of Thysville)].

(44) **KuT + OII, SUBST., (ku)** (Part II, p. 381): *kutu*, plur *makutu*; MB *matu*; oreille [ear] ...

4.2 On grammar (comment on form)

The microstructure of most dictionary entries may be divided into a comment on form and a comment on semantics (Hausmann and Wiegand 1989: 353-357), directed at the items which are the focus of the lexicographical description. The comment on form in Polis's manuscript mainly deals with pronunciation, morphology and additional word-class information.

In our manuscript, pronunciation information typically concerns tone, and could for instance already be seen for the noun *nzanzaanga* in Table 6 (at **ZaZ + OI, V.DER., NGA**), or for the ideophones *nuu* in Table 7 (at **Nu + Y**). Another example is shown in (45):

(45) **Bok + OII** (Part I, p. 31): ... *nwa uboka* (*o*, ton très bas), la bouche gronde: bruit de l'eau dans la bouche [(*o*, very low tone), the mouth roars: sound of water in the mouth] ...

Examples of morphological information could also already be seen in Table 6, in the column headed by 'Grammar ...', where the perfect forms are listed (following '>'), or where causatives are explicitly mentioned (following '*caus. of*'). Another example is shown in (46):

(46) **PuP + OII, V.DER., KA** (Part II, p. 32): ... *pupuka*, passif de *pupula* [passive of *pupula*]

While the main word classes are taken care of at Step 2 of the Horizontal Base, less-frequent word classes are mentioned in the microstructure every now and then, especially following the signpost VARIA. These word classes include: '*adverbe*' [adverb], '*adjectif*' [adjective], '*pronom personnel / pron. pers.*' [personal pronoun], '*pronom interrogatif / pron. interrog.*' [interrogative pronoun], '*pronom*

relatif' [relative pronoun], '*pronom démonstratif / pron. dém.*' [demonstrative pronoun], '*préposition / prépos. / prép.*' [preposition], '*possessif / poss.*' [possessive], '*conjonction / conj.*' [conjunction], '*locatif*' [locative], '*particule honorifique*' [honorific particle], etc., as well as '*verbe / v.*' [verb] and '*auxiliaire*' [auxiliary]. Examples are shown in (47) and (48):

(47) **Na + O, VARIA** (Part II, p. 120): ... *na* conjonction et prépos. [conjunction and preposition]; *bu kikala na ntaangu masiinsa*, quand il était environ midi [when it was about midday], *na bumbuta tala ye mana*, depuis que mes parents m'ont engendré [since my parents begot me], *na tee ye kuna nsoongi*, tout le long jusqu'au bout [all the way until the end]

(48) **NeN + O, VARIA** (Part II, p. 124): ... *neene*, adjectif; grand (dans tous les sens) [adjective; big (in all senses)]

The abbreviation '*v.*' is mostly found in Part I, where verbs derived from the pre-grammatical forms are listed, as seen in (49). These verbs are not normally repeated in Part II, though (50) shows an exception:

(49) **Ful + OII** (Part I, p. 39): *fulu*, remplissement [of filling up]; v. *fulusa*, *fulwasa*, *fuluka*

(50) **FuD & FuL + OII, V.DER., uKA** (Part II, p. 277): *fuluka*, plénitude [fullness (i.e., be full)]; *nzo ifulukidi na ndyaa*, la maison était pleine [the house was full]; - *baantu baf. ye kyeese*, les gens étaient remplis de joie [the people were full of joy] ...

4.3 On meaning (comment on semantics) and etymology

The dictionary's semantics is actually not as undeveloped as the flat microstructure tends to suggest. While it is for instance true that the numbering of senses is rare, an interesting exception is found in Part II, p. 83 "*disasana*: 1) se donner mutuellement de la nourriture. 2) se manger l'un l'autre" [1) to give one another food; 2) to eat each other]. In Part II, pp. 106-107, up to seven uses/ senses are even listed (and numbered as such) for *ta*.

Polis also often tries to move from core meanings to peripheral ones when ordering his microstructural material. A small selection of short examples includes: "*nzeefo vwamvala*, barbe touffue, d'où poilu, velu" [bushy beard, hence hairy] (Part I, p. 10), "*nua swe swe*, boire en aspirant, d'où prégustation" [drink while inhaling, hence pre-tasting] (Part I, p. 29), "*mmfu*, un trépassé, d'où revenant, spectre" [a deceased, hence ghost] (Part II, p. 47), "*lungungu*, roue, d'où bicyclette" [wheel, hence bicycle] (Part II, p. 60), etc.

In order to support the meanings, the French translation equivalents are at times complemented by Dutch equivalents, such as "flauwkens gaan" [to stroll] (Part II, p. 13), "gaarne groot" [be pretentious] (Part II, p. 32), "reus van'ne mensch" [tall, strong man] (Part II, p. 40), "hij loopt op eiren" [he walks very carefully] (Part II, p. 64), "mond vol tanden" [not know what to say] (Part II, p. 117), "stoefferken" [a pocket square] (Part II, p. 165), or "hij kent den hoofdman"

[he recognises the authority of the chief] (Part II, p. 632).[15]

Loanwords are labelled. For Portuguese with '*du portugais*', for instance: *dimpa* < pão [bread] (Part II, p. 26), *papelo* < papel [paper] (Part II, p. 28 bis), or *ndoona* < dona [lady] (Part II, p. 488); and for French with '*du français*', for instance: *buni* < bonnet [bonnet, hat] (Part II, p. 215), *moti* < mont [mountain] (Part II, p. 236), or *ntoma* < automobile [car] (Part II, p. 437).

There are also over 50 instances of the label '*vieux mot*' [old word], found at items such as: "*mmpu*, nom récent du chapeau, couvre-chef; vieux mot *mmpeewo*" [recent word for hat, headgear; old word *mmpeewo*] (Part II, p. 27), "*kimpuumpu*, vieux mot pour *kitaansi* machette" [old word for *kitaansi* machete] (Part II, p. 31), or "*mmbangi*, témoin; le vieux mot est *kyeesi*" [witness; the old word is *kyeesi*] (Part II, p. 154). References to lexical material that goes back to the 19th century are especially valuable in this regard, as contrary to the Southern and Western KLC varieties (for which lexical material exists that dates from the mid-17th century, respectively late-18th century, cf. Bostoen and De Schryver (2015)), there are no pre-20th-century sources for the Eastern KLC varieties. Glimpses such as those seen in (51) to (53) are thus highly welcome:

(51) **BoB + OII, V.DER., LA (, zi)** (Part II, p. 19): ... *mboobila*, petite torsade de perles d'il y a 40 ans [small spiral-shaped pearl necklace from 40 years ago] ...

(52) **Se + Y, VERBE** (Part II, p. 135): ... *sye nge!* vieille formule de salut (d'il y a plus de 75 ans) [old greeting formula (from over 75 years ago)]

(53) **NaB + YI, SUBST., bi** (Part II, p. 496): *kinyaambi kimana osi*, une épidémie meurtrière d'il y a plus ou moins 100 ans [a fatal epidemic from roughly 100 years ago]; = *kibwaaka*; *nitu zibweke*; le corps devenait (pâle) [the body turned 'pale']; cf *nyaangi*; *ngaambu*. - le fétiche causant cette maladie [the fetish causing this sickness]

4.4 On usage (in the extended microstructure) and labels

Quite a number of items are labelled for register, using mostly '*péjoratif*' [pejorative] vs. '*terme honnête*' [decent term], as seen in (54) and (55):

(54) **PeK + OII, VERBE** (Part II, p. 175): *peka*; courir de tous côtés, à généralement sens péjoratif de courir les femmes [run in all directions, generally has the pejorative meaning of chasing women]

(55) **TaK + OII, SUBST., ma** (Part II, p. 467): ... *taku*, fesse [arse]; grossier [rude]; terme honnête [decent term], *kito*, ou [or] *sina di kuulu* ...

Recurrent text boxes are hard to draw on a typewriter, but had the manuscript under consideration been professionally set, the contents of the various Usage Notes would have been a prime candidate to feature in such text boxes. According to Gouws and Prinsloo (2010) text boxes are a device employed in an extended microstructure:

> Instead of a compulsory and consistently applied homogeneous article structure the lexicographer has the liberty to opt for a less rigid heterogeneous article structure in which the compulsory microstructure, the default version of the specific dictionary, can be supplemented by items representing an extended microstructure. [...] One lexicographic device frequently employed in the presentation of data [is] the lexicographic text box [for which the] default presentation seems to be as article-internal microstructural entries within a typical relation of lemmatic addressing. (Gouws and Prinsloo 2010: 501)

In Polis's work such lexicographic devices are introduced by the label '*Nota:*' [Note:], and can be directed at both the comment on form, as in (56), or the comment on semantics, as in (57):

(56) **BoB + OI, VERBE** (Part II, p. 17): ... *boomba muuntu mu maambu*, engager qqun à son insu dans une affaire [draw someone into an affair without their knowing], *kotisa mu maambu, bu kena mu maambu ko*, le faire entrer dans l'affaire alors qu'il n'y est pour rien [make him join the affair while he has nothing to do with it].

(Nota: *boomba* ne prend pas deux régimes) [(Note: *boomba* does not take two concords)]

(57) **TuM + SUBST., zi** (Part II, p. 437): ... *ntumu iyisidi*, une (invitation) est arrivée [an 'invitation' has arrived], *kindaanda* [(i.e., a message)].

Nota: le concept d'invitation (incluant liberté) n'est pas de mise ici, il s'agit d'annonce d'événements qu'on doit annoncer et pour lesquels on doit répondre à l'annonce par une visite [Note: the concept of an invitation (without obligations) is out of place here; this refers to the announcement of events which have to be announced and to which one must reply by means of a visit]

Observe that the Note presented at (5) above has, in contrast to (56) and (57), a synoptic assignment, as it is directed at all entries where the label '*MB.g*' is used. While the single assignment at (56) and (57) employs immediate addressing, the synoptic assignment seen at (5) does not employ any immediate addressing, only distant addressing.

4.5 On culture and history

As is the case for any reference work that seeks to present the lexicon of a language which is foreign to the reader, culture-specific information stands out, both with regard to concrete objects, as in (58), and with regard to the speakers' world view, as in (59):

(58) **DaZ & DaS & LaZ & LaS + OI, SUBST., mi** (Part II, p. 490): *nnlaanzi*, bout d'étoffe pendu par devant et par derrière; cache-sexe [piece of cloth hung up in front and at the back; sexual-organ cover, 'G-string'] ...

(59) **KoK + OII, V.DER., uLA** (Part II, p. 78): *kiilo ki nnkookolo mi nsusu*, le sommeil du temps des chants des coqs, c'est le troisième sommeil, celui du petit matin [that

part of sleep during the singing of the cocks, this is the third sleep, the one of the early morning]

Other aspects of the culture, while being rather universal, are certainly marked among the Bakongo/Bantandu, as may for instance be deduced from (60) and (61), which both deal with language:

(60) **NiM + O, VARIA** (Part II, p. 499): *goga lunima*, ou *goga nseka noomba*, parler en permutant l'ordre des syllabes, *ba-ngu* pour *ngu-ba*, langue secrète des enfants [talk in which the order of the syllables is swapped, *ba-ngu* for *ngu-ba*, secret language of children]

(61) **SaL & SaD + OII, V.DER., iSA** (Part II, p. 583): *sadisa muuntu*, travailler avec qqn., l'aider; faire pour qqn. [work with someone, help him/her; do for someone] (*sadila muuntu*); - *bansadisa mu kunsaansa*, ils m'aident à l'élever, le soigner [they help me to raise him/her, to take care of him/her]; - *nyiimpi kena, kaansi bafweete kunsadisila mu maambu maakulu*, il est bien portant, mais on doit l'aider pour tout; (devinette; le petit enfant) [he is in good health, but needs to be helped for everything; (riddle; (answer:) the small child)] ...

Finally, Polis's manuscript also contains quite a number of snippets comprising historical information, with for instance references to and descriptions of the course of trade networks passing through the region (62), the location of former local markets and their goods (63), as well as explanations of 'indigenous' customs which turn out to be anything but, being merely the result of historical encounters (64):

(62) **Fa + W, SUBST., mi** (Part II, p. 45): ... *ffwa, unene ngutu*; le Stanley Pool est très grand; jadis les gens d'ici faisaient la navette entre le Pool (là où est l'emplacement actuel de la CITAS, parait-il) et le port de Mbala (probablement Ambrizette) [the Stanley Pool is very large; in times past the people from here shuttled between the Pool (there where CITAS is currently located, so they say) and the port of Mbala (probably Ambrizette)]

(63) **ZeM + O, V.DER., iSA** (Part II, p. 536): *zeemisa matadi mu kiseengo*, fondre du minerai de fer; il y avait un marché de lingots de fonte à Koonso u Ndeele, l'actuel Boko Kifulama, près de Kisantu [to smelt iron ore; there used to be a market of cast-iron ingots at Konso u Ndeele, the current Boko Kifulama, close to Kisantu]

(64) **DaD + OI, SUBST., zi** (Part II, p. 87): *ndaandu*, shake-hands indigène qui se fait en frottant paume contre paume; les ancêtres avaient vu les Blancs à l'Angola et les imitaient avec un peu d'adaptation [indigenous handshake produced by rubbing palm against palm; the ancestors had seen the Whites in Angola and copied them with a bit of adaptation]

In contrast, explicit references to Polis's missionary work are not very frequent, example (65) shows a rare instance:

(65) **YiK + II, V.DER., aMA** (Part II, p. 615): *bakristu mu yikama*, les chrétiens augmentent (en nombre) [there are an increasing number of Christians]

5. Metalexicographical analysis of Polis's *Lexique*

5.1 On the nature of the macrostructure

I have argued elsewhere that a dictionary's macrostructure is not merely the list of lemma signs but also necessarily includes information on the word class and morphology of that lemma sign (De Schryver 2013: 1384). This drew harsh criticism from Rufus Gouws:

> *Dit is 'n standpunt wat teoreties nie steek hou nie. De Schryver verwar hier die makrostruktuur met die vormkommentaar wat 'n funksioneel en posisioneel segmenteerbare komponent van die mikrostruktuur van 'n woordeboekartikel is.* [This is a point of view which doesn't make any sense theoretically. Here, De Schryver confuses the macrostructure with the comment on form, which is a segmentable component — both functionally and positionally — of the microstructure of a dictionary article.] (Gouws 2014: 485-486)

So, what is a dictionary's macrostructure then? According to Wiegand and Gouws's most recent definition (their definition has changed a number of times over the past few decades), in Volume 4 of the *International Encyclopedia of Lexicography*:

> The macrostructure of a printed dictionary is that textual structure that presents the ordering of all those elements of the data memories that contribute to the dictionary type specific macrostructural coverage. (Wiegand and Gouws 2013: 78)

Unfortunately, the concept 'data memory/-ies' is not defined anywhere in the *International Encyclopedia of Lexicography*. Even in the largest lexicographic documentation corpus available to us (De Schryver 2012), it is only found once, namely in the systematic introduction included in Volume 1 of the *Dictionary of Lexicography and Dictionary Research*:

> All lexicographical partial texts of a dictionary with outer access structure and therefore all the lexicographical partial texts in which an informed user can find lexicographical data in a purposeful way, form the data memory of a printed dictionary. If there is only one word list in a dictionary, this word list is identical to the lexicographical data memory. (Wiegand et al. 2010: 147-148)

A definition which avoids the concept of 'data memory/-ies' has also been provided in Volume 4 of the *International Encyclopedia of Lexicography*:

> A new understanding of macrostructure is introduced: It is the structure which is responsible for the order of all elements of a printed dictionary which make a dictionary type-specific contribution to the macrostructural coverage; the cardinality of its structure-carrying set is identical to the extent of the macrostructural coverage with the result that the macrostructure indicates the macrostructural coverage. (Wiegand et al. 2013: 31)

No doubt, all of this makes perfect sense within Herbert Ernst Wiegand's General Theory of Lexicography (Gouws 2012), but the outright dismissal of my points seems regrettable, especially given that they also admit:

> The term *macrostructure* and its equivalents in other modern languages of culture has been used in dictionary research for the past more or less six decades. The use of this term by different researchers shows smaller or bigger differences according to the specific reference object. Currently there is no unified metalexicographic use of *macrostructure* and consequently a unified concept of *macrostructure* also lacks within dictionary research. ¶ From the perspective of the philosophy of science it is not imperative to have only one concept in a specific discipline for a scientific object. A scientific pluralism could rather lead to findings, especially when the competing theoretical concepts constituting an object from different perspectives are clearly defined [...] (Wiegand and Gouws 2013: 74)

My initial point was that in order to be able to differentiate between, say, *record* as a noun, *record* as verb, and *record* as an adjective in a dictionary's macrostructure, knowledge about the word class and/or the morphology is also needed. Considering (part of) the comment on form as microstructural is only a convention. But, point taken, I guess that what I wrote was not based on conventional practice in any of the 'modern languages of culture', but rather the result of a lifetime of work in Bantu lexicography. In that regard, Polis's dictionary provides me with another good example of why I wish to stand by my point of view. Let us start with the traditional definition of the lemma sign:

> Strictly speaking, the information contained in the microstructure does not refer to the lemma but to the linguistic sign which the lemma arbitrarily represents in the macrostructure. The sign may be called the lemma sign [...] (Hausmann and Wiegand 1989: 329)

If we now look back at Polis's dictionary, then we see that the arbitrary linguistic sign is simply what we termed the Vertical Base + Horizontal Base. Those two together, which we presented in bold type throughout this article, thus make up each lemma sign. Given that the sequence of all such lemma signs constitutes the macrostructure, it follows from Polis's approach that word class information (at Step 2 of the Horizontal Base) and morphology (at Step 3 of the Horizontal Base) are thus indeed *part and parcel* of a dictionary's *macro*structure. There is (assuming purposeful dictionary consultation and not random browsing) in other words no way that one can arrive at any item in the dictionary's microstructure without passing through the macrostructure, and that macrostructure has word-class information as well as morphological information *built into* the very lemma signs!

This observation does not mean that no other word class or morphological information cannot appear in the microstructure — and in effect, we saw that it can — the point is that in order to get at the data, (some) word classes and (some) morphology have to be obligatory taken into account at the macro-

structural level. Assuming that all comments on form are necessarily segmentable and distinct from the macrostructure, as suggested by Gouws (2014: 485-486), is thus a faulty theoretical construct which is not rooted in actual practice. Or to paraphrase Michael Rundell's wonderful aphorism, this is yet another case of a lexicographic matter that works in practice but doesn't work in theory (Rundell 2012).

5.2 On the nature of the access structure

A dictionary's macrostructure is intimately linked with its outer access structures, as defined by Wiegand:

> Informally one can distinguish as follows between macro- and access structures: each macrostructure contains as partial structures n outer access structures (with $n \geq 1$); when $n = 1$ a special case prevails, i.e. that the macro- and the only outer access structure (regarding the word list) of one and the same dictionary coincide, so that a mono-accessible dictionary with one (outer, index-external) access structure prevails. (Wiegand 1989: 393, Wiegand and Gouws 2013: 77)

In Polis's dictionary there is just one outer access structure, which means that the outer access structure and macrostructure coincide. The outer access structure itself may be defined as follows (again with reference to the latest definitional incarnation of this concept):

> An *outer access structure* is a linear textual structure, whose elements are outer access text elements, which can be accessed externally through the knowledge of general dictionary-external and specifically defined dictionary-internal principles for the linear sequence of the access text elements. (Wiegand and Beer 2013: 113)

As was shown in Sections 3.1.2 (on the Vertical Base) and 3.1.3 (on the Horizontal Base) Polis's dictionary makes especially use of dictionary-internal principles for the linear sequence of the access text elements. Once a lemma sign is reached, the inner access structure takes over. In contrast to the outer access structure, however, there isn't much structure in the inner access structure of Polis's work. The various items to which the lemma signs lead are simply listed in no particular order, and as we saw in Section 3.2 (on the microstructure) both the exact position and the form of those items vary.

5.3 On the nature of the access route

Considering the entire access route which a dictionary user can follow to reach the actual lexicographical treatments of each item, one can conclude from the previous section that the innovative outer access structure provides for a rather fool-proof guidance, while the inner access structure is haphazard.

The outer access structure, while rather fool-proof, is indeed not fully fool-proof: During ideal dictionary consultation, a user should be able to look up a word without the need to know anything about that word. In Polis's work, this is not really the case. If one wishes to look up *lufwá*, for instance, should one go to the cluster **L-f**, **LuF**, or **Fa** in the Vertical Base? Reformulated, if one doesn't already know whether one is dealing with, say, an ideophone, a verb or a noun, each of those starting positions in the Vertical Base is 'correct'. Given that *lufwá* 'death' is a noun in class 11, the lemma sign is **Fa + W, SUBST., tu** (Part II, p. 45). No sooner has one mastered this as a user that one runs into trouble for nouns in other noun classes. The plural noun class 6 is a case in point. The noun *maanga*, for instance, shouldn't be looked up under the cluster **Ga**, but rather under the lemma sign **MaG + OI, SUBST., ma** (Part II, p. 228), as was seen in (36). And then there are the genuine doublets. Take for instance the word *nyalu*, should it be looked up under the cluster **NaL** or **YaL**? It turns out that it can be found under both: as *siidi nyalu yiingi* 'he/she disappeared hiding' under the lemma sign **NaD & NaL + YII, SUBST., zi** (Part II, p. 514), and as *nyalu* 'row, layer' under the lemma sign **YaD & YaL + II, SUBST., zi** (Part II, p. 617). The second outer access route here is actually the correct one. Another bone of contention concerns collocates, and the question of where they should be looked up (Bogaards 1990). Take for instance *nzo tubu*: It turns out that Polis covered this particular set under both constituents, as seen in (66) and (67), but this is by no means the general rule:

(66) **Zo + O, SUBST., zi** (Part II, p. 126): ... *nzo tubu*, ensemble de 2 pièces d'étoffe cousues ensemble (d'environ 12 m.) [outfit consisting of two lengths of cloth sewn together (of about 12 m)] ...

(67) **TuB + OII, SUBST., ?** (Part II, p. 429): *nzo tubu*, deux pagnes cousus en un seul [two loincloths sewn together as one]; *twiiku*.

The strict separation between the type of lexicographical information covered in the pre-grammatical section vs. the main section is also not always adhered to, as may be deduced from (68) vs. (69), where the verbal information (here regarding *sweetama*) is expounded on both in the pre-grammatical section (Part I), and in the main section (Part II):

(68) **Set + WII, 4** (Part I, p. 59): *sweetika*, resserrement [of tightening]; v. *sweetama*; au moral aussi, de qqun "in 't nauw gebracht", serré de près dans une affaire [also figuratively, of someone (in Dutch) 'be put in a tight spot', be forced into an affair]; v. *swetika*. (*swatata*, mince [thin]; *swetete*, s'amincir avec effort [to get thinner with effort])

(69) **SeT + WII, V.DER., aMA** (Part II, p. 591):[16] *nnti usweetama ga nsuka boonso mbaambi nkay*, le bois s'amincit à l'extrémité comme une corne (d'antilope) [the wood thins at the tip, like a horn (of an antelope)]; - *kinkutu fiswetamene*, le veston est étroit, serrant [the jacket is tight].
nnkaanda uwiidi sweetama, le papier s'est resserré, il reste peu de place pour écrire après tout ce qu'on a déjà écrit [the paper has narrowed, little space remains to

write after everything we already wrote]; - *bisona biswetamene*, les lettres sont serrées, l'écriture est serrée [the letters are packed, the writing is compressed]; *na swetika*; - *muuntu swetamene*, qqn (est rétréci), on lui coupe la parole [someone 'is shortened', he is cut off (i.e., interrupted, stopped from speaking)]: *susa*; - *muuntu usweetama mu maambu*, qqn. est serré dans les affaires, à quia, pas d'issue! [someone is forced into the affairs, nonplussed, with no way out!]

By contrast, the route a user must follow to navigate the inner access structure of Polis's work is definitely challenging at times. Whenever the microstructural description of a certain item is limited to a single example, without the provision of a direct translation equivalent, the user will have to go back and forth between the Kintandu text and its French translation in order to extract the meaning. Doing so for the entry shown in (70), for instance, will hopefully lead to the meaning 'be able to be pulled' for the verb *golakana*:

(70) **GoD & GoL + OII, V.DER., aKA** (Part II, p. 339): *myo nnsiinga ka migolakana ko*, ces lianes ne peuvent se tirer [these lianas cannot pull themselves] ...

Quite often, however, a substantial amount of cultural information needs to be negotiated in addition, as in (71), where the meaning for the verb *suunguta* 'to limp' has to be extracted from the wider context:

(71) **SuG + OI, V.DER., uTA** (Part II, p. 561): *nsusu go yadi syeetuka, muuntu mpi yadidi suunguta*, (pour soigner une jambe cassée, le féticheur met dans la maison du patient une poule à qui il a cassé et bandé la patte; les deux guériront ensemble), la poule quand elle commencera à se traîner, l'homme aussi commencera (à se mouvoir) [(in order to treat a broken leg, the fetish-priest puts in the house of the patient a chicken whose leg he broke and taped; the two will heal together), the chicken when it will start to drag itself, the person too will start 'to move'].

On occasion, Polis helps the reader by placing the (approximate) translation equivalent, rather idiosyncratically, between brackets, as seen in (71) — an approach which could also be seen in for instance (36), (53), (57) and (69) above. However, without any real structural markers in the microstructure — whether typographical (e.g., various typefaces and font sizes, the use of bold print vs. italics, etc.) or non-typographical (e.g., the use of numbering, various symbols, etc.) — the inner access route is bound to be more convoluted than it needs to be. The fact that there are also quite a number of typos when it comes to all matters punctuation, quotation marks and brackets doesn't help in this regard.[17]

The hardest part to work out in the microstructure remains the task to extract generic meanings when only a series of example phrases and/or collocates is presented, which thus implies going back and forth between each line of a microstructural entry. Some of those meanings will turn out to be related, others won't (cf., e.g., the Meaning column in Table 6). If a comparison with modern corpus-based/driven lexicography is allowed here, Polis thus provides us with the corpus evidence (i.e., collocations as well as short example

sentences that his language consultants provided for each formula), and tends to leave it to the reader to deduce the meanings from this data. Those deductions are known to be the hard part of the work of a lexicographer (Hanks 2002), so this may also be one of the reasons why Polis does not refer to his work as a dictionary yet: both the items and the translation equivalents are still hidden and often only implicit in his work's microstructure. In a way, this state of affairs could be viewed as the reverse problem of the situation whereby dictionaries typically offer context-free answers to real-world context-sensitive questions:

> I [...] demand that the future dictionary pay attention to the contexts in which the professional user needs dictionary help. In other words, I suggest that the only way to overcome the present user dilemma of general (context-free) answers to context-sensitive questions is to try to predict the reasons why the user looks up a particular headword and then try to provide a set of adequate answers. (Varantola 2002: 33)

Of course, Polis was never part of this metalexicographical discussion, but it is nonetheless interesting to note that his and his collaborators' intuition was to present actual data in context, rather than attempt to systematically abstract away that context. Whereas a typical bilingual dictionary's focus is the translation equivalent, that of Polis's dictionary is the illustrative material. That the access route, in the absence of pre-packaged ready-made answers, suffers in the process seems to be merely a trade-off. But does it work? Does Polis's dictionary describe items (in context) that a typical user may need? Does it also describe items (in context) that an advanced user may need? This is the topic of the next section.

6. Practical evaluation of Polis's *Lexique*

6.1 Billions, millions or hundreds of thousands of combinations?

The overarching question which remains to be answered is whether Polis's method and presentation actually succeeded in capturing and sharing the lexicon of Kintandu with any reasonable success. If someone were given the 26 letters of the Latin alphabet, and told that words may, say, contain up to eight letters, a fully blind approach would generate over 217 billion possibilities (i.e., $26 + 26^2 + 26^3 + 26^4 + 26^5 + 26^6 + 26^7 + 26^8$). Realizing that Bantu languages have a CV structure — and assuming C is a single letter also covering cases like CS, NC, NCS, CH, etc. (with S the glides, N the nasals, and H aspiration) — this could immediately be reduced to about 123 million (i.e., $26 + (21 \times 5) + (21 \times 5)^2 + (21 \times 5)^3 + (21 \times 5)^4$). Of course, on the one hand words can be longer than eight letters, while on the other hand there are of course numerous combinatorial constraints in any language, which enable one to bring down that number

substantially. This is exactly what Polis did. First, he realised that he doesn't need all the letters of the alphabet for an approximate description of Kintandu: *c*, *h*, *j*, *q*, *r* and *x* were not used. This is in line with Butaye (1910: 10-12), except that Butaye also uses *r*. Although Polis doesn't list *r* as part of his alphabet, he does use it sporadically, especially for foreign words, as in *Balaari* (under the cluster **LaD**). A comparison with the phonemes in the various Kikongo varieties tabled by Laman (1936: xc-xci) indicates that *h* could have been kept to indicate aspiration. In any case, with just 20 letters, of which 15 consonants and five vowels, and with the same assumption about C, the number of theoretically possible single-letter words and CV-sequences up to eight letters goes down to about 32 million (i.e., $20 + (15 \times 5) + (15 \times 5)^2 + (15 \times 5)^3 + (15 \times 5)^4$). Second, Polis also took Bantu morphology into account, and as we saw he builds everything around single CV(C) clusters — his Vertical Base. In reality, Bantu word roots are not limited to CV(C) clusters and may be longer, but he takes care of the longer ones in his Horizontal Base. Indeed, his CV(C) structures of the Vertical Base do not normally take noun class prefixes into account, nor any suffixes (i.e., verbal extensions, and others). These are taken care of in paradigmatic fashion in Steps 2 and 3 of the Horizontal Base, which further brings down the random variation. Disregarding diphthongs and vowel length to simplify the present thought experiment, the maximum number of CV(C) structures in the pre-grammatical and main sections would then be $2 \times ((15 \times 5) + (15 \times 5 \times 15))$ or two series of 1 200 CV(C) clusters. Polis actually ended up with 1 592 clusters in the Vertical Base (cf. Section 3.1.2), 527 of which in Part I and 1 065 in Part II. Each of the 1 065 CV(C) clusters in Part II was subjected to the six Step 1 combinations of palatalization, labialization and prenasalization (cf. Code 2 in Table 2), and for each of these also the Step 3 variation, thus the seven main noun prefixes on the one hand (cf. Table 3) and the more than 30 main extensions on the other (cf. Table 4). Together with the base verbs and varia from Step 2, that's still about 300 000 combinations that must have been tried out in Part II. Given the absence of base verbs and noun classes in Part I, as well as generally fewer extensions and no varia, the number of pre-grammatical combinations tried out may have been around 30 000. Of the total of 330 000 combinations, it is estimated that no more than about 5% resulted in actual words and were subsequently treated in Polis's work.

In order to evaluate the eventual selection, one would ideally be able to compare it with Kintandu corpus data. This is problematic for a number of reasons, chief among them the idiosyncratic spelling adopted by Polis. Amongst others Polis tends to double vowels in environments of compensatory lengthening (i.e., V:NC and CSV:) whereas others don't mark this as it's predictable, he uses *y* for glides whereas others use *i*, he doesn't mark tone whereas several others do, etc., but in contrast he for instance differentiates between the syllabic vs. non-syllabic homorganic nasals (at noun level) whereas others don't.[18] Nonetheless, what is feasible to do is to compare Polis's work with Butaye's dictionary, and vice versa. Two tests to this end are discussed next.

6.2 Test 1: Polis vs. Butaye

Butaye's (1909) dictionary consists of a number of sections; the part which interests us here is the first, i.e., the Kintandu–French direction. Compared to that section, Polis's work is about three times larger. Given that Butaye brought together around 8 300 lemma signs for Kintandu, but also that the treatment for each lemma sign in Polis is much more extensive, with most of Polis's lemma signs leading to multiple items in the microstructure, the assumption is that Polis covers more of the Kintandu lexicon than Butaye does. As a first test to check this hypothesis, we took every tenth page in Polis's work, selected the top item on each of those pages, and checked whether or not it is also covered by Butaye. The data for this test is shown in Addendum 5.

Of the 71 test items from Polis, as many as 27 have no equivalent in Butaye. For 6 of those who have, the correspondence is furthermore only approximate. Extrapolating from this random sample, anything between 54% and 62% of the material from Polis is also covered by Butaye, while 38% up to 46% isn't. For the pre-grammatical items, the correspondence is even lower: only two (one of which dubious) of the seven Polis items is also included in Butaye. Here the overlap is thus a mere 14% to 29%.

While some of the missing items clearly belong to the fringes of common vocabulary, such as the ideophones *kyu* and *zwe* for the sound the throat makes when one swallows, respectively the instance of a thorn entering the body, or the verb *kedingiinza* used for the pounding in small double beats such as that of one's heart, others seem much more common, such as the ideophone *zubu* to refer to intelligence or vigilance, the noun *ntuuta* for a quarrel, or even the verb *vidika* used to say something in a round-about way. If anything, this clearly indicates that Polis's dictionary contains a good number of unique items, especially ideophones and words derived from them, not covered by the otherwise excellent dictionary of Butaye.

6.3 Test 2: Butaye vs. Polis

A comparison of Butaye's dictionary with Polis's work has the potential to be even more revealing, as Butaye's dictionary can be treated as a corpus whose main lexical material ought to be in Polis's work. In this regard we are rather fortunate that Butaye took the trouble to mark all the main lemma signs in his dictionary with tramlines (i.e., two vertical parallel lines ‖),[19] leaving the common ones unmarked, while also marking dialectal/local forms with an asterisk (*). About one fifth of Butaye's lemma signs are preceded by tramlines. In order to obtain a sufficiently large and balanced sample the following procedure was followed: we selected the first lemma sign from each page, but exploiting Butaye's frequency indications, we consistently took the first unmarked word (i.e., lemma sign) from an even-numbered page and the tramline-marked word from an odd-numbered page. The focus was also on true words rather than

formatives and affixes, and those words also had to have a (recoverable, e.g. through cross-referencing) meaning. This procedure resulted in a set of 154 main (i.e., supposedly frequent) words, and 154 common (i.e., supposedly less frequent) words. Each of these was then looked up in Polis's dictionary, where we searched for 'the same word' (with either the same or an adapted spelling) having 'the same meaning' (broadly defined) and 'the same word class' (glossing over erroneous class assignations in either work). In the process, the practicability of Polis's innovative outer access structure was also tested. The full data set for this extensive test is shown in Addendum 6, which immediately reveals that Polis's dictionary can indeed successfully be used as a proper dictionary. From a user's perspective, this is an important result, one which can only be considered valid following such an extensive test.

Overall, 73% of the lemma signs from Butaye could be found as items in Polis. A further 4% were used throughout the dictionary but were not listed following their logical position in Polis's macrostructure, while for 7% of the lemmata a related item could be found in the macrostructure. Overall, then, just 16% of the material from Butaye could not be found in Polis at all. This stands in sharp contrast with the 38% to 46% of the material from Polis which could not be found in Butaye. A quick back-of-the-envelope calculation may now even suggest the number of items in Polis's dictionary. If up to 84% of the 8 300 lemmata in Butaye can eventually be found in Polis, these 8 300 x 0.84 or about 7 000 items must also be in Polis. If 7 000 items can be seen, the first test tells us that this represents only 54% to 62% of what is really covered, which suggests that Polis contains about 12 000 items.

These are the overall figures. When the data is split up according to main lemmata vs. common lemmata, as done in Table 9 and Figure 2, a highly interesting pattern emerges. Indeed, while as many as 73% of the Butaye lemmata can be looked up directly in Polis, there is a marked difference between the findability of frequent vs. less-frequent parts of the lexicon, with Polis covering as many as 87% of the frequent ones, but only 60% of the less-frequent ones. From the point of view of the material that cannot be found at all, the difference is even more marked, with just 5% of the main material not covered in Polis vs. 27% of the common material.

Table 9: Butaye lemmata vs. Polis items: Statistics for the full sample

Butaye lemma is ...	**Full sample**		**Main lemmata**		**Common lemmata**	
	N	**%**	**N**	**%**	**N**	**%**
in Polis as item	226	73%	133	87%	93	60%
used in Polis	12	4%	8	5%	4	3%
related item in Polis	21	7%	5	3%	16	10%
not in Polis	49	16%	8	5%	41	27%
SUM	308	100%	154	100%	154	100%

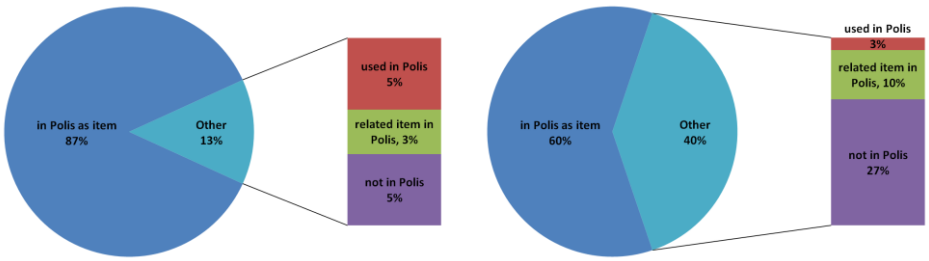


Figure 2: Butaye lemmata vs. Polis items: Main lemmata (left) vs. common lemmata (right)

While the macrostructural coverage of no two dictionaries can ever be the same (unless blatant copying is involved), one can expect any general language dictionary to cover at least the largest part of the frequent section of a language's lexicon. The left pie diagram shown in Figure 2 indicates that Polis succeeded well on this level.

Given that the sample extracted from Butaye also comes with part-of-speech (POS) information, a further breakdown is possible in order to answer questions like: Which types of words did Polis typically miss, or conversely, which types did he tend to include at the expense of others? In other words, did Polis pay particular attention to nouns, or verbs, or perhaps any other word class? The answers to these questions are summarised in Figure 3 for the main lemmata and in Figure 4 for the common lemmata.

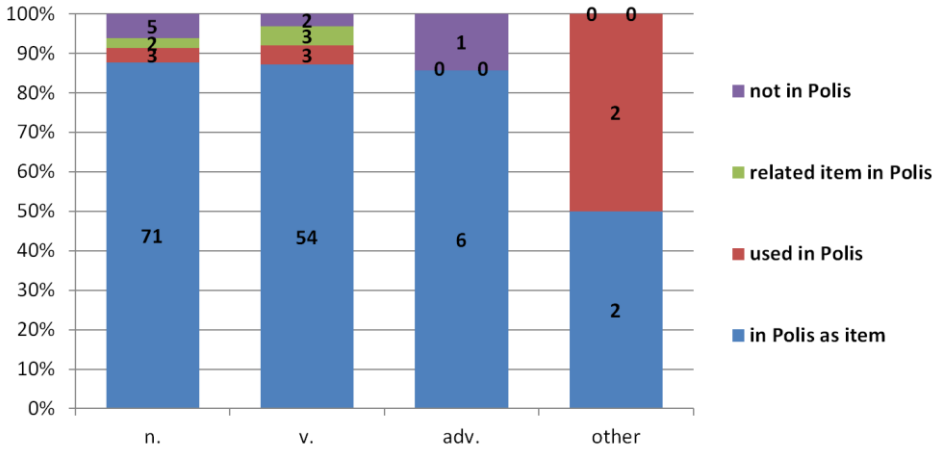


Figure 3: Main Butaye lemmata vs. Polis items: POS distribution (with the number of actual occurrences in the sample shown in the histograms)

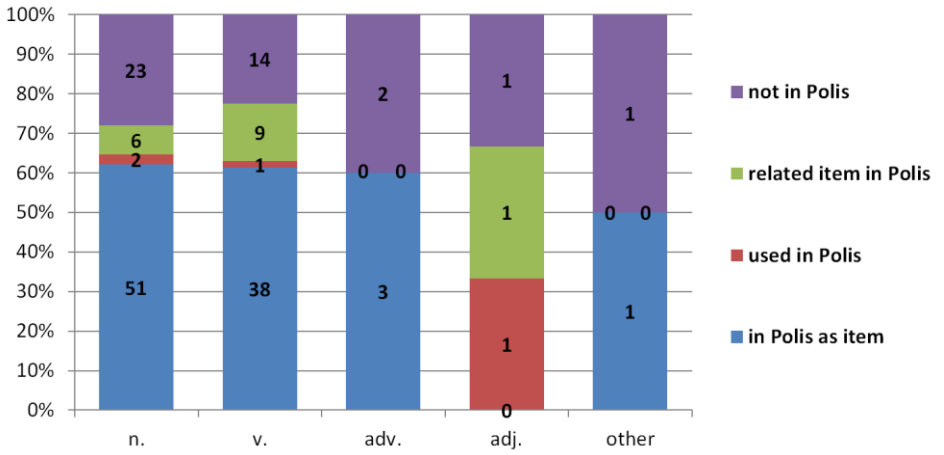


Figure 4: Common Butaye lemmata vs. Polis items: POS distribution (with the number of actual occurrences in the sample shown in the histograms)

Astonishingly, when it comes to the main word classes — nouns, verbs and adverbs — Polis seems to have been exceptionally consistent. While 87% of Butaye's material was directly included for the main lemmata overall, the distribution for these main lemmata across nouns, verbs and adverbs is respectively 88%, 87% and 86%. Likewise, while 60% of Butaye's material was directly included for the common lemmata overall, the distribution for these common lemmata across nouns, verbs and adverbs is respectively 62%, 61%, and 60%. The values for the other word classes, as may be derived from the number of actual occurrences shown in the histograms of Figure 3 and Figure 4, are too low to have any statistical value. These tests, then, not only show that Polis approached the compilation of his dictionary with military precision, but also and more importantly that he truly managed to capture the general-language lexicon of Kintandu as well as the more peripheral one (in the pre-grammatical section).

The answer to the question "Does it work?" is thus a resounding "Yes." At the same time, it is clear that his approach is thoroughly user-unfriendly for any real-world dictionary user. So what could have been Polis's underlying motives? We try to answer this in the next section.

6.4 Polis's underlying motives

Polis begins the introduction to his dictionary with the following paragraph, which in truth should be seen as a teaser, as its contents only make sense once one has painstakingly studied his dictionary:

> *Les présentes listes ne sont pas un dictionnaire; elles n'en sont que la préparation. Elles constituent un instrument de travail; elles permettent la recherche systématique des formes manquantes; elles rendent plus aisée la comparaison des formes entre elles, en comparaison verticale des mots d'une même racine, en comparaison horizontale d'un même échelon dans les racines différentes; plus aisée aussi la comparaison avec les correspondants dans des dialectes et des langues apparentés; lesquelles comparaisons permettront de préciser le sens des éléments.* [The present lists are not a dictionary; they are but their groundwork. They constitute a tool; they allow for the systematic search of the missing forms; they facilitate the comparison of the forms amongst one another, by means of a vertical comparison for the words with the same roots, and by means of a horizontal comparison for the same paradigm applied to different roots; they also facilitate the comparison with cognates in related dialects and languages; all of which will enable one to clarify the sense of the items.] (Polis 1938: i)

Even though Polis himself prefers not to see his work as a dictionary (yet), we have chosen to do so, as any modern dictionary definition would recognise his work to be exactly that; for instance: "a dictionary may be defined as a lexicographic reference work that has been designed to fulfil one or more functions, contains lexicographic data supporting the function(s), and contains lexicographic structures that combine and link the data in order to fulfil the function(s)" (Nielsen and Mourier 2007: 121). The main function of Polis's dictionary is to arrive at the best possible clarification of the various senses of an item. The lexicographic structures Polis developed for this goal have been expounded on at length in this article. With them, Polis firstly claims that "they allow for the systematic search of the missing forms". Although Polis has not succeeded in pinpointing *all* the items of Kintandu, his approach could certainly have revealed virtually all — probably given more time and a larger team, as it is known that no one person knows all the words and all their senses of a language. It is now indeed entirely feasible to run through the different formulas, to complete them where necessary, and to add the missing items.

Polis's second claim in the quoted opening paragraph summarises his core motive, namely to "facilitate the comparison of the forms amongst one another, by means of a vertical comparison for the words with the same roots, and by means of a horizontal comparison for the same paradigm applied to different roots". Unfortunately, the concept 'root' should not be understood in a true linguistic sense here, but even his blind, mechanical approach to vary CV(C) clusters and to combine them with affixes will obviously bring related material together. Take for instance the **ZaZ** cluster shown in Addendum 2 and summarised in Table 6. Going up and down vertically indeed reveals items that are likely related morphologically, such as *-zyanzamene* 'be straight' and *nnzyaasi* 'a straight thing', *dinzanza* 'small boat' and *nzaanza* 'whale boat, steel boat', *nzaanza* 'line' and *zaanzama* 'be in line, to make a line', *zaanzaba* 'to crawl (snakes, insects, …)', *zaanzala* 'to crawl; …?' and *bizaanzala* 'that which crawls (insects, ants)', etc. Recalling that the pre-grammatical material actually belongs with the main material, one should also extent the vertical discovery path from **ZaZ** to **Zaz**,

shown in Addendum 4 and summarised in Table 8, whereupon one may for instance want to link *mazaaza* 'hoarse' in Part II with *zazaza* 'of trembling (arrow, lightning, dance, voice)' in Part I. The horizontal discovery path Polis suggests simply refers to the application of Steps 1 to 3 of the Horizontal Base: these steps are brought into position for each and every CV(C) structure, so cross-comparing the results will ultimately lead to the discovery of the morphological rules of the language.

Polis's third claim in support for his approach refers to the fact that his lists "facilitate the comparison with cognates in related dialects and languages". This aspect is actually not well developed, and is certainly not unique to his approach. References to other KLC varieties and even other languages have indeed been included (see the paragraph following the variation discussed in (5) to (8) in Section 2 above), but they are not the result of treating those other varieties and languages in the way Polis does.

Should one now take Polis's suggestion seriously and proceed with the compilation of a 'classical' dictionary based on his 'raw material'? This could certainly be a worthwhile (and challenging) venture, but it is our contention that a digitised and searchable version of his work as it stands is far more intellectually stimulating to interact with. Already, this unique source joins the ranks of only a few other lexicographical reference works which challenge preconceived metalexicographical notions. This in itself makes Polis's manuscript one to cherish.

6.5 Test 3: Keywords in Polis as compared to Butaye

Tests 1 and 2 dealt with macrostructural issues, so in this last section of the practical evaluation we propose an attempt to qualify the microstructure by computational means. Due to the idiosyncratic orthography used by Polis, any automated method faces an uphill struggle. However, the microstructure is also full of French (whose orthography has eagerly been guarded by the French Academy since the mid-seventeenth century), and thereby provides a way out. We propose to look at the French translations in order to know more about the Kintandu contents throughout the dictionary. In actual fact we will also look at the Kintandu in the process, but we do not expect the results for it to be very revealing. We are especially interested in knowing in what way the microstructural contents of Polis differ from those of Butaye.

We proceeded as follows: a text corpus was prepared containing the full text of Polis, bar the introduction, and another text corpus was prepared consisting of the full text of the Kintandu–French direction from Butaye. Both corpora thus consisted of material in Kintandu and French. Wordlists were drawn up for each text corpus, the result of which was a list of all the Kintandu and French words in the respective texts, together with the frequencies for each of those words as used in the respective dictionaries. The two frequency lists were then compared to one another, using the KeyWords function of WordSmith

Tools (Scott 1996–2015). The overall statistics for these two text corpora are given in Table 10, from which it may be seen that both behave similarly: for each 1 000 words added, Polis contains an average of 55.65 unseen words, while Butaye contains an average of 56.64 unseen words. The difference is just a single word, so Polis is only slightly more repetitive (or less original in its variation of word choice).

Table 10: Overall statistics for the Polis and Butaye text corpora

	Items (Polis) **Lemma signs** (Butaye)	**Tokens**	**Types**	**STTR** (1 000)	**STTR std. dev.**
Polis	± 12,000	319,552	47,603	55.65	45.26
Butaye	± 8,300	107,662	24,108	56.64	42.49

(with STTR = standardised type-token ratio; std. dev. = standard deviation)

For a word to be considered a keyword in Polis, it had to occur at least three times in his dictionary. WordSmith Tools's KeyWords function calculates so-called keyness values, for which we used the log-likelihood statistical test and set the probability to ≤ 0.000001, meaning that each keyword's appearance had a chance of only one in a million of not being statistically significant (Taljard and De Schryver 2002: 52). Abbreviations and the dictionaries' metalanguage were pruned from the resulting keyword list, as were function words, and for Kintandu also the keywords which are the result of the different spellings. The keyword list was divided into a French and Kintandu section, and each of these was again divided according to positive and negative keyness values. Positive keywords occur more often in Polis than would be expected by chance in comparison with Butaye; the occurrence of negative keywords is the reverse. In short, then, we trust that this method will allow us to characterise the microstructure of Polis's dictionary in comparison to Butaye's.

Rather than to now present long lists of words with their respective frequencies in the two text corpora, together with their actual keyness values and probabilities, we have opted for four word clouds generated with Wordle (Feinberg 2008–14), as what interests us here is to get an instant picture of the differences.[20] The relative size of the keywords shown in the word clouds is in proportion to their keyness values, and they have mostly been arranged in alphabetical progression (the standard A-to-Z that is, not Polis's ordering).

From Figure 5 it is immediately clear that the most outstanding concepts in Polis are actually nondescript, as they refer to generic objects, animals and people: 'thing' (*chose*), 'animal, beast' (*bête*), 'child' (*enfant*), 'kid, child' (*gamin*), 'people' (*gens*), 'issues' (*affaires*), and 'kind, type, species (of)' (*sorte (de)*). When the concepts become descript they refer to non-verbal communication, strong feelings, or issues to do with face: 'look, appearance' (*allure*), 'feeling' (*sensation*), 'heart' (*coeur*), 'gesture' (*geste*), 'intensely' (*intensément*), 'body' (*corps*), 'rhythm' (*rythme*), 'empty' (*vide*), 'eyes' (*yeux*), 'mouth' (*bouche*), and 'strongly' (*fortement*).

Figure 5: French positive keywords in Polis as compared to Butaye

Next up are those concepts that connect the Bantandu to mother earth, their social structures, and their close or cult animals: 'earth' (*terre*), 'stream, brook' (*ruisseau*), 'wall' (*paroi*), 'clan' (*clan*), 'fetish-priest' (*féticheur*), 'fetish' (*fétiche*), 'dog' (*chien*), 'rat' (*rat*), and 'leopard' (*léopard*). The words which refer to daily activities that stand out in Polis compared to Butaye include: 'job, work' (*travail*), 'slap, smack' (*giffle*), 'blow' (*coup*), 'fall' (*tombe*), 'hold' (*tient*), and 'shout, scream' (*crie*), as well as 'truck' (*camion*) and 'cooking pot' (*marmite*).

Figure 6: French negative keywords in Polis as compared to Butaye

Figure 6 is the inverse of Figure 5 in that it shows us the concepts and words which Polis does not often use compared to Butaye. The nondescript words favoured by Butaye are simply the higher-register versions of some of those of Polis: 'species' (*espèce*), 'game, prey' (*gibier*), 'names' (*noms*) and 'any, some' (*quelconque*). The connection with mother earth is represented in Butaye by 'wood' (*bois*) and 'grass' (*herbe*). The activities underrepresented by Polis compared to Butaye are: 'to give back' (*rendre*), 'act' (*action*), 'to spill' (*répandre*), 'to put' (*mettre*), 'to move away' (*éloigner*), 'to break, to crush' (*briser*), 'to burn' (*brûler*), 'to spread out' (*étendre*), 'to muddle up, to tangle up' (*embrouiller*), and 'to cover' (*couvrir*). Characterisations are also often more specific and more individualistic in Butaye: 'a lot' (*beaucoup*), 'native, indigenous' (*indigène*), 'individual' (*individu*), 'today' (*aujourd'hui*), 'side' (*côté*), and 'better' (*mieux*), as well as '-self, each other' (*se*) and 'able (to)' (*pouvoir (être)*).

Overall, then, one may state that the material in Polis is closer to nature; wilder, less-polished and less-precise in its word use and activities described; deals with the deeper culture; and places more weight on the need to live in harmony, while condoning violence, stressing the importance to keep face, and avoiding the singling out of the individual.

Considering the limitations placed on an analogous exercise for the Kintandu words used throughout both dictionaries, Figure 7 confirms rather than disputes these findings.

Figure 7: Kintandu positive (left) and negative (right) keywords in Polis as compared to Butaye

The positive keywords in Kintandu include: 'palm wine' (*malafu*), 'animal; meat' (*mbisi*), 'job, work' (*kisalu*), 'body; colour; form' (*nitu*), 'house' (*nzo*), 'dog' (*mbwa*), 'oil' (*mafuta*), 'fetish' (*nkita*), and 'peanut' (*nguba*). The fact that the ideophones 'of nothing' (*kena*) and 'of sound' (*po*) have a strong showing here, once more indicates the virtual absence of this word class from Butaye.

Kintandu words underused in Polis compared to Butaye include: 'to walk, to go' (*kuenda*), 'fire; heat' (*tiya*), 'to blame' (*bela*), 'knife' (*mbele*), 'material, fabric' (*nlele*), 'Matadi' (*Matadi*), 'things' (*bima*), 'to lie down' (*leka*), 'Nzeza (an indigenous 'saint')' (*Nzeza*), 'to put on' (*sila*), and 'illness' (*kimbefo*).

These few Kintandu keywords suggest a far more local concern in Polis than in Butaye, and related to this, Polis seems to refrain from over-comparing with the outside world: he simply lives among the Bantandu and describes

what he notices from within the culture. Butaye, on the other hand, tries to bridge cultures, and looks at the Bantandu as an outsider. From the perspective of the user, Butaye's is the more user-friendly approach. From the perspective of the original culture, Polis's is the truest approach.

7. Polis's *Lexique* and the most innovative outer access structure of any Bantu dictionary

We set out to present a little-known dictionary manuscript for Kikongo, and showed that its contents are so unique that it more than deserves to be known and consulted as a dictionary in its own right. The material for this dictionary was brought together about a century ago, and typed up on over 700 pages in 1938 in Leuven, by the Jesuit missionary Charles Polis. Only a few of the stencil duplicated copies are still extant, and their readability is variable, but the data has now been digitised and is available for computerised searches. Polis worked in the Lower Congo region, east of the Inkisi River, and all the evidence — both dictionary-external and especially dictionary-internal — converges to pinpoint the particular variety described in the work as Kintandu, an eastern KLC (i.e., Kikongo Language Cluster) variety.

Our analysis of the structure of Polis's dictionary revealed a highly complex approach to the macrostructure, consisting of a Vertical Base + Horizontal Base. In simple terms, the Vertical Base is the sequence of all possible CV(C) clusters in the language. These clusters are listed in a non-standard ordering, which makes an index to the work highly desirable (for which see Addendum 1). These clusters vaguely resemble word roots. The Vertical Base consists of three steps. Step 1 is concerned with combinations of palatalization, labialization and prenasalization of each CV(C) cluster (for which codes are used, as seen in Table 2). Step 2 signposts the main word classes (base verbs, base nouns, derived forms (verbs, nouns, and others), and varia), which leads to differential variation in Step 3: noun class prefixes for the nouns (for which see Table 3), and extensions for the derived forms (for which see Table 4). The whole work is also divided into a short pre-grammatical section, Part I, which mainly deals with ideophones and derivations thereof, and the main section for all the other word classes, Part II. The macrostructure in Part I uses a simplified version of that from Part II. The same is true for the microstructure, where Part I mimics and simplifies what is done in Part II. On the whole, the focus in the microstructure is on presenting evidence (example phrases, collocations, etc.) together with translations into French, rather than on generic translation equivalents. Compared to the highly structured macrostructure, the microstructure is rather flat. The mediostructure mainly connects the relevant bits from Part II with those from Part I.

In the lexicographical appreciation we detailed how Polis must have interacted with his native-speaker collaborators, noting their every comment and hesitation, and also how he tried to differentiate between his main variety

Kintandu and the neighbouring variety Kimbata. We further indicated how the main components of the traditional dictionary were nonetheless also covered in Polis's idiosyncratic dictionary: comments on form, comments on semantics, etymology, usage notes (in the extended microstructure), labels, and also encyclopaedic excursions into the culture and history of the Bantandu.

In the metalexicographical analysis we argued that even the latest definition for the concept 'macrostructure' that has emanated from the General-Theory-of-Lexicography school may not be correct. The sequence of lemma signs which constitute the macrostructure of Polis's dictionary consist of the combinations of the Horizontal Base + Vertical Base in each case, and thus necessarily imbed word class information and morphological information right into the very heart of the lemma signs, and hence also into the very heart of the macrostructure. There is no escaping this, and the insistence on assigning the comments on form to the microstructure is untenable. We further showed that in Polis's dictionary, the outer access structure and macrostructure coincide, and that while Polis makes especially use of dictionary-internal principles for the linear sequence of the access text elements, there isn't much structure in the inner access structure. As to the nature of the access route, once a user succeeds in reaching the entry they are looking up, the need to map meaning onto use takes over — a very modern lexicographic concept indeed (cf. Hanks 2002).

In the practical evaluation we first suggested that Polis and his team must have tried out well over 300 000 combinations as candidate word forms, to eventually keep about 12 000 'valid' items (i.e., real words) as the focus of their lexicographical descriptions. Three tests accompanied this evaluation, all involving Polis's dictionary on the one hand, and a contemporary Kintandu dictionary by Butaye (1909) on the other. These tests confirmed Polis to be an absolutely unique source: just 16% of the material from Butaye could not be found in Polis at all, while as much as 38% to 46% of the material from Polis could not be found in Butaye. Polis is especially strong on the so-called pregrammatical forms (i.e., ideophones, but also words for sounds and exclamations), and generally focuses more on the Bantandu in situ, while Butaye tends to contrast the Bantandu with the outside world. These subtle differences were revealed by means of a keyword study of the full Kintandu–French texts of these two dictionaries, the results of which were illustrated in the form of word clouds (for which see Figures 5 through 7).

From all of the above it should thus be clear that in addition to its intrinsic value as a highly informative lexicographic resource about a Kikongo variety for which there is otherwise very little data available, Polis's most important contribution to the lexicographic community at large is that he designed a truly unique outer access structure which forces all those interested in lexicographic theorising to go back to their drawing boards. In more practical terms, and returning to our three opening questions, we are now in a position to provide answers to them: (1) "Could this outer access structure be applied to all Bantu languages?" Yes, no question; but it can immediately be pointed out that if it

were done with more linguistic insight and if the usual alphabetical ordering were used, a far better product would be the result. (2) "Is this outer access structure perhaps universal?" No, not in its present form, given the approach assumes open syllable structures, more specifically CVCV sequences, and includes Bantuesque features like a system of noun classes and (verbal) extensions. (3) "Could such an outer access structure perhaps be a solution for all non-corpus-based lexicographic endeavours?" The question is thus whether an — obviously language-dependent — atomic approach could be designed whereby all types of permutations are tried out in order to 'find' all the words of a language's lexicon. This certainly seems feasible theoretically, but in this day and age of large electronic corpora we see no need for it. Polis's attempt, however, is testimony to what some were capable of in the pre-corpus era, including in the absence of the tens of thousands of paper slips which were otherwise common in the lexicographic backrooms of the time. Polis truly came up with a system whereby he could, without access to any draft material, work through the alphabet and hardly miss any word — literally conjuring up a language out of thin air.

Acknowledgements

The research for this article was funded by Starting Grant No. 284126 of the European Research Council and by the Special Research Fund of Ghent University.

Endnotes

1. Each of the sections of the manuscript is dated, as is the introductory material, with dates ranging from March 21, 1938 up to October 9, 1938 — see Table 1. Three of those dates are also accompanied by "Louvain", French for the city of Leuven.
2. Van de Casteele (1968: col. 827-828) mentions eight fascicules, as does the earlier Daeleman (1966: 7), but it must have been nine based on the current binding of the instalments as seen in the extant copies.
3. For readability of all cited material from Polis (1938), we have added formatting (bold, italics, underlining, proper spacing, etc.) as well as translations into English (between square brackets). Obvious typos have also silently been corrected. Some inconsistencies were left as is, however, such as the inconsistent notation of vowel length (e.g., *mpyaata* vs. *mpyata*), the use of 'qqun', 'qqn.' and 'qqn' as abbreviations for 'quelqu'un', etc. Sample pages from the actual manuscript may be seen in Addenda 2 and 3.
4. Frequency counts are lower-bound values, meaning that there may be more occurrences (which we haven't spotted).
5. 'BLR' stands for the Bantu Lexical Reconstructions database (Bastin et al. 2002), and the integer following it is the index number in that database. The abbreviation Ṇ stands for a syllabic homorganic nasal.

6. Likely using the heavily reworked version by Van Wing and Penders (1928), cf. De Kind et al. (2012) for more information on this source.
7. By considering the subsequent steps of the outer access route, this infelicity may at least be explained, however.
8. However, for the plural class 10 the pronominal prefix (PP) zi- is used as label, rather than N-, which leads to an alternative interpretation: the labels at the nouns show the plural PPs. This is indeed useful information, as PPs are used to form "connectives, substitutives, possessives, precessives, determinants, interrogatives, demonstratives and numerals" (Daeleman 1966: 212, our translation) with those nouns, but why show plurals, especially considering the fact that nouns are typically presented and illustrated in their singular forms in the text that follows?
9. For the variation seen in Kikongo/Kintandu at classes 1, 3 and 4, as well as class 7, the reader is referred to the detailed exposition in Bostoen and De Schryver (2015).
10. Unfortunately, Polis does not draw this convention to its logical conclusion, as he doesn't use double nasals for the object concord of class 1, as in "*bantombwele kuna gaata mu kufuundisa*, on le fit monter au village pour l'accuser (chez le chef)" [they made him go up to the village to accuse him (at the chief's place)] (Part II, p. 427), where it should have been *banntombwele* for *bantombwele*.
11. There are some small errors on this level of the ordering, however. Question marks are also used to indicate uncertainty with regard to class membership.
12. Observe that there are some small errors on this level of the ordering as well.
13. Except for the term 'imperfective' (Nurse and Philippson 2006: 190-192), the terminology used here is that of Daeleman (1966), with the 'rollative' our coinage based on his discussion on p. 179.
14. Although there is a cross-reference at the reference position **YI, PREGR.** in the **ZaZ** cluster, there is no corresponding reference address in the **Zaz** cluster. Recall (see Table 1) that **ZaZ** was typed in before **Zaz**, which may explain this infelicity. Other similar dead cross-references from fascicule 4 onwards must be seen as errors though.
15. In at least one instance Daeleman (1966: 175, 190) copied over such a Dutch paraphrase: "spartelen om iets te bekomen" [to flounder in order to obtain something] (Part II, p. 336).
16. Observe that there is no cross-reference from the cluster **SeT** to the pre-grammatical cluster **Set**. This seems to be part of a more general issue; as such explicit cross-references (using *cfr antea* [cf. previous]) seem to suddenly stop being used after **KeF** (Part II, p. 322).
17. Recall, however, cf. note 3, that such typographical markers were added for the quoted material in this article, and that (most) errors were also corrected.
18. The differences in spelling will become clear from the data shown in Addenda 5 and 6.
19. Observe that this use is the opposite of their function in the early OED, where they indicate words which are "alien or not yet naturalized" (Murray 1888: xxvi). See also Ogilvie (2008).
20. For another use of word clouds in metalexicography, see De Schryver (2012: 471-473). In practical lexicography, word clouds have been proposed (e.g., Bowker 2012: 384-385), but they have not actually been used to date. The closest to them are the wordmaps found in Wordnik (McKean et al. 2009–15), an extension of Dream no. 17 (De Schryver 2003: 165-166).

References

Bastin, Y., A. Coupez, E. Mumba and T.C. Schadeberg (Eds.). 2002. *Bantu Lexical Reconstructions 3 / Reconstructions lexicales bantoues 3*. Tervuren: Royal Museum for Central Africa. Available online at http://www.africamuseum.be/collections/browsecollections/humansciences/blr.

Bentley, W.H. 1887. *Dictionary and Grammar of the Kongo Language, as Spoken at San Salvador, the Ancient Capital of the Old Kongo Empire, West Africa / Compiled and Prepared for the Baptist Mission on the Kongo River, West Africa*. London: Baptist Missionary Society; Trübner & Co.

Bentley, W.H. 1895. *Appendix to the Dictionary and Grammar of the Kongo Language, as spoken at San Salvador, the Ancient Capital of the Old Kongo Empire, West Africa / Compiled and Prepared for the Baptist Mission on the Kongo River, West Africa*. London: Baptist Missionary Society; Kegan Paul, Trench, Trübner & Co.

Bittremieux, L. 1922. *Mayombsch idioticon, Deel I-II* (Congo-Bibliotheek 10-11). Gent: Drukkerij Erasmus.

Bittremieux, L. 1927. *Mayombsch idioticon, Deel III: Verbeteringen en Aanvullingen, Plantenkundige Woordenlijsten en Zakenregister* (Congo-Bibliotheek 21). Brussels: Drukkerij Essorial.

Bogaards, P. 1990. Où cherche-t-on dans le dictionnaire? *International Journal of Lexicography* 3(2): 79-102.

Boone, O. 1973. *Carte ethnique de la République du Zaïre, quart sud-ouest* (Annales — Série in-8° — Sciences humaines 78). Tervuren: Royal Museum for Central Africa.

Bostoen, K. and G.-M. de Schryver. 2015. Linguistic Innovation, Political Centralization and Economic Integration in the Kongo Kingdom: Reconstructing the Spread of Prefix Reduction. *Diachronica* 32(2): 139-185 + 13 pages of supplementary material online.

Bostoen, K., O. Ndonda Tshiyayi and G.-M. de Schryver. 2013. On the Origin of the Royal Kongo Title *ngangula*. *Africana Linguistica* 19: 53-83.

Bowker, L. 2012. Meeting the Needs of Translators in the Age of e-Lexicography: Exploring the Possibilities. Granger, S. and M. Paquot (Eds.). 2012. *Electronic Lexicography*: 379-397. Oxford: Oxford University Press.

Butaye, R. 1909. *Dictionnaire kikongo–français, français–kikongo*. Roulers: Jules De Meester.

Butaye, R. 1910. *Grammaire congolaise*. Roulers: Jules De Meester.

Daeleman, J. 1966. *Morfologie van naamwoord en werkwoord in het Kongo (Ntandu) met ontleding van het foneemsysteem*. Unpublished Ph.D. dissertation. Leuven: Katolieke Universiteit te Leuven.

Daeleman, J. 1983. Les réflexes du proto-bantu en ntándu (dialecte kóongo). Faïk-Nzuji Madiya, C. and E. Sulzmann (Eds.). 1983. *Mélanges de culture et de linguistique africaines publiés à la mémoire de Leo Stappers*: 331-397. Berlin: Dietrich Reimer Verlag.

De Kind, J., G.-M. de Schryver and K. Bostoen. 2012. Pushing Back the Origin of Bantu Lexicography: The *Vocabularium Congense* of 1652, 1928, 2012. *Lexikos* 22: 159-194.

Delplace, E. 1895. *Éléments de la langue congolaise, suivis d'un choix de phrases graduées et de deux vocabulaires*. Bruges: Desclée, De Brouwer et Cie.

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

De Schryver, G.-M. 2008. A New Way to Lemmatize Adjectives in a User-friendly Zulu–English Dictionary. *Lexikos* 18: 63-91.

De Schryver, G.-M. 2009. The Lexicographic Treatment of Ideophones in Zulu. *Lexikos* 19: 34-54.

De Schryver, G.-M. 2010. Revolutionizing Bantu Lexicography — A Zulu Case Study. *Lexikos* 20: 161-201.

De Schryver, G.-M. 2012. Trends in Twenty-five Years of Academic Lexicography. *International Journal of Lexicography* 25(4): 464-506 + 42 pages of supplementary material online.

De Schryver, G.-M. 2013. Tools to Support the Design of a Macrostructure. Gouws, R.H., U. Heid, W. Schweickard and H.E. Wiegand (Eds.). 2013. *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography* (Handbooks of Linguistics and Communication Science, HSK 5.4): 1384-1395. Berlin: Walter de Gruyter.

De Schryver, G.-M., R. Grollemund, S. Branford and K. Bostoen. 2015. Introducing a State-of-the-art Phylogenetic Classification of the Kikongo Language Cluster. *Africana Linguistica* 21: 87-162.

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): 291-309.

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.

De Schryver, G.-M. and D.J. Prinsloo. 2001. Towards a Sound Lemmatisation Strategy for the Bantu Verb through the Use of Frequency-based Tail Slots — With Special Reference to Cilubà, Sepedi and Kiswahili. Mdee, J.S. and H.J.M. Mwansoko (Eds.). 2001. *Makala ya kongamano la kimataifa Kiswahili 2000. Proceedings*: 216-242, 372. Dar es Salaam: TUKI, Chuo Kikuu cha Dar es Salaam.

De Schryver, G.-M. and A. Wilkes. 2008. User-friendly Dictionaries for Zulu: An Exercise in Complexicography. Bernal, E. and J. DeCesaris (Eds.). 2008. *Proceedings of the XIII EURALEX International Congress (Barcelona, 15–19 July 2008)* (Sèrie Activitats 20): 827-836. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.

Doke, C.M. 1935. *Bantu Linguistic Terminology*. London: Longmans, Green & Co.

Feinberg, J. 2008–14. *Wordle — Beautiful Word Clouds.* Available online at: http://www.wordle. net/.

Gillet, J. and É. Pâque. 1910. *Plantes principales de la région de Kisantu : Leur nom indigène, leur nom scientifique, leurs usages* (Annales du musée du Congo belge / Botanique — Série V — Bas- et Moyen-Congo / Notes botaniques sur la région du Bas- et Moyen-Congo — Fascicule 1). Brussels: Ministère des Colonies.

Gouws, R.H. 2012. Theoretical Lexicography and the *International Journal of Lexicography*. *International Journal of Lexicography* 25(4): 450-463.

Gouws, R.H. 2014. Makrostruktuuraanpassings vanaf gedrukte na e-woordeboeke. *Tydskrif vir Geesteswetenskappe* 54(3): 481-504.

Gouws, R.H. and D.J. Prinsloo. 1998. Cross-referencing as a Lexicographic Device. *Lexikos* 8: 17-36.

Gouws, R.H. and D.J. Prinsloo. 2010. Thinking out of the Box — Perspectives on the Use of Lexicographic Text Boxes. Dykstra, A. and T. Schoonheim (Eds.). 2010. *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6–10 July 2010)*: 501-511. Leeuwarden: Fryske Akademy.

Hanks, P. 2002. Mapping Meaning onto Use. Corréard, M.-H. (Ed.). 2002. *Lexicography and Natural Language Processing. A Festschrift in Honour of B.T.S. Atkins*: 156-198. s.l.: Euralex.

Hausmann, F.J. and H.E. Wiegand. 1989. Component Parts and Structures of General Monolingual Dictionaries: A Survey. Hausmann, F.J., O. Reichmann, H.E. Wiegand and L. Zgusta (Eds.). 1989. *Wörterbücher / Dictionaries / Dictionnaires. Ein internationales Handbuch zur Lexikographie / An International Encyclopedia of Lexicography / Encyclopédie internationale de lexicographie. Erster*

Teilband / First Volume / Tome Premier: 328-360. Berlin: Walter de Gruyter.

Kalonji, M.T.Z. 1993. *Le lexicographie bilingue en Afrique francophone : l'example français-cilubà*. Paris: Éditions L'Harmattan.

KongoKing Fieldwork. 2012. Linguistic fieldwork undertaken by Proff. K. Bostoen and G.-M. de Schryver, as well as their students J. De Kind and S. Dom, from August 4 to September 14, 2012, on ten Kikongo varieties spoken in the Lower Congo Province of the Democratic Republic of the Congo. Available online at: http://kongoking.org/fieldwork-linguistics-2012.html.

Laman, K.E. 1936. *Dictionnaire kikongo–français, avec une étude phonétique décrivant les dialectes les plus importants de la langue dite kikongo* (Institut royal colonial belge, Section des sciences morales et politiques, Mémoires — Collection in-8°, Tome II). Brussels: Georges van Campenhout.

Makaya Lutumba, C. 1999. *Éléments différentiels du ntandu et du manianga, dialectes kongo*. Unpublished M.A. dissertation. Kinshasa: Institut Pédagogique National.

Makolo Miaka, J. 2000. *Éléments de dialectologie kongo : esquisse comparative du kintandu, kiyombe et kimanyanga*. Unpublished M.A. dissertation. Kinshasa: Institut Pédagogique National.

Mbatha, M.O. 2006. *Isichazamazwi sesiZulu*. Pietermaritzburg: New Dawn Publishers.

McKean, E., G. Barrett and O. Montoya. 2009–15. Wordnik — The World's Biggest Online English Dictionary. Available online at: https://www.wordnik.com/.

Murray, J.A.H. 1888. *A New English Dictionary on Historical Principles: Founded Mainly on the Materials Collected by the Philological Society ... with the Assistance of Many Scholars and Men of Science. Volume I. A and B*. Oxford: Clarendon Press.

Nielsen, S. and L. Mourier. 2007. Design of a Function-based Internet Accounting Dictionary. Gottlieb, H. and J.E. Mogensen (Eds.). 2007. *Dictionary Visions, Research and Practice: Selected Papers from the 12th International Symposium on Lexicography, Copenhagen 2004* (Terminology and Lexicography Research and Practice 10): 119-135. Amsterdam: John Benjamins.

Nurse, D. and G. Philippson. 2006. Common Tense-Aspect Markers in Bantu. *Journal of African Languages and Linguistics* 27(2): 155-196.

Ogilvie, S. 2008. The Mysterious Case of the Vanishing Tramlines: James Murray's Legacy and the 1933 OED Supplement. *Dictionaries: Journal of the Dictionary Society of North America* 29: 1-22.

Polis, C. 1938. *Lexique kikongo–français* (manuscript in nine fascicules, stencil duplicated). Leuven.

Polis, C. 1942. Bij de Bakongo van Mbata. *Aequatoria* 5(3-4): 85-87.

Prinsloo, D.J. 2014. A Critical Evaluation of the Paradigm Approach in Sepedi Lemmatisation — The *Groot Noord-Sotho Woordeboek* as a Case in Point. *Lexikos* 24: 251-271.

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.

Rundell, M. 2012. It Works in Practice but Will it Work in Theory? The Uneasy Relationship between Lexicography and Matters Theoretical. Fjeld, R.V. and J.M. Torjusen (Eds.). 2012. *Proceedings of the 15th EURALEX International Congress, 7–11 August, 2012, Oslo*: 47-92. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo.

Scott, M. 1996–2015. *WordSmith Tools*. Available online at: http://www.lexically.net/wordsmith/.

Swartenbroeckx, P. 1973. *Dictionnaire kikongo et kituba–français : vocabulaire comparé des langages Kongo traditionnels et véhiculaires* (Ceeba Publications Série III, Vol. 2). Bandundu: Centre d'études ethnolinguistiques Bandundu.

Taljard, E. and G.-M. de Schryver. 2002. Semi-automatic Term Extraction for the African Lan-

guages, with Special Reference to Northern Sotho. *Lexikos* 12: 44-74.

Van de Casteele, J. 1968. POLIS (Charles), Jésuite (Anvers, 1.11.1885 — Léopoldville, 27.6.1943). Jadot, J.-M. (Ed.). 1968. *Biographie belge d'outre-mer, Vol. VI*: col. 826-828. Brussels: Académie royale des sciences d'outre-mer.

Van Gheel, J. 1652. *Vocabularium Latinum, Hispanicum, e Congense. Ad Usum Missionariorû transmittendorû ad Regni Congo Missiones* (manuscript kept in the Biblioteca nazionale centrale di Roma: Fondi minori 1896, Mss. Varia 274).

Van Wing, J. and C. Penders. 1928. *Le plus ancien dictionnaire bantu / Het oudste Bantu-woordenboek. Vocabularium P. Georgii Gelensis* (Bibliothèque Congo / Congo Bibliotheek 27). Louvain: J. Kuyl-Otto.

Varantola, K. 2002. Use and Usability of Dictionaries: Common Sense and Context Sensibility? Corréard, M.-H. (Ed.). 2002. *Lexicography and Natural Language Processing. A Festschrift in Honour of B.T.S. Atkins*: 30-44. s.l.: Euralex.

Vermeersch, G. 1922. *Dictionnaire tshiluba–français*. Brussels: Librairie Albert Dewit.

Wiegand, H.E. 1989. Aspekte der Makrostruktur im allgemeinen einsprachigen Wörterbuch: Alphabetische Anordnungsformen und ihre Probleme. Hausmann, F.J., O. Reichmann, H.E. Wiegand and L. Zgusta (Eds.). 1989. *Wörterbücher / Dictionaries / Dictionnaires. Ein internationales Handbuch zur Lexikographie / An International Encyclopedia of Lexicography / Encyclopédie internationale de lexicographie. Erster Teilband / First Volume / Tome Premier*: 371-409. Berlin: Walter de Gruyter.

Wiegand, H.E. 1996. Über die Mediostrukturen bei gedruckten Wörterbüchern. Zettersten, A. and V.H. Pedersen (Eds.). 1996. *Symposium on Lexicography VII. Proceedings of the Seventh International Symposium on Lexicography, May 5–6, 1994 at the University of Copenhagen* (Lexicographica Series Maior 76): 11-43. Tübingen: Max Niemeyer.

Wiegand, H.E. and S. Beer. 2013. Access Structures in Printed Dictionaries. Gouws, R.H., U. Heid, W. Schweickard and H.E. Wiegand (Eds.). 2013. *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography* (Handbooks of Linguistics and Communication Science, HSK 5.4): 110-149. Berlin: Walter de Gruyter.

Wiegand, H.E., S. Beer and R.H. Gouws. 2013. Textual Structures in Printed Dictionaries: An Overview. Gouws, R.H., U. Heid, W. Schweickard and H.E. Wiegand (Eds.). 2013. *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography* (Handbooks of Linguistics and Communication Science, HSK 5.4): 31-73. Berlin: Walter de Gruyter.

Wiegand, H.E., M. Beißwenger, R.H. Gouws, M. Kammerer, A. Storrer and W. Wolski (Eds.). 2010. *Wörterbuch zur Lexikographie und Wörterbuchforschung / Dictionary of Lexicography and Dictionary Research. Band 1: Systematische Einführung, A–C mit englishen Übersetzungen der Umtexte und Definitionen sowie Äquivalenten in neun Sprachen / Volume 1: Systematic Introduction, A–C with English Translations of the Outer Texts and Definitions as well as Equivalents in Nine Languages*. Berlin: Walter de Gruyter.

Wiegand, H.E. and R.H. Gouws. 2013. Macrostructures in Printed Dictionaries. Gouws, R.H., U. Heid, W. Schweickard and H.E. Wiegand (Eds.). 2013. *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography* (Handbooks of Linguistics and Communication Science, HSK 5.4): 73-110. Berlin: Walter de Gruyter.

Addendum 1: Vertical Base, or CV(C) sequence, of the macrostructure in the *Lexique kikongo–français* by Charles Polis (1938) [corrections and erroneous positions are marked in red]

Page	Vertical Base
0.	**Introduction**
i	on word order
ii	example matrix
iii	on layout
PART I	
Stretch 0 – CV(c), CVc' V, varia WV(c,w,y), CVw YV(c), CVy, YVy VV	
1.	**Pre-grammatical forms**
1	Ba, Be, Bi, Bo
2	Bu
3	Bab, Beb, Bib, Bob
4	Bub, Pa, Pe
4bis	Pa ctd.
5	Pi, Po
6	Pu, Ma, Me
7	Mi, Mo, Mu, M-m, Va, Ve
8	Vi, Vo, Vu
9	Fa, Fe, Fi, Fo
10	Fu, Vav, Vev, Viv, Vov, Vuv
11	Faf, Fef, Fif, Fof, Fuf
12	Ga, Ge, Gi, Go
13	Gu, Gag, Geg, Gig, Gog, Gug
14	Ka, Ke
15	Ki, Ko
16	Ku, Kak, Kek
17	Kik, Kok, Kuk, Da
18	De, Di, Do, Du, La, Le, Li
19	Lo, Lu, Dad, Ded, Did, Dod
20	Dud, Dal, Del, Dil, Dol, Dul, Lal, Lel
21	Lil, Lol, Lul, Lad, Led, Lid, Lod, Lud, Ta, Te
22	Ti, To, Tu, Tat, Tet
23	Tit, Tot, Tut, Na
24	Ne, Ni, No, Nu, N diphthong, Nan
25	Nen, Nin, Non, Nun, Za, Ze
26	Zi, Zo, Zu, Zaz
27	Zez, Ziz, Zoz, Zuz, Sa
28	Se, Si, So, Su, Sas
29	Ses, Sis, Sos, Sus, B-p, Pab, Peb, Pib, Pob, Pub, B-m, Pam, B-v, Bof
30	P-v, Paf, Pof, Puf, Bag, Beg, Big, Bog, Bug, Bak, Bek
31	Bik, Bok, Buk, Pag, Peg, Pig, Pog, Pug, Pak
32	Pek, Pik, Pok, Puk, Bad, Bed, Bid, Bod, Bud, Bal
33	Bel, Bil, Bol, Bul, Bat, Bet
34	Bit, Bot, But, P-d, Pal, Pel, Pil, Pol, Pul, Pat, Pet
35	Pit, Pot, Put, B-n, P-n, Bez, Boz, Buz, Bis, Paz, Pas, Pos, Pus, M-b, M-p, M-v, M-f, M-g
36	Mak, Mek, Mok, Muk, Mad, Mat, Mit, Mot, M-n, M-z, M-s, Vab, Veb
37	Vib, Vob, Vub, V-p, V-m, V-f, Vag, Veg, Vug, Vak, Vik
38	Vuk, Vid, Val, Vid, Vil, Vul, Vat, Vet, Vit, Vut, Vez, Viz
39	Vis, Fab, Fob, Fop, Fup, F-m, F-v, Fog, Fak, Fek, Fok, Fuk, Fid, Fud, Fal, Fol, Ful, Fat
40	Fet, Fit, Fot, Fut, Fin, F-z, Fus, Gab, Geb, Gob, Gub
41	Gep, Gam, Gom, G-v, Gof, Guf, Gok, Guk, Gad, Ged, God, Gud, Gal, Gel, Gil, Gol
42	Gul, Gat, Got, G-n, G-z, G-s, Kab
43	Keb, Kib, Kob, Kub, K-p, K-m
44	K-f, K-g, Keg, Kog, Kug, Kad, Ked, Kid
45	Kod, Kud, K-l, Kel
46	Kil, Kol, Kul, Kat, Ket, Kit
47	Kot, Kut, Kan, Ken, Kon, Kun, Kaz, Kez, Kiz, Kuz
48	Kas, Kes, Kos, Dab, Deb, Dib, Dub, Leb, Lob, D-p, L-p
49	Tab, Teb, Tib, Tob, Tub, Tap, Tep, Top, Tup, D-m
50	L-m, T-m, D-v, L-v, D-f, L-f, T-v, T-f, Dag, Deg, Dog
51	Dug, Lag, Leg, Log, Dak, Dek, Dik, Dok, Duk, Lak
52	Lek, Lok, Luk, Tag, Teg, Tig, Tog, Tug, Tak
53	Tek, Tik, Tok, Tuk, D-t, L-t, T-d, T-l
54	D-n, T-n, D-z, D-s, L-z, L-s, T-z, T-s, N-b, N-p, N-m, N-v, N-f, N-g
55	N-k, N-d, N-l, N-t, N-z, N-s, Zab, Zeb, Zib, Zob

56	Zub, Sab, Seb, Sib, Sob, S-p, Z-m, S-m, S-v, S-f, Zag
57	Zeg, Zug, Zak, Seg, Sig, Sak, Sek, Sik, Sok, Suk
58	Zad, Zed, Zid, Zud, Zal, Zel, Zi, Zol, Zul, Zat, Zit, Zot, Zut, Sad
59	Sed, Sod, Sud, Sal, Sel, Sol, Sul, Sat, Set, Sot, Z-n
60	S-n, Z-s, S-s, a, e, i, o
61	u, varia, Wa, Wi, Wo, Wu
62	W-b, Wag, Weg, Wog, W-k, W-d, Wal, Wil, Wol, Wul, W-t, W-n, W-z
63	W-s, W-w, W-y, B-w, Ya, Ye, Yi, Yo, Yu
64	Yab, Yeb, Yub, Y-p, Y-m, Y-v, Y-f, Yag, Yeg, Yig, Yug
65	Yak, Yek, Yik, Yok, Yuk, Yad, Yed, Yid, Yod, Y-l, Yil
66	Yol, Yul, Y-t, Y-n, Y-z, Y-s, B-y, V-y, K-y, L-y, Y-y
67	diphthongs
PART II	
Stretch 1 – CV(C)	
2.	**Simple & double labials**
1	Ba
2	…
3	BaB
4	…
5	…
6	…
7	…
8	…
9	…
10	…
11	Be
12	BeB
13	…
14	Bi, BiB
15	…
16	Bo
17	BoB
18	…
19	…
20	Bu, BuB
21	…
22	…
23	…
24	…
25	…
26	Pa
27	Pe, Pi, Po, Pu
28	PaP
28bis	…
29	PeP, PiP
30	PoP
31	PuP
32	…
33	Ma
34	Me, Mi, Mo, Mu
35	MaM, MeM
36	MiM, MoM, MuM
37	Va
38	Ve
39	Vi, Vo, Vu
40	VaV, VeV
41	ViV, VoV, VuV
42	…
43	Fa
44	…
45	…
46	Fe
47	Fi, Fo, Fu, Foy
48	FaF, FeF
49	FiF, FoF
50	FuF
51	Ga, Ge
52	Gi, Go, Gu, Gay
53	GaG
54	…
55	GeG
56	…
57	GiG, GoG
58	…
59	GuG
60	…
61	Ka
62	…
63	Ke, Ki
64	Ko
65	Ku, Kau, Kay
66	KaK
67	…
68	…
69	…
70	…
71	KeK
72	…
73	KiK
74	…
75	KoK
76	…
77	…
78	KuK
79	…
80	…
3.	**Simple & double linguals**
81	Da
82	…
83	…
84	De, Di
85	Do, Du, La
86	Le, Li, Lo, Lu, Lau, DaD
87	…
88	DaL
89	LaL
90	LaD
91	…
92	…
93	LaD, DeD
94	DeL
95	LeL
96	LeD
97	…
98	DiD
99	DiL
100	LiL, LiD, DoD
101	DoL, LoL, LoD
102	DuD
103	DuL, LuL
104	LuD
105	…
106	Ta
107	…
108	Te
109	Ti, To
110	Tu
111	TaT
112	…
113	TeT
114	TiT
115	ToT
116	TuT
117	…
118	…
119	Na
120	Ne
121	Ni, No, Nu, N-N
122	…
123	NeN
124	NiN, NoN, NuN
125	Za, Ze
126	Zi, Zo, Zu, Zau
127	ZaZ
128	ZeZ

129 …
130 ZiZ, ZoZ
131 …
132 ZuZ
133 Sa
134 …
135 Se
136 Si
137 So, Su, Sau, Say
138 SaS
139 …
140 SeS
141 SiS
142 …
143 SoS
144 SuS

Stretch 2 – CVC'

4. B/P with other consonants

145 B-P, PaB
146 PeB, PiB
147 PoB, PuB, BaM
148 …
149 BeM, BiM, BoM
150 BuM, PaM, PeM
151 PoM, PuM, BaV, BeV, BuV, BaF
152 BeF, BiF, BoF
153 BuF, PaV, PeV
154 PuV, PaF, PoF, PuF, BaG
155 …
156 …
157 BeG, BiG
158 BoG
159 BuG
160 …
161 BaK
162 …
163 …
164 BeK
165 BiK
166 BoK
167 …
168 BuK
169 …
170 …
171 PaG
172 PeG
173 PiG, PoG, PuG
174 PaK
175 PeK, PiK, PoK
176 PuK
177 BaD
178 …
179 …
180 BaL
181 …
182 BeD
183 BeL
184 …
185 BiD
186 …
187 BiL
188 …
189 BoD, BoL
190 …
191 BuD
192 …
193 BuL
194 …
195 BaT
196 …
197 BeT
198 …
199 BiT
200 BoT
201 …
202 BuT
203 PaD
204 PaL
205 PeD, PeL
206 PiD
207 PiL, PoD, PoL
208 PuD, PuL
209 PaT
210 …
211 PeT
212 PiT
213 PoT, PuT
214 BaN, BeN
215 BoN, BuN, PaN
216 PeN, PiN, PoN, PuN, BaZ & BaS
217 …
218 …
219 BeZ & BeS
220 BiZ & BiS
221 BoZ & BoS
222 BuZ & BuS
223 PaZ & PaS
224 PeZ & PeS
225 PiZ & PiS, PoZ & PoS, PuZ & PuS

5. M and V/F with other consonants

226 MaB, MeB, MiB, MoB, MaV & MaF, MeV & MeF, MiV & MiF, MoV & MoF, MuV & MuF
227 MaG
228 MeG
229 MiG, MoG
230 MuG, MaK
231 MeK
232 MiK, MoK
233 MuK, MaD & MaL
234 MeD & MeL, MiD & MiL, MoD & MoL, MuD & MuL
235 MaT, MeT
236 MiT, MoT, MuT
237 MaN
238 MeN, MiN
239 MoN
240 MuN, MaZ & MaS
241 MeZ & MeS, MiZ & MiS, MoZ & MoS
242 VaB, VeB
243 ViB, VoB, VuB
244 …
245 VaP, VeP, ViP, VoP, VuP, FaB
246 FeB, FiB, FoB
247 FuB
248 FaP, FeP, FiP, FoP, FuP
249 VaM, VeM, ViM, VoM, VuM
250 FaM, FeM, FiM, FoM
251 FuM
252 VaF, ViF, VuF, FuV, VaG
253 …
254 VeG, ViG, VoG
255 VuG, VaK
256 VeK, ViK, VoK
257 VuK
258 FaG, FeG
259 FiG, FoG
260 FuG, FaK
261 FeK, FiK
262 FoK, FuK
263 …
264 VaD
265 VaL
266 VeL, ViD
267 ViL
268 VoD & VoL, VuD, VuL
269 …
270 VaT, VeT
271 ViT
272 VoT, VuT
273 FaD & FaL
274 FeD & FeL, FiD & FiL

275 FoD & FoL, FuD & FuL
276 …
277 …
278 FaT
279 …
280 FeT, FiT
281 …
282 FoT, FuT
283 …
284 …
285 VaN, VeN, ViN, VoN, VuN
286 FaN, FeN, FiN
287 FoN, FuN
288 VaZ & VaS, VeZ & VeS
289 ViZ & ViS, VoZ & VoS, VuZ & VuS
290 FaZ & FaS, FeZ & FeS
291 FiZ & FiS, FoZ & FoS, FuZ & FuS
292 …

6. G/K with other consonants

293 GaB
294 GeB, GiB, GoB
295 GuB
296 GaP, GeP, GoP
297 GuP, KaB
298 …
299 …
300 …
301 …
302 KeB
303 …
304 KiB
305 KoB
306 …
307 KuB
308 …
309 …
310 …
311 …
312 KaP, KeP, KoP
313 KuP, GaM
314 GeM, GoM, GuM
315 KaM
316 KeM, KiM, KoM
317 KuM
318 …
319 GaV & GaF, GeV & GeF
320 GiV & GiF, GoV & GoF, GuV & GuF, KaV, KeV, KiV, KoV
321 KuV, KaF
322 KeF, KoF, KuF
323 GaK
324 GeK, GiK, GoK
325 GuK
326 KaG
327 …
328 …
329 …
330 KeG
331 KiG
332 KoG
333 …
334 KuG
335 GaD
336 GaL & GaD
337 GeD & GeL, GiD
338 GoD & GoL
339 …
340 GuD & GuL
341 GaT
342 GeT
343 GiT, GoT
344 GuT
345 KaD & KaL
346 …
347 …
348 …
349 …
350 …
351 …
352 …
353 KeD & KeL
354 …
355 …
356 …
357 …
358 …
359 KiD & KiL
360 …
361 KoD & KoL
362 …
363 …
364 …
365 …
366 …
367 KuD & KuL
368 …
369 …
370 …
371 …
372 KaT
373 …
374 KeT
375 …
376 KiT
377 …
378 KoT
379 …
380 KuT
381 …
382 GaN
383 GeN, GiN
384 GoN, GuN, KaN
385 …
386 …
387 KeN
388 KiN
389 KoN
390 KuN, GaZ & GaS
391 GeZ & GeS
392 GiZ & GiS, GoZ & GoS
393 GuZ & GuS
394 KaZ & KaS
395 …
396 …
397 …
398 …
399 KeZ & KeS
400 …
401 KiZ & KiS
402 KoZ & KoS
403 …
404 KuZ & KuS
405 …
406 …

7. D~L/T with other consonants

407 DaB & LaB
408 …
409 …
410 DeB & LeB
411 …
412 …
413 DiB & LiB
414 …
415 DoB & LoB
416 DuB & LuB
417 DaP & LaP, DeP & LeP
418 DiP & LiP, DoP & LoP, DuP & LuP
419 TaB
420 …
421 …
422 …
423 …
424 TeB
425 TiB
426 ToB
427 TuB
428 …
429 …
430 TaP, TeP
431 TiP, ToP, TuP

432	DaM & LaM	467	TaK	496	NaB, NeB, NiB	528	…
433	LeM, DiM, DoM & LoM	468	…	497	NoB, NuB, NaP, NeP, NiP, NoP, NuP	529	SeB
434	DuM & LuM, TaM	469	TeK			530	SiB
435	TeM	470	…			531	SoB
436	TiM, ToM	471	TiK, ToK	498	NaM, NeM, NiM	532	SuB
437	TuM	472	…	499	NoM, NuM, NaV & NaF, NeV & NeF, NoV & NoF	533	SaP
438	D-V & L-V, DaF & LaF, DeF & LeF	473	TuK			534	SeP, SiP, SoP, SuP
		474	DaT & LaT				
		475	DeT & LeT, DiT, DoT & LoT	500	NuV & NuF, NaG	535	ZaM
						536	ZeM, ZiM
439	DiF, DoF & LoF, DuF & LuF, TaV & TaF	476	DuT & LuT	501	…	537	ZoM, ZuM, SaM
		477	TaD & TaL	502	…	538	SeM, SiM
		478	…	503	NeG	539	SoM, SuM
		479	…	504	…	540	…
440	TeV & TeF, TiV & TiF, ToV & ToF	480	…	505	NiG	541	ZaV & ZaF & SaV & SaF, ZoV & ZoF & SoV & SoF
		481	…	506	NoG		
		482	TeD & TeL	507	NuG		
441	TuV & TuF, DaG & LaG	483	…	508	…		
		484	TiD & TiL, ToD & ToL	509	NaK	542	ZuV & ZuF & SuV & SuF, ZaG
442	…			510	NeK		
443	DeG & LeG	485	…	511	NiK	543	…
444	…	486	TuD & TuL	512	NoK	544	ZeG
445	…	487	…	513	NuK	545	ZiG
446	…	488	DaN & LaN, DeN & LeN, DoN & LoN, DuN & LuN, TaN, TeN, TiN	514	NaD & NaL, NeD & NeL, NoD & NoL, NuD & NuL, NaT	546	ZoG
447	DiG					547	ZuG
448	DoG & LoG					548	ZaK
449	…					549	…
450	DuG & LuG					550	ZeK
451	DaK & LaK	489	ToN, TuN	515	…	551	ZiK
452	…	490	DaZ & DaS & LaZ & LaS	516	NeT, NiT	552	ZoK
453	DeK & LeK			517	NoT, NuT, NaZ & NaS	553	ZuK
454	…	491	DeZ & DeS & LeZ & LeS			554	SaG
455	…			518	NeZ & NeS, NiZ & NiS	555	…
456	DiK	492	DiZ & DiS, DoZ & DoS & LoZ & LoS			556	SeG
457	DoK & LoK			519	NoZ & NoS	557	…
458	…			520	ZaB	558	SiG
459	DuK & LuK	493	DuZ & DuS & LuZ & LuS	521	…	559	SoG
460	…			522	ZeB	560	SuG
461	TaG	494	TaZ & TaS, TeZ & TeS	522bis	ZiB	561	SaK
462	…			523	ZoB	562	…
463	TeG, TiG	495	TiZ & TiS, ToZ & ToS, TuZ & TuS	524	ZuB	563	…
464	ToG			525	SaB	564	…
465	TuG	**8. N and Z/S with other consonants**		526	…	565	SeK
466	…			527	…	566	…

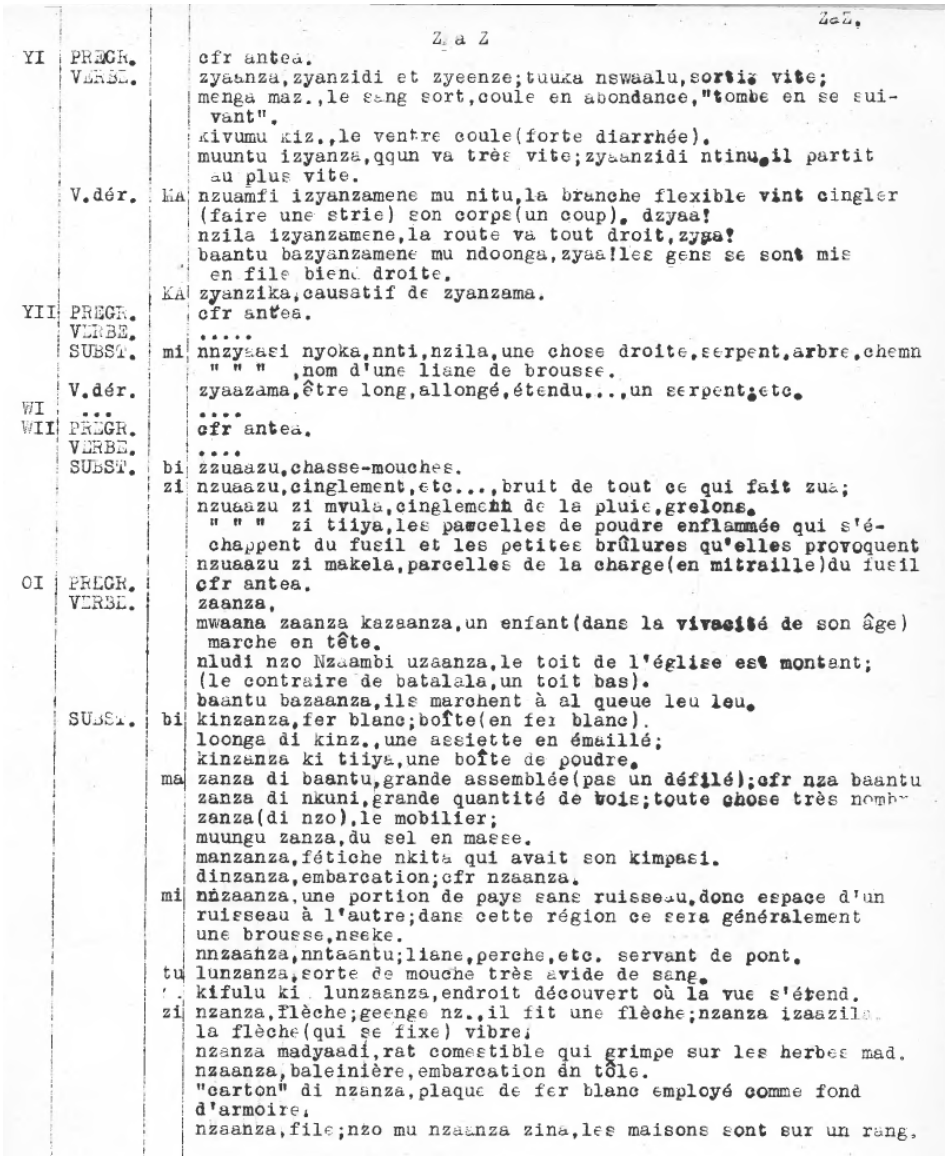
567 SiK
568 SoK
569 …
570 SuK
571 …
572 ZaD & ZaL
573 …
574 ZeD & ZeL
575 ZiD & ZiL
576 ZoD & ZoL
577 ZuD & ZuL, ZaT
578 ZeT, ZiT
579 ZoT
580 ZuT, SaD & SaL
581 …
582 …
583 …
584 SeD & SeL
585 SiD & SiL
586 …
587 SoD & SoL
588 SuD & SuL
589 SaT
590 SeT
591 SiT
592 SoT, SuT, ZaN
593 ZeN, ZiN
594 ZoN, ZuN, SaN
595 SeN, SiN
596 SoN, SuN
597 ZaS & SaZ
598 ZeS & SeZ, ZiS & SiZ
599 ZoS & SoZ, ZuS & SuZ

Stretch 3 – YV(Y), YVC,W, CVY WV(W), WVC, CVW VV

9. Words with Y, W and diphthongs

600 Ya, Ye, Yi
601 Yo, Yu, YaY
602 YeY, YiY, YoY, YuY
603 YaB
604 …
605 YeB, YiB
606 YoB, YuB
607 YaP, YeP, YiP, YoP, YuP, YaM
608 YeM, YiM, YoM, YuM
609 YaV & YaF, YuV & YuF, YaG
610 …
611 YeG
612 YiG, YoG, YuG
613 YaK
614 YeK
615 YiK
616 YoK, YuK
617 YaD & YaL
618 …
619 YeD & YeL, YiD & YiL
620 YoD & YoL
621 YuD & YuL
622 YaT
623 YeT, YiT, YoT, YuT
624 YaN, YeN, YiN, YoN
625 YuN, YaZ & YaS, YeZ & YeS, YiZ & YiS
626 YoZ & YoS, YuZ & YuS, YaW, BaY
627 BeY, BiY, BoY, PaY, PeY, PiY, PoY, PuY, MaY, MeY, MiY
628 MoY, VaY, VeY, ViY, VoY, VuY, FaY, FeY
629 FiY, GaY, GeY, GiY, GoY, GuY, KaY
630 KeY
631 KiY, KoY, KuY, DaY & LaY, DeY & LeY, DiY & LiY
632 DoY & LoY, DuY & LuY, TaY, TeY, TiY, ToY, TuY, N-Y, ZaY
633 ZeY, ZiY, ZoY, ZuY, SaY, SeY, SoY, SuY, Y-W
634 Wa, We, Wo
635 Wu, WaW, WeW, WiW, WoW, WuW, WaB
636 WeB, WiB, WoB, WuB, WuP, WuM, WuF, WanG
637 WeG, WiG, WoG, WuG, WaK, WeK, WiK, WoK
638 WuK, WaD & WaL, WeD & WeL
639 WiD & WiL, WoD & WoL, WuD & WuL, WaT
640 WeT, WiT, WoT, WuT, WaN, WeN, WiN, WuN
641 WaZ & WaS, WeZ & WeS, WiZ & WiS, WoZ & WoS, WuZ & WuS, BaW & PaW, BeW & PeW
642 BiW & PiW, BoW & PoW, PuW & PuW, M-W, KaW, KeW, KiW, KoW
643 KuW, D/L-W, DoW & LoW, DuW & LuW, TaW, ToW, Z-W, S-W
644 ay, au (monosyllabic)
645 au (polysyllabic)
646 ey, eu, oy

Addendum 2: Sample pages from Part II of the *Lexique kikongo–français* by Charles Polis (1938) [cluster **ZaZ**; Part II, pp. 127-128]



ZaZ.

Z a Z

YI PREGR. cfr antea.
VERBE. zyaanza,zyanzidi et zyeenze;tuuka nswaalu,sortir vite;
menga maz.,le sang sort,coule en abondance,"tombe en se suivant".
kivumu kiz.,le ventre coule(forte diarrhée).
muuntu izyanza,qqun va très vite;zyaanzidi ntinu,il partit au plus vite.
V.dér. ka nzuamfi izyanzamene mu nitu,la branche flexible vint cingler (faire une strie) son corps(un coup), dzyaa!
nzila izyanzamene,la route va tout droit,zyga!
baantu bazyanzamene mu ndoonga,zyaa!les gens se sont mis en file bien droite.
KA zyanzika,causatif de zyanzama.

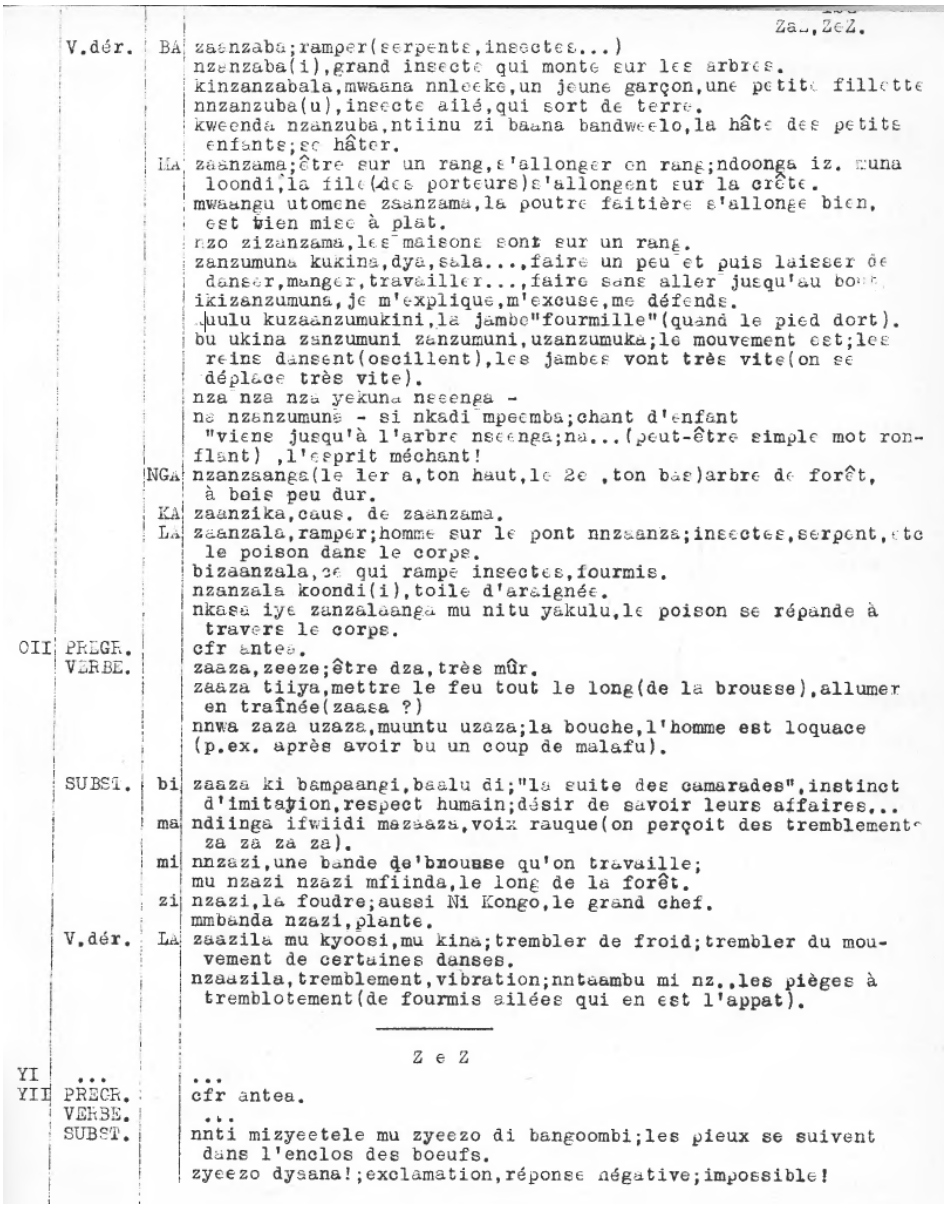
YII PREGR. cfr antea.
VERBE. .....
SUBST. mi nnzyaasi nyoka,nnti,nzila,une chose droite,serpent,arbre,chemn
" " " ,nom d'une liane de brousse.
V.dér. zyaazama,être long,allongé,étendu...,un serpent;etc.

WI ...

WII PREGR. cfr antea.
VERBE. ....
SUBST. bi zzuaazu,chasse-mouches.
zi nzuaazu,cinglement,etc...,bruit de tout ce qui fait zua;
nzuaazu zi mvula,cinglement de la pluie,grelons.
" " " zi tiiya,les parcelles de poudre enflammée qui s'échappent du fusil et les petites brûlures qu'elles provoquent
nzuaazu zi makela,parcelles de la charge(en mitraille)du fusil

OI PREGR. cfr antea.
VERBE. zaanza,
mwaana zaanza kazaanza,un enfant(dans la vivacité de son âge) marche en tête.
nludi nzo Nzaambi uzaanza,le toit de l'église est montant;
(le contraire de batalala,un toit bas).
baantu bazaanza,ils marchent à al queue leu leu.
SUBST. bi kinzanza,fer blanc;boîte(en fer blanc).
loonga di kinz.,une assiette en émaillé;
kinzanza ki tiiya,une boîte de poudre.
ma zanza di baantu,grande assemblée(pas un défilé);cfr nza baantu
zanza di nkuni,grande quantité de bois;toute chose très nomb
zanza(di nzo),le mobilier;
muungu zanza,du sel en masse.
manzanza,fétiche nkita qui avait son kimpasi.
dinzanza,embarcation;cfr nzaanza.
mi nnzaanza,une portion de pays sans ruisseau,donc espace d'un ruisseau à l'autre;dans cette région ce sera généralement une brousse,nseke.
nnzaanza,nntaantu;liane,perche,etc. servant de pont.
tu lunzanza,sorte de mouche très avide de sang.
. kifulu ki lunzaanza,endroit découvert où la vue s'étend.
zi nzanza,flèche;geenge nz.,il fit une flèche;nzanza izaazila
la flèche(qui se fixe) vibre;
nzanza madyaadi,rat comestible qui grimpe sur les herbes mad.
nzaanza,baleinière,embarcation en tôle.
"carton" di nzanza,plaque de fer blanc employé comme fond d'armoire.
nzaanza,file;nzo mu nzaanza zina,les maisons sont sur un rang.

(ZaZ continued, and beginning of ZeZ)



Za.., ZeZ.

V.dér. BA zaanzaba;ramper(serpents,insectes...)
nzanzaba(i),grand insecte qui monte sur les arbres.
kinzanzabala,mwaana nnleeke,un jeune garçon,une petite fillette
nnzanzuba(u),insecte ailé,qui sort de terre.
kweenda nzanzuba,ntiinu zi baana bandweelo,la hâte des petits enfants;se hâter.
MA zaanzama;être sur un rang,s'allonger en rang;ndoonga iz. muna loondi,la file(des porteurs)s'allongent sur la crête.
mwaangu utomene zaanzama,la poutre faîtière s'allonge bien, est bien mise à plat.
nzo zizanzama,les maisons sont sur un rang.
zanzumuna kukina,dya,sala...,faire un peu et puis laisser de danser,manger,travailler...,faire sans aller jusqu'au bout
ikizanzumuna,je m'explique,m'excuse,me défends.
kuulu kuzaanzumukini,la jambe"fourmille"(quand le pied dort).
bu ukina zanzumuni zanzumuni,uzanzumuka;le mouvement est;les reins dansent(oscillent),les jambes vont très vite(on se déplace très vite).
nza nza nza yekuna nseenga -
na nzanzumuna - si nkadi mpeemba;chant d'enfant "viens jusqu'à l'arbre nseenga;na...(peut-être simple mot ronflant) ,l'esprit méchant!
NGA nzanzaanga(le 1er a,ton haut,le 2e ,ton bas)arbre de forêt, à bois peu dur.
KA zaanzika,caus. de zaanzama.
LA zaanzala,ramper;homme sur le pont nnzaanza;insectes,serpent,etc le poison dans le corps.
bizaanzala,ce qui rampe insectes,fourmis.
nzanzala koondi(i),toile d'araignée.
nkasa iye zanzalaanga mu nitu yakulu,le poison se répande à travers le corps.

OII PREGR. VERBE. cfr antea.
zaaza,zeeze;être dza,très mûr.
zaaza tiiya,mettre le feu tout le long(de la brousse),allumer en traînée(zaasa ?)
nnwa zaza uzaza,muuntu uzaza;la bouche,l'homme est loquace (p.ex. après avoir bu un coup de malafu).

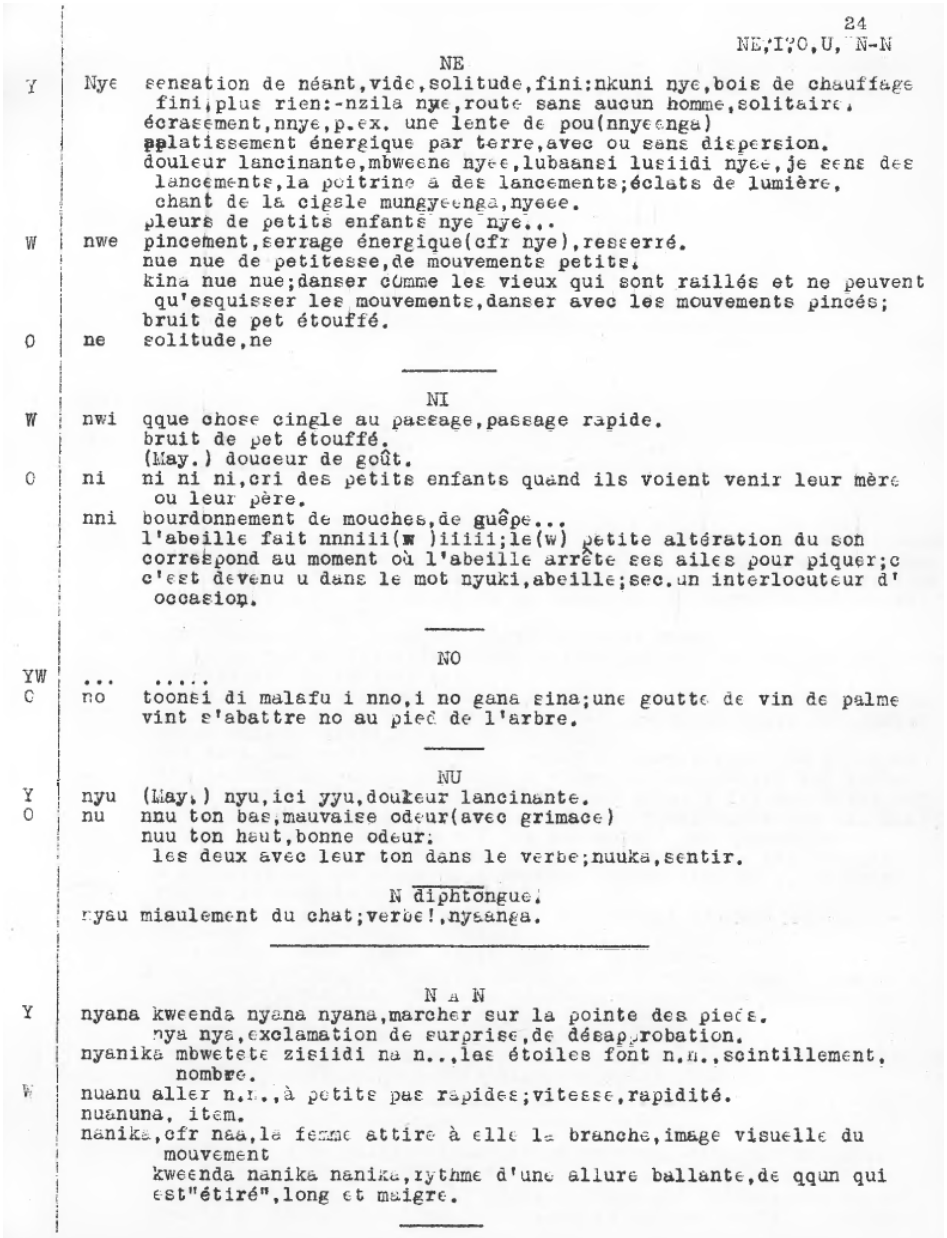
SUBST. bi zaaza ki bampaangi,baalu di;"la suite des camarades",instinct d'imitation,respect humain;désir de savoir leurs affaires...
ma ndiinga ifwiidi mazaaza,voix rauque(on perçoit des tremblements za za za za).
mi nnzazi,une bande de'brousse qu'on travaille;
mu nzazi nzazi mfiinda,le long de la forêt.
zi nzazi,la foudre;aussi Ni Kongo,le grand chef.
mmbanda nzazi,plante.

V.dér. LA zaazila mu kyoosi,mu kina;trembler de froid;trembler du mouvement de certaines danses.
nzaazila,tremblement,vibration;nntaambu mi nz.,les pièges à tremblotement(de fourmis ailées qui en est l'appat).

Z e Z

YI ...
YII PRECR. VERBE. SUBST. ...
cfr antea.
...
nnti mizyeetele mu zyeezo di bangoombi;les pieux se suivent dans l'enclos des boeufs.
zyeezo dysana!;exclamation,réponse négative;impossible!

Addendum 3: Sample page from Part I of the *Lexique kikongo–français* by Charles Polis (1938) [clusters **Ne**, **Ni**, **No**, **Nu**, **Nau**, and **Nan**; Part I, p. 24]



24
NE;I;O,U, N-N

NE

Y Nye sensation de néant,vide,solitude,fini:nkuni nye,bois de chauffage fini,plus rien:-nzila nye,route sans aucun homme,solitaire, écrasement,nnye,p.ex. une lente de pou(nnyeenga) aplatissement énergique par terre,avec ou sans dispersion. douleur lancinante,mbweene nyee,lubaansi lusiidi nyee,je sens des lancements,la poitrine a des lancements;éclats de lumière, chant de la cigale mungyeenga,nyeee. pleurs de petits enfants nye nye...

W nwe pincement,serrage énergique(cfr nye),resserré. nue nue de petitesse,de mouvements petits. kina nue nue;danser comme les vieux qui sont raillés et ne peuvent qu'esquisser les mouvements,danser avec les mouvements pincés; bruit de pet étouffé.

O ne solitude,ne

NI

W nwi qque chose cingle au passage,passage rapide. bruit de pet étouffé. (May.) douceur de goût.

O ni ni ni ni,cri des petits enfants quand ils voient venir leur mère ou leur père.

nni bourdonnement de mouches,de guêpe... l'abeille fait nnniii(w)iiiii;le(w) petite altération du son correspond au moment où l'abeille arrête ses ailes pour piquer;c c'est devenu u dans le mot nyuki,abeille;sec.un interlocuteur d' occasion.

NO

YW ... .

O no toonsi di malafu i nno,i no gana sina;une goutte de vin de palme vint s'abattre no au pied de l'arbre.

NU

Y nyu (May.) nyu,ici yyu,douleur lancinante.

O nu nnu ton bas,mauvaise odeur(avec grimace) nuu ton haut,bonne odeur; les deux avec leur ton dans le verbe;nuuka,sentir.

N diphtongue.

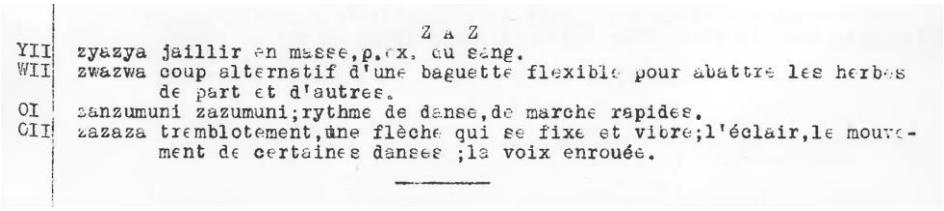
nyau miaulement du chat;verbe!,nyaanga.

N a N

Y nyana kweenda nyana nyana,marcher sur la pointe des pieds. nya nya,exclamation de surprise,de désapprobation. nyanika mbwetete zisiidi na n..,las étoiles font n.n..,scintillement, nombre.

W nuanu aller n.n..,à petits pas rapides;vitesse,rapidité. nuanuna, item. nanika,cfr naa,la femme attire à elle la branche,image visuelle du mouvement kweenda nanika nanika,rythme d'une allure ballante,de qqun qui est"étiré",long et maigre.

Addendum 4: The cluster **Zaz** in the *Lexique kikongo–français* by Charles Polis (1938) [Part I, p. 26]



Z a Z

YII zyazya jaillir en masse, p. ex. du sang.
WII zwazwa coup alternatif d'une baguette flexible pour abattre les herbes de part et d'autres.
OI zanzumuni zazumuni; rythme de danse, de marche rapides.
OII zazaza tremblotement, une flèche qui se fixe et vibre; l'éclair, le mouvement de certaines danses ; la voix enrouée.

Addendum 5: Data for Test 1: Polis (1938) vs. Butaye (1909) [with for Butaye: ‖ = main word; * = dialectal form; grey = the form, the meaning, or both are only approximately that/those of Polis; all translations into English are ours]

POLIS			**BUTAYE**	
Part, p.	**Item**	**English translation**	**Lemma sign**	**English translation**
I, 6	*pyu*	pitch-black, very dirty	‖*pilu, piu*	black, blue
I, 16	*kyu*	sound of throat movement when swallowing	—	—
I, 26	*zwe*	of thorn that enters	—	—
I, 36	*myaka*	sparkling, bright like varnish	*miaka*	itching
I, 46	*kilaa*	together!	—	—
I, 56	*zubu*	of intelligence; of vigilance	—	—
I, 66	*yololo*	of being soaked, of being drenched	—	—
II, 9	*kibabu*	*usa* ~ = he puts his zeal in …	—	—
II, 19	*booba*	to beat	*buba*	to beat
II, 28bis	*-papumuka*	to beat (the wings) heavily	*pápumúka*	to flutter about
II, 38	*-vwiidi* Mb., *-wiidi* Kis.	be finished	* *uwa, vwa* B.K. *(-widi)*	be completed
II, 48	*mfyaafi*	armpit	*mfímfiáfi & nsimfiafi*	armpit
II, 58	*ngoongi*	internode (of bamboo, fingers, etc.)	*ngongi, ngonge*	... part of a reed between the nodes
II, 68	*kaka*	to close	‖*kaka*	to obstruct; to close; to hit
II, 78	*nnkookolo*	song	‖*kékila & kokila*	to cackle, to sing like cocks & to sing (said of a cock)

II, 88	*daada*	to beat	* *dada*	to die a sudden death
II, 98	*nnledi*	educators	—	—
II, 108	*ta*	?	*ta*	[various]
II, 118	*ntuuta*	quarrel	—	—
II, 128	*zaanzaba*	to crawl, to slither	‖*zánzala*	to crawl, to walk like insects
II, 138	*nsyaasi, ntsyaasi*	noise of a small thing	—	—
II, 148	*bama*	do quickly (and badly); *b. nnsiinga* = make a bad knot	*bama*	to shout; to tighten a knot (*babamini nsinga* = they have tightened the knot)
II, 158	*biinga*	to call, to summon	‖*binga*	to go and meet
II, 168	*-bokila*	to proclaim (cf. p. 75)	‖*bókila*	to call after
II, 178	*baanda*	to hit	‖*banda*	to hit
II, 188	*mbila*	call	‖*mbila*	call
II, 198	*-bweetama*	to crush	‖*bweta*	to crush
II, 208	*kimpodi*	return obligation	*mpodi*	credit
II, 218	*baasa*	to cut, to split, to tear	‖*basa*	to split, make jump
II, 228	*maanga*	fetish to discover the magical cause of an illness; charm to indicate the guilty person	*manga*	charm to uncover the guilty person
II, 238	*mweena*	for 'to see'	*muene*	preterite of 'mona'
II, 248	*mfubu*	*makeenge ma* ~ = kind of pandanus with which mats are made	*fubu*	soft stem of the sugar cane, tender part of the palm leaf; old basket
II, 258	*mmfyaangu*	the tissues close to the waist, the kidneys	*mfiangu*	bundle of muscles in the back, from top to bottom
II, 268	*vidika*	to say something in a round-about way	—	—
II, 278	*fulukidi*	regained one's senses	‖*fúluka (-kidi)*	be full; dial. to come back to life
II, 288	*mmfunu*	utility	*mfunu*	utility
II, 298	*-kwabama*	to hit while passing	—	—
II, 308	*kuumbidi*	opened the mouth	‖*kumba (-bidi)*	to murmur, to make noise
II, 318	*kuma*	to put on	*kuma*	to erect, to place, to fix
II, 328	*keenge*	~ *kyaafu* = to make a bridge	‖*kanga (kenge, kangidi)*	to link up, to attach, to close, to bandage, to conclude
II, 338	*gyyodi gyyodi*	bush herb	*kiodi-kiodi*	herb that grows on stony terrain
II, 348	*-kaandama*	be prohibited	*kandama*	be withheld
II, 358	*kedingiinza*	to pound in small double beats (of the heart)	—	—

II, 368	*kunduba*	to move unwieldly	—	—
II, 378	*kitumuna*	to make someone change their mind	*kítumuna*	to change radically, to bribe, to seduce
II, 388	*kena*	to discover	* *gena*	to curl up
II, 398	*kasu*	ring to keep something open	—	—
II, 408	*laamba*	to knead	‖*lamba*	to cook, to steam/braise; to prepare food
II, 418	*-leepama, leempama?*	to be lying in wait, glued to the floor	—	—
II, 428	*tuumba*	to impose, to dedicate, to install	‖*tumba*	to initiate, to confer, to dedicate
II, 438	*bundyaafu*	gluttony, greed	—	—
II, 448	*dingutuna*	to cut off a big piece	—	—
II, 458	*-looka*	to become darker (of fruit)	‖*lóka*	to become dry
II, 468	*taki*	black	—	—
II, 478	*ntwaadi*	possessed jointly (by two or more)	‖*ntwadi*	association; in common
II, 488	*dyaana*	to make a loud noise	—	—
II, 498	*nyama*	animal, meat [recent word]	—	—
II, 508	*nyuunguta*	confusion of a happy crowd [sic]	* *niunguta*	to delight, to dance out of happiness
II, 518	*manyaanza-nyaanza*	mist, droplets of rain	*maniánga-nianga*	a few drops of rain, light and passing rain
II, 527	*nsaamba*	palm wine drawn from the male flower of this tree	*nsamba*	palm wine
II, 537	*-zoomene*	be ripe	*zoma (-mene)*	to be yellow or soft like a ripe fruit
II, 547	*zyoonguna*	to twist, to pull out	*zonguna*	to prick
II, 557	*-swengila*	to hurry, to hasten	—	—
II, 567	*swiika*	to gobble up (in large quantities)	—	—
II, 577	*kinzola*	snobbery, vanity	—	—
II, 587	*nsila*	strip	‖*nsila*	line, vein, groove
II, 597	*masuni*	canine teeth	*nsunda, nsundi*	boar tusk, canine tooth
II, 607	*mayùba*	old word for *mantusi* = 'meat maggots'	—	—
II, 617	*dyukusu, yukusu*	to have eaten one's fill	—	—
II, 627	*biiya, biiyidi*	to become bad	—	—
II, 637	*wengeenge*	cartilage	—	—

Addendum 6: Data for Test 2: Butaye (1909) vs. Polis (1938) [with for Butaye: ‖ and no shading = main words; grey = common words; for Polis: green = extensions to the lemma-sign system [sic]; red = errors in the lemma-sign system [sic]; all translations into English are ours]

BUTAYE				POLIS	
Page	**Lemma sign**	**POS**	**English translation**	**Lemma sign**	**Part, p.**
1	‖*aka*	adv.	yes	**Ka + O, Varia**	II, 63
2	*aritmetika*	n.	mathematics	—	—
3	‖*baba*	n. cl. 5	person who is mute or stutters	**BaB + OII, SUBST., bi**	II, 8
4	*bákila*	v. rel.	to take for	**BaK + OII, V.DER., iLA**	II, 163
5	‖*bámbuka*	v. m.	be reminded	**BaB + OI, V.DER., KA** [as *baambuka*]	II, 6
6	*bángamísa*	v. caus.	to cause an oppressed state	—	—
7	‖*bau*	pron. & adj. cl. 2	they; their	**au + B, --**	II, 644
8	*bémbama*	v. i.	be bent, folded	**BeB + OI, V. DER., MA**	II, 13
9	‖*beto, betu*	pron. & adj. cl. 2	we; our	—	Y [as *beeto*]
10	*bikáku*	n. cl. 8	congestion	**KaK + OII, SUBST., bi**	II, 68
11	‖*bindama*	v. m.	be closed	**BiD + OI, V.DER., aMA** [as *biindama*]	II, 186
12	*bítana*	v. rec.	to catch one another	—	—
13	‖*bôla*	n. cl. 5	onion	**BoL + OII, SUBST., ma** [as *boola*]	II, 190
14	*bonsikila, bonsukila*	v. i.	to insist	**BoZ & BoS + WII, V.DER., iKA**	II, 221
15	‖*buau*	adv.	like this	**au + B, bu** [as *bwau;* but not with main meaning]	II, 644
16	*buingi*	adj. & adv.	a lot	—	Y [as *bubwiingi*]
17	‖*búkumúka*	v. m.	be destroyed; be overturned	**BuK + OII, V.DER., uMA**	II, 170
18	*bumbangu*	n. cl. 14	craft knowledge	—	—
19	‖*buna*	v. tr.	to skin	**BuN + O, VERBE**	II, 215
20	*bundumba*	n. cl. 14	standing of a girl	—	—
21	‖*busi*	n. cl. 5	wad of a gun	**BuZ & BuS + OII, SUBST.** [as *buusi*]	II, 222

22	*buyúmbulu*	n. cl. 14	big stupidity	**Yub + I, 5** [as *yuumbulu*]	I, 64
23	‖*dâka*	v. i.	to split/crack; to shatter	**DaK & LaK + OII, VERBE**	II, 451
24	*diadia*	n. cl. 5	tall bush	—	Y [as *dyadya*]
25	‖*diata*	v. i.	to walk	**DaT & LaT + YII, VERBE** [as *dyaata*]	II, 474
26	*diéngasa*	v. tr.	to make turn	**DeG & LeG + YI, V.DER., aSA** [as *dyengasa*]	II, 443
27	‖*dila*	v. i.	to cry/weep	**DiL + O, VERBE**	II, 99
28	*dinkúndi(a)*	n. cl. 5	sp. of climbing plant	—	—
29	‖*dóngumúka*	v. m.	to emerge	—	—
30	*e*	quest. part.; affirm. part.	?; !	—	—
31	‖*fi-*	dim. pref.	a little bit, a small amount	**Fi + O, SUBST., FI**	II, 47
32	*fídila*	v. rel.	to lead to	**FiD & FiL + OII, V.DER., iLA**	II, 275
33	‖*fioti*	adv.	a little	**FoT + YII, VARIA** [as *fyote*]	II, 282
34	*fókuka*	v. m.	be folded; be multiplied	**FoK + OII, V.DER., uKA**	II, 262
35	‖*fula*	v. tr.	to conclude; to forge; to blow	**FuD & FuL + OII, VERBE** [as *fula* and *fuula*]	II, 276
36	*fúmbama*	v. m.	be curved	**FuB + OI, V.DER., aMA** [as *-fuubama*; a typo]	II, 247
37	‖*fúnguna*	v. tr.	to confess	**FuG + OI, V.DER., uNA**	II, 260
38	*futana*	v. tr.	be vexed	—	—
39	‖*fwanasa*	v. caus.	to make equal	**FaN + W, V.DER., aSA**	II, 286
40	*ga*	v. i.	to produce (fruit)	**Ga + O, VERBE**	II, 51
41	‖*gala-nti, gadi-nti*	n. cl. 7	carpenter	—	—
42	*gangama*	v. m.	to be in order	**GaG + OI, V.DER., MA**	II, 54
43	‖*gedila*	v. rel.	be clean/pure	**GeD & GeL + OII, V.DER., iLA**	II, 337

44	*gelele, gelele-gelele*	inv.	purity	**Gel**	I, 41
45	‖*go, gogo*	adv.; conj.	here/there; if, when; or	—	—
46	*gólakána*	v. poss.	be able to be pulled; to slip	**GoL + OII, V.DER., aKA**	II, 339
47	‖*gongi*	n. cl. 5	ball (of food)	**GoG + O, SUBST., ma** [as *goongi*]	II, 57
48	*guga*	n. cl. 5	space between two objects	—	—
49	‖*gulusa*	v. caus.	to save	**GuD + OII, V.DER., uSA**	II, 341
50	*Iezu Kristu*	n.	Jesus Christ	—	—
51	‖*kaba*	v. tr.	to give, to share	**KaB + OII, VERBE**	II, 300
52	*kakala*	v. i.	to slip, to worm/inch	—	—
53	‖*kádila*	v. rel.	to stay for; to be with	**KaD & KaL + OII, V.DER., iLA**	II, 352
54	*kambalala*	v. i.	be at fault; to over-shoot	**KaB + OI, V.DER., aLA**	II, 299
55	‖*kándikíla*	v. tr.	to prohibit	—	Y [at *kadila*, p. II, 352]
56	*kani*	adv.	not yet; not; or	**KaN + O, VARIA**	II, 385
57	‖*kati-kati*	adv.	in the middle	**KaT + OII, Varia** [as *kati kati*]	II, 373
58	*kekumuna*	v. tr.	be thirsty	**KeK + OII, V.DER., uMA** [as 'to clear one's throat']	II, 72
59	‖*kenga*	v. i.	to stop doing, to renounce	**KeG + OI, VERBE** [as *keenga*]	II, 330
60	*kétalála*	v. i.	to hold on; to insist	**KeT + OII, V.DER., aLA**	II, 376
61	‖*kia, kiya*	v. i.	to clear up	**Ka + Y, VERBE** [as *kya*]	II, 61
62	*kiatumuna*	v. tr.	to place a crowd in rows	—	—
63	‖*kibota*	n. cl. 7	club, bludgeon	—	—
64	*kidíla*	n. cl. 7	stock of goods	**DiL + O, SUBST., bi**	II, 99
65	‖*kifu*	n. cl. 7	quality, character; vice, error	**Fu + O, SUBST., bi**	II, 47
66	*kigana*	v. refl.	to sacrifice oneself	—	—
67	‖*kikálulu*	n. cl. 7	residence; character	**KaD & KaL + OII, V.DER., uLA**	II, 352
68	*kikongi*	n. cl. 7	small duck	**KoG + OI, SUBST., bi**	II, 332

69	‖*kikwa*	n. cl. 7	potato; big tuber	**Ka + W, SUBST., bi**	II, 61
70	*kileso ki muini*	n. cl. 7	dazzling ray of light	—	—
71	‖*kimbefo*	n. cl. 7	illness	**BeF + OII, SUBST., bi** [as *kimbeefo*]	II, 152
72	*kimbundu*	adv.	as a whole	**BuD + OI, SUBST., bi** [as *(ki)buundu*]	II, 191
73	‖*kimosi*	n. cl. 7	unity	**MoZ & MoS + OI, VARIA**	II, 241
74	*kimpólokósu, kimpolongoso*	n. cl. 7	cavity	**PoL + OII, V.DER., NGA**	II, 208
75	‖*kindíku*	n. cl. 7	friendship	**DiK + OII, SUBST., bi**	II, 456
76	*kingándi*	n. cl. 7	whatchamacallit	**GaD + OI, VARIA**	II, 335
77	‖*kinkála*	n. cl. 7	kick	**KaD & KaL + OII, SUBST., bi**	II, 350
78	*kinkutula*	adj.	that opens by itself	—	— [but verb is in, p. II, 382]
79	‖*kintete*	n. cl. 7; adv.	priority; first	—	Y [as *ntete*]
80	*kintwema*	n. cl. 7	breathlessness	—	—
81	‖*kiôlólo*	n. cl. 7	cry; applause	**KoD & KoL + YII, V.DER., uLA** [as *kyololo*]	II, 361
82	*kisania*	n. cl. 7	small tree with beautiful white flowers and good wood	**SaN + O, SUBST., bi**	II, 594
83	‖*kisína*	n. cl. 7	origin	**SiN + O, SUBST., bi**	II, 595
84	*kiswamu*	n. cl. 7	hiding-place	**SaM + W, V.DER., uNA** [as *kiswaamunu*]	II, 538
85	‖*kiteso*	n. cl. 7	measure, model	**TeZ & TeS + OII, SUBST., bi** [as *kiteeso*]	II, 495
86	*kítumúka*	v. m.	to undergo a radical change	**KiT + OII, V.DER., uMA**	II, 378
87	‖*kivúmu*	n. cl. 7	stomach	**VuM + O, SUBST., bi**	II, 249
88	*kiyaka*	n. cl. 15	assumption	—	—
89	‖*kizanu, kizalu*	n. cl. 7	insubordination	—	— [but verb is in, p. II, 593]
90	*kobe*	n. cl. 5	strong man	—	—
91	‖*kodila*	v. rel.	to pull for; be strong at	**KoD & KoL + OII, V.DER., iLA**	II, 366

92	*kómakána*	v. i.	be additionally added	—	—
93	‖*kondisa*	v. tr.	to make chase; to subtract	**KoD & KoL + OI, V.DER., iSA** [as *-koondisa*]	II, 363
94	*konko*	n. cl. 7	angle	**KoK + OI, SUBST., bi** [as *ikoonko*]	II, 75
95	‖*ku, kuku, kuna*	adv.; dem.; prep.	here/there; this/that; by, for, towards	**Ku + O, VARIA; KuK + OII, VARIA**	II, 65; II, 80
96	*kúbakána*	v. poss.	to be able to err	**KuB + OII, V.DER., aKA**	II, 310
97	‖*kúfama*	v. m.	be short	**KuF + OII, V.DER., aMA**	II, 323
98	*kukusu*	adv.	*kituka* ~ = be full of mud	—	—
99	‖*kúlumúka*	v. m.	to go down	**KuD & KuL + OII, V.DER., uMA**	II, 370
100	*kúmbalála*	v. i.	to be abundant	**KuB + OI, V.DER., aLA**	II, 309
101	‖*kundá*	adv.	far, high, deep	**KuD & KuL + OI, SUBST., mi** [as n. *nnkuunda*]	II, 367
102	*kúnkúfi*	adv.	very/too close, too short	—	—
103	‖*kútuka*	v. m.	to undo	**KuT + OII, V.DER., uKA**	II, 381
104	*kwámína*	v. rel.	to persevere in	**KaM + W, V.DER., iNA**	II, 315
105	‖*kwika*	v. tr.	to switch on, to arrange	**KiK + WII, VERBE** [as *kwiika*]	II, 73
106	*labidika*	v. i.	to launch, to throw	**BaB & LaB + OII, V.DER., iLA**	II, 409
107	‖*lalánsa*	n. cl. 5	orange (tree)	**LaL + OII, SUBST., ma** [as *mlaala*]	II, 89
108	*lâmbidíka*	v. tr.	to lay on its side	**BaB & LaB + OI, V.DER., aLA**	II, 408
109	‖*lángidíla*	v. rel.	to watch, to keep an eye on	**DaG & LaG + OI, V.DER., iLA**	II, 442
110	*laza*	n. cl. 5	empty words	—	—
111	‖*lela*	v. i.	be slippery	**LeL + OII, VERBE**	II, 95
112	*lembanzau*	n. cl. 7	wild pomegranate tree; fetish palm	—	—
113	‖*lenga*	v. i.	to walk, be on the way	**DeG & LeG + OI, VERBE**	II, 444

114	*lensi*	n. cl. 5	bitter taste	**DeZ & DeS & LeZ & LeS + OI, SUBST., ma**	II, 491
115	‖*lolula*	v. tr.	to forgive	—	— [but *loluka* 'be pardoned' is in, p. II, 101]
116	*lóngakana*	v. i.	to be prone to instruction	—	— [but *longuka* 'be instructed' is in, p. II, 449]
117	‖*lotó*	n. cl. 11	spoon	**To + O, SUBST., tu**	II, 109
118	*lubasa lu nima*	n. cl. 11	backbone	—	Y [pp. II, 139; II, 432; II, 615]
119	‖*lufwá*	n. cl. 11	death	**Fa + W, SUBST., tu**	II, 45
120	*lugambuku*	n. cl. 11	space, distance; retreat	—	— [but verb *-gaambuku* is in, p. II, 293]
121	‖*lukáya*	n. cl. 11	leaf	**KaY + O, SUBST., tu**	II, 630
122	*lukúba*	n. cl. 11	pillow	**KuB + OII, SUBST., tu**	II, 310
123	‖*lumba*	v. tr.; v. i.	to drop violently; to move forward	**DuB & LuB + OI, VERBE**	II, 416
124	*lumoko*	n. cl. 11	chatter, gossip	**MoK + OII, SUBST., bi & ma** [as *moko* & *mamoko*]	II, 232
125	‖*lunga*	v. tr.; v. i.	to keep; be correct, be perfect	**DuG & LuG + OI, VERBE**	II, 450
126	*lunsamba*	n. cl. 11	edible mushroom	**SaB + OI, SUBST., bu** [as *bunsambi*]	II, 526
127	‖*lusendo*	n. cl. 11	thorn	**SeD & SeL + OI, SUBST., tu** [as *luseende*]	II, 584
128	*lutangu*	n. cl. 11	reading, enumeration	**TaG + OI, SUBST., tu**	II, 462
129	‖*luziku*	n. cl. 11	funeral	**ZiK + OII, SUBST., tu**	II, 551
130	*lwákila*	v. rel.	to arrive at	—	— [but base verb *-lwaaka* 'to arrive' is in, p. II, 451]
131	‖*mádia*	n. cl. 15	food	**Da + Y, SUBST., bi** [as *kindya*, *kidya*]	II, 82
132	*mafubu*	n. cl. 6	marrow of palm leaves	**FuB + OII, SUBST., ma**	II, 247

133	‖*makábu*	n. cl. 6	present, gift	**KaB + WII, SUBST., ma** [as *makwaabu*]	II, 297
134	*makésa*	n. cl. 6	soldiers, army	**KeS + OII, SUBST., ma** [as *kesa*]	II, 401
135	‖*makutu*	n. cl. 6	ears	**KuT + OII, SUBST., (ku)**	II, 381
136	*maléla*	n. cl. 6	small pagnes (front and back)	**LeL + OII, SUBST., ma** [as *maleela*]	II, 95
137	‖*mampa*	n. cl. 6	bread	**Pa + O, SUBST., ma** [as *dimpa*]	II, 26
138	*manganana*	v. i.	to hold one's chest in front	**MaG + OI, V.DER., aNA**	II, 228
139	‖*manta, mata*	v.	to climb	**MaT + OII, VERBE**	II, 235
140	*masekwasa*	n. cl. 6	sparrows	**ZoK + OII, V.DER., aSA** [as *mazokaasi, masekwasi*]	II, 553
141	‖*maté*	n. cl. 6	saliva, mucus	**Te + O, SUBST., ma** [as *te*]	II, 108
142	*mavwangi*	n. cl. 6	bushes in a wood	**VaG + WI, SUBST., ma, i**	II, 253
143	‖*mazóno*	n. cl. 6	yesterday, an earlier day	**ZoN + O, SUBST., ma** [as *zono, mazoono*]	II, 594
144	*mbamu*	n. cl. 9	abuser of power, exploiter	**BaM + O, SUBST., zi** [as *mbaamu,* and meaning shift]	II, 148
145	‖*mbasi, mbazi*	n. cl. 7	tomorrow, a future day	**BaZ & BaS + OII, VARIA**	II, 218
146	*mbemba-mbemba*	n. cl. 9	butterflies	**BeB + OI, SUBST., tu** [as *lumbemba mbemba*]	II, 13
147	‖*mbeni*	n. cl. 9	enemy	**BeN + O, SUBST., zi** [as *mbeeni*]	II, 214
148	*mbílama*	n. cl. 9	abundance, multi-plication	**BiL + OII, V.DER., aMA** [as perfect verb *-bilamene*]	II, 188
149	‖*mboko*	n. cl. 9	bought peace	**BoK + OII, SUBST., zi, o**	II, 167
150	*mbota*	n. cl. 9	hard-wood tree	**BoT + OII, SUBST., zi**	II, 201
151	‖*mbulu-mbulu*	n. cl. 9	small black fly; black soldier	**BuL + SUBST., zi**	II, 194
152	*mbundu*	n. cl. 3	short grass; big crowd	**BuD + OI, SUBST., ma, IIu** [as *buundu*]	II, 191
153	‖*mbweta*	n. cl. 9	ball, sphere	**BeT + WII, SUBST., zi** [as *mbweeta*]	II, 197

154	*mengi*	adj. cl. 6	a lot	—	—
155	‖*mfiku*	n. cl. 9	age of wine; power of food; low price	**FiK + OII, SUBST., mi & zi** [as *mmfiku* & *mfiku*]	II, 261-2
156	*mfuki*	n. cl. 3	small civet; pain radiance	**FuK + OII, SUBST., mi & zi**	II, 263
157	‖*mfundu*	n. cl. 3; n. cl. 9	lawsuit; secret	**FuD & FuL + OI, SUBST., mi** [as *mmfuundu*]	II, 275
158	*mfwa*	n. cl. 9	death	**Fa + W, SUBST., zi**	II, 45
159	‖*miáya*	n. cl. 4	yawning	**MaY + WII, SUBST., mi** [as *mmwaayi*]	II, 627
160	*minuta*	n. cl. 3	minute	—	—
161	‖*mongo*	n. cl. 3	mountain	**MoG + OI, SUBST., mi** [as *moongo*]	II, 230
162	*mpambu*	n. cl. 9	bifurcation	**PaB + OI, SUBST., u** [as *mpaambu*]	II, 145
163	‖*mpemba*	n. cl. 9	white clay (used to whitewash)	**PeB + OI, SUBST., zi** [as *mpeemba*]	II, 146
164	*mpímbidi*	n. cl. 9	flower of the banana tree	**PiB + OI, SUBST., zi**	II, 146
165	‖*mpuku*	n. cl. 9	rat	**PuK + OII, SUBST., zi, u**	II, 177
166	‖*mputulukesi*	n. cl. 1	Portuguese	**PuT + OII, SUBST., zi**	II, 214
167	‖*muánga*	n. cl. 3	sting, venom	**MaG + WI, SUBST., mi** [as *mmwaanga*]	II, 227
168	*muila*	n. cl. 3	estuary	**MiD & MiL + OII, V.DER., iLA** [as *mwiidila*, *mwiila*]	II, 234
169	‖*mumpani*	n. cl. 3	pagan, heathen	**PaN + O, SUBST., mi**	II, 215
170	*mungwalanga*	n. cl. 3	any tall tree	—	—
171	‖*muéso*	n. cl. 3	whistling	—	— [as *mu–mpyooso* on p. II, 225]
172	*mvubi, mvuya*	n. cl. 3	excellent rattan to weave baskets	**VuY + SUBST., zi**	II, 628
173	‖*mvutu*	n. cl. 9	return, answer	**VuT + OII, SUBST., zi**	II, 272
174	*mvwototo*	n. cl. 9	plain (food)	—	—
175	‖*nánama*	v. m.	be tight, tall and thin	**NaN + O, V.DER., aMA**	II, 123
176	*nangimisa*	v. caus.	to bother, to offend, to annoy	**NaG + OI, V.DER., uMA**	II, 502

177	‖*ndala*	n. cl. 9	palm leaf; sleep; alarm	**DaL + O, SUBST., zi**	II, 88
178	*ndédila*	n. cl. 9	strip of cloth/paper	—	— [other word: *nzyaadi* on p. II, 572]
179	‖*ndíkila*	n. cl. 9	action to feed with; poison	**DiK + OII, V.DER., iLA**	II, 457
180	*ndoka*	n. cl. 9	rain	**DoK & LoK + OII, SUBST., zi**	II, 458
181	‖*ndûka*	n. cl. 9	caution, care; stink	**DuK & LuK + OII, SUBST., zi** [as *nduuka*]	II, 460
182	*ndwanisa*	n. cl. 9	attack	—	—
183	‖*ngandu*	n. cl. 9	crocodile	**GaD + OI, SUBST., zi** [as *ngaandu*]	II, 335
184	*ngau, ngawa*	n. cl. 9	palm tree bird	**au (monosyllabic)**	II, 644
185	‖*ngo*	n. cl. 9	leopard	**Go + O, SUBST., zi**	II, 52
186	*ngonda*	n. cl. 9; adj.	murder; bloody	—	— [but *-goonda* 'to kill' is in, p. II, 338]
187	‖*nguka*	n. cl. 3	caterpillar	**GuK + OII, SUBST., mi** [as *nnguka*]	II, 325
188	*ngunsa*	n. cl. 1	prophet	—	—
189	‖*niakuna*	v. tr.	to chew	**N-k + YII**	I, 55
190	*niémita*	v. tr.	to pinch	**NeM + Y, V.DER., iTA** [as *nyeemita*]	II, 498
191	‖*nioka*	n. cl. 9	snake	**NoK + YII, SUBST., zi** [as *nyoka*]	II, 512
192	*nkâdidi*	< v. (pret. < *kala*)	I denied; I don't want	**KaD & KaL + OII, VERBE**	II, 349
193	‖*nkama*	n. cl. 9	hundred; act of dyking up; husband/spouse; dyke	**KaM + O, SUBST., zi & bi & mi**	II, 315
194	*nkanga*	n. cl. 9	hand of bananas; act of binding; quail	**KaG + OI, SUBST., zi** [as *nkaanga*]	II, 329
195	‖*nkási*	n. cl. 9	fraction of a number	**KaZ & KaS + OII, SUBST., mi** [as *nnkasi*]	II, 398
196	*nkefo, nkefua*	n. cl. 9	spicy (pepper)	**KeF + OII, SUBST., zi**	II, 322
197	‖*nkeni(a)*	n. cl. 9	corn	**KeN + O, SUBST., zi**	II, 388

198	*nkila*	n. cl. 3	tail	**KiD & KiL + OII, SUBST., mi** [as *nnkila*]	II, 361
199	‖*nkókila*	n. cl. 9	evening; cock's singing; act of attracting with a hook	**KoK + OII, V.DER., iLA** [as *nkookila*]	II, 77
200	*nkondi*	n. cl. 1 & 9; n. cl. 9	hunter; fetish	**KoD & KoL + OI, SUBST., zi** [as *nkoondi*]	II, 363
201	‖*nkufi*	n. cl. 9	salutation	**KuF + OII, SUBST., ma** [as *makufi, kufi*]	II, 322
202	*nkumba*	n. cl. 3	navel	**KuB + OI, SUBST., mi** [as *nnkumba*]	II, 308
203	‖*nkusu*	n. cl. 9	parrot	**KuS + OII, SUBST., zi**	II, 405
204	*nkwamu*	n. cl. 9; adj.	duration; continuous	**KaM + W, SUBST., tu** [as *lukwaamu*]	II, 315
205	‖*nlele*	n. cl. 3	cloth, pagne made out of cloth	**LeL + OII, SUBST., mi** [as *nnleele*]	II, 95
206	*nluku*	n. cl. 3	young breasts	**DuK & LuK + OII, SUBST., ?** [as *nnluku*]	II, 460
207	‖*nsa*	n. cl. 9	greyish antelope	**Sa + O, SUBST., zi**	II, 135
208	*nsaku*	n. cl. 9	old thing	—	—
209	‖*nsampatu, nsampítu, nsamputu*	n. cl. 9	shoe	—	Y [and *nsabaatu* is also in, p. II, 529]
210	*nsau*	n. cl. 9	pilot, ferryman	**Sau + mi** [as *nnsau*]	II, 137
211	‖*nsi*	n. cl. 9	floor, region, country	**Si + O, SUBST., zi**	II, 136
212	*nsila mvula*	n. cl. 9	heavy rain	—	—
213	‖*nsoki*	n. cl. 3; n. cl. 9	mistake; ramrod; long type of grass	**SoK + OII, SUBST., mi & tu** [as *nnsoki* & *lusoki*]	II, 569
214	*nsóngisa*	n. cl. 9	act of redressing	—	— [but base verb *soonga* 'to be straight' is in, p. II, 559]
215	‖*nsuka*	n. cl. 9	end	**SuK + OII, SUBST., zi**	II, 571
216	*nsungu*	n. cl. 9	worry	—	— [but base verb *suunga*

					'to worry about' is in, p. II, 560]
217	‖*nswalu*	n. cl. 3	speed	**SaL & SaD + OII, SUBST., mi** [as *nnswaala*]	II, 581
218	*ntamba*	n. cl. 9	offer	—	— [but verb is in, p. II, 419]
219	‖*nteba*	n. cl. 9	mud	**TeB + OII, SUBST., zi**	II, 425
220	*nteti-mbisi*	n. cl. 1	butcher	—	—
221	‖*ntomboka, ntombuka*	n. cl. 9	ascension	—	Y [p. II, 314; and verb is in, p. II, 426]
222	*ntoya*	n. cl. 9	bird of the banana tree	**ToY + O, SUBST., zi**	II, 632
223	‖*ntûtu*	n. cl. 3	bottle; opening	**TuT + OII, SUBST., mi** [as *nntuutu*]	II, 117
224	*nungumuka, nungumuna*	v. tr.	to push hard	**NuG + OI, V.DER., uMA**	II, 508
225	‖*nyalu*	n. cl. 9	row, layer	**YaD & YaL + II, SUBST., zi**	II, 617
226	*nzalala*	n. cl. 9	haste	**ZaD & ZaL + OII, V.DER., aLA** [as *zalala*]	II, 574
227	‖*nzaza*	n. cl. 9	vessel, ship	**ZaZ + OI, SUBST., zi** [as *nzaanza*]	II, 127
228	*nzíkisa*	n. cl. 9	check, proof	—	— [but verb is in, p. II, 552]
229	‖*nzo*	n. cl. 9	house	**Zo + O, SUBST., zi**	II, 126
230	*nzo zi matubu tatu*	n. cl. 10	two pagnes sewn together	**Zo + O, SUBST., zi** [as *nzo tubu*] **& TuB + OII, SUBST., ?** [as *nzo tubu*]	II, 126 & II, 429
231	‖*nzuzi*	n. cl. 9	"tiger cat"	**ZuZ + OII, SUBST., zi**	II, 132
232	*pala*	v. i.	to make jealous	**PaL + OII, VERBE**	II, 204
233	‖*pii, pidi*	n. cl. 5	silence	**Pi + O ; P—d, OII**	I, 5 ; I, 34
234	*pupula*	v. i.	to stay up	—	—
235	‖*sabi*	n. cl. 5	cork	**SaB + OII, SUBST., zi** [as *nsabi*]	II, 528
236	*sakula*	v. tr.	to weed	**SaK + OII, V.DER., uLA**	II, 564

237	‖*sampula*	v. tr.	to carry over	**SaP + OI, V.DER., uLA** [as *saampula*]	II, 533
238	*sangalakasa*	v. i.	to mix	—	— [but base verb *-saanga* 'to mix' is in, p. II, 554]
239	‖*sasuka*	v. m.	to hurry	**SaS + OII, V.DER., u--**	II, 140
240	*seke, seki*	n. cl. 5	small bush sparrow	**SeK + OII, SUBST., ma**	II, 466
241	‖*sénguka*	v. m.	be put face-up, be discovered	**SeG + OI, V.DER., uKA** [as *seenguka*]	II, 557
242	*siatunina*	v. rel.	to inject in	—	— [but verb *-syatula* 'to spurt out' is in, p. II, 590]
243	‖*simba*	v. tr.	to hold	**SiB + OI, VERBE**	II, 530
244	*síngamísa*	v. caus.	to make sure st. is straight	—	— [but verb *-siingama* 'to be straight' is in, p. II, 558]
245	‖*sókama*	v. m.	be rare	—	—
246	*sómbuka*	v. m.	to jump	**SoB + OI, V.DER., uKA** [as *soombuka*]	II, 531
247	‖*sótuka*	v. m.	to get lost	**SoT + OII, V.DER., uKA**	II, 592
248	*sukulu*	n. cl. 5	corner behind the door	—	—
249	‖*sumuka*	v. m.	to transgress, to commit a sin	**SuM + O, V.DER., uKA**	II, 540
250	*sunsimika*	v. tr.	to swallow whole	—	—
251	‖*swétama*	v. m.	be thin, be narrow	**SeT, WII, V.DER., aMA** [as *sweetama*]	II, 591
252	*taka*	n. cl. 7	fork (used to hold thieves)	**TaK + OII, SUBST., ma**	II, 467
253	‖*táluka*	v. m.	be spread	**TaL + OII, V.DER., uKA**	II, 481
254	*tanda-tanda*	n. cl. 7	extreme thinness, wasting away	**TaD & TaL + OI, SUBST., mu**	II, 478
255	‖*tâta*	n. cl. 1	father; chief	**TaT + OII, SUBST., ba** [as *taata*]	II, 111
256	*tebisa*	v. caus.	to fill to the brim	—	Y [p. II, 574]
257	‖*teluka*	v. m.	be removed from the fire	**TeL + OII, V.DER., uKA**	II, 483

258	*tengama*	v. m.	to accumulate (in a pool of water)	**TeG + OI, V.DER., aMA**	II, 463
259	‖*tiaka*	v. tr.	to spill, to pour	**TaK + YII, VERBE** [as *tyaaka*]	II, 467
260	*tikita*	n. cl. 5	knife	**TiK + OII, V.DER., iTA**	II, 471
261	‖*titila*	v. rel.	to shake because of, to be afraid of	—	— [but base verb *-tiita* 'to tremble' is in, p. II, 115]
262	*tokolo*	n. cl. 5	jaw	**ToK + OII, V.DER., aLA** [as *matookolo*]	II, 472
263	‖*tona*	v. i.	to understand	—	Y [p. II, 490]
264	*tótakána*	v. i.	to converge	—	—
265	‖*tubuka*	v. m.	be pierced	**TuB + OII, V.DER., uKA**	II, 429
266	*túkula, dukula*	v. i.	to move, to shake	**TuK + OII, V.DER., uLA**	II, 474
267	‖*tumbuka*	v. m.	to manifest; to stop dripping	**TuB + OI, V.DER., uKA** [as *tuumbuka*]	II, 428
268	*tungalakala*	v. i.	to interlace, to intertwine	—	— [but verb *tungalaka* is in, p. II, 466]
269	‖*tuti*	n. cl. 5	cloud	**TuT + OII, SUBST., ma**	II, 117
270	*twenga*	n. cl. 5	young shoot	**TeG + WI, SUBST., bi** [as *tweengi*]	II, 463
271	‖*uele*	< v.	he/she is gone	—	Y
272	*vika*	v. aux.	(marks an imminent action)	**ViK + OII, VERBE**	II, 256
273	‖*vítika*	v. tr.	to lower, to bow	**ViT + OII, V.DER., iKA**	II, 272
274	*vúmana*	v. rec.	to respect one another	—	— [but noun *luvuma* 'respect' is in, p. II, 249]
275	‖*vunga*	n. cl. 5	any cover	**VuG + OI, SUBST., ma** [as *vuunga*]	II, 255
276	*vúzana*	v. i.	~ *mbundu* = to feel nauseous	**VuZ & VuS + OI, V.DER., aNA** [as *vuunzana*]	II, 290
277	‖*vwansa, vwanza*	v. i.	to damage, to obliterate, to dirty	**VaZ & VaS, WI, VERBE** [as *vwaanza*]	II, 288

278	*wa*	v. tr.	to hear; to understand; (completed action)	**Wa, VERBE**	II, 634
279	‖*wena*	v. i.	to stop, to leave	**WeN + VERBE** [as *weena*]	II, 640
280	*yabakana*	v. rec.	to talk a lot amongst one another	—	—
281	‖*yálumuna*	v. tr.	to unroll, to spread	**YaD & YaL + II, V.DER., uMA**	II, 618
282	*yanga*	v. tr.	to heat	**YaG + I, VERBE**	II, 609
283	‖*yaya*	n. cl. 1	grandmother	**YaY + OII, SUBST., ba**	II, 601
284	*yekikila*	v. tr.	to stand/lean against st.	**YeK + II, V.DER., iKA** [as 'to stop; to keep']	II, 614
285	‖*yemba*	v. tr.	to steal	**YeB + I, VERBE** [as *yeemba* 'to leave']	II, 605
286	*yensula*	v. tr.	to open wide	—	—
287	‖*yikama*	v. m.	be added	**YiK + II, V.DER., aMA**	II, 615
288	*yindumuna*	v. rev.	to make forget	—	—
289	‖*yobila*	v. i.	to take a bath	**YoB + II, V.DER., iLA**	II, 606
290	*yukula*	v. i.	to make light with a torch	—	— [but a homonym is in, p. II, 616]
291	‖*yúngana*	v. i.	to wander	**YuG + I, V.DER., aNA**	II, 613
292	*zakila, zakimina*	v. rel.	to tremble	**ZaK + OII, V.DER., aMA**	II, 549
293	‖*zaula*	v. tr.	to draw, to pump	**au (polysyllabic) + Z-L**	II, 645
294	*zeleniana, zelengana, zelangana*	v. i.	be exhausted, be weak	**ZeD & ZeL + OII, V.DER., aNGA** [as *-zelangana*]	II, 575
295	‖*zíama*	v. m.	be buried	**ZaM + Y, VERBE** [as *zyaama*]	II, 535
296	*ziku-ziku*	adv.	just	**ZiK + OII, VARIA** [as *ziku ziku*]	II, 551
297	‖*zingisa*	v. caus.	to make last, to make live	—	— [but base verb is in, p. II, 545]

298	*ziúngalakána*	v. i.	to whirl around, to surround	**ZuG + YI, V.DER., aLA** [as *–zyunga-lakana*]	II, 547
299	‖*zole*	num.	two	—	Y
300	*zundu*	n. cl. 5	giant frog	**ZuD & ZuL + OI, SUBST., ma** [as *zuundu*]	II, 577
301	‖*babila*	v. tr.	to flap, to flutter; to dance	**BaB + OII, V.DER., iLA**	II, 10
302	*bondengila*	v. i.	to resonate far away	—	—
303	‖*dombo*	n. cl. 5	a single bar of brass (which serves as money)	—	—
304	*gakula*	v. tr.	to deceive	—	— [but a homonym is in, p. II, 323]
305	‖*kibidi*	n. cl. 7	place where one rets (soaks) manioc	—	—
306	*kintuku*	n. cl. 7	charm, fetish	—	—
307	‖*sunguta*	v. i.	to limp	**SuG + OI, V.DER., uTA** [as *suunguta*]	II, 561
308	*vonga, fonga*	v. i.	to sit oneself, be seated	**VoG + OI, VERBE**	II, 254

— Abbreviations in the POS (part of speech) column, using Butaye's terminology: < = derived from ...; *adj*. = adjective; *adv*. = adverb; *affirm*. = affirmative; *aux*. = auxiliary; *caus*. = causative (-is-); *cl*. = class; *conj*. = conjunction; *dem*. = demonstrative; *dim*. = diminutive; *i*. = intransitive; *inv*. = invariable; *m*. = -uk-, -am-; *n*. = noun; *num*. = numeral; *part*. = particle; *poss*. = possibility (-akan-); *pref*. = prefix; *prep*. = preposition; *pret*. = preterite; *pron*. = pronoun; *quest*. = question; *rec*. = reciprocal (-an-); *refl*. = reflexive (ki-, i-); *rel*. = relative (-il-, -in-); *rev*. = reversive (-un-); *tr*. = transitive; *v*. = verb.

— The noun class numbers in the POS column reflect those of Butaye (and include errors), but were transposed to their modern equivalents.

— In the last column Y = yes, meaning that while the word may not be an item, it is still used in Polis.

http://lexikos.journals.ac.za

Towards Linking User Interface Translation Needs to Lexicographic Theory*

André du Plessis, *Department of Afrikaans and Dutch, Stellenbosch University, Stellenbosch, South Africa, (ahdp@sun.ac.za)*
and
Marius Swart, *Department of Afrikaans and Dutch, Stellenbosch University, Stellenbosch, South Africa (mariusswart@sun.ac.za)*

Abstract: In a time of proliferating electronic devices such as smartphones, translators of user interfaces are faced with new challenges, such as the use of existing words in new contexts or in their obtaining new meanings. In this article, three lexicographic reference works available to translators in this field are compared: the *Kuberwoordeboek/Cyber Dictionary* (Viljoen 2006), the Pharos Afrikaans–Engels/English–Afrikaans Dictionary (Du Plessis et al. 2010) and the Microsoft Language Portal (www.microsoft.com/Language 2015). A list of selected examples (in English) is used to determine the extent to which each of these three works fulfils the needs of the user in terms of meaning discrimination for translating into Afrikaans. After determining this, an attempt is made to indicate whether the use of meaning discriminators such as part-of-speech markers, punctuation, paraphrases of meaning, and contextual and co-textual guidance (as indicated by Beyer 2009: 11) may have contributed to the success or failure of the given reference work, in order to arrive at a conclusion about the link between lexicographic theory and usability.

Keywords: TRANSLATION, TRANSLATORS, BILINGUAL DICTIONARIES, MOBILE DICTIONARIES, LEXICOGRAPHIC THEORY, MEANING DISCRIMINATION, EQUIVALENT DISCRIMINATION, PARTS OF SPEECH, SMARTPHONES, USER INTERFACE

Opsomming: 'n Ondersoek na die verband tussen leksikografieteorie en die behoeftes van koppelvlakvertalers. In 'n tydvak waar die gebruik van elektroniese toestelle soos slimfone hand oor hand toeneem, word vertalers deur nuwe uitdagings in die gesig gestaar, soos die gebruik van bestaande woorde in nuwe kontekste of die verwerwing van nuwe betekenisse. In hierdie artikel word drie leksikografiese naslaanbronne wat vir vertalers in hierdie veld relevant is, vergelyk: die *Kuberwoordeboek/Cyber Dictionary* (Viljoen 2006), Pharos se Afrikaans–Engels/English–Afrikaans-woordeboek (Du Plessis et al. 2010) en Microsoft se taalportaal (www.

* This article was presented as a paper at the LSSA/SAALA/SAALT Joint Annual Conference 2015, which was held at the North-West University, Potchefstroom Campus, Potchefstroom, South Africa, 24–26 June 2015. It was also presented at the Twentieth Annual International Conference of the African Association for Lexicography (AFRILEX), which was hosted by the University of KwaZulu-Natal, Durban, South Africa, 6–8 July 2015.

microsoft.com/Language 2015). 'n Lys gekose voorbeelde (in Engels) word gebruik om te bepaal in welke mate elk van die drie bronne die gebruiker se behoeftes in terme van betekenisonderskeiding vervul wanneer daar in Afrikaans vertaal word. Vervolgens word daar gepoog om aan te dui of die gebruik van betekenisonderskeiers soos woordsoortmerkers, punktuasie, betekenisopsommings en kontekstuele en ko-tekstuele leiding (soos deur Beyer 2009: 11 aangedui) bygedra het tot die geslaagdheid al dan nie van die betrokke bron ten einde 'n afleiding te maak oor die skakel tussen die leksikografieteorie en bruikbaarheid.

Sleutelwoorde: VERTALING, VERTALERS, TWEETALIGE WOORDEBOEKE, SELFOON-WOORDEBOEKE, LEKSIKOGRAFIESE TEORIE, BETEKENISDISKRIMINASIE, EKWIVALENT-DISKRIMINASIE, WOORDSOORTE, SLIMFONE, GEBRUIKERSKOPPELVLAK

Introduction

Technological innovations such as smartphones result not only in the creation of new words, but also in the use of existing ones in new contexts or in their obtaining new semantic values (Fontenelle 2013: 1097). Furthermore, users of these innovations need the language these devices speak to be their own (Kelly and Zetzsche 2012: 250). Translators are therefore faced with unique challenges when translating the text in the user interface (UI) of these devices.

The global market for smartphones is dominated by two operating systems — Apple iOS and Google Android, sharing 96.3% of the market as of 2014 (Apple with 14.8% vs. Google with 81.5%). In the fourth quarter of 2014, Android hit the one billion mark, selling 1.06 billion units for the year eventually. More phones with Android as their operating system were sold in 2014 than total smartphone sales combined in the previous year (IDC 2015).

Smartphone usage in South Africa, as in the rest of the world, is also experiencing this kind of exponential growth. In 2014, approximately 19 million people in South Africa (out of a total of 42 million phone users) used smartphones, with this number set to rise to above 23 million in 2015 (Van Zyl 2015). Especially "low-cost Android phones" are seen to be an important contributor to this growth and these kinds of figures (ibid.). The South African smartphone market is "becoming an Android market" (Goldstuck in Van Zyl 2015).

The Android operating system supports 44 international languages in its UI, along with Australian, American and British English (https://support.google.com/googleplay/android-developer/table/4419860?hl=en). Afrikaans is one of these languages and along with Amharic, Swahili and Zulu, form the African contingent in the group. Judging by the proliferation of devices like smartphones, as well as Android's overwhelming market share and its multitude of UI language options, it would stand to reason that this field is one of the most prolific and important types of technical translation undertaken in the world today. These are also the reasons why we have decided to use text from the Android UI for this investigation.

Translators of UI text in these languages are faced with numerous problems, one of which is the expansion in meaning of existing words (mostly nouns and verbs) when used in a UI environment. Byrne (2006: 4) posits that a failure to comply with target language text conventions can undermine not only the credibility of the text itself, but, by implication, also that of the author and the information in the text. To reach the goal of having the device speaking the user's language, it is therefore imperative that translators have access to sources that can help guide them through this veritable minefield.

Therefore, in this preliminary investigation, this problem is used as a springboard for comparing three authoritative lexicographic reference works commonly used by UI translators working from English into Afrikaans. These three works are the *Kuberwoordeboek/Cyber Dictionary* (Viljoen 2006), the Pharos Afrikaans–English/English–Afrikaans dictionary (2010) and the Microsoft Language Portal (www.microsoft.com/Language 2015). The first is the only explicitly named specialised bilingual dictionary in the computer-related field for this language pair. The second is the most popular comprehensive Afrikaans/English English/Afrikaans dictionary readily available in South Africa. The third is a free online terminology list owned by Microsoft, described in literature as a "remarkably multilingual" company that is "represented on the majority of the world's computers" (Kelly and Zetzsche 2012: 250). This list caters for the target users translating numerous types of UI text, although it originated specifically for use when translating UI text for Microsoft.

It should be mentioned at this point that for the purposes of this article we will not be looking at the use of other online forums, chat groups and term lists/databases also utilised by translators, which should form a future investigation in its own right, but only at these three existing lexicographic reference works.

A list of selected examples (in English), each with more than one potential part-of-speech possibility and, therefore, at least two senses of meaning, is used to determine the extent to which each of the three works mentioned above fulfils the needs of this target user group when translating into Afrikaans by not being inadequate or inconsistent in the recognition of the manifold semantic values of the given words. The purpose of this article is not to formulate value judgements on the acceptability or adequacy of a given translational equivalent, but to determine the extent to which each of the three lexicographic reference works succeeds in indicating different senses of a given word. We have attempted to determine whether the application of lexicographic principles to a given reference work can be indicated to have an effect on its utility for the translator of UI text.

Theoretical overview

According to Gouws (2006: 85), a dictionary should never be compiled at random — the user must be the central consideration in all processes of lexico-

graphical production and the lexicographer has to remain aware of the specific situations in which the dictionary will be used. Potgieter (2011: 3) furthermore points out that discussions relating to the compilation of specialised dictionaries have traditionally not paid significant attention to translators as target users, despite their being a group with particularly challenging requirements. These requirements can range from paraphrases of meaning, exact and recognised terminology, to contextual and co-textual guidance (De Foglio and Lubbe 2002: 121). Burkhanov (2004: 22) states that the ideal translation-orientated bilingual dictionary should be able to satisfy the "translator's attempts to produce an adequate target text that conforms to the requirements of translational norms." Therefore, in order to enable translators to extract maximum value from a dictionary, it is imperative that data must be presented in the most useful and accessible fashion.

Zgusta (1984: 147) points out that "[a bilingual] dictionary should offer not explanatory paraphrases or definitions, but real lexical units of the target language which, when inserted into the context, produce a smooth translation." Zgusta's statement has merit, but it is easier said than done. In order to produce a smooth translation, a target user might require more contextual information than just lexical units of the target language. In the case of translators, the target user is one who has a good linguistic knowledge of both source and target languages, but does not necessarily have specialised knowledge as it pertains to the subject field. In such a case it becomes evident that bilingual dictionaries or translation-orientated dictionaries must provide more information than just the translation equivalent. A bilingual dictionary should therefore strive towards semantic-pragmatic and communicative equivalence (Beyer 2009: 2).

From the abovementioned it can be derived that a bilingual dictionary must guide the user through functional dictionary entries by means of equivalent discrimination so that communicative equivalence can be reached (Beyer 2009: 2-3). The importance of equivalence discrimination is given by Al-Kasimi (1977: 63):

(i) No word (or semantic unit) ever has exactly the same meaning in two different utterances.
(ii) There are no complete synonyms within a language.
(iii) There are no exact correspondences between related words in different languages.
(iv) Absolute equivalents, which have exactly the same semantic and grammatical function in both languages, are rare.

Subsequently, a bilingual dictionary should provide "meaning discriminations which enable the user to select the appropriate equivalent or the proper sense of an equivalent" (Al-Kasimi 1977: 68). The need for clearly distinguishable equivalents cannot be denied. However, this can only be achieved if, as Robert (1990: 219) states, "semantic and stylistic discrimination of equivalents, detailed

grammatical information, and collocational specifications for each headword and [...] for each sense division of the headword" are also provided. Here Robert suggests the idea of facilitating contextual and semantic data in the pursuit of semantic and communicative equivalence.

Apart from the dictionary or lexicographer, the user must also play an active role in the achievement of communicative equivalence (Beyer 2009: 4). This relates to the encoding of lexicographic data in the dictionary and the decoding of the data or information by the user (Beyer 2009: 4). The encoding and decoding of lexicographic data ultimately facilitates successful or unsuccessful equivalent discrimination, i.e. a dictionary user selects an appropriate or inappropriate communicative translation equivalent in a given translation equivalent paradigm (Beyer 2009: 5). This is where discrepancies come to the fore. Translators rely on dictionaries to provide accurate data so that communicative equivalence can be achieved (Gouws 1992: 38), but, as alluded to earlier, there is an existing trend that bilingual or translation orientated dictionaries do not meet these requirements or satisfy translator needs (Burkhanov 2004: 26).

Moreover, Beyer (2009: 5) notes that there are cases where bilingual dictionary entries provide adequate facilitating translation or meaning discriminators that lead to successful equivalent discrimination (called functional facilitated successful equivalence), but that this can also occur in the absence of such discriminators. Beyer (2009: 5) refers to this as incidental successful equivalent discrimination. In a translator's case, it would be more accurate to link this to the translator's apt dictionary usage skills, language proficiency or translation experience, rather than an accidental success. Therefore on this basis we propose another distinction when the success of equivalence discrimination is measured in the case of translators — successful intuitional equivalent discrimination. Regardless of how proficient or skilled the translator or user is, no dictionary user should be left to rely solely on his/her intuition, or fate, in the pursuit of communicative equivalence. Following from this, Beyer (2009: 5) correctly notes that it is lexicographically unjustifiable to apply or rely on unsuccessful, incidental or even intuitional, equivalent discrimination.

Potgieter (2011: 97) states that most South African translation-orientated dictionaries succeed in helping users achieve semantic equivalence, but very few of these dictionaries also succeed in helping users achieve communicative equivalence. Beyer (2009: 4) echoes this point when he argues that Afrikaans bilingual dictionaries are often inadequate for the accomplishment of communicative equivalence, as they contain inadequately encoded lexicographic data and/or the data is inadequately decoded. For translators [and other user groups — AdP and MS] the problems stem from inadequate equivalence discrimination in bilingual dictionaries (Crafford 2005: 27). This issue again relates to contextual data or guidance. In different contexts, translation equivalents can be semantically equivalent, but if these different contexts aren't given to the user, communicative equivalence cannot be achieved (Gouws 1992: 37).

Likewise Beyer (2009: 6) focuses on the need for contextual data and distinguishes between primary and secondary contexts: Semantic equivalence can be regarded as the primary context in which translation equivalents are presented in the bilingual dictionary, but the broader discourse situation, i.e. communicative equivalence, must also be taken into account as a secondary context. Crystal's (1991: 78-79) definition of context emphasises the need for extra information to make meaning transparent and shows the important distinction between linguistic context or co-text, and pragmatic context:

> A general term used in linguistics and phonetics to refer to specific parts of an utterance (or text) near or adjacent to a unit which is the focus of attention [...] The everyday sense of the term is related to this, as when one 'puts a word in context', in order to clarify the meaning intended, as in dictionary entries. Providing a context in this way is referred to as contextualisation. [...][Furthermore it refers] to the features of the non-linguistic world in relation to which linguistic units are systematically used [also referred to as situational context]. In its broadest sense, situational context includes the total non-linguistic background to a text or utterance, including the immediate situation in which it is used. […] Further distinctions are usually made in semantics and stylistics, distinguishing, for example, referential and emotive meaning from contextual meaning, i.e. information is signalled about the kind of use a linguistic unit has in its social context, e.g. whether it has a 'restricted' use (as in social pleasantries, or religious settings), or how it relates to such factors as age, sex or class of the speakers.

From Crystal and Beyer's definitions it becomes clear that contextualisation should be an important consideration when working with lexicographic data, as it can facilitate successful communicative equivalence in bilingual dictionaries. Beyer (2009: 8) also refers to contextualisation when a lexicographer encodes primary or secondary contextual data in an equivalent discriminatory dictionary entry. It must be noted that contextualisation does not automatically lead to functional facilitated successful equivalence, as the data should still be encoded and decoded in an adequate and functional manner (cf. Beyer 2009). Herein lies the possible key to success for translators or other users when using a bilingual dictionary. As with any dictionary, bilingual and translation-orientated dictionaries' data must be provided in a consistent and accurate manner so that the user can be led to achieve semantic and communicative equivalence.

Contextual guidance can be achieved through the use of different types of equivalent discriminators. Iannucci (1967), Al-Kasimi (1977) and Beyer (2009; 2013) propose a range of context-giving and meaning/equivalent discriminators, which we have adapted into four different types:

— Part-of-speech indicators, which differentiate syntactic functions;
— Punctuation, which provides negative discrimination by using different punctuation marks to separate different equivalents;
— Contextual and co-textual guidance, like (lexicographic or subject-field) labels, glossaries, collocations, example sentences and translation compo-

nents, which highlight the secondary context in which the particular equivalent operates, i.e. its pragmatic potential; and
— Paraphrases of meaning, which differentiate between polysemous values in order to determine the primary context.

Using these four types of meaning discriminators, we have performed a practical investigation into the extent to which each is used in the translation equivalent discrimination for words from the Android UI.

Practical analysis

For the comparison of the three texts in terms of the abovementioned four types of meaning or equivalent discriminators, we selected a list of 30 words based on a group of UI translators' indicating difficulty in finding suitable translations for a word, or for all its different senses. The translators comprised a group of six individuals all working as freelancers in the field of computer-related (including UI) translation. Due to contractual requirements and professional reasons, they asked to remain anonymous. They provided a list of words they have had difficulty translating, with reasons, and from this list we selected 30 words where the difficulty arose from the fact that these words can act in more than one part-of-speech. It should be stated at this point that the translators did not limit their reporting to words from the Android UI as such, and most of these words are also found in the UI of numerous other environments. We have, however, limited this investigation to words which are all found in the Android interface. Furthermore, we have specifically selected those words that are not computer terminology as such, but rather general words mostly indicating actions and their products. This list is provided alphabetically below:

English word	**Part-of-speech**
access	n. & v.
add	n. & v.
archive	n. & v.
bookmark	n. & v.
cache	n. & v.
caption	n. & v.
comment	n. & v.
crash	n. & v.
display	n. & v.
dock	n. & v.
download	n. & v.

draft	n. & v.
edit	n. & v.
flag	n. & v.
install	n. & v.
launch	n. & v.
like	n. & v.
overlay	n. & v.
preview	n. & v.
remote	n. & adj.
request	n. & v.
search	n. & v.
share	n. & v.
sign up	n. & v.
support	n. & v.
tag	n. & v.
text	n. & v.
update	n. & v.
upgrade	n. & v.
upload	n. & v.

The words were looked up manually in each of the three reference works and then scored according to whether they appear in the given text at all, as well as the extent to which their meaning discrimination is presented.

In the case of the *Cyber Dictionary*, it was found that 28 of the 30 words are indeed included in the dictionary (please refer to Appendix A). In 3 cases, only one part-of-speech is indicated; in 11 cases no indication of this kind is provided, and in the remaining 16 cases the parts-of-speech are indicated more comprehensively. Punctuation as a meaning discriminating aid is used in 5 out of the 30 cases and in 11 out of the 30 there is some form of contextual and co-textual guidance. Only in a single case, the somewhat unusually treated "flag", is there a paraphrase of meaning:

> **flag (*sentinel*)** vlag (*'n veranderlike wat 'n toestand verteenwoordig*); merker
>
> **flag** [*v*] merk, vlag

It should be noted that this dictionary is explicitly marked as a dictionary related to this particular subject field, so it is to be expected that the use of punctuation would be relatively low, because these words should all fall within the related field, broadly speaking. Nonetheless, it is quite telling that almost half of the words show inadequate treatment of their part-of-speech possibili-

ties — most basically, this comes down to incomplete assistance to the user. Of course, this dictionary is approximately 10 years old and many of the semantic extensions may be of more recent vintage, but this is merely speculative.

For the Pharos dictionary it must be noted that the discriminating elements were scored both for explicating the computer-related use of the word, for example by marking it with a label such as "rek." (or "comp." in English, indicating that the term is in some way computer-related), and for merely distinguishing between different usages in general, since this distinction could also very often aid the translator in discriminating between different senses, albeit by means of applying intuitional knowledge to entries which do not provide suitable equivalents explicitly. In the Pharos dictionary, every example is included in the text (please refer to Appendix B). 5 words have no indication of part-of-speech, 6 have only one form indicated, and the remaining 19 are treated more comprehensively. Punctuation is used in all but a single case (the word "download", which is arguably an especially subject-specific word and therefore doesn't need as thorough discriminating treatment). 25 out of the 30 words have some form of co-textual or contextual guidance, which would implicitly or explicitly aid the translator even if by process of elimination or substitution. Paraphrases of meaning are not included, this being a bilingual dictionary.

The Microsoft list is, of course, not a dictionary as such. The results page is not structured like a dictionary article; rather, it makes use of columns containing the English word, its Afrikaans equivalent, and a definition:

Microsoft Terminology Collection

Showing 1 - 4 of 4

English	Translation	Definition
archive	argief	A compressed file.
archive	argiveer	To move selected items to another location for long term storage.
Web archive	Webargief	A presentation saved in MHTML format that integrates all supporting information, including graphics and other files, into a single file.
Online Archive	Aanlyn argief	A repository that is separate from the user's mailbox and that allows items to be archived on the server.

Therefore, the punctuation category and the co-text/context category were both excluded in the case of this work. 29 of the 30 words do, however, appear in the list and all of these have definitions included. When scoring for parts-of-speech in the case of this work, the criterion used was whether more than one part-of-speech is included for the user, whether in the form of different equivalents or different paraphrases of meaning, though not explicitly marked as such. Using this guideline, 18 out of the 30 words are treated in such a way that the translator would easily see both forms in the results of a single search query, thus getting essentially the same information as they would get from a part-of-speech label.

Three further points of interest stem from the Microsoft list. Firstly, closer investigation revealed that in 7 out of the 30 cases, the word itself is actually not included in the list on its own in both forms, but found in compounds. This kind of compound formation in this subject field and in this language pair could form the basis of a following investigation, although it falls beyond the scope of this article. Secondly, it became clear that the Microsoft list seemingly requires the greatest amount of intuitional knowledge from the translator, but also yields very comprehensive results. Thirdly, there is the case of the verb form "sign up". When searching for "sign up", only the verb form is found in Afrikaans in the Microsoft list. However, when removing the space and searching instead for "signup", a nominal form is actually found at the end of a seemingly unrelated semicolon-separated list of other words. This kind of contraction indicating the difference between a nominal form and a verbal form is also found in a number of other UI-related words, such as "log in"/"login", and deserves further investigation in a separate study.

Conclusion

What becomes clear from this preliminary investigation, is that in all probability, no single reference work out of the three mentioned would be adequate in terms of the requirements described in the theoretical framework presented above. Some words are not included in one or more of the works. Often, manifold meanings are not included, whether implicitly or explicitly. In a number of cases, the word itself is not included, but only appears as part of a compound or another structure. All of these shortcomings could cause translators difficulty.

Notably, the Microsoft list, which is the one that ostensibly is the "least lexicographical" of the three, seems to be potentially the most useful to the translators if judged in terms of the criteria of indicating, whether implicitly or explicitly, parts-of-speech, as well as meaning paraphrases. However, the user must apply significant intuitional knowledge to unlock this utility. This is not an ideal situation, lexicographically speaking, as indicated in our theoretical framework.

Equally unjustifiable, although the work scores well in our framework, is the fact that the Pharos dictionary is so inconsistent and often borders on confusing in the way it treats entries themselves. It is troubling that the bilingual dictionary that is widely regarded as the most authoritative one for this language pair is suffering from such persistent and serious problems.

Similar issues are also seen in the explicitly marked subject dictionary in our selection, in that 28 out of 30 words are included, but just about half of them are not treated comprehensively or in a way that could aid the translator in discriminating usages or meanings. Considering that this is both a bilingual dictionary and a dictionary for specialised purposes (and here it must be noted that although this may or may not be the intention of the compilers, it is still a

source that translators indicate they turn to for guidance), this situation is quite dire. One could surmise that the fact that this dictionary is almost 10 years old at this point is one of the root causes of its unsatisfying performance, considering the rapid technological advancement inherent in this field. However, it should be reiterated that 28 of the 30 words are actually there. The problem is that they are not always treated in a way that is lexicographically justifiable. The user needs to perform a significant amount of incidental or even intuitional equivalent discrimination — pointed out by Beyer (2009: 5) as problematic.

From a lexicographical standpoint, there is room for improvement in all three of the works referenced. There are numerous cases where the meaning discriminations are either insufficient or completely absent. What becomes clear, is that the print dictionaries, especially the subject dictionary in this case, have trouble keeping up with developments in this highly dynamic and ever-evolving subject field, whilst the electronic resource, although not a traditional dictionary, is best suited to this practical reality. Therefore, in a follow-up article, we will undertake an investigation into which other online resources translators use and whether their degree of adherence to lexicographic principles makes a difference in their usability.

References

Al-Kasimi, A.M. 1977. *Linguistics and Bilingual Dictionaries*. Leyden: E.J. Brill.

Beyer, H.L. 2009. 'n Teoretiese basis vir kontekstualisering in tweetalige woordeboeke. *Lexikos* 19: 1-22.

Beyer, H.L. 2013. Glosse en vertaalkomplemente in tweetalige woordeboeke. Botha, W., P. Mavoungou and D. Nkomo (Eds.). 2013. *Festschrift Rufus H. Gouws.* Stellenbosch: SUN PReSS.

Burkhanov, I. 2004. Requirements for an "Ideal" Bilingual $L_1 \rightarrow L_2$ Translation-oriented Dictionary. *Lexikos* 14: 17-32.

Byrne, J. 2006. *Technical Translation: Usability Strategies for Translating Technical Documentation.* Dordrecht: Springer.

Crafford, M.F. 2005. *Vertalers en hul bronne: Die behoefte aan 'n vertaalwoordeboek met Engels en Afrikaans as behandelde taalpaar*. Unpublished M.A. thesis. Stellenbosch: University of Stellenbosch.

Crystal, D. 1991. *A Dictionary of Linguistics and Phonetics*. Third Edition. Oxford/Cambridge, Mass.: Blackwell Publishers.

De Foglio, F. and H.J. Lubbe. 2002. 'n Praktiese benadering tot die samestelling van 'n tweerigting-vakwoordeboek met Afrikaans en Italiaans as taalpaar. *Lexikos* 12: 117-132.

Du Plessis, M., F. Pheiffer, W. Smith-Müller and J. Luther (Eds.). 2010. *Pharos Afrikaans–Engels/ English–Afrikaans Woordeboek/Dictionary.* Cape Town: Pharos Dictionaries.

Fontenelle, T. 2013. Electronic Dictionaries for Text Processing and Correction Systems. Gouws, R.H., U. Heid, W. Schweickard and H.E. Wiegand (Eds.). 2013. *Dictionaries. An International Encyclopedia of Lexicography: Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography.* HSK 5.4: 1096-1102. Berlin/New York: Walter de Gruyter.

Google Play Developer Console Help. 2015. *Supported Languages.* [Online]. Available: https://support.google.com/googleplay/android-developer/table/4419860?hl=en.

Gouws, R.H. 1992. Kommunikatiewe ekwivalensie in vertalende woordeboeke — 'n probleem vir die vertaler. *South African Journal of Linguistics* 10(1): 37-42.

Gouws, R.H. 2006. Discipline, Dilemmas, Decisions and Data Distribution in the Planning and Compilation of Monolingual Dictionaries. *Lexikos* 16: 84-94.

Iannucci, J.E. 1967. Meaning Discrimination in Bilingual Dictionaries. Householder, F.W. and S. Saporta (Eds.). 1967. *Problems in Lexicography*: 201-216. The Hague: Mouton. (19:2009)

IDC (International Data Corporation). 2015. *Android and iOS Squeeze the Competition, Swelling to 96.3% of the Smartphone Operating System Market for Both 4Q14 and CY14.* [Online]. Available: http://www.idc.com/getdoc.jsp?containerId=prUS25450615.

Kelly, N. and J. Zetzsche. 2012. *Found in Translation. How Language Shapes our Lives and Transforms the World.* New York: Perigee.

Microsoft Language Portal. 2015. [Online]. Available: http://www.microsoft.com/Language/. [Accessed: 15 June 2015]

Potgieter, L. 2011. *Die ontwerp van vertalende vakwoordeboeke met vertalers as teikengebruikers: 'n teoretiese model.* Unpublished PhD dissertation. Stellenbosch: University of Stellenbosch.

Van Zyl, G. 2015. *Digital Tsunami to Hit SA Workplace.* [Online]. Available: http://www.fin24.com/Tech/News/Digital-tsunami-to-hit-SA-workplace-20150319.

Viljoen, H.C. (Ed.). In collaboration with N.F. du Plooy and S. Murray. 2006. *Kuberwoordeboek Afrikaans–Engels, Engels–Afrikaans/Cyber Dictionary Afrikaans–English, English–Afrikaans.* Under the auspices of Vriende van Afrikaans. Pretoria: Protea Boekhuis.

Zgusta, L. 1984. Translational Equivalence in the Bilingual Dictionary. Hartmann, R.R.K (Ed.). 1984. *LEXeter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9–12 September 1983*: 147-154. Lexicographica. Series Maior 1. Tübingen: Max Niemeyer.

Appendix A: Scoring for *Cyber Dictionary*

Item	Included	Part-of-speech	Punctuation	Context/co-text	Paraphrase
access	1	2	0	0	0
add	1	1	0	0	0
archive	1	2	0	0	0
bookmark	1	2	0	0	0
cache	1	2	0	0	0
caption	1	0	0	0	0
comment	1	2	0	1	0
crash	1	0	0	0	0
display	1	2	0	1	0
dock	1	0	0	0	0
download	1	0	0	0	0
draft	1	1	0	1	0
edit	1	0	0	1	0
flag	1	1	1	1	1
install	1	0	0	1	0
launch	1	2	1	1	0
like	0	0	0	0	0
overlay	1	3	0	0	0
preview	1	2	0	0	0
remote	1	0	0	0	0
request	1	2	0	0	0
search	1	2	1	0	0
share	1	2	0	1	0
sign up	0	0	0	0	0
support	1	2	0	0	0
tag	1	2	1	1	0
text	1	0	0	0	0
update	1	2	0	1	0
upgrade	1	2	1	1	0
upload	1	0	0	0	0

Appendix B: Scoring for Pharos Dictionary

Item	Included	Part-of-speech	Punctuation	Context/co-text
access	1	1	1	1
add	1	0	1	1
archive	1	0	1	0
bookmark	1	1	1	0
cache	1	2	1	1
caption	1	1	1	1
comment	1	2	1	1
crash	1	2	1	1
display	1	2	1	1
dock	1	2	1	1
download	1	2	0	1
draft	1	2	1	1
edit	1	1	1	1
flag	1	2	1	1
install	1	2	1	1
launch	1	2	1	1
like	1	2	1	1
overlay	1	2	1	0
preview	1	2	1	0
remote	1	0	1	1
request	1	2	1	1
search	1	2	1	1
share	1	2	1	1
sign up	1	1	1	1
support	1	2	1	1
tag	1	2	1	1
text	1	0	1	1
update	1	0	1	0
upgrade	1	2	1	1
upload	1	1	1	1

Appendix C: Scoring for Microsoft Language Portal

Item	Included	Part-of-speech	Paraphrase
access	1	1	1
add	1	1	1
archive	1	1	1
bookmark	1	0	1
cache	1	1	1
caption	1	0	1
comment	1	0	1
crash	1	1	1
display	1	1	1
dock	1	1	1
download	1	1	1
draft	1	0	1
edit	1	1	1
flag	1	1	1
install	1	1	1
launch	1	0	1
like	0	0	0
overlay	1	0	1
preview	1	0	1
remote	1	0	1
request	1	0	1
search	1	1	1
share	1	1	1
sign up	1	0	1
support	1	0	1
tag	1	1	1
text	1	1	1
update	1	1	1
upgrade	1	1	1
upload	1	1	1

Ekwivalentverhoudings in tweetalige woordeboeke: Implikasies vir die databasis van 'n elektroniese tweetalige woordeboek van Suid-Afrikaanse Gebaretaal en Afrikaans*

Hanelle Fourie Blair, *Buro van die WAT, Stellenbosch, Suid-Afrika* en *Departement Afrikaans en Nederlands, Universiteit Stellenbosch, Stellenbosch, Suid-Afrika (hfourie@sun.ac.za)*

Opsomming: Hierdie artikel is 'n uittreksel uit 'n vollediger teoretiese raamwerk waarin 'n konsepmodel vir 'n elektroniese tweetalige grondslagfasewoordeboek van Suid-Afrikaanse Gebaretaal en Afrikaans voorgestel word, spesifiek vir die De la Bat Skool vir Dowes in Worcester, Suid-Afrika (Fourie 2013). Die doel van die voorgestelde woordeboek is om die leerlinge van die De la Bat Skool elektroniese toegang tot geskrewe woordelyste te gee, met gebare en voorbeelde in gebaretaal wat in videoformaat vertoon word — dus in die vorm van 'n elektroniese woordeboek. 'n Geskrewe woordelys voldoen glad nie in hierdie geval aan die behoeftes van skoolleerders nie en 'n woordeboek wat vir die Dowe teikengebruikers by die skool ontwerp is, sal van veel groter waarde wees in die leerders se omgang met gebaretaal sowel as die geskrewe taal. Sekere komponente van die model kan op die ou end ook deur gebruikers buiten die oorspronklike teikengebruikers toegepas word, sodat dit ook vir die breër Dowe gemeenskap van nut sou kon wees.

Hierdie artikel ondersoek die tipiese ekwivalentverhoudings in tweetalige woordeboeke en die kern van die artikel maak op grond daarvan voorstelle oor hoe die leksikograaf ekwivalentverhoudings tussen gebaretaal en Afrikaans in die voorgestelde woordeboek sou kon hanteer.

Sleutelwoorde: AFRIKAANS, DATABASIS, DOWE TEIKENGEBRUIKERS, EKWIVALENSIE, ELEKTRONIESE WOORDEBOEKE, GEBARETAAL, GEBARETAALWOORDEBOEKE, GLOSSERING, GRONDSLAGFASE, LEKSIKOGRAFIE, TWEETALIG, WOORDEBOEK

Abstract: **Equivalent Relations in Bilingual Dictionaries: Implications for the Database of an Electronic Bilingual Dictionary of South African Sign Language and Afrikaans.** This article is an extract from a more comprehensive theoretical framework in which a concept model for an electronic bilingual foundation phase dictionary of South African Sign Language and Afrikaans is suggested, specifically for the De la Bat School for

* Hierdie artikel is ook aangebied as lesing by die Twintigste Jaarlikse Internasionale Konferensie van die African Association for Lexicography (AFRILEX), Universiteit van KwaZulu-Natal, Durban, Suid-Afrika, 6–8 Julie 2015.

the Deaf in Worcester, South Africa (Fourie 2013). The purpose of the proposed dictionary is to give the learners of the De la Bat School electronic access to written school word lists, with signs and examples in sign language displayed in video format — therefore in the form of an electronic dictionary. A written word list does not meet the needs of school learners in this case and a dictionary that is designed for the Deaf target users at the school will be of far greater value in the learners' communication in both sign language and the written language. Certain components of the model can in fact also be implemented by users other than the original target users, so that it may also be of use to the broader Deaf community.

This article investigates the typical equivalent relations in bilingual dictionaries and on that basis the core of the article makes suggestions about how the lexicographer could treat equivalent relations between sign language and Afrikaans in the proposed dictionary.

Keywords: AFRIKAANS, BILINGUAL, DATABASE, DEAF TARGET USERS, DICTIONARY, ELECTRONIC DICTIONARIES, EQUIVALENCE, FOUNDATION PHASE, GLOSSING, LEXICOGRAPHY, SIGN LANGUAGE, SIGN LANGUAGE DICTIONARIES

1. Inleiding

Manley, Jacobsen en Pedersen (1988) is van mening dat beide eentalige en tweetalige woordeboeke eintlik 'n spul leuens is — eentalige woordeboeke begewe hulle op die weg om die indruk oor te dra dat woorde "betekenisse" het, eerder as sekere vermoëns om tot betekenisvolle kontekste toe te tree; tweetalige woordeboeke probeer weer aandui dat woorde "ekwivalente" in ander tale het, eerder as sekere verhoudings van gedeeltelike ekwivalensie wat verder gekompliseer word deur die reeks linguistiese en ekstralinguistiese kontekste waartoe die woorde en hul ekwivalente mag toetree (1988: 281). Die stelling is sekerlik ironies of komies bedoel, maar dui wel op die komplekse verhoudings tussen woorde en hul betekenis(se) in konteks in eentalige woordeboeke, en woorde en hul ekwivalente in konteks in tweetalige woordeboeke. Hierdie artikel ondersoek die tipiese ekwivalentverhoudings in tweetalige woordeboeke en die kern van die artikel maak op grond daarvan voorstelle oor hoe die leksikograaf ekwivalentverhoudings tussen gebaretaal en Afrikaans in die voorgestelde woordeboek sou kon hanteer. Daar is nie 'n brontaal of doeltaal *per se* in die voorgestelde woordeboek nie omdat die bewerkingseenhede parallel in albei tale gebied word, maar die ekwivalentverhoudings tussen die twee tale is steeds belangrik en sal veral 'n impak hê op hoe gebare en woorde in die databasis hanteer word.

2. Ekwivalentverhoudings in tweetalige woordeboeke

Die semantiese kommentaar in 'n ("gewone") tweetalige woordeboek moet ook voorsiening maak vir 'n behandelingsprosedure wat al die polisemiese betekenisonderskeidinge in sy bestek het. Polisemie is 'n taalspesifieke eienskap wat impliseer dat vir 'n polisemiese woord in die brontaal daar nie noodwendig 'n

vertaalekwivalent in die doeltaal gevind sal word wat presies dieselfde polisemiese betekenisonderskeidinge het nie (Gouws en Prinsloo 2005: 151, 152). Dit kom dus daarop neer dat die leksikograaf dikwels 'n aparte vertaalekwivalent vir elk van die polisemiese betekenisonderskeidinge van 'n lemma moet verskaf. Die versameling vertaalekwivalente, hetsy een of meer as een, wat in die semantiese kommentaar van 'n enkele artikel aangebied word, staan bekend as die vertaalekwivalentparadigma van die gegewe artikel. Die leksikograaf moet verseker dat die teikengebruiker van 'n bepaalde woordeboek inligting suksesvol uit 'n vertaalekwivalentparadigma kan bekom. Die leksikograaf mag in die geval waar daar meer as een vertaalekwivalent in die paradigma verskyn, nie op die intuïsie van die gebruiker staatmaak om die korrekte keuse(s) te maak nie (Gouws en Prinsloo 2005: 152).

Gebruikers wat 'n tweetalige woordeboek raadpleeg, verwag dikwels dat hierdie tipe woordeboek data oor die betekenis van die brontaalitem aanbied, soos wat deur die lemma verteenwoordig word, en sien die vertaalekwivalente as die "betekenis van die brontaalwoord in die ander taal". Hulle besef selde dat die data wat aangebied word, nie 'n betekenisparafrase is nie maar eerder 'n lys vertaalekwivalente (Gouws en Prinsloo 2005: 153). Vertaalekwivalente mag egter nie gesien word as inskrywings wat die betekenis van die lemma gee nie, maar as leksikale items in die doeltaal wat gebruik mag word om die brontaalitem in 'n spesifieke situasie te vervang. Die leksikografiese behandeling in 'n tweetalige woordeboek mag dus nie vertaalekwivalente van hul tipiese kontekste en kotekste isoleer nie, soos die volgende voorbeeld uit die *Groot Woordeboek* (Kritzinger et al. 1972) toon:

vrug, (-te), fruit; result, effect; embryo, foetus ...

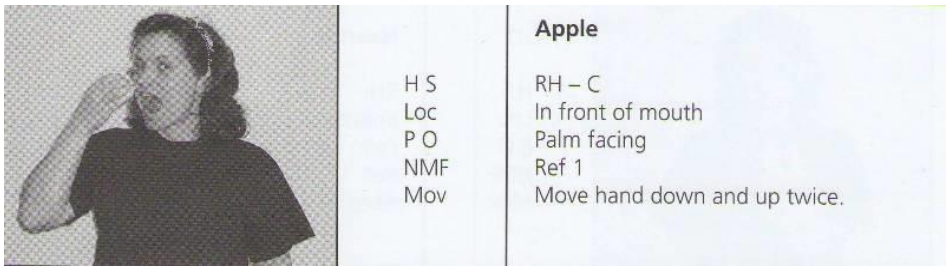
Vanuit 'n semantiese perspektief is dit verkeerd om te redeneer dat enige van die ekwivalente soos *fruit*, *result*, *effect, embryo* en *foetus* as die betekenis van die woord *vrug* gesien kan word; *vrug* **beteken** nie *effect* nie, maar kan in 'n spesifieke konteks met die woord *effect* **vertaal** word.

3. Verskillende tipes ekwivalentverhoudings

Leksikograwe het 'n plig teenoor hul gebruikers om 'n aanbieding en behandeling van vertaalekwivalente te verseker wat 'n ondubbelsinnige verkryging van inligting moontlik te maak uit die data wat in die semantiese kommentaar van 'n tweetalige woordeboek aangebied word. Die behoorlike aanbieding en behandeling van vertaalekwivalente voorvereis 'n deeglike begrip van die verskillende tipes ekwivalentverhoudings. Juis vanweë die feit dat so baie woordeboekartikels nie 'n een-tot-een-verhouding tussen die brontaal- en doeltaalitems toon nie, maak dit vir die leksikograaf noodsaaklik om addisionele inskrywings as ondersteunende materiaal aan te bied sodat die gebruiker 'n ingeligte besluit kan maak wanneer 'n vertaalekwivalent vir 'n gegewe geval van die brontaalitem gekies word (Gouws en Prinsloo 2005: 154).

3.1 Volledige ekwivalensie

Volledige ekwivalensie heers waar daar 'n een-tot-een-verhouding bestaan tussen die brontaalitem, soos verteenwoordig deur die lemma, en doeltaalitem, soos verteenwoordig deur die vertaalekwivalent, op beide die leksikale en semantiese vlak. Hierdie tipe volledige ekwivalensie staan ook as kongruensie bekend: die brontaal- en doeltaalitems het presies dieselfde betekenis, funksioneer op dieselfde sintaktiese vlak en verteenwoordig dieselfde register, wat impliseer dat die doeltaalitem sonder enige beperkings as 'n vertaalekwivalent van die brontaalitem gebruik kan word (Gouws en Prinsloo 2005: 154):



Figuur 1: Die gebaar APPEL, wat in kongruensie met die woord "appel" is (Howard 2008: 103)

Volledige ekwivalensie hoef nie noodwendig 'n een-tot-een-verhouding tussen bron- en doeltaal te verg nie, aangesien die doeltaal ook twee absolute sinonieme as ekwivalente vir 'n gegewe brontaalitem kan hê. Dit is dan nie meer 'n verhouding van kongruensie nie, maar eerder 'n subtipe divergensie (Gouws en Prinsloo 2005: 155), soos in die onderstaande bespreek sal word.

3.2 Gedeeltelike ekwivalensie

Gedeeltelike ekwivalensie heers waar daar nie 'n een-tot-een-verhouding tussen bron- en doeltaalitems bestaan nie. Dit kan op die leksikale vlak of die semantiese vlak of op albei leksikale en semantiese vlakke voorkom. Dit is ook van toepassing indien 'n een-tot-een-verhouding tussen die bron- en doeltaal slegs op leksikale vlak bestaan maar nie op semantiese vlak nie, soos wanneer 'n gegewe doeltaalitem die enigste vertaalekwivalent van die brontaalitem is maar die twee items nie presies dieselfde betekenis gemeen het nie. Die rede hiervoor kan wees dat die brontaalitem monosemies is waar die doeltaalitem polisemies mag wees en as 'n ekwivalent vir die brontaalitem funksioneer omdat een van sy betekenisonderskeidinge met die enigste betekenis van die brontaalitem oorvleuel (Gouws en Prinsloo 2005: 155).

'n Tweede rede vir gedeeltelike ekwivalensie is wanneer 'n een-tot-een-verhouding op die leksikale vlak nie 'n parallelle ooreenkoms op die seman-

tiese vlak het nie — die bron- en doeltaalitems het dus wel dieselfde betekenis maar val nie in dieselfde register nie, wat impliseer dat die doeltaalitem nie die brontaalitem in elke geval of situasie kan vervang nie aangesien dit vir 'n spesifieke register gemerk is (Gouws en Prinsloo 2005: 155, 156).

3.2.1 Divergensie

'n Een-tot-meer-as-een-verhouding tussen items in die bron- en doeltaal staan bekend as divergensie en impliseer nie noodwendig gedeeltelike ekwivalensie nie, maar kan ontstaan wanneer die doeltaal twee sinonieme as vertaalekwivalente vir die taalitem in die brontaal het. Hierdie een-tot-meer-as-een-verhouding kan 'n subkategorie van volledige ekwivalensie verteenwoordig, alhoewel die meerderheid van sulke gevalle 'n vorm van gedeeltelike ekwivalensie verteenwoordig (Gouws en Prinsloo 2005: 156). In 'n artikel wat 'n ekwivalente verhouding van divergensie vertoon, kan twee subtipes onderskei word, nl. leksikale divergensie en semantiese divergensie.

Leksikale divergensie heers wanneer 'n monosemiese leksikale brontaalitem wat as lemma funksioneer, een of meer vertaalekwivalente het wat gewoonlik gedeeltelike sinonieme in die doeltaal is en daarom 'n verhouding van gedeeltelike ekwivalensie toon. Indien die ekwivalente egter absolute of volkome sinonieme is, stel hulle 'n verhouding van volledige ekwivalensie daar. Leksikale divergensie word gewoonlik aangedui d.m.v. 'n komma as nietipografiese struktuurmerker wat gebruik word om hierdie ekwivalente te skei. Die voorbeeld *grootliks* uit die *Groot Woordeboek* (Kritzinger et al. 1972) is so 'n voorbeeld en het drie vertaalekwivalente:

groot'liks, greatly, to a great extent, to a high degree.

Die woorde *greatly, to a great extent* en *to a high degree* kan mekaar dus in enige vertaling van die woord *grootliks* vervang. Die meeste gevalle van leksikale divergensie toon ekwivalente wat gedeeltelike sinonieme is en die gebruiker behoort daarop attent gemaak te word dat alhoewel die gedeeltelike sinonieme die brontaalitem kan vervang, hulle dit nie in alle situasies kan doen nie. Die leksikograaf moet dus een of ander kontekstuele of kotekstuele leiding verskaf om die tipiese omgewing aan te dui waarin die gemeenskaplike semantiese waarde van die vertaalekwivalente geaktiveer word. Leksikale divergensie vereis nie noodwendig 'n ingewikkelde sisteem om die beste toegang tot inligting te verseker nie, maar benodig wel die konsekwente toepassing van 'n weldeurdagte sisteem (Gouws en Prinsloo 2005: 156, 157).

Semantiese divergensie word beskou as die mees tipiese verskynsel van gedeeltelike ekwivalensie en heers waar die lemmateken 'n polisemiese leksikale item verteenwoordig. Polisemie is taalspesifiek en die kanse is daarom skraal dat 'n doeltaalitem dieselfde semantiese lading as die polisemiese brontaalitem sal hê. Die semantiese kommentaar bevat daarom 'n semantiese subkommentaar vir elk van die polisemiese betekenisonderskeidinge van die bron-

taalvorm. Deur polisemiese betekenisonderskeidinge met 'n poliseemmerker soos 'n syfer aan te dui, vorm die poliseemmerkers deel van die kitstoegang-struktuur wat die gebruiker op sy soekroete na die gewenste inskrywing lei. Die gebruik van nommers as poliseemmerkers is 'n goed gevestigde en gestandaardiseerde leksikografiese gebruik wat in eentalige sowel as sommige tweetalige woordeboeke gebruik word. Gedrukte tweetalige woordeboeke maak ook dikwels gebruik van kommas en kommapunte as struktuurmerkers om 'n verhouding van semantiese divergensie aan te toon (Gouws en Prinsloo 2005: 157). In 'n elektroniese woordeboek, waar spasie nie dieselfde probleem as in 'n gedrukte woordeboek is nie, sou die gebruik van nommers eerder as kommas en kommapunte waarskynlik 'n beter en meer eksplisiete merker van semantiese divergensie wees. Soos wat in die bespreking hieronder in paragraaf 5 aangedui word, sou die grondslagfasewoordeboek kon kies om selfs polisemiese of homonimiese brontaalitems op 'n eenvoudige 'monosemiese' wyse per betekenisonderskeiding aan te bied, soortgelyk aan ander beginnerwoordeboeke, wat die gebruik van struktuurmerkers om semantiese divergensie aan te dui, in hierdie spesifieke geval oorbodig maak. Voorstelle oor die uitleg van die mikrostruktuur sal in in 'n volgende publikasie in meer besonderhede bespreek word.

3.3 Nul-ekwivalensie

Nul-ekwivalensie heers wanneer die doeltaal nie oor 'n item beskik wat met 'n lemma wat die brontaalitem verteenwoordig, gekoördineer kan word nie. Die afwesigheid van woorde in 'n vergelyking tussen twee tale staan as leksikale gapings bekend en is taalspesifiek. Dagut (1981) identifiseer twee soorte semantiese gapings in die oordrag van 'n teks van een taal na 'n ander: gapings wat ontstaan vanweë taalkundige faktore en gapings wat ontstaan vanweë ekstralinguistiese faktore. Hierdie twee kategorieë staan onderskeidelik as linguistiese en referensiële gapings bekend (Gouws en Prinsloo 2005: 158, 159). 'n Linguistiese gaping kan geïdentifiseer word wanneer sprekers van beide die tale wat vergelyk word, bekend is met 'n sekere konsep maar waar die een taal nie 'n spesifieke woord het om daarna te verwys nie en die ander taal wel. Engelssprekende sowel as Afrikaanssprekende Suid-Afrikaners is byvoorbeeld vertroud met woorde en konsepte soos *bakkie* (beide die voertuig en 'n plastiekhouer waarin iets gebêre word), *kloof*, *koeksister*, *kombi*, *bobotie*, *samoesa*, ens. Hierdie woorde word algemeen in albei tale gebruik hoewel (of dalk juis omdat) Engels nie 'n woord daarvoor het nie. 'n Referensiële gaping ontstaan wanneer die sprekers van taal B glad nie vertroud is met 'n referent of konsep wat in taal A 'n naam het nie, bv. sprekers van 'n Europese taal wat nie weet wat die woord *lobola* beteken nie, al is dit welbekend aan al die sprekers van die Ngunitale. Die brontaalitem verteenwoordig dus 'n referensiële gaping aangesien dit 'n kultuurgebonde leksikale item is en die sprekers van die doeltaal nie daardie kultuur deel nie (Gouws en Prinsloo 2005: 159). Die gebrek aan vertaalekwivalente waar leksikale gapings bestaan, skep 'n verhouding van nul-ekwivalen-

sie en lei dikwels tot die gebruik van **surrogaatekwivalente**, m.a.w. 'n doeltaalinskrywing wat 'n vertaalekwivalent vervang. Die aard van die leksikale gaping in die doeltaal sal ook die omvang bepaal van die leiding wat die leksikograaf in die behandeling van die doeltaalitem gee: in die geval van 'n linguistiese gaping weet die leksikograaf dat die sprekers van die doeltaal ook bekend is met die spesifieke konsep in die brontaal en 'n kort beskrywing of omskrywing van betekenis sal voldoende wees; in die geval van 'n referensiële gaping is 'n meer omvattende behandeling nodig aangesien die gebruikers van die doeltaal nie bekend is met die konsep wat deur die brontaallemma voorgestel word nie (Gouws en Prinsloo 2005: 159, 160). Artikels waarin nul-ekwivalensie heers, is die ideale kandidate vir die insluiting van prentjies of illustrasies in woordeboeke wat van prentjies/illustrasies in die mikrostruktuur gebruik maak — 'n prentjie sê immers meer as 'n duisend woorde.

3.4 Konvergensie, of omgekeerde ekwivalensie

Konvergensie ontstaan wanneer die konsep wat onderliggend aan die een taal (as doeltaal) se uitdrukking of lemma is, dus minder onderskeidende eienskappe het as dié wat onderliggend aan die ander taal (as brontaal) se lemmas is. Daar ontstaan gevolglik konvergensie van die brontaal na die doeltaal wanneer 'n enkele doeltaalitem ooreenkom met die som van die betekenis van twee (of meer) brontaalitems (Svensén 2009: 259):

Figuur 2: Konvergensie van die brontaal na die doeltaal (Svensén 2009: 259)

3.5 Kommunikatiewe ekwivalensie

Die funksie(s) van 'n spesifieke woordeboek, bv. 'n teksresepsie- of 'n teksproduksiefunksie, behoort 'n invloed te hê op die aard en omvang van die addisionele inskrywings wat gegee word om die vertaalekwivalente te ondersteun en aan te vul. Die meerderheid tweetalige woordeboeke word as bifunksionele produkte saamgestel en moet dus hul gebruikers in minstens die funksies van teksresepsie of –produksie ondersteun. Tog gee die data wat in die semantiese kommentaar van tweetalige woordeboeke aangebied word, dikwels die gebruiker min geleentheid om die regte ekwivalent te kies — leksikograwe beperk te dikwels hul pogings tot die blote lysting van 'n aantal doeltaalitems om vertaalekwivalensie te verseker. In sulke gevalle ontvang die gebruikers geen leiding om hulle in die keuse van ekwivalente by te staan of om hulle te help om die doeltaalitems op die regte manier te gebruik nie. Dit kan voldoende wees in gevalle van teksresepsie, maar nie vir teksproduksie nie — **kommunikatiewe**

ekwivalensie kan slegs bereik word indien die behandeling van die vertaalekwivalentparadigma nie net tot 'n lysting van ekwivalente beperk word nie, maar as hierdie ekwivalente deur konteks- en koteksinskrywings aangevul word wat die gebruiker kan help om die korrekte ekwivalent vir 'n gegewe geval van die brontaalitem te kies en om hierdie ekwivalent op die regte manier te gebruik (Gouws en Prinsloo 2005: 161, 162). Voorbeelde van vertaalekwivalente wat sonder konteks- of koteksinskrywings gelys word en die gebruiker dus nie met teksproduksie kan bystaan nie, lyk soos volg:

> **form** ... vorm; maak; fatsoeneer; formeer; set; kweek; oprig, stig; in gelid stel, rangskik ... (*Tweetalige Woordeboek*) (Bosman et al. 1984)

> **op'staan, (-ge-),** stand up, rise, arise, get up; get on one's legs; rebel, revolt ... (*Groot Woordeboek*) (Kritzinger et al. 1972)

Beide hierdie artikels toon **polidivergensie** en het dus 'n kombinasie van leksikale en semantiese divergensie. Semantiese ekwivalensie is wel tussen die lemmas en hul onderskeidelike vertaalekwivalentparadigmas vasgestel, maar nie kommunikatiewe ekwivalensie nie (Gouws en Prinsloo 2005: 162). 'n Oplossing vir hierdie probleem is te vinde in die voorsiening van bykomende inskrywings sodat ruimte geskep word vir nielemmatiese adressering met die vertaalekwivalente as adresse van die konteks- en koteksinskrywings, soos in die volgende paragraaf bespreek word.

Die bostaande voorbeelde maak slegs voorsiening vir lemmatiese adressering, alhoewel die adresseringstruktuur van tweetalige woordeboeke vir beide **lemmatiese** en **nielemmatiese adressering** voorsiening kan maak. Nielemmatiese adressering kan die prosedures van lemmatiese adressering aanvul en die leksikograaf in staat stel om die gebruiker met teksproduksie te help. Nielemmatiese adressering impliseer dat die lemma nie die enigste behandelingseenheid in 'n woordeboekartikel behoort te wees nie maar dat die status van vertaalekwivalente as doeltaalitems verhoog moet word om ook behandelingseenhede te wees. Addisionele inskrywings kan dus by die vertaalekwivalente gevoeg word om die gebruiker te help om hulle beter te verstaan en om 'n idee te vorm van watter ekwivalent die brontaalitem in 'n spesifieke geval kan vervang (Gouws en Prinsloo 2005: 162). 'n Voorbeeld van nielemmatiese adressering wat kommunikatiewe ekwivalensie help bereik, kan in die volgende voorbeeld uit die *Tweetalige Aanleerderswoordeboek* (Du Plessis 1993) gesien word:

> **figure 1** figuur **[a]** *Sy het ′n mooi **figuur** en lieflike lang bene*. She has a beautiful **figure** and lovely long legs. **[b]** *Die skoolhoof is ′n belangrike **figuur** op ′n klein dorpie*. The headmaster is an important **figure** in a small village. **2** syfer "*Ek kan nie dié syfer uitmaak nie. Is dit ′n 1 of ′n 7?*" "I can't make out this figure. Is it a 1 or a 7?"...

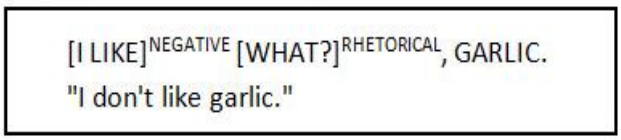
In hierdie behandeling van die polisemiese Engelse woord *figure* is die verskillende semantiese subkommentare met nommers as struktuurmerkers gemerk,

wat uitdruklik aandui dat die lemma 'n polisemiese leksikale item verteenwoordig. Die geïntegreerde mikrostruktuur laat elke vertaalekwivalent toe om dadelik deur 'n beskrywende voorbeeld gevolg te word wat aan hierdie vertaalekwivalent geadresseer is, en dui dus aan hoe die doeltaalitem gebruik word om die spesifieke polisemiese betekenisonderskeiding van die lemma as brontaalitem uit te druk (Gouws en Prinsloo 2005: 162). Konteksinskrywings kan ook in die toepassing van nielemmatiese adresseringsprosedures gebruik word om kommunikatiewe ekwivalensie te bevorder. In die meeste gedrukte woordeboeke word van woorde tussen hakies as konteksinskrywings gebruik gemaak, soos in die volgende voorbeeld uit die *Oxford French Dictionary and Grammar* (OFD) (Rowlinson 2001) gesien kan word:

> **line** ... ligne ...; (track) voie ...; (wrinkle) ride ...; (row) rangée ..., file ...; (of poem) vers ...; (rope) corde ...; (of goods) gamme ...; (queue: US) queue ...

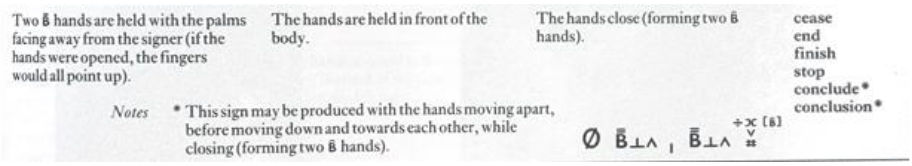
4. Glosse en ekwivalensie in gebaretaalwoordeboeke

Gebaretale word tipies gebaar-vir-gebaar in die geskrewe taal getranskribeer d.m.v. 'n geskrewe glos wat in hoofletters geskryf word. Prosodie word dikwels in boskrif geglos met die omvang daarvan tussen hakies aangedui:



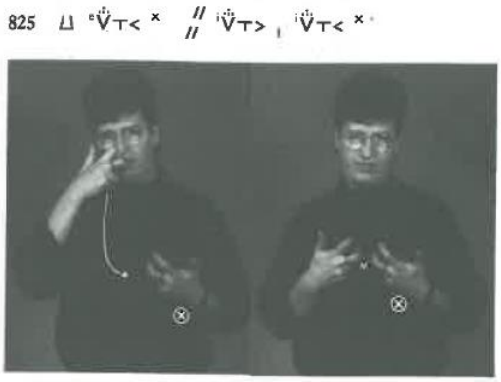
Figuur 3: Voorbeeld van die glossering van 'n gebare-uiting d.m.v. geskrewe woorde en prosodiese inligting in boskrif gevolg deur 'n vertaling van die uiting in idiomatiese Engels (Wikipedia 2012)

Primêre glosse word gewoonlik beskou as die naaste vertaalekwivalente uit die geskrewe taal vir 'n spesifieke gebaar, terwyl sekondêre glosse beskou word as enige ander betekenisse wat aan die gebaar toegeken sou kon word in die geval van enkele of minimale veranderinge aan die gebaar, soos 'n verandering in handpalmoriëntasie of beweging:



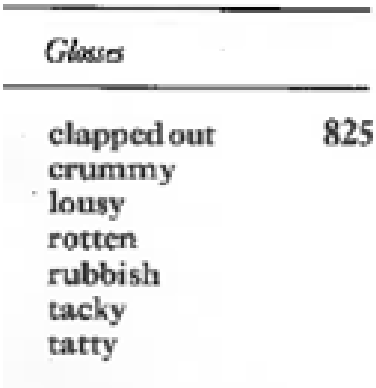
Figuur 4: Voorbeeld van die gebruik van sekondêre glossering in die geval waar 'n verandering in die beweging van die gebaar 'n sekondêre betekenis kan meebring (Brien 1992: 749)

In die tradisie van gebaretaalnavorsing word die woord **glos** gebruik om na die transkripsie van die betekenis van 'n gebaar te verwys (lekewoordeboekmakers maak ook dikwels van hierdie term gebruik, waarskynlik uit onkunde, maar dit is nie nou hier ter sprake nie). Leksikograwe van gesproke/geskrewe tale vind dié gebruik egter vreemd en verkies om van 'n vertaalekwivalent te praat. Dit is egter belangrik om te onthou dat gebaretale nie sonder meer met gesproke tale vergelyk kan word nie, al betwis niemand die feit dat gebaretale nié om daardie of enige ander rede minderwaardige tale is nie. Hulle is eerder *unieke* tale wat in sekere opsigte op unieke maniere funksioneer. Om 'n gebaretaal in terme van 'n gesproke taal te beskryf, is in baie opsigte problematies, aangesien dit maklik is om sekere kenmerke of beperkinge van die gesproke taal (in vergelyking met die gebaretaal) op die gebaretaal te projekteer. Geen gebaretaalleksikograaf het 'n maklike taak wanneer besluite geneem moet word t.o.v. leksikalisering, aanhalingsvorme en selfs woordsoorte nie, omdat (i) daar nie 'n bepaalde tradisie of voorskrif hiervoor (kan) bestaan nie (ii) aangesien gebaretale hulle nie summier in dieselfde kassies as gesproke tale laat groepeer nie — gebaretale "is prone to synthesising localised meanings that are not strictly lexicalised" (McKee en McKee 2012). Wanneer 'n gebaretaal met 'n gesproke/geskrewe taal in 'n tweetalige woordeboek vergelyk of gekontrasteer word, gebeur dit baie selde dat 'n enkele gebaar-tot-woord-ekwivalent in die geskrewe taal gevind kan word. Dit is selfs gevaarlik en misleidend om die indruk te skep dat so 'n gebaar-tot-woord- of woord-tot-gebaar-verhouding sonder meer bestaan. Feitlik alle gebaretaalwoordeboeke het te make met hierdie unieke sintetiserende eienskap van gebaretale en die enigste algemene hantering hiervan is om die moontlike betekenisse van 'n gebaar d.m.v. glosse aan te dui. Neem die onderstaande voorbeeld uit die *Dictionary of British Sign Language/English* (Brien 1992: 456-7):



Figuur 5: Brien 1992: 456

'n Uittreksel uit die woordeboekartikel vir dié gebaar lyk soos volg:


Glosses

clapped out 825
crummy
lousy
rotten
rubbish
tacky
tatty

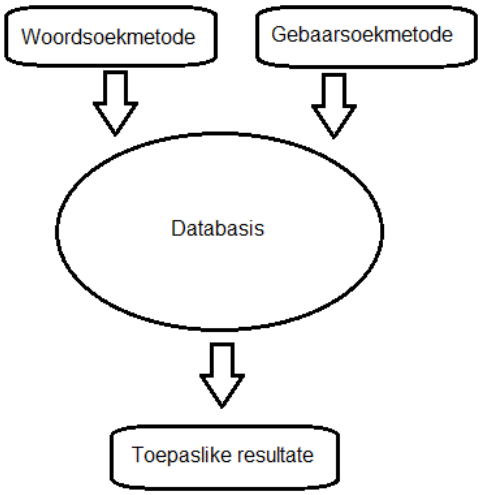
Figuur 6: Brien 1992: 457

Onder die opskrif "Glosses" word 'n lys van die gebaar se moontlike "vertaalekwivalente" as glosse aangedui, nl. "clapped out", "crummy", "lousy", "rotten", "rubbish", "tacky" en "tatty". In die gebruiksaanwysings van hierdie woordeboek word die volgende oor die glosse gesê: "the glosses column contains a number of entries that *can be thought of as possible parallels to the sign* given opposite" (xv) (eie beklemtoning). Dit is duidelik dat die woordeboek nie probeer om aan te voer dat die inskrywings in die glossekolom die vertaalekwivalente van die betrokke gebaar is nie, maar dat hulle "moontlike parallelle" is in 'n taal waar die lyne van grammatiese sinsontleding en beskrywing soms 'n bietjie wasig is in vergelyking met dié van gesproke/ geskrewe tale.

5. Die impak van ekwivalentverhoudings op die databasis en voorstelle vir die woordeboek

Die ekwivalentverhoudings tussen gebaretaal en Afrikaans het sekere implikasies vir die manier(e) waarop gebare en woorde in die databasis hanteer sal word. Beide gebare asook geskrewe woorde is parallelle lemmakandidate wat lede van die databasis is en daarin bewerk word.

Die voorgestelde woordeboek is gebaseer op een elektroniese databasis van gebare én geskrewe woorde wat d.m.v. 'n elektroniese soekprosedure (waarskynlik via 'n webblaaier) toegang tot 'n gebaar parallel aan 'n geskrewe woord sal verleen sodat beide gebare én geskrewe woorde soekbaar is en die woordeboek volledig **tweetalig** is.



Figuur 7: Beide die woordsoekmetode en die gebaarsoekmetode lei tot toepaslike resultate

'n Databasis van gebare word saamgestel wat bestaan uit (a) gebare waarvoor die **woorde** in die geskrewe klaswoordelyste vertaalekwivalente sou wees, en (b) enige bykomende **gebare** wat deur 'n raadgewende Dowe gebaretaalkomitee gekies of aanbeveel word tydens die uiteindelike praktiese samestelling van die voorgestelde woordeboek. Gebare en woorde word dus per *betekeniswaarde* gedupliseer en onderling as gidselemente aan mekaar gekoppel.

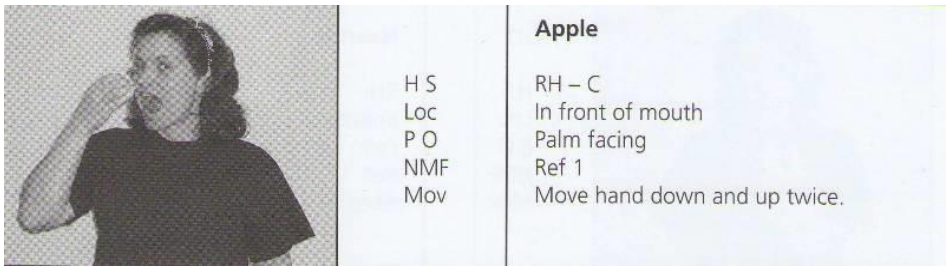
Die woorde wat in die voorgestelde woordeboek as geskrewe lemmas aangebied word, behoort ooreen te stem met die doeltaalvlak van die gebruiker en sal dus nie noodwendig alle moontlike glosse van die gebaar (sien paragraaf 4) as lemmas in die geskrewe taal insluit nie.

'n Soekopdrag gaan nie noodwendig dadelik die volle mikrostrukturele bewerking van 'n bepaalde meervoudige lemma oplewer nie maar slegs 'n deel van die mikro- en artikelstruktuur. Die data wat in hierdie deelstruktuur gebruik word, kan meebring dat dit in sommige gevalle selfs nie vir die gebruiker nodig sal wees om die soekproses voort te sit deur op een van die soekresultate te klik nie, amper soos wanneer die gebruiker van 'n gedrukte woordeboek 'n artikel vinnig opsoek net om die vorm of spelling te kontroleer en nie enige behoefte daaraan het om verdere inligting aan die artikel te onttrek nie. Hipotetiese voorbeelde van so 'n deelstruktuur word hieronder in figuur 12 en 13 voorgestel. Die volledige mikrostrukturele bewerking van die lemma — of, meer spesifiek, 'n spesifieke *betekeniswaarde* van die lemma — word bekom deur een van die soekresultate te selekteer.

Vanweë die parallelle aanbieding van albei tale en die meervoudige lemmas wat daaruit spruit, kan die verhouding tussen die twee tale ook nie in terme van gewone ekwivalentverhoudings in tweetalige woordeboeke bespreek

word nie. Dit sal wel in terme van een-tot-een-, een-tot-meer- en een-tot-nul-verhoudings bespreek en geïllustreer word.

5.1 Een gebaar tot een woord, en omgekeerd



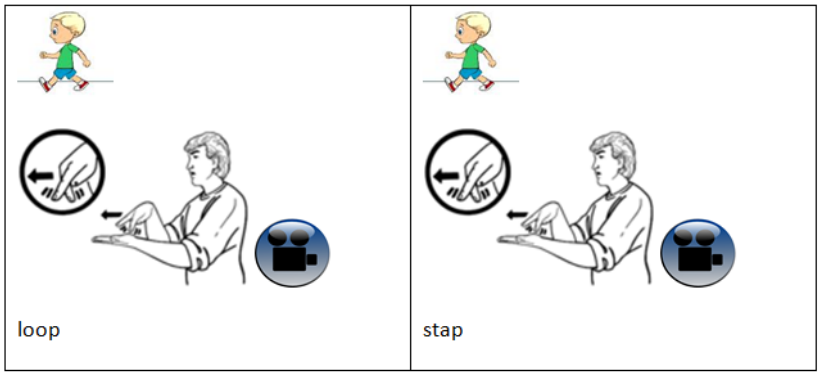
Figuur 8: Die gebaar APPEL, wat in kongruensie met die woord "appel" is (Howard 2008: 103)

In gevalle van 'n een-tot-een-verhouding tussen die twee tale word geïmpliseer dat daar dus net een geskrewe woord aan 'n gebaar (en omgekeerd) in die databasis toegeken of gekoppel sal word en hierdie gevalle behoort geen probleme te veroorsaak nie.

5.2 Een gebaar tot meer as een (sinonieme) Afrikaanse woord

Figuur 9: Die gebaar LOOP/STAP wat, soos uit die glossering afgelei kan word, met die sinonieme *loop* of *stap* vertaal kan word

In die geval van 'n een-tot-meer-verhouding van gebaretaal na Afrikaans, soos in die geval van die gebaar wat met die Afrikaanse sinonieme "loop" en "stap" vertaal kan word, sal beide woorde in die databasis aan dieselfde gebaar toegeken of as gidselemente vir dieselfde gebaar opgeneem moet word sodat 'n soektog na enige van die twee woorde 'n suksesvolle en toepaslike soekresultaat lewer:



Figuur 10: Hipotetiese soekresultaat vir 'n gebaar wat in 'n een-tot-meer-verhouding met die geskrewe taal staan

Let wel: in figuur 10 word die geskrewe woorde nie in hoofletters gegee soos wat gewoonlik in die glossering van gebare gedoen word nie, aangesien die geskrewe woorde nie meer as vertaalekwivalente of glosse beskou word nie maar as **lemmas**. Die gebruik van 'n meervoudige lemma word in 'n komende publikasie bespreek.

5.3 Een gebaar tot meer as een woord

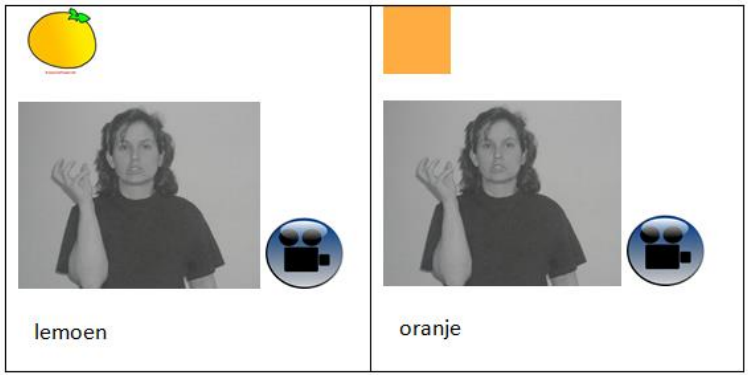
'n Voorbeeld van 'n polisemiese gebaar wat voorkom in die spesifieke dialek of variant van SASL wat by De la Bat gebruik word, is die gebaar wat "lemoen" of "oranje" kan beteken (homonieme of poliseme is egter baie skaars in gebaretaal en hierdie voorbeeld is die enigste een wat by navraag gevind kon word). Hierdie gebaar is besonder gepas in hierdie bespreking aangesien dit geskrewe woorde verteenwoordig wat in die grondslagfasewoordelys verskyn en wat dus ook gelemmatiseer sal word:

Figuur 11: Die polisemiese gebaar wat "oranje" of "lemoen" kan beteken, soos dit by De la Bat gebruik word (Howard 2008: 104)

Die deelstruktuur van die soekresultaat wat deur die gebaarsoekmetode bekom word, sou soos in figuur 12 aangedui word.

Die teikengebruiker van die *Longman Grondslagfasewoordeboek* (Gouws et al. 2010) wat die woord "lig" in die alfabetiese lys opsoek (Gouws et al. 2010: 55, 56) en sien dat daar drie lemmas vir daardie woord aangebied word, nl. 'n selfstandige naamwoord, 'n byvoeglike naamwoord (teenoor donker) en nog 'n byvoeglike naamwoord (teenoor swaar), beskik nie oor die insig dat die woord twee maal homonimies en ook twee maal polisemies is nie — dit word ook nie van die gebruiker verwag nie. Die betekeniswaardes vir "lig" word dus nie soos in algemene tweetalige woordeboeke van getalle in boskrif voorsien om homonieme te onderskei nie, en ook nie van Romeinse syfers om polisemiese betekenisonderskeidings te onderskei nie, maar eerder in 'n heeltemal streng alfabetiese makrostruktuur, sodat elkeen as 'n aparte lemma in die makrostruktuur verskyn.

Die gebare en woorde wat in die databasis is en wat polisemies of homonimies is, sal dus op 'n soortgelyke wyse vir die grondslagfasewoordeboek 'gedupliseer' kon word sodat elke *betekeniswaarde* apart in die soekresultate van onderskeidelik die gebaarsoekmetode en woordsoekmetode aangebied word — trouens, 'n volkome tweetalige woordeboek waarin sowel gebaretaal as 'n geskrewe taal as bron- óf doeltaal kan optree, vereis hierdie tipe bewerking. Die soekresultaat vir *een gebaar waaraan meer as een woord met verskillende betekeniswaardes* toegeken word, sou dus soos in figuur 12 kon lyk (steeds met prentjies as aanduiders om die onderskeid so duidelik moontlik te maak sodat die gebruiker deur visuele leidrade waarvan die betekenis deursigtig is, gelei word):



Figuur 12: Hipotetiese soekresultaat vir 'n gebaar waaraan meer as een woord met verskillende betekeniswaardes in die databasis toegeken word en wat, deur die gebaarsoekmetode te gebruik, as aparte soekresultate vertoon

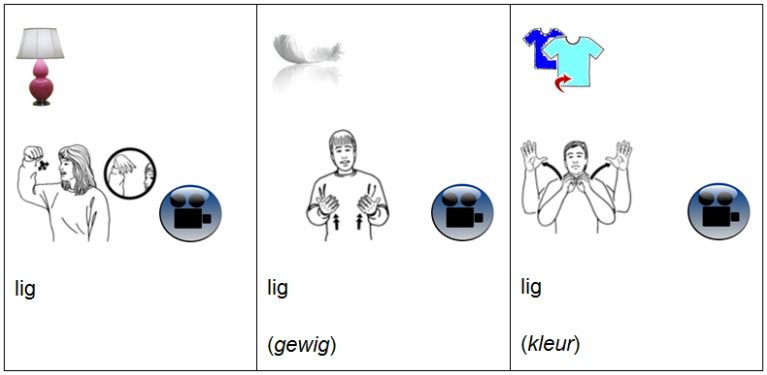
Die herhaling of duplisering van gebare én woorde per betekeniswaarde in die databasis sou ook verseker dat die tematiese soekmetode (waar gebare volgens

referent en dus in een spesifieke betekeniswaarde vertoon word) sowel as die woord- en gebaresoekmetodes gepaste resultate lewer. Dit sou verhoed dat die tematiese kategorieë uit 'n aparte databasis moet bestaan en laat die tematiese soekmetode op 'n uniforme en bevredigende manier met die woord- en gebare-soekmetodes sluit.

5.4 Een woord tot meer as een gebaar

In gevalle van *een woord waaraan meer as een gebaar* toegeken word, sal dit waarskynlik in die meerderheid gevalle wees a.g.v. een woord met meer as een betekeniswaarde, tensy daar natuurlik hoegenaamd gevalle van sinonieme gebare (waarvan geen voorbeelde gevind kon word nie) of variante van 'n gebaar teëgekom word. In die geval van variante sou die meer of mees gebruiklike variant wat op 'n proskriptiewe wyse deur die woordeboek aanbeveel kon word, met 'n etiket gemerk moet word om die voorgestelde keuse aan te dui. Gegewe die beperkte woordeskat wat in die woordeboek op leerders in die grondslagfase gerig sal word, is dit redelik onwaarskynlik dat daar enige gevalle van variasie sal voorkom. Dit sal waarskynlik eers in woordeboeke vir meer gevorderde leerders in die intermediêre en senior fases na vore kom en kan dan op die spesifieke gebruikersvlak hanteer word.

'n Soektog na die polisemiese of homoniese woord sal dus al die gebare wat daardie woord kan "vertaal", oplewer. Die deelstruktuur wat in die soekresultate gereflekteer word, is van die uiterste belang om die gebruiker te help om die regte keuse te maak deur te verseker dat genoeg kontekstuele leiding verskaf word. Die deelstruktuur van sulke soekresultate sou soos volg kon lyk, bv. in die geval van die geskrewe woord "lig" wat beide homonimies en polisemies is:



Figuur 13: Hipotetiese soekresultaat van die woord "lig"

Die uitleg t.o.v. die gebare en geskrewe woorde in figuur 13 is slegs hipoteties en is ter wille van bladsyruimte met die gebaar bo en die geskrewe woord

daaronder gedoen — daar is geen rede hoekom die gebaar en die geskrewe woord as parallelle lemmas in 'n volledig tweetalige woordeboek nie *langs* mekaar geplaas kan word nie. Die feit dat gebare en woorde per betekeniswaarde in die databasis gemerk word, lei tot meervoudige lemmas en die mikrostruktuur behoort ook aan te dui dat daar nie voorkeur aan die een bo die ander gegee word nie.

Soos gesien kan word in figuur 13, sou die deelstruktuur van die soekresultate ook van aanduiders (die prentjies) gebruik kon maak om die referent waarna verwys word, te beklemtoon en te illustreer. Dit sal die gebruiker help om die korrekte lemma te kies en kommunikatiewe ekwivalensie ondersteun.

Soos reeds genoem, bied die deelstruktuur nie net leiding aan die gebruiker in die korrekte keuse van 'n lemma (of die betekeniswaarde van 'n lemma) nie, maar kan dit reeds aan sommige basiese navrae van die gebruiker voldoen, amper soos wanneer die gebruiker van 'n gedrukte woordeboek 'n artikel net vlugtig besoek om die vorm of spelling na te gaan. Watter datatipes ook al as inskrywings vir die deelstruktuur van soekresultate gekies word, behoort só gekies te word dat dit die gebruiker se basiese navrae reeds op hierdie vlak kan beantwoord.

5.5 Een-tot-nul-verhoudings

Daar is nie (baie) inligting beskikbaar oor een-tot-nul-verhoudings van SASL na Afrikaans nie, waarskynlik omdat 'n gebaretaal nog nooit (in Suid-Afrika en slegs selde elders) as brontaal in 'n woordeboek aangebied is nie maar meer dikwels as doeltaal. Daar is dus heelwat meer gerapporteerde gevalle van een-tot-nul-verhoudings van Afrikaans na SASL, en gebruikers kla dikwels dat daar nie 'n gebaar is wat 'n spesifieke konsep of term in Afrikaans kan vertaal nie. Sulke gevalle kom veral in gespesialiseerde woordeskat voor, soos wetenskap, landbou, en dergelike. Daar is selfs binne die Dowe gemeenskap nie ooreenstemming oor die verhouding van een-tot-nul-verhoudings van Afrikaans na SASL nie: sommige Dowes sal aandui dat daar 'n linguistiese gaping bestaan, terwyl ander sal sê dat dit glad nie waar is nie en dat die gebaar eenvoudig nie baie bekend is nie, waarskynlik juis omdat dit tot 'n gespesialiseerde veld van die woordeskat behoort. Aangesien SASL nie geskryf kan word nie, is dit dikwels moeilik om woordeskat (of 'gebareskat') wat dalk in een deel van die land ontwikkel (word) na ander dele van die land te laat versprei. Daar bestaan vermoedelik min of geen gevalle van referensiële gapings tussen SASL en Afrikaans nie, aangesien daar nie dieselfde tipe kultuurverskille ter sprake is wat dikwels tussen die sprekers van ander taalpare voorkom nie — Dowes leef steeds 'binne' die kultuur van die mense rondom hulle. Die mees ooglopende verskille is kultuurspesifieke dinge wat tot die Dowe kultuur behoort, soos 'n deurklokkie wat uit 'n flikkerende lig in die Dowe persoon se huis bestaan eerder as 'n apparaat wat 'n geluid maak soos in die huis van 'n horende persoon. Hierdie referente sou in 'n tematiese kategorie saamgegroe-

peer kon word om die horende gebruiker 'n oorsig van al dié tipe items te gee sonder om individueel daarvoor te gaan soek, veral in gevalle waar die horende gebruiker dalk nie eens daarvan bewus is dat daar sulke verskille bestaan nie. Die gebare vir sulke referente in die Dowe kultuur sou dan ook met die naaste woord in die horende kultuur "vertaal" of daaraan gekoppel kon word, bv. in die geval van 'n gebaar soos LIG FLIKKER wat dui op die visuele sein wat aan 'n Dowe persoon wys dat daar iemand by die deur is, sou die woord "deurklokkie" ook in die databasis aan die gebaar gekoppel kon word sodat 'n woordsoektog na "deurklokkie" ook hierdie gebaar as soekresultaat sal lewer. 'n Skakel tussen sulke artikels en 'n tematiese kategorie van al die toepaslike items uit die Dowe kultuur, of 'n buiteteks oor Dowe kultuur, sou die gebruiker dan kon lei na 'n vollediger oorsig van kultuurspesifieke items.

6. Algemene slotopmerkings

Hierdie artikel het die ekwivalentverhoudings tussen gebaretaal en Afrikaans ondersoek en aangedui watter impak hierdie verhoudings op die hantering van die gebare en woorde in die databasis sou hê. Al het die voorgestelde woordeboek nie 'n spesifieke bron- of doeltaal nie, maar eerder twee tale wat parallel aan mekaar aangebied word, sal die voorgestelde hantering van die ekwivalentverhoudings tussen woorde en gebare in die databasis verseker dat die woordsoekmetode sowel as die gebaarsoekmetode (in aansluiting met die tematiese soekmetode) suksesvol is en die toepaslike resultate lewer.

Bronnelys

Bosman, D.B. et al. (Reds.). 1984[8]. *Tweetalige Woordeboek/Bilingual Dictionary*. Kaapstad: Tafelberg.

Brien, D. (Red.). 1992. *Dictionary of British Sign Language/English*. Londen: Faber&Faber.

Dagut, M. 1981. Semantic Voids as a Problem in the Translation Process. *Poetics Today* 2(4): 61-71.

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

Fourie, H. 2013. *'n Leksikografiese model vir 'n elektroniese tweetalige grondslagfasewoordeboek van Suid-Afrikaanse Gebaretaal en Afrikaans.* Ongepubliseerde D.Litt.-proefskrif. Stellenbosch: Universiteit Stellenbosch.

Gouws, R.H. en D.J. Prinsloo. 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.

Gouws, R.H., L. Potgieter en S. Burgess (Reds.). 2010. *Longman Grondslagfasewoordeboek Afrikaans/Engels*. Kaapstad: Maskew Miller Longman.

Howard, S. 2008. *Finger Talk: South African Sign Language (SASL) Dictionary*. Gillitts: Fulton School for the Deaf.

Kritzinger, M.S.B., P.C. Schoonees en U.J. Cronjé (Reds.). 1972[11]. *Groot Woordeboek Afrikaans–Engels/Major Dictionary English–Afrikaans*.Pretoria: J.L. van Schaik.

Manley, J., J. Jacobsen en V.H. Pedersen. 1988. Telling Lies Efficiently: Terminology and the Microstructure in the Bilingual Dictionary. Hyldgaard-Jensen, K. en A. Zettersten (Reds.).

1988. *Symposium on Lexicography III. Proceedings of the Third International Symposium on Lexicography, May 14–16, 1986, at the University of Copenhagen*: 281-302. Tübingen: Max Niemeyer.

McKee, R. en D. McKee. 2012. *New Zealand Sign Language Online: Insights from the Making of a Modern, Multimedia Dictionary.* Programrede aangebied tydens die 17de jaarlikse Internasionale Afrilex-konferensie van 2 tot 5 Julie 2012 by die Universiteit van Pretoria, Pretoria, Suid-Afrika.

Rowlinson, W. 2001². *Oxford French Dictionary and Grammar*. Oxford/New York: Oxford University Press.

Svensén, B. 2009. *A Handbook of Lexicography: The Theory and Practice of Dictionary-Making.* Cambridge: Cambridge University Press.

Wikipedia. 2012. Glossing Signed Languages. http://en.wikipedia.org/wiki/Gloss#Glossing_signed_languages op 1 Junie 2012.

The Presentation and Treatment of Collocations as Secondary Guiding Elements in Dictionaries

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

Abstract: Collocations represent an item type that is frequently presented and treated in an insufficient way in dictionaries. It is cumbersome for the user to distinguish between collocations and example sentences, and dictionaries often include no additional treatment directed exclusively at collocations. In this contribution the need is shown for the recognition of collocations as an item type. It is suggested that they should be presented in such a way in dictionary articles that they can function as secondary guiding elements. Besides their inclusion in dictionaries collocations also need to receive a treatment that makes provision for, in many instances, at least an example sentence but that could also include, when necessary, stylistic and other labels and even a brief paraphrase of meaning. It is shown how collocations are presented and treated in a few existing dictionaries. These presentations are criticised and proposals are made for an improved presentation. In this regard Wiegand's model for a semi-integrated microstructure is adapted to provide for the presentation of collocations in their own non-integrated search zone. Besides the explicit presentation of single collocations proposals are made for an implicit presentation of complex collocations. From a single item form users should then be able to retrieve different items. Although the discussion primarily focuses on printed dictionaries proposals are also made for the presentation of collocations in online dictionaries.

Keywords: ARTICLE STRUCTURE, COLLOCATION, COMPLEX COLLOCATION, COTEXT, EXAMPLE SENTENCES, INTEGRATED MICROSTRUCTURE, NON-GROUPED ORDERING, SEARCH ZONE, SEMI-INTEGRATED MICROSTRUCTURE, SINGLE COLLOCATION, SUB-ZONE

Opsomming: Die aanbieding en bewerking van kollokasies as sekondêre gidselemente in woordeboeke. Kollokasies is 'n aanduidertipe in woordeboeke waarvan die aanbieding en bewerking dikwels onvoldoende is. Dit is vir die gebruiker moeilik om kollokasies van voorbeeldsinne te onderskei en woordeboeke bied dikwels geen bykomende bewerking wat uitsluitlik op kollokasies gerig is nie. In hierdie bydrae word die behoefte aan 'n volwaardige erkenning van kollokasies as aanduidertipe bepleit. Daar word voorgestel dat die aanbieding in woordeboekartikels sodanig moet wees dat hulle as sekondêre gidselemente kan optree. Naas die opname van kollokasies moet daar ook 'n bewerking gebied word wat vir minstens 'n voorbeeldsin maar waar nodig ook vir styl- of ander etikette en selfs vir 'n bondige betekenisparafrase voorsiening maak. Daar word gewys op die manier waarop kollokasies in enkele bestaande woordeboeke

opgeneem en bewerk word. Kritiek word hierop uitgespreek en voorstelle word aan die hand gedoen vir 'n verbeterede aanbieding. In hierdie verband word Wiegand se model van 'n semi-geïntegreerde mikrostruktuur aangepas om voorsiening te maak vir die aanbieding van kollokasies in 'n eie niegeïntegreerde soeksone. Naas die eksplisiete aanbieding van enkelkollokasies word voorstelle gemaak vir 'n implisiete aanbieding van komplekse kollokasies. Vanuit een aanduidervorm moet gebruikers daartoe in staat wees om verskillende aanduiders te kan onttrek. Die bespreking is primêr op gedrukte woordeboeke gerig, maar voorstelle word ook gemaak vir die aanbieding van kollokasies in aanlynwoordeboeke.

Sleutelwoorde: ARTIKELSTRUKTUUR, ENKELKOLLOKASIE, GEÏNTEGREERDE MIKROSTRUKTUUR, KOLLOKASIE, KOMPLEKSE KOLLOKASIE, KOTEKS, NIEGEGROEPEERDE ORDENING, SEMIGEÏNTEGREERDE MIKROSTRUKTUUR, SOEKSONE, SUBSONE, VOORBEELDSINNE

1. Introduction

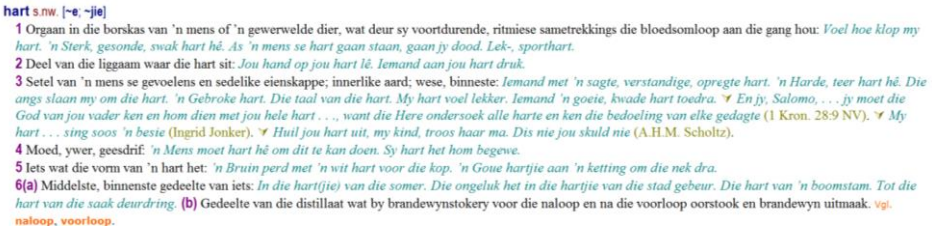
In general language dictionaries with a text production and text reception function lexicographers often include items to illustrate the typical use of the word represented by the lemma sign of a given article. These items from which cotextual information can be retrieved, fall into different categories. Example sentences, the typical citations from corpus material or even some made up by the lexicographer, are one type and collocations are the second type. In many dictionaries both these types of text segments are included in the same search zone, i.e. a general article slot for illustrative material. In this search zone dictionaries too often do not make a distinction between the different types of cotextual items. As a result the user has no clear guidance to assist him/her in identifying collocations and to ensure a successful retrieval of information from entries representing this item type. In printed dictionaries this inferior way of presentation is partially due to a lack of space but also to an insufficient focus on collocations as fully-fledged items in dictionary articles. This insufficient focus also prevails in many online dictionaries. Collocations are significant units in the syntactic patterns of a language, cf. Hausmann (2008). Lexicographers have to negotiate the best possible ways of presenting and treating them in their dictionaries.

For text production purposes, for both mother-tongue and foreign speakers of a given language, collocations have to be regarded as necessary items, and they should be presented in a clearly identifiable way. Their lexicographic significance may never be underestimated. Collocations need to be presented as part of the cotextual guidance of a given lemma. However, their importance exceeds their occurrence as illustrative material of the word represented by the lemma sign. Consequently they should not only be entered as items addressed at the lemma sign but should also be presented as addresses, by functioning as items with their own treatment, albeit of a restricted nature.

In this paper reference is made to aspects of some of the prevailing ways in which collocations are presented and treated. Proposals are then made for a more salient lexicographic presentation and a more comprehensive treatment of collocations, especially by employing evidence from speech corpora.

2. Collocations in existing dictionaries

Unfortunately the presentation and treatment of collocations in many existing dictionaries, both printed and online dictionaries, eschew the extent of the need of users consulting the dictionary to solve a text production problem. These dictionaries often do not enable their users to achieve an unambiguous identification of collocations included in the articles and to retrieve the information they require to use these collocations in a proper way. This can be seen in the following article of the lemma sign *hart* (=heart) in the 5th edition of the Afrikaans monolingual dictionary HAT:

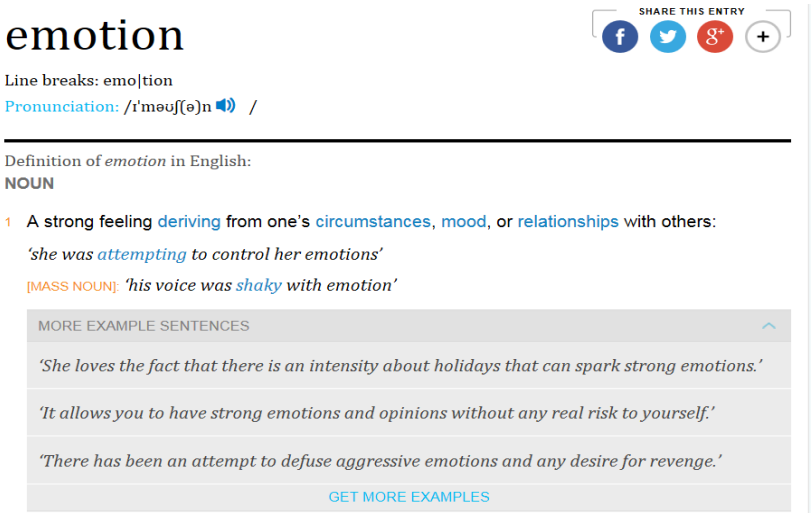


hart s.nw. [~e; ~jie]
1 Orgaan in die borskas van 'n mens of 'n gewerwelde dier, wat deur sy voortdurende, ritmiese sametrekkings die bloedsomloop aan die gang hou: *Voel hoe klop my hart. 'n Sterk, gesonde, swak hart hê. As 'n mens se hart gaan staan, gaan jy dood. Lek-, sporthart.*
2 Deel van die liggaam waar die hart sit: *Jou hand op jou hart lê. Iemand aan jou hart druk.*
3 Setel van 'n mens se gevoelens en sedelike eienskappe; innerlike aard; wese, binneste: *Iemand met 'n sagte, verstandige, opregte hart. 'n Harde, teer hart hê. Die angs slaan my om die hart. 'n Gebroke hart. Die taal van die hart. My hart voel lekker. Iemand 'n goeie, kwade hart toedra.* ▼ *En jy, Salomo, . . . jy moet die God van jou vader ken en hom dien met jou hele hart . . ., want die Here ondersoek alle harte en ken die bedoeling van elke gedagte* (1 Kron. 28:9 NV). ▼ *My hart . . . sing soos 'n beste* (Ingrid Jonker). ▼ *Huil jou hart uit, my kind, troos haar ma. Dis nie jou skuld nie* (A.H.M. Scholtz).
4 Moed, ywer, geesdrif: *'n Mens moet hart hê om dit te kan doen. Sy hart het hom begewe.*
5 Iets wat die vorm van 'n hart het: *'n Bruin perd met 'n wit hart voor die kop. 'n Goue hartjie aan 'n ketting om die nek dra.*
6(a) Middelste, binnenste gedeelte van iets: *In die hart(jie) van die somer. Die ongeluk het in die hartjie van die stad gebeur. Die hart van 'n boomstam. Tot die hart van die saak deurdring.* **(b)** Gedeelte van die distillaat wat by brandewynstokery voor die naloop en na die voorloop oorstook en brandewyn uitmaak. Vgl. **naloop**, **voorloop**.

Example 1 from HAT 5

Although each subcomment on semantics contains a search zone populated by cotextual items, presented in italics, no distinction is made between example sentences and collocations. The user does not know whether a given cotextual item is merely an illustration of the use of the word, e.g. *Jou hand op jou hart lê* (= to put your hand on your heart), in the second subcomment on semantics, or the use of the word as component of a collocation, e.g. *As 'n mens se hart gaan staan, gaan jy dood* (= when your heart stops one dies), in the first subcomment on semantics, where *hart* collocates with *gaan staan* (= to stop). This presentation does not contribute at all to a user's need with regard to collocational guidance because the user does not know whether *gaan staan* only incidentally combines with *heart* or whether it is a frequent combination.

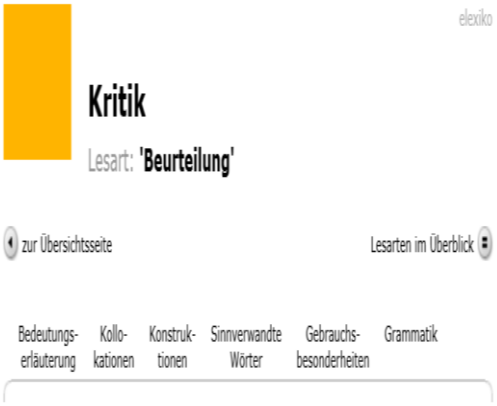
Similar criticism applies to the treatment of the lemma sign *emotion* in the *Oxford Online Dictionaries* where users are not informed that the cotext item *she was attempting to control her emotions* contains the collocation *to control emotions* or that *spark strong emotions* is actually a complex collocation, i.e. a combination of the two collocations *spark emotions* and *strong emotions*.



Example 2 from *Oxford Online Dictionaries*

In both these dictionary articles the collocations are addressed at the lemma but they have no items explicitly addressed at them to elevate them to treatment units in their own right. Users have no idea that these example sentences illustrate the use of both the word represented by the lemma sign and the specific collocation or collocations.

The *elexiko* dictionary, as seen in the following two screenshots from the article of the lemma sign *Kritik* (= criticism), gives users the opportunity to click on the button "Kollokationen" (= collocations). They are then directed to a list of cotext partners:



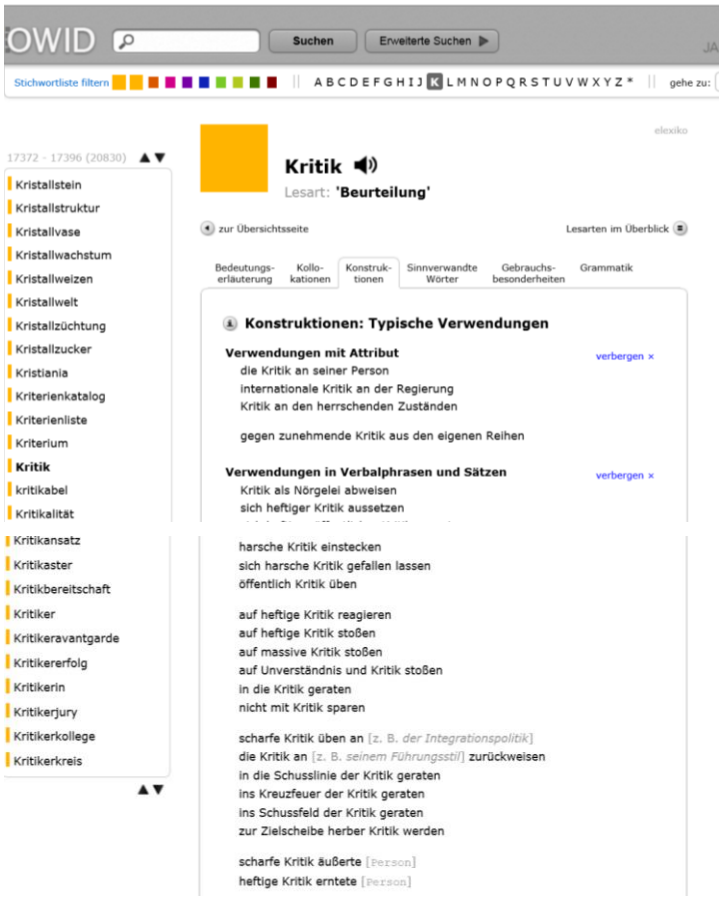
Example 3 from *elexiko*

Example 4 from *elexiko*

Here different collocational contexts are given albeit that the entries given as collocations are not the addresses of items presenting any form of further treatment. Questions could be raised regarding the real collocational status of these entries. A click on the button "Konstruktionen" (= constructions) guides the user to the screenshot, seen in Example 5.

These entries include some real collocations, e.g. *harsche Kritik einstecken, auf heftige Kritik reagieren,* etc. For the typical user it is not that easy to find these collocations.

It is a distressing fact that collocations typically occur in dictionaries exclusively as part of the treatment of the word represented by the lemma sign. Their significance as well as their potential as items that can enhance the dictionary consultation experience and can help to solve the real lexicographic needs of the target users of many dictionaries are eschewed. This might be due to a prevailing lemma-bias and insufficient attempts to present data that can ensure a more comprehensive information retrieval by the dictionary user. It may also among others be due to an insufficient theoretical approach to the selection and treatment of items in dictionaries.



Example 5 from *elexiko*

A much better lexicographic presentation and treatment can be found in the *MacMillan Dictionary*, as seen in the reduced version of the article of the lemma sign *money* in example 6.

Following the paraphrase of meaning a single example sentence is given. This is followed by a text block containing a non-grouped listing of collocations, cf. Bergenholtz, Tarp and Wiegand (1999) for a discussion of the notion of grouped and non-grouped entries, complemented by example sentences addressed at the respective collocations in procedures of non-lemmatic addressing. In this dictionary the collocations are elevated to secondary treatment units and consequently also become secondary guiding elements within the respective dictionary articles. Unfortunately the ordering of the collocations in the text block seems to have been done in a haphazard way.

money - definitions and thesaurus ★★★

NOUN [UNCOUNTABLE] Pronunciation /'mʌni/

what you earn, save, invest and use to pay for things. Money can be kept in a bank, where it can earn interest. If you have a bank account, you can pay for things with a cheque
No, I can't come – I haven't got any money.
make/earn money: *The business has made more money this year.*
spend money (on something): *We've spent a lot of money on this house.*
cost (someone) money: *It would have cost us a lot of money to cancel the event.*
borrow money: *I have had to borrow money from my family.*
save money (=avoid spending money): *You can save money by taking your own lunch.*
save money (=put money somewhere so that you can spend it later): *They're trying to save money so that they can have a holiday.*
money coming in (=money being earned and available to spend): *He had no job so there was no money coming in.*
have money on you (=have money in your pocket etc): *Have you got any money on you?*
lose money (=earn less money than you spend): *The industry is losing money and the government wants to sell it.*
birthday/Christmas money (=money received as a birthday/Christmas present): *I'm going to spend my birthday money on some new clothes.*
raise money (=collect money for a particular purpose): *Her bike ride*

Example 6 from the *MacMillan Dictionary*

An improved approach, compared to the 5th edition, is found in the recently published 6th edition of HAT. Collocations are allocated a separate search zone where they are presented in a grouped text block. Each collocation, given in bold, is preceded by a "|" and is the address of a brief paraphrase of meaning as well as an example sentence.

ma·te *s.nw.* ❖ **in *dié* mate** so erg: *Die getuie het in dié mate gesnik dat die hofverrigtinge tydelik gestaak moes word.* | **'n *groot* mate van** baie: *'n groot mate van onsekerheid.* | **in ('n) *hoë* mate** baie; grotendeels: *Veiligheid op die pad word in hoë mate bepaal deur die toestand van 'n voertuig se bande.* | ***in/tot* 'n mate** 'n bietjie; effens: *Peter het die beoordelaars tot 'n mate beïndruk, maar hulle het op die ou end meer van die ander deelnemer gehou.* | **in *meerdere* of mindere mate** *Ook* **in mindere of meerdere mate** tot 'n sekere mate; in wisselende graad: *Enige medisyne is in meerdere of mindere mate giftig as 'n mens te veel daarvan gebruik.* | ***na* die mate van** in verhouding tot: *belas word na die mate van jou inkomste.* | **in *ruim* mate** ruimskoots. | **'n *sekere* mate van** 'n bietjie: *Ek bespeur 'n sekere mate van jaloesie by jou.* Vgl. MAAT²

Example 7 from HAT 6

3. Guiding elements in dictionary articles

Within the article stretches of a dictionary each individual article is introduced by a lemma sign functioning as guiding element of that specific article, cf. Wiegand and Gouws (2013: 275). Data included as part of the treatment of a given word can be reached via the lemma sign as guiding element of the article. The function of a lemma sign as guiding element is not only restricted to main lemmata but it also applies to sublemmata where these sublemmata guide the users to those entries of which they are the addresses.

In printed dictionaries a consultation process typically proceeds via the lemma sign. However, the lemma is not necessarily the only address, treatment unit or even guiding element in such a dictionary article. Procedures of non-lemmatic addressing often occur, e.g. where a cotext entry is given for a translation equivalent whilst a label directed at a microstructural item also participates in a procedure of non-lemmatic addressing. Where a synonym is given for a translation equivalent in a bilingual dictionary there is a dual addressing procedure with the second equivalent directed at the lemma in a procedure of lemmatic addressing but also at the preceding equivalent in a procedure of non-lemmatic addressing.

In monolingual general language dictionaries items giving the cotext, both example sentences and collocations, typically have a lemmatic address. Where example sentences are included for text production purposes, they help to ensure a better use of the word represented by the lemma sign. Collocations can serve the same purpose but the need for collocations in dictionary articles also goes further than merely assisting the user in the proper use of the single word. For text production purposes users do not only need guidance regarding the choice of collocations but also about the proper use of these collocations. This demands the inclusion of more items addressed at collocations — a procedure that establishes the status of collocations as fully-fledged treatment units.

In printed dictionaries, but even more so in online dictionaries with less space restrictions adaptations to the article structure could lead to the inclusion of search zones that will exclusively accommodate collocations and their treatment, cf. Example 9 in section 5.2.2. Within such a search zone these collocations will be elevated to treatment units and to secondary guiding elements, because the user can access data regarding a collocation via that collocation as an entry in a dictionary article. In an online dictionary the access process to a collocation needed for text reception purposes does not even have to proceed via the lemma sign but the user can identify a specific collocation as search string and enjoy rapid and direct access to this secondary guiding element — where e.g. syntactic information needed for text production purposes can then be retrieved.

4. Data presented in the search zone for collocations

Different types of items can be included as part of the treatment of collocations.

As in the case of the microstructure of any dictionary lexicographers need to make provision for both an obligatory and an extended search zone structure. This needs to be done in accordance with the needs of the envisaged target user and the functions of the dictionary. In a general language dictionary with text production as a function the obligatory search zone structure for collocations and their treatment could include at least one or more items giving example sentences. These examples may also help to support the text reception function of a dictionary. Relying on a corpus the lexicographer needs to distinguish in the selection of example sentences between those illustrating the single word and those that have the collocation and not only the word represented by the lemma sign as significant component. Example sentences in the default slot for cotext entries as well as in a text block allocated to collocations could support the users in their text production endeavours where collocations have to be employed in actual language use. Currently too few dictionaries assist the user with regard to the proper use of those collocations included as part of the treatment of a given word.

4.1 Labels

Dictionaries have to mark items for stylistic, geographical, chronolectic or other deviations from the default standard language variety of the dictionary. This also applies to collocations as items that are targets of addressing procedures. The corpus used for a given dictionary should also be consulted to determine these different restrictions, including morphosyntactic restrictions, that apply to collocations so that the required lexicographic labels, used elsewhere in the dictionary article, could also be employed. This establishes the use of labels not only as entries directed at lemmata but also at other types of microstructural items. Looking at the possible labelling of collocations it is important to realise that there may be e.g. stylistic differences between the collocation and the single word that is the guiding element of the article and that functions as either the base or the collocate of the specific collocation, e.g. whereas the word *invitation* is stylistically neutral the collocation *extend an invitation* belongs to formal language use. Yet again, as is the case with the retrieval of information with regard to single words, the nature of the texts where collocations are found in a given corpus will assist lexicographers to select and employ the necessary lexicographic labels.

4.2 Semantic guidance

Collocations are not independent lexical items with an own meaning, but the meaning of a collocation is usually a combination of the meanings of the collocating lexical items, e.g. *win a match/lose a match.* The search zone in which collocations are presented in the article of the lemma sign *match* does not have to

include an item giving the paraphrase of meaning of the collocation as item in the obligatory search zone structure because this collocation has a transparent meaning, constituted by the meaning of the collocating words. The retrieval of the meaning of a collocation is not always effortless. Tutin and Grossmann (2002: 8) regard collocations as semi-compositional and distinguish between regular, transparent and opaque collocations, cf. also Heid (2011). Because of these varying degrees of transparency the data on offer from a corpus may suggest to a lexicographer the need for an extended obligatory search zone structure to accommodate some semantic guidance as part of the treatment of a given collocation. In the article of the lemma sign *abandon* the *Oxford Advanced Learner's Dictionary* has the following item giving a paraphrase of meaning of this word:

> 2 .. to leave a thing or place, especially because it is impossible or dangerous to stay.

This dictionary does not have a search zone allocated to collocations but includes collocations embedded into example sentences where they are marked as items by a procedure of typographical enrichment according to which these sentence-internal collocations are given in bold. Within the second subcomment on semantics that accommodates the above-mentioned paraphrase of meaning, the following example sentence is included as an item giving cotextual data. This example sentence is complemented by a brief paraphrase of meaning:

> *He gave the order to* ***abandon ship*** (= to leave the ship because it was sinking).

This example sentence does not only illustrate the use of the single word but also that of the collocation. The lexicographer furthermore realised that the typical user of this dictionary needs additional semantic guidance to ensure a successful comprehension of the collocation *to abandon ship.* This brief paraphrase of meaning is addressed at the embedded collocation, resulting in that collocation becoming an addressed item and therefore also a treatment unit. A procedure of typographical enrichment has been employed to identify a part of an item, i.e. the item giving the example sentence, as an item in its own right, a collocation, and to identify this embedded item as the address of a subsequent item giving a paraphrase of meaning. This presentation and treatment help with the comprehension of the collocation, in a text reception situation, and with the proper word choice in text production situations.

In the above-given example 5 from the *MacMillan Dictionary* collocations are presented in a single text block where they are elevated to secondary treatment units and consequently also become secondary guiding elements within the respective dictionary article. Enriching a single part of a more comprehensive item is therefore not needed in order to display the collocations as fully-fledged items. In the article of the lemma sign *money* some collocations are also

complemented by items giving semantic guidance, e.g.:

save money (= avoid spending money)
money coming in (= money being earned and available to spend).

When collocations are entered as secondary guiding elements, as is the case in this dictionary, the need for a brief paraphrase of meaning addressed at a given collocation may become apparent for some items. Working with collocations as they occur in the actual language as retrieved from a corpus will assist the lexicographer in deciding whether this data type should be included for a specific collocation.

4.3 Syntactic guidance

Finding example sentences that illustrate the typical use of collocations will also help the lexicographer to identify those instances where additional syntactic guidance needs to be given for a specific collocation. Compare in this regard the restricted guidance found in the article of the lemma sign *money* in the *MacMillan Dictionary*:

spend money (on something)
cost (someone) money

Although collocations typically consist of two components, cf. Heid (1994: 228), Tutin (2008: 1445) shows that some collocations like *to pay close attention to* include more than two elements. She regards this as a merging of two or three binary collocations, identified as *collocational chains* by Hausmann (2004) or *collocational clusters* by Spohr (2005). From the different occurrences of a given collocation in a corpus the lexicographer will quite often find the occurrence of the combination of two collocations with a mutual element, cf. Zinsmeister and Heid (2003), Spohr (2005), Heid and Gouws (2006). Such a combination of two collocations with one mutual element can be regarded as a complex collocation, compared to a single collocation which is the default type. In the search zone with a collocation as guiding element these different types of cotext items should be included where applicable, i.e. an item giving the cotext where a single collocation occurs as item form and an item giving the cotext where a complex collocation is included as item form.

The noun *habit* combines with the adjective *bad* to form the collocation *a bad habit*, and also with the verb *break* to form the collocation *to break a habit*. These two single collocations can combine, resulting in the complex collocation *to break a bad habit*. A corpus should be used to find the relevant occurrences of both single and complex collocations; yet again, looking beyond features relevant to a single word as treatment unit.

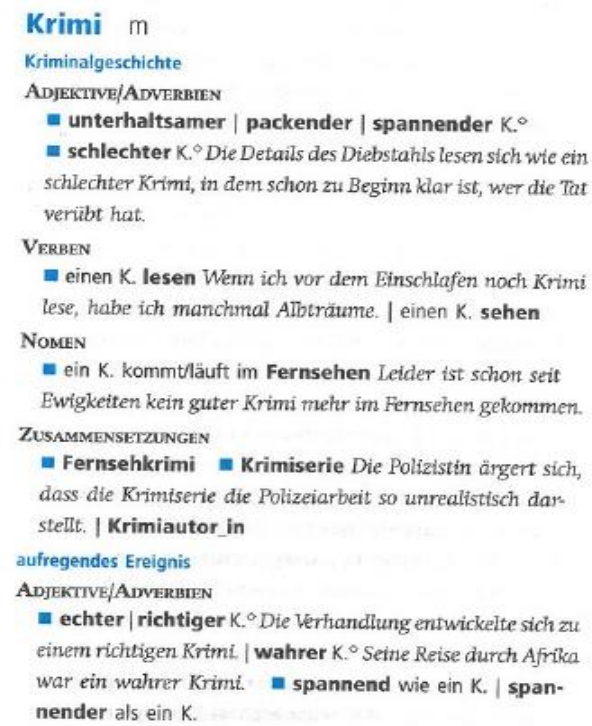
Dictionary users need to be made aware of single and complex collocations as two separate item types and the presentation of the items giving the

cotext of collocations in the dictionary article should enable the user to distinguish between them. This will be discussed in a subsequent section.

5. The presentation of collocations

5.1 Single collocations

The following discussion only focuses on the presentation of collocations in general language dictionaries. A different approach would be employed in a dictionary of collocations. Lexicographers of general language dictionaries should pay attention to the way in which dictionaries of collocations present and treat their collocations because some of these features could also be introduced in general language dictionaries. The following article for the lemma sign *Krimi* (= crime story/film) from *Feste Wortverbindungen des Deutschen* (Häcki Buhofer et al. 2014), a German collocation dictionary, illustrates the occurrence of the word *Krimi* in different collocations, arranged in the article firstly according to the different senses of the word, given in subsequent subcomments on semantics, and within each one of these subcomments on semantics according to the part of speech of the partner word:



Krimi m

Kriminalgeschichte

ADJEKTIVE/ADVERBIEN

■ **unterhaltsamer** | **packender** | **spannender** K.°

■ **schlechter** K.° *Die Details des Diebstahls lesen sich wie ein schlechter Krimi, in dem schon zu Beginn klar ist, wer die Tat verübt hat.*

VERBEN

■ einen K. **lesen** *Wenn ich vor dem Einschlafen noch Krimi lese, habe ich manchmal Albträume.* | einen K. **sehen**

NOMEN

■ ein K. kommt/läuft im **Fernsehen** *Leider ist schon seit Ewigkeiten kein guter Krimi mehr im Fernsehen gekommen.*

ZUSAMMENSETZUNGEN

■ **Fernsehkrimi** ■ **Krimiserie** *Die Polizistin ärgert sich, dass die Krimiserie die Polizeiarbeit so unrealistisch darstellt.* | **Krimiautor_in**

aufregendes Ereignis

ADJEKTIVE/ADVERBIEN

■ **echter** | **richtiger** K.° *Die Verhandlung entwickelte sich zu einem richtigen Krimi.* | **wahrer** K.° *Seine Reise durch Afrika war ein wahrer Krimi.* ■ **spannend** wie ein K. | **spannender** als ein K.

Example 8 from *Feste Wortverbindungen des Deutschen*

This arrangement according to part of speech within each subcomment on semantics resembles the arrangement proposed in Wiegand (1996) for the non-integrated component of a dictionary article where a semi-integrated micro-structure is used, and could be very helpful to satisfy text production needs. Compare in this regard the arrangement suggested by Wiegand (1996: 36) for the article of the lemma sign *Kranz* where the search zones in the non-integrated component, following the 6th subcomment on semantics, are ordered according to the part of speech of the words combining with the word represented by the lemma sign. The structure of this article is relevant to this discussion; not the contents.

Kranz, der; -es; Krän·ze

1. *in der Form eines Rings geflochtene od. gebundene Blumen, Zweige o.ä.* | In der Gärtnerei konnte man Kränze aller Art, z.B. Blumen-, Efeu-, Lorbeer- und Adventskränze bestellen.

2. *etw., das in einer kreisförmigen Form erscheint* | ein ~ von Haaren wuchs rings um seinen kahlen Schädel.

3. *Anzahl von etw. od. jmdm., die um eine Art Mittelpunkt gruppiert sind* | Der Kanzler präsentierte sich mit einem ~ von Ministern vor der Presse.

4. (schweiz.) *einer der ersten drei Plätze bei einem Wettbewerb* | Sie setzten sich nach den Kränzen auf die Ehrentribüne.

5. (Jägerspr.) *Abdruck der äußeren Kante der Hufe von Rotwild auf trockenem Boden* | Er konnte die Fährten deutlich erkennen, war sich aber nicht sicher, ob es sich um Kränze handelt.

6. (Kegeln) *Wurf, bei dem alle acht um den König stehenden Kegel fallen* | Ein Kranz war ihm heute abend nicht gelungen

Sub. ein ¬¹ ***Feigen*** | ein bunter ¬² tanzender ***Kinder*** | ein ¬² junger ***Mädchen*** (geh. *Gruppe, Kreis*) | ein schmackhaftes ***Stück*** ¬² (kurz für ↑Kranzkuchen)

V. einem Mädchen einen ¬¹ ***aufsetzen*** | einen ¬² ***backen*** (*Kranzkuchen*) | einen ¬¹ ***binden*** | einen ¬¹ ***flechten*** | einen ¬¹ ***heraushängen*** (*bes. als Zeichen, daß Wein ausgeschenkt wird*) | einen ¬¹ ***niederlegen*** (*am Grab [Ehrenmal] als Zeichen der Trauer*) | den Hinterbliebenen einen ¬¹ ***schicken*** (*als Zeichen der Anteilnahme*) | dem Sieger einen ¬¹ ***überreichen*** [***aushändigen***] | einen ¬⁶ ***werfen*** | einen ¬¹ ***winden***

Adj. bunter [grüner, verwelkter] ¬¹

Präp. einen ¬¹ ***auf*** das Grab legen | einen ¬¹ ***auf*** dem Kopf tragen | Kerzen ***auf*** einen ¬¹ stecken (*bes. einen Advents~*) | ***in*** die Kränze kommen (schweiz.; umg. *unter den Siegern sein*) | das kommt nicht ***in*** die Kränze (schweiz.; umg. *hat keine Aussicht auf Erfolg [auf Verwirklichung]*) | den Sarg ***mit*** Kränzen (¬¹) schmücken | ein ¬¹ ***mit*** Schleifen [Kerzen] | einen ¬² ***von*** Blüten um den Teller des Ehrengastes legen | ein ¬¹ ***von*** Rosen | ein ¬² ***von*** Seen umringt die Stadt | die Blumen ***zu*** einem ¬¹ winden | das Haar flechten und ***zum*** ¬² aufstecken

Example 9 from Wiegand (1996)

The model used in *Feste Wortverbindungen des Deutschen* and the proposals made in Wiegand (1996) with regard to the article structure could also to be considered by lexicographers of general language dictionaries with regard to the presentation and ordering of collocations. For lexicographers who wish to give more prominence to collocations, especially for text reception purposes, their presentation as clearly identifiable items is non-negotiable. The mere inclusion of a collocation as item form within a more comprehensive item, albeit that the collocation has been identified by a procedure of typographical enriching, will no longer suffice. In a general language dictionary, especially a printed dictionary, a separate subzone for each combination of the given word with partners from different part of speech categories may perhaps seem to be less advisable due to space restrictions, but the text production function of a dictionary may eventually compel lexicographers to distinguish between different categories of collocations and to order the collocations according to these categories. The typical criterion for a distinction between categories of collocations could be the part of speech of the collocates. The typological classification of the collocation, i.e. as single or complex collocation could also play a role. The inclusion and presentation of complex collocations will be discussed in a subsequent section.

As a first step to ensure a successful system where collocations are included in a specific search zone as secondary guiding elements, a thorough planning of a data distribution structure is needed with an article structure that allows the inclusion of a search zone dedicated to collocations. This could be a fully independent search zone dedicated exclusively to the presentation and treatment of collocations or it could be a subsection of a larger search zone reserved for items giving cotextual guidance, but within such a broader search zone the subzones for collocations and for example sentences should be clearly distinguishable. As is the case with example sentences, collocations need to be considered for each subcomment on semantics of an article in which the lemma sign represents a polysemous lexical item. Seeing that the inclusion of collocations will not be compulsory in each dictionary article because all words do not participate in collocational combinations this subzone can be regarded as part of an extended obligatory article structure.

The positioning of the search zone for items giving cotextual guidance, and therefore also of the zone or subzone giving collocations, needs to be determined by the nature of the microstructure opted for in a given dictionary. A semi-integrated microstructure, cf. Wiegand (1996), could be an ideal way of presenting and treating cotextual data, with the example sentences given as items in the integrated part and collocations occupying the text block presented in the non-integrated component. In the example article from Wiegand (1996) the cotextual items in the non-integrated section are arranged according to the part of speech of the words combining with the word represented by the lemma sign. This grammatical feature of the combining words is regarded as more significant than semantic features of the core word, because it determines the presentation of a new category whereas the different senses of the word represented by the lemma sign are not allocated to different semantically deter-

mined categories. Guidance regarding the applicable sense is conveyed by a superscript number given above the tilde functioning as place-keeping symbol to indicate where the word represented by the lemma sign occurs in a given cotextual item. The superscript numbers coincide with the numbers of the subcomments on semantics in the integrated component of the article where paraphrases of meaning for the corresponding senses are accommodated. If a lexicographer wants to adapt the Wiegand example of a semi-integrated microstructure for the inclusion and treatment of collocations, the arrangement in the non-integrated section needs to be done in such a way that the collocations are clearly identifiable as secondary guiding elements.

An adapted version of the approach followed by Wiegand could lead to the inclusion of a search zone for items giving example sentences in each one of the subcomments on semantics in the integrated section and to phase all collocations out to the non-integrated section. This section could display a text block containing a subzone for each polysemous sense of the word represented by the lemma sign, with each subzone populated by the relevant non-grouped ordering of collocations, typographically marked as treatment units. Such a non-grouped ordering will enhance the visibility of collocations as fully-fledged secondary guiding elements that function as targets of procedures of non-lemmatic addressing. Whereas the lemma sign functions as primary guiding element of the article, each collocation in the text block is a secondary guiding element because it guides the user to items addressed at the specific collocation. The treatment directed at the collocations will at least demand the inclusion of a typical example sentence, given as cotextual guidance, or some specific morphosyntactic features. If needed a stylistic or other label may be included and even a brief paraphrase of meaning — if the meaning of the collocation deviates from the combined meaning of the constituent words.

5.2 Complex collocations

5.2.1 Guiding elements

The introduction of complex collocations as a subcategory of secondary guiding elements in the search zone for collocations could also be considered. Although such a presentation will certainly add value to the dictionary and could be considered as the most desired way of presenting complex collocations because of the explicit acknowledgement of these combinations as fully-fledged collocations, it might perhaps, in a general language dictionary, be regarded as belabouring this data type.

The inclusion of complex collocations remains important and lexicographers should negotiate the best possible way of presenting them and of making users aware of their existence. This could be done either as guiding elements in a subcategory of the search zone for collocations or in a more implicit way as part of the treatment of single collocations, typically within an example sentence illustrating the use of the single collocation but also its occurrence as component

of a complex collocation. If the latter route is followed the lexicographer has to ensure that the complex collocations can be identified as such by the target users of the specific dictionary. To achieve a presentation that will result in an unambiguous comprehension of such a cotextual item the lexicographer should embark on a system where different types of information can be retrieved from a single item form. This will demand presentation procedures not typically employed in dictionaries. This confronts both the metalexicographer who has to devise such a system and the practical lexicographer who has to implement it with definite challenges. The following section will deal with this aspect.

5.2.2 Embedded in cotextual items

The word *preference* combines with *personal* to form the collocation *personal preference*, and it combines with *express* to form the collocation *express a preference*. These two collocations can combine to form the complex collocation *to express a personal preference*. In the article of the lemma sign *preference* the search zone or subzone for collocations should include both these single collocations as secondary guiding elements and treatment units. The treatment of these collocations can include different items but needs to include at least one or more items illustrating their use. The following item giving an example sentence in which the complex collocation *to express a personal preference* is embedded could also be presented as part of the treatment of each one of these single collocations:

She tried to be impartial and did not want to express a personal preference.

Had this sentence been given in the search zone for example sentences, and not for collocations, in the article of the lemma sign *preference*, it would have been an illustration of the use of the single word. If this sentence is given in the search zone for collocations it can be an item illustrating the use of the collocation but also that of the word represented by the lemma sign by means of a procedure of remote addressing, cf. Louw and Gouws (1996). In the search zone for collocations this item is immediately addressed at the preceding secondary guiding element and it is therefore in the first instance an item illustrating the use of that specific collocation, i.e. either *express a preference* or *personal preference*. When given as a cotextual item addressed at either of the two collocations this item also illustrates the use of the complex collocation *to express a personal preference.* This item participates in a procedure of dual addressing, i.e. both lemmatic and non-lemmatic addressing, because it is addressed at the lemma sign by means of a procedure of remote addressing and it is addressed at the preceding collocation — a procedure of non-lemmatic addressing. In its non-lemmatic addressing it has both the single and the complex collocations as address.

The user of a dictionary that includes collocations and their treatment needs to be made aware of the fact that this example sentence contains two single collocations as well as a complex collocation. The presentation of an exam-

ple sentence from which the user can retrieve all this information demands a new type of microstructural entry. Such a single entry has to display a variety of items, i.e. an item giving the collocation *express a preference*, an item giving the collocation *a personal preference,* an item giving the complex collocation *express a personal preference,* an item giving an example of the use of the word *preference*, an item giving an example of the use of each one of the single collocations, i.e. *express a preference*, and *a personal preference* and an item giving an example of the use of the complex collocation *express a personal preference.* To include all these items in a single entry demands a high degree of textual condensation, perhaps too high for the average dictionary user. It is an entry devised to respect the space-saving endeavours of a given dictionary. The knowledgeable user who is familiar with the system of the dictionary will be able to retrieve the necessary information from the following data presentation:

She tried to be impartial and did not want to ***express*** *a* <u>personal</u> PREFERENCE.

The word represented by the lemma sign, i.e. *preference*, functions as base of the collocations and is given in small caps. The collocates are distinguished from each other by means of enrichment (the use of bold) and downward expansion (the use of underlining) — with the underlining functioning as an upwardly addressed item addition. The entry as a whole is given in italics to identify it as an illustrative example. The small caps, bold and underlining function as structural indicators that identify collocation components. This entry includes different items with different genuine purposes, combined into a single item form. The presentation could be indicated as follows — where single collocations in the example sentence do not have the word represented by the lemma sign in small caps and the collocate in bold, whereas this is done in the example sentences illustrating complex collocations. These sentences also have the second collocate underlined:

preference
1. The selection of something or somebody over something or somebody else
He has a preference for designer clothes.
2. The state of being selected over somebody or something else
Her preference for luxury hotels is no secret.
3. ….
4. ….
COL:
1. ▪decided preference: *They have a decided preference for classical music.*
▪express a preference (formal)**:** *She tried to be impartial and did not want to* ***express*** *a* <u>personal</u> PREFERENCE
▪personal preference: *She tried to be impartial and did not want to* ***express*** *a* <u>personal</u> PREFERENCE
2. ▪ …
3. ▪ ….

Example 10

In the non-integrated component, introduced by the marker "COL." single collocations are given as guiding elements in different subcomments. Each single collocation is the address of at least an item giving an example sentence. Complex collocations are not entered as secondary guiding elements but are embedded in the example sentence given for a single collocation. In this example sentence the use of small caps, bold and underlining will identify the complex collocation.

This arrangement of collocations in a non-grouped text block does not differ much from the presentation in example 6 from the *MacMillan Dictionary*. However, the *MacMillan Dictionary* includes collocations in a text block within an integrated microstructure and not in a separate non-integrated text block.

This is more clearly noticeable in the following reduced version of the partial article of the lemma sign *defence*.:

defence ….
NOUN …
1
[UNCOUNTABLE] actions that you take to protect someone or something that is being attacked
come/rush to someone's defence: *Two of his friends came to his defence.*
.
.
a.
[UNCOUNTABLE] the system of weapons, equipment, and people that is used to protect a country
Huge amounts of government money are spent on defence.
an attempt to reduce the defence budget
The President has promised to increase defence spending.
defence cuts (=spending less money on defence): *More defence cuts are expected next year.*
b. …
2
[COUNTABLE/UNCOUNTABLE] something that you say to support someone or something that is being criticized
defence of: *The Minister responded with a strong defence of the government's policy.*
speak/say something in someone's defence: *Several people spoke in my defence.*
in defence of: *Monroe has been fierce in defence of the present system.*
come/spring/leap to someone's defence: *Martin immediately sprang to her defence, saying she was innocent.*
.
:
b.
the defence
LEGAL the people in a court case who try to prove that someone is not guilty. The people who try to prove that someone is guilty are called the prosecution. The defence can be followed by a singular or plural verb
The defence will begin giving evidence tomorrow.

Example 11 from the *MacMillan Dictionary*

It should however be stated that in some articles the *MacMillan Dictionary* does give an additional search zone for collocations, as can be seen in the following

reduced partial article of the lemma sign *attack:*

attack …
1
[INTRANSITIVE/TRANSITIVE] to use violence to harm a person, animal, or place
It was shortly before midnight when the terrorists attacked.
attack someone with something: *Two prison officers were brutally attacked with a knife.*
a.
to use weapons to try to defeat an enemy
The enemy attacked at dawn.
US jets attacked five air defence sites in the north.
.
.
2
[TRANSITIVE] to strongly criticize someone or something for their ideas or actions
Opponents attacked the government's plan to increase road tax.
attack someone/something for (doing) something: *Parliament has been attacked for failing to take action.*
attack something as something: *The deputy Prime Minister last night attacked the decision as foolish.*
.
.
Collocates: attack
attack
▪ bitterly, fiercely, openly, sharply, strongly, vigorously

Example12 from the *MacMillan Dictionary*

Here the search zone is clearly marked but the collocations are presented without any further treatment.

The inclusion of collocations in a separate non-integrated text block, separated from the example sentences, ensures a more salient position for this item type and clearly identifies the collocations as secondary guiding elements within a dictionary article. The treatment also conveys data regarding complex collocations — a data type lacking in the most general language dictionaries. In online dictionaries where space restrictions are not as stringent the search zone for collocations could make provision for a separate subzone for complex collocations. This will diminish the high degree of textual condensation prevailing in the proposed presentation for printed dictionaries.

6. In conclusion

Collocations are part and parcel of the daily language use of all speakers. Dictionaries often still fail to give a satisfactory presentation and treatment of collocations. The nature, extent and quality of the selection, presentation and treatment of collocations need to be improved. This demands allocating an elevated status to collocations, i.e. as guiding elements within their own search zones and as fully-fledged treatment units.

Acknowledgement

1. This research is (a) conducted within the SeLA project (Scientific e-Lexicography for Africa), supported by a grant from the German Ministry for Education and Research, administered by the DAAD and (b) supported in part by the National Research Foundation of South Africa (grant specific unique reference number (UID) 85434).
2. The Grantholder acknowledges that opinions, findings and conclusions or recommendations expressed in any publication generated by the NRF supported research are that of the author, and that the NRF accepts no liability whatsoever in this regard.
3. I appreciate Ulrich Heid's valuable comments on a first draft of this paper.

7. Bibliography

7.1 Dictionaries

elexiko = http://www.owid.de/wb/elexiko/start.html

Häcki Buhofer, Annelies et al. (Eds.). 2014. *Feste Wortverbindungen des Deutschen.* Tübingen: Narr Francke.

HAT 5 = Odendal, F.F. and R.H. Gouws (Eds.). 2005. *Handwoordeboek van die Afrikaanse taal (e-HAT).* Fifth edition. Cape Town: Pearson Education South Africa.

HAT 6 = Luther, J., F. Pheiffer and R.H. Gouws (Eds.). 2015. *Handwoordeboek van die Afrikaanse Taal.* Cape Town: Pearson Education South Africa.

MacMillan = http://www.macmillandictionary.com/open-dictionary/

Oxford Online Dictionaries = http://www.oxforddictionaries.com/

7.2 Other literature

Bergenholtz, H., S. Tarp and H.E. Wiegand. 1999. Datendistributionsstrukturen, Makro- und Mikrostrukturen in neueren Fachwörterbüchern. Hoffmann, L. et al. (Eds.). 1999. *Fachsprachen. Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft/Languages for Special Purposes. An International Handbook of Special-Language and Terminology Research, Bd./Vol. 2*: 1762-1832. Berlin/New York: Walter de Gruyter.

Hausmann, F.J. 2004. Was sind eigentlich Kollokationen? Steyer, K. (Ed.). 2004. *Wortverbindungen — mehr oder weniger fest.* Institut für Deutsche Sprache Jahrbuch 2003, 2004: 309-334. Berlin/New York: De Gruyter.

Hausmann, F.J. 2008. Kollokationen und darüber hinaus. Einleitung in den thematischen Teil "Kollokationen in der europäischen Lexikographie und Wörterbuchforschung". *Lexicographica* 24: 1-8.

Heid, U. 1994. On Ways Words Work Together — Topics in Lexical Combinatorics. Martin, W. et al. (Eds.). 1994. *Euralex '94 Proceedings. Papers Submitted to the 6th EURALEX International Congress on Lexicography in Amsterdam, The Netherlands*: 226-257. Amsterdam: Vrije Universiteit.

Heid, U. 2011. German Noun+Verb Collocations in the Sentence Context: Morphosyntactic Proper-

ties Contributing to Idiomaticity. Herbst, T., S. Faulhaber and P. Uhrig (Eds.). 2011. *The Phraseological View of Language*: 283-307. Berlin: Mouton de Gruyter.

Heid, U. and R.H. Gouws. 2006. A Model for a Multifunctional Dictionary of Collocations. Corino, E., C. Marello and C. Onesti (Eds.). 2006. *Proceedings XII EURALEX International Congress, Turin, Italy, September 6th–9th, 2006*: 979-988. Alessandria: Edizioni dell'Orso.

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

Spohr, D. 2005. *A Description Logic Approach to Modelling Collocations.* Diploma Thesis. Stuttgart: Institut für Maschinelle Sprachverarbeitung (IMS).

Tutin, A. 2008. For an Extended Definition of Lexical Collocations. Bernal, E. and J. DeCesaris (Eds.). 2008. *Proceedings of the XIII EURALEX International Congress, Barcelona, 15–19 July 2008*: 1453-1460. Sèrie Activitats 20. Barcelona: Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada.

Tutin, A. and F. Grossmann. 2002. Collocations régulières et irrégulières: esquisse de typologie du phénomène collocatif. *Revue française de Linguistique appliquée*. Lexique: recherches actuelles VII(1): 7-25.

Wiegand, H.E. 1996. Das Konzept der semiintegrierten Mikrostrukturen. Ein Beitrag zur Theorie zweisprachiger Printwörterbücher. Wiegand, H.E. (Ed.). 1996. *Wörterbücher in der Diskussion II. Vorträge aus dem Heidelberger lexikographischen Kolloquium*: 1-82. Lexicographica. Series Maior 70. Tübingen: Max Niemeyer.

Wiegand, H.E. and R.H. Gouws. 2013. Addressing and Addressing Structures in Printed Dictionaries. Gouws, R.H. et al. (Eds.). 2013. *Dictionaries. An International Encyclopedia of Lexicography*. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography: 273-314. Berlin/New York: De Gruyter.

Zinsmeister, H. and U. Heid. 2003. Significant Triples: Adjective+Noun+Verb Combinations. Kiefer, F. and J. Pajzs. (Eds.). 2003. *Proceedings of COMPLEX 2003: 7th Conference on Computational Lexicography and Text Research*: 92-101. Budapest: Linguistics Institute, Hungarian Academy of Sciences.

Using an Online Dictionary for Identifying the Meanings of Verb Phrases by Chinese EFL Learners

Lingling Li, *Department of Literature and Communication, City College of Dongguan University of Technology, Dongguan, P.R. China (rowena_lee@126.com)*
and
Hai Xu, *Centre for Linguistics and Applied Linguistics, Guangdong University of Foreign Studies, Guangzhou, P.R. China (xuhai1101@gdufs.edu.cn)*

Abstract: This article reports on the results of a study which investigated the use of an online dictionary by Chinese EFL learners in identifying the meanings of verb phrases. Thirty-two students with English as major subject participated in a meaning determination task with and without the help of the *Macmillan English Dictionary Online* (MEDO). Introspective and retrospective questionnaires were used to establish the participants' perception of the usefulness of the online dictionary, and how they used the online dictionary to search for their needed information. The results of the study show that learners improved in doing the task after consulting the online dictionary, but they also encountered some problems which led to incorrect choices. It is suggested that dictionary skills training should be provided, and improvements to online dictionaries should also be made.

Keywords: ONLINE DICTIONARY, *MACMILLAN ENGLISH DICTIONARY ONLINE*, VERB PHRASES, MEANING IDENTIFICATION, DICTIONARY USE, DICTIONARY SKILLS, ENTRY STRUCTURE AND INFORMATION

Opsomming: Die gebruik van 'n aanlyn woordeboek deur Chinese EVT-leerders vir die identifisering van werkwoordelike frases. Hierdie artikel doen verslag oor die resultate van 'n studie waarin 'n aanlyn woordeboek deur Chinese EVT-leerders gebruik is in die identifisering van werkwoordelike frases. Twee en dertig studente met Engels as hoofvak het deelgeneem in 'n betekenisbepalingsopdrag met en sonder die hulp van die *Macmillan English Dictionary Online* (MEDO). Introspektiewe en retrospektiewe vraelyste is gebruik om die deelnemers se persepsie van die nuttigheid van die aanlyn woordeboek te bepaal, asook om te bepaal hoe hulle die aanlyn woordeboek gebruik het om te soek na die verlangde inligting. Die resultate van die studie toon dat leerders verbeter het in die voltooiing van die taak nadat hulle die aanlyn woordeboek geraadpleeg het, maar dat hulle ook sommige probleme teëgekom het wat gelei het tot verkeerde keuses. Daar word voorgestel dat opleiding in woordeboekvaardighede verskaf moet word en dat verbeterings aan aanlyn woordeboeke ook aangebring moet word.

Sleutelwoorde: AANLYN WOORDEBOEK, *MACMILLAN ENGLISH DICTIONARY ONLINE*,

WERKWOORDELIKE FRASES, BETEKENISHERKENNING, WOORDEBOEKGEBRUIK, WOORDEBOEKVAARDIGHEDE, INSKRYWINGSTRUKTUUR EN -INLIGTING

1. Introduction

Sparked by technological advancements, electronic dictionaries (EDs) have developed rapidly in the last three decades. From the first human readable electronic dictionary (ED) available for the public in the late 1980s, EDs have been available on: CD-ROM, small hand-held calculator-type reference works (i.e. pocket ED), and online. De Schryver (2003: 146) defined EDs as 'collections of structured electronic data that can be accessed with multiple tools, enhanced with a wide range of functionalities, and used in various environments'.

Research shows that EDs are in many respects valuable. Compared with the bulkiness of paper dictionaries and slow retrieval, the increasing popularity of EDs can be attributed to better readability and their ease of use which makes the consultation process less time-consuming. The most innovative aspect of EDs is probably the retrieval system (Nesi 2000a: 839). It encourages exploratory browsing which results in a large number of words looked up. It is believed that looking up more words might be helpful in word acquisition (Komuro et al. 2006). However, some researchers doubt the efficiency of EDs. Laufer and Hill (2000: 72) found that 'the number of times the word is looked up during a learning session bears almost no relation to its retention'. Hulstijn and Laufer (2001) further point out that deeper and more elaborate processing resulted in better word retention. There is a 'possibility that the most easily extracted information may require least thought, and be soonest forgotten' (Nesi 2000a: 844).

By investigating the usefulness of paper and CD-ROM versions of the *Oxford Advanced Learner's Dictionary of Current English* (5th edition), Nesi (2000b) concluded that there were no significant differences in comprehension scores between the two dictionary formats. Studies which investigated the use of hand-held devices reveal a complicated picture concerning the efficiency of EDs. Osaki et al. (2003) observed that compared with the paper dictionary use group, the portable ED consultation group performed much better in choosing the appropriate meaning of underlined target words, and gained better reading comprehension scores. Chen (2010) claimed that portable EDs could be as effective as paper dictionaries for new word acquisition for EFL learners. In stark contrast to the previous findings, some results suggest that the use of portable EDs contributed little to reading comprehension (Koyama and Takeuchi 2007), and compared with paper dictionaries, no statistically significant differences in reading comprehension scores could be observed (Kobayashi 2007).

In addition to CD-ROM and hand-held devices, researchers also began to show interests in online dictionaries. Due to the rapid development of the Internet, smart phones and iPads, ways of obtaining information have been diversified. With the easy and widespread access to the Internet, dictionary

users can consult online dictionaries by hovering the mouse over a word on computer screens, or clicking on their smart phones and iPads wherever they go. Nearly all major traditional dictionaries have provided online versions with lexicographic information which can be as useful as their traditional hard copies (Li 2005: 16). Some publishers even go far beyond. *Macmillan English Dictionary*, one of the 'Big Six', has ceased the publication of paper dictionaries in 2013 and decided in favour of an online version only. Together with the multi-functions provided by online dictionaries, dictionary users can learn more about the latest revisions and new words as editors can update entry information every few months.

The advent of online dictionaries has greatly influenced users' preferences and patterns of dictionary use. In Dziemianko (2010), which concerns the paper and online versions of a monolingual dictionary, the online dictionary was found to be more useful in dealing with both receptive and productive tasks. Lew and Tokarek (2010) also examined the efficiency of entry menus in bilingual dictionaries in electronic format, but it is noted that the online dictionary used in their study was an experimental rather than an authentic one. It deserves further enquiry how useful an authentic online dictionary is for users and what problems they may encounter during their online dictionary consultation.

This study focuses on how users look up verb phrases in an authentic online dictionary. Verb phrases consist of a verb and one or two additional elements (generally called particles). They constitute a syntactic or lexical unit functioning as a whole like a single lexical verb, and they have to be acquired, stored and retrieved from memory as a holistic unit (Wray and Perkins 2000). They can be viewed as a type of formulaic sequence (Wray 2002). It is evident that the verb phrase is an important element of language learning and usage. Normal discourse, both written and spoken, contains a large percentage of verb phrases, but they are especially problematic for EFL learners (Howarth 1996; Nesselhauf 2005). The meanings of many verb phrases can hardly be decoded literally but have to be considered as a whole. One single verb can connect with several different particles to form different verb phrases with totally different meanings, and EFL learners often get confused. Moreover, the polysemy of many verb phrases confounds EFL learners. It is hard for them to know all the meanings of a polysemous verb phrase, and they may not have the ability to distinguish the subtle nuances of its many meanings, and mistakenly perceive the dominant meaning as the only meaning when they put it into practical use.

Compared with paper dictionaries that mainly organize entries around a single word, online dictionaries allow users to access multi-words by consulting more than one component of a multi-word (Lew 2012: 349-351). Thus, users can search for a certain verb phrase as a whole to identify its meaning. Furthermore, online dictionaries provide a search engine which links to other relevant information. Nevertheless, when faced with a lexical request, an online dictionary is rather less helpful (Lew and De Schryver 2014: 349). While search-

ing polysemous entries, users themselves have to decide which sense is relevant in a particular context. It is therefore worthwhile to examine how EFL learners look up the meanings of polysemous verb phrases in an online dictionary, and what the problems they may have in choosing one sense over the others according to a certain context.

2. Research questions

The present study addresses the following questions:

(1) How useful is an online dictionary to determine the meaning of verb phrases?
(2) Do Chinese EFL learners perceive *Macmillan English Dictionary Online* as useful?
(3) What are the general problems Chinese EFL learners encounter in consulting an online dictionary?

3. Participants

One hundred and seven students with English as major subject at a local university in China took the Oxford Quick Placement Test in thirty minutes, and their scores were marked. Thirty-two students who scored between 30 and 39 were regarded as lower intermediates, and chosen as the participants. The implementation of the Oxford Quick Placement Test was to make sure that the participants were at the same English proficiency level. There were 7 males and 25 females, and their ages ranged from 18 to 20 years. Although advanced learners are considered to be those who often use a dictionary for self-learning purposes (Chan 2005), few studies focus on lower intermediate students. The results of studies on advanced learners may not be applicable to lower intermediates. Thus their dictionary usage should be examined.

4. Instruments

4.1 Meaning determination task of verb phrases

Twenty-six polysemous verb phrases were chosen, and were piloted with 33 students who were excluded from those 107 students. All the verbs in these phrases were simple words familiar to the learners, but the phrases as a whole might have been familiar or unfamiliar to them. The testing material was adapted from http://www.usingenglish.com/reference/phrasal-verbs/. These students had to determine the most appropriate meaning of an underlined verb phrase in each sentence by choosing the correct answer from a list of four choices without consulting dictionaries. The ten verb phrases with the highest

error rate (between 81.82% and 54.55%) were collected as the target phrases. The following is one example of the target verb phrases (see the Appendix 1 for a complete list).

(1) He had to break down their opposition to his ideas.
A. separate into parts
B. stop working
C. remove a barrier or obstacle
D. end negotiations unsuccessfully

The meaning determination task consisted of two parts. The first part examined the participants' prior knowledge of the target verb phrases, and the second part investigated how they used the online dictionary to help them determine the meanings of the target phrases.

4.2 Questionnaires

One introspective questionnaire and one retrospective questionnaire were adopted in this research. The introspective questionnaire aimed at eliciting participants' instant and detailed feedback on how they used a certain dictionary entry to help them determine the meaning of a verb phrase (Chan 2012), and the retrospective questionnaire aimed at figuring out the participants' overall evaluation of the usefulness of the online dictionary and its different elements in the meaning determination task (ibid.). Chan (2012, 2013) used her instant introspective questionnaire and retrospective questionnaire to investigate the participants' thinking processes of how they determined the meanings of familiar English words used in less familiar contexts. Since the results of both studies identified different kinds of consultation problems, the questionnaires can be viewed as effective. As Chan's studies concerned paper dictionaries only, some modifications were made due to the special features of online dictionaries. Extra column and code/abbreviation were deleted in the entry information as there were no such parts in an online dictionary, while thesaurus and menu were added. The participants had to fill in both introspective and retrospective questionnaires. All the questionnaires were given in English and the participants responded in English. Though the chosen participants are lower intermediate, they are students with English as major subject. Using English to express themselves is one basic requirement in China, and all of them claimed that they could answer the questionnaires in English in spite of some grammatical errors found in their answers.

4.3 *The Macmillan English Dictionary Online*

The Macmillan English Dictionary Online (MEDO) was chosen because it only provides an online version nowadays, and it can be viewed as a leading online

dictionary. MEDO claims in http://www.macmillandictionary.com/about.html that it is 'the perfect free online dictionary — a one-stop reference for English speakers around the world and is the smart choice for teachers and students, and makes learning, using and understanding English a lot easier!' This statement indicates that MEDO should be useful to advanced learners as well as learners with a lower proficiency. The special thesaurus feature in MEDO is unique among the online dictionaries. The introduction of MEDO reads 'MacmillanDictionary.com contains a thesaurus with a difference. Many online thesauruses sit completely independently from a dictionary and generate an unsystematic and often overwhelming list from the word you have looked up. Faced with this list, it is often difficult to know which synonym to choose, and many words in the list might not relate to the meaning of the word you have searched. The free online thesaurus from Macmillan thesaurus is fully integrated into the dictionary entries, giving you a targeted list of synonyms and words related exactly to the meaning you are looking for.' (http://www.macmillandictionary.com/about_thesaurus.html)

5. Method

Thirty-two participants were gathered in a computer lab to do the whole task. In the first part, 10 sentences containing 10 target verb phrases respectively were provided. The participants had to decide on the most appropriate meaning without consulting anything. Their answers were collected before they moved to the second part. To begin the second part, every participant had access to the webpage of www.macmillandictionary.com. Some guidance on how to use MEDO was provided: the verb phrase *eat up* was given to the participants as a sample. The researcher explained all the parts of this entry, especially the function of the thesaurus. As there is no menu box information under the entry phrase *eat up*, the single word *eat* was searched to illustrate the function of the menu to the participants. Participants were then given five extra minutes to practise how to use MEDO. Finally, the same sentences with the same verb phrases underlined and the same meaning options were given. Participants had to do the same task again but with the help of the online dictionary consultation. At the same time, they were required to do the introspective questionnaire immediately after finishing each question. After collecting all the introspective questionnaires and answers of Part Two, a retrospective questionnaire was handed out. Participants were allowed to leave with a gift after they handed in their retrospective questionnaires.

6. Results

6.1 Participants' performance in meaning determination of verb phrases

To answer the first research question, participants' performance in meaning

determination of verb phrases was analyzed. Their performance would be measured in terms of accuracy rates.

Without any dictionary consultation, participants' performance was rather poor with an overall accuracy rate of 26.52%. Of the ten verb phrases included in the task, only *feed on* and *run up* received an accuracy rate of over 40%. Over half (6) of the phrases received an accuracy rate ranging from 20% to 30%, and two phrases below 15%. The verb phrase with the lowest accuracy rate was *bring forth*: only three participants could accurately determine the meaning (Table 1).

Table 1: A comparison between the accuracy rate of each verb phrase with and without the use of MEDO

Verb phrases	Percentage of participants who could determine the meanings of the verb phrases correctly	
	Without MEDO	With MEDO
wait on	28.13% (9/32)	37.5% (12/32)
break down	12.5% (4/32)	31.25% (10/32)
bring forth	9.38% (3/32)	31.25% (10/32)
call on	28.13% (9/32)	21.88% (7/32)
feed on	46.88% (15/32)	75% (24/32)
fight off	21.88% (7/32)	43.75% (14/32)
get around	21.88% (7/32)	56.25% (18/32)
kick in	28.13% (9/32)	65.63% (21/32)
run up	43.75% (14/32)	68.75% (22/32)
ring up	21.88% (7/32)	65.63% (21/32)
Total	26.25% (84/320)	49.69% (159/320)

After consulting MEDO, participants' performance improved markedly with a total accuracy rate of 49.69%. Half of the phrases (five) received an accuracy rate of over 50%, and the highest one even reached 75%. Four phrases received an accuracy rate ranging from 30% to 45%. Compared with the previous performance, the participants gained a higher accuracy rate in these nine phrases after consulting MEDO. The verb phrase with the lowest accuracy rate was *call on*, receiving an accuracy rate of only 21.88%, and it also showed regression from its original rate of 28.13% (The discussion for this will be presented in Section 6.3). To check whether there was any difference between participants' performance in these two tasks, the results were analyzed through a two-tailed, paired samples *t*-test by using SPSS 21.0. It showed that there was statistically significant difference between the overall meaning determination of the target

verb phrases before and after the dictionary consultation at the significance level of 0.05 (t = -10.522, df = 31, p = 0.000) (Table 2).

Table 2: A comparison between individual participants' performance in determining the meanings of the verb phrases with and without the use of MEDO

Participants	Percentage of meaning determination accuracy	
	Without MEDO	With MEDO
Participant 1	30% (3/10)	40% (4/10)
Participant 2	20% (2/10)	50% (5/10)
Participant 3	40% (4/10)	60% (6/10)
Participant 4	20% (2/10)	40% (4/10)
Participant 5	40% (4/10)	70% (7/10)
Participant 6	40% (4/10)	30%(3/10)
Participant 7	10% (1/10)	40% (4/10)
Participant 8	50% (5/10)	50% (5/10)
Participant 9	30% (3/10)	30% (3/10)
Participant 10	20% (2/10)	40% (4/10)
Participant 11	10% (1/10)	30% (3/10)
Participant 12	30% (3/10)	70% (7/10)
Participant 13	0% (0/10)	30% (3/10)
Participant 14	20% (2/10)	50% (5/10)
Participant 15	30% (3/10)	50% (5/10)
Participant 16	10% (1/10)	50% (5/10)
Participant 17	40% (4/10)	60% (6/10)
Participant 18	10% (1/10)	50% (5/10)
Participant 19	50% (5/10)	80% (8/10)
Participant 20	30% (3/10)	60% (6/10)
Participant 21	10% (1/10)	40% (4/10)
Participant 22	30% (3/10)	70% (7/10)
Participant 23	30% (3/10)	40% (4/10)
Participant 24	20% (2/10)	50% (5/10)
Participant 25	30% (3/10)	60% (6/10)
Participant 26	20% (2/10)	50% (5/10)
Participant 27	20% (2/10)	50% (5/10)

Participant 28	50% (5/10)	60% (6/10)
Participant 29	20% (2/10)	40% (4/10)
Participant 30	50% (5/10)	70% (7/10)
Participant 31	10% (1/10)	40% (4/10)
Participant 32	20% (2/10)	50% (5/10)
Total	26.25% (84/320)	49.69% (159/320)

Based on the results, it can be concluded that MEDO is very useful to Chinese EFL learners to determine the meaning of verb phrases.

6.2 Perception of usefulness of MEDO

To answer the second research question, the participants' retrospective questionnaire was explored.

Definitions and examples were the only two which were regarded as extremely or very useful by most participants, receiving 93.76% and 90.63% respectively (Table 3). These were also the items which most participants consulted, with definitions 75% and examples 84.38% (Table 4). Moreover, examples ranked the highest in information from which the answers to most of the questions were found, reaching 87.5% with 28 participants having chosen them. Definitions just ranked behind examples: 65.63% participants (21) chose them. Less than one-third (9) of the participants consulted the menu most; only 7 participants could find the answers they needed, and less than half (14) of the participants thought they were useful. Very few consulted the thesaurus, more dictionary definitions, and special features: their percentages were all below 15%. None of the participants used the thesaurus and special features to find the answers to the questions. Fifty percent of the participants considered the thesaurus and special features not very useful, and 40.63% participants thought that more dictionary information was not very useful. Overall, 18 participants out of 32 deemed MEDO useful; 10 thought the dictionary was very useful; 2 considered it extremely useful; and the other 2 did not view it very useful (Tables 3 and 4).

Table 3: Degree of usefulness of different parts of MEDO

Dictionary information	Extremely useful	Very useful	Useful	Not very useful	Not useful at all
Definitions	28.13% (9/32)	65.63% (21/32)	6.25% (2/32)	0% (0/32)	0% (0/32)
Examples	56.25% (18/32)	34.38% (11/32)	9.38% (3/32)	0% (0/32)	0% (0/32)

Thesaurus	0% (0/32)	12.5% (4/32)	31.25% (10/32)	50% (16/32)	6.25% (2/32)
Menu	0% (0/32)	28.13% (9/32)	43.75% (14/32)	12.5% (4/32)	15.63% (5/32)
More dictionary definitions	0% (0/32)	12.5% (4/32)	18.75% (6/32)	40.63% (13/32)	28.13% (9/32)
Special features	0% (0/32)	6.25% (2/32)	28.13% (9/32)	50% (16/32)	15.63% (5/32)
Others	0% (0/32)	0% (0/32)	0% (0/32)	0% (0/32)	0% (0/32)
Overall usefulness	6.25% (2/32)	31.25% (10/32)	56.25% (18/32)	6.25% (2/32)	0% (0/32)

Table 4: Perception of the usefulness of different parts of MEDO

Dictionary information	Information which participants consulted most	Information which participants found the answers to most of the questions
Definitions	75% (24/32)	65.63% (21/32)
Examples	84.38% (27/32)	87.5% (28/32)
Thesaurus	6.25% (2/32)	0% (0/32)
Menu	28.13% (9/32)	21.88% (7/32)
More dictionary definition	12.5% (4/32)	6.25% (2/32)
Special features	3.13% (1/32)	0% (0/32)
Others	0% (0/32)	0% (0/32)

It can be deduced from the results that Chinese EFL learners perceive MEDO as useful in general.

6.3 Problems encountered in dictionary consultation

Although MEDO consultation resulted in significant improvements in meaning determination of the verb phrases, all the participants encountered certain difficulties which rendered their false judgments, as no one achieved 100% accuracy rate after consulting the dictionary. In order to answer the third research question, the participants' introspective questionnaire was scrutinized.

Deficient definitions in the dictionary. As for the phrase *wait on*, 16 participants chose 'sure that I didn't get the right information', and 13 participants chose 'not sure whether my decision was correct' on ending the dictionary search. Most of them claimed they just couldn't find the answer to the question; the definitions provided in MEDO were not what they need. It might be argued

that *butler* is quite a rare word and distracts participants' choice, but this was designed on purpose. If the participants had known that *butler* means the most important male servant in a wealthy house, they would have guessed the meaning of *wait on* without consultation.

> I only get two meanings, but they are not right to my feeling. One meaning is 'to serve people in a restaurant'. I know this is not the answer, but I feel it's better than the other one, so I guess choice B, because B is close to this one. (Participant 2)

> I can't find my answers in the dictionary, so I look for *butler*. After I know the meaning, I just choose *B*. But I'm not sure. (Participant 10)

The situation with the phrase *bring forth* was quite different. Twenty-three participants chose 'sure that my decision was correct' after consulting the dictionary, but only 10 chose the right answer. It could be attributed to the single definition provided in the dictionary. The sole definition 'to produce something, or to cause a particular reaction' is very close to the given option *C* 'produced something, make it known or visible'. Some participants just attempted to match the key words given in the definition with those provided in the answers. Some participants doubted option *C*, but they still chose it for they believed in the dictionary's authority.

> *C* is a little strange, but the dictionary only provides this definition. I think dictionary should be correct, so I choose *C*. (Participant 7)

> I only get one meaning; it is the same with *C*. Then *A, B, D* must be incorrect. (Participant 11)

Inadequate examples in the dictionary. The accuracy rate of *fight off* showed significant gains after the dictionary consultation, and 13 participants were sure they did not get the right information. Two definitions were provided in MEDO: while one definition has one example, the other one has no example at all. Some participants complained that one example was far from adequate to help them obtain the right information, and more examples should be provided. Though some of them knew 'to stop someone who is trying to attack you' was the definition they needed, but they had difficulty in differentiating the subtle differences between the options *B* and *D*. If more examples were provided, they might be able to figure out the right answer. More often than not, definitions together with examples could help them gain a better understanding, and some viewed examples as an essential part during the dictionary consultation.

> I like to read examples to help me, but there is only one here. So I choose *B* according to the context. (Participant 10)

> I know the first meaning is the right one, but I still don't know which one to choose. There is only one example under this meaning; I think it should have more examples. (Participant 14)

> Examples are very important, but almost every meaning has only one example, I don't know how to choose: *B* or *D*? (Participant 29)

Unable to distinguish meanings by learners. *Call on* is the one which received the lowest accuracy rate even with the help of the dictionary. Only 7 participants made the correct judgment. Though 11 participants were sure about their answers, more than half (17) of the participants were sure they did not get the right information. Surprisingly, *call on* is the only phrase which showed a decreased accuracy rate after the dictionary consultation. The reason might be ascribed to the participants' inability to judge which definition was relevant to the context of the given sentence. Some were daunted by the numerous definitions in the entry and felt confused about their choices. Most of them complained that they could not distinguish so many definitions, especially when this phrase impeded the understanding of the whole sentence. Though definitions were different from one another, participants still found difficulty in seeking the needed information.

> There are so many different meanings which make me confused. I don't know how to choose. (Participant 3)

> There are five meanings in the dictionary; I just don't know which one is right. (Participant 7)

> I am not sure about the meaning of this phrase; there are so many meanings here. (Participant 12)

Preoccupied by the familiar meaning. Results of the phrase *break down* showed that 21 participants were quite sure they supplied the right answer. But in fact, only 10 participants made the right judgment. Participants' preconception of this phrase might mislead them into selecting an incorrect answer. The preoccupied meaning of this phrase to most participants was 'separate something into parts', probably because 'their opposition' could be separated into parts which could be inferred as shattered or stopped. Most participants chose option *A* without further consideration, and they paid no attention to the rest of the definitions and examples. If they looked closer at the sixth definition's example '*The aim of the agreement is to break down barriers to trade.*', they could find that this example had the same sentence pattern 'break down something to something or somebody' as the given sentence.

> There are so many meanings here, but when I see 'separate something into parts', I feel this is the answer. After I put this meaning into the sentence, I can understand the sentence. (Participant 27)

Ignorance of immediate cross-reference. *Feed on* is the one which showed the highest accuracy rate. Two definitions were provided in the dictionary entry, though one example was provided in the second one, dictionary users have to

cross reference **feed off** to get its definition. It means participants have to click on the hyperlink of **feed off**. The screenshot is shown below.

Figure 1: Entry for feed on in MEDO

Once participants clicked on **feed off**, they would get the exact definition as was shown in option *C*. Twenty-four participants made the right judgment, but still 8 participants missed it. Their questionnaires were examined, and some problems were identified. The main problem was due to their ignorance of the hyperlink. It was quite common for most participants to use the hyperlink, but not everyone was familiar with it.

> I click **feed off**, then I got 'to use something in order to continue to exist or become stronger'. I compared the two meanings; I feel 'become stronger' is the one I need. (Participant 18)
>
> I doubt my answer, because there is no definition. (Participant 6)
>
> I cannot find what I want, I just guess. (Participant 31)

To recapitulate, the participants encountered at least four problems. First, they found that they could not find sufficient definitions of certain phrases which hindered their judgements. Secondly, faced with numerous definitions in certain entries, they could not distinguish between them. Thirdly, while some participants complained that no dictionary example or inadequate examples resulted in their false judgements, some were preoccupied by the familiar meaning of a target phrase. Lastly, not every participant knew the function of a hyperlink in an online dictionary.

7. Discussion and Implications

Based on this study, it can be argued that the online dictionary is, to some extent, useful in helping Chinese EFL learners to determine the meanings of verb phrases, for there is a significant difference between learners' performance

before and after the dictionary consultation. However, despite the general usefulness of the online dictionary, no one achieved 100% accuracy rate. Participants still encountered several problems during the consultation.

An online dictionary is flexible and dynamic; it can offer a layered, hierarchical inner access structure (Dziemianko 2012: 321). Take MEDO for example: if the learner selects a particular sense from the menu which is displayed at the top of the entry, direct access to a specific definition will be provided (Lew and Tokarek 2010). The menu box has the same function, but it only appears under single words with provision of related dictionary definitions. The thesaurus is a special feature of MEDO, but no participant could identify the meaning of this word before it had been explained to them, and participants were unfamiliar with its function. Even though some guidance on how to use the menu and thesaurus was provided, many participants still made use of definitions and examples in the entry most to guide their meaning determination, very rare for the rest part of entry information. One reason for the rare usage of the menu and thesaurus is that learners are quite unfamiliar with these functions, and they are still quite used to the way of using a paper dictionary. Another reason is concerned with the design of the task. Since the task is to determine the appropriate meanings of verb phrases, participants would definitely focus more on definitions. Checking examples at the same time also helps them gain a better understanding and assists them in differentiating semantic differences. If the task is designed in a different way, the result will probably be different.

Participants' perception of the online dictionary may be affected by their language proficiency and their dictionary skills. The participants are all lower intermediates in this study. Compared to advanced learners, they have more difficulty in identifying the correct definition to the corresponding sentences due to their language proficiency. In addition, their preconception of the verb phrases may affect their judgments to a large degree. With no critical thinking or ignorance of other information parts, they usually made an inappropriate decision by guessing or just by matching the definition with the given options when they encountered the situation of finding no relevant answer in the entry's definitions. As for their dictionary skills, there is often a gap 'between the sophistication of the typical dictionary structure and the inadequacy of the reference skills possessed by the average dictionary users' (Lew and Galas 2008: 1273). Despite the gap, learners' dictionary skills are taken for granted. We had assumed that every participant would probably click on the hyperlink of **feed off** when they looked up **feed on**. But the factual situation was that not everyone acquired this simple skill. It is quite common in China that dictionary skills are seldom taught, and some teachers may lack adequate skills themselves, especially online dictionary consultation skills.

In general, an online dictionary has many advantages, such as simplified outer access structure, hierarchical inner access structure, clear and concise layout, and eye-catching special features. But to Chinese EFL learners, the inadequacy of definitions and examples are main problems they have encountered.

Of the ten verb phrases given in the task, only one definition under the entry **get around** has two examples. Most of the definitions have only one example, and three have no example at all. Several definitions can be found under entry **bring forth** in other dictionaries, but only one is provided in MEDO. It is a paradox as sometimes learners complain that there are so many definitions and examples which confuse them, whereas sometimes definitions and examples are just inadequate to their use. It is not an easy task to provide enough information without redundancy.

This study has some pedagogical and lexicographic implications. Firstly, dictionary skills should be taught to EFL learners as they are of vital importance. Müller-Spitzer et al. (2012) did a survey on what users expect of online dictionaries. Almost identical to that study, the present study found that online dictionary content received the highest ratings while features typical of online dictionaries received the lowest. As can be found in this study, not every learner possesses adequate skills. Learners have to learn how to make good use of the other information parts of an online dictionary in addition to the definitions and examples. If learners can take advantage of these features and functions, they will be more likely to choose the appropriate answer of the phrases they look up. Secondly, it is important to consider contextual clues during dictionary consultation. Faced with a list of definitions under a certain phrase, it is one of the best ways for learners to resort to the context information to select the appropriate meaning. Training learners to exploit contextual clues may help them reduce their uncertainty of the multiple definitions and increase their accuracy rate. Finally, it should be recognized that learners' success of dictionary look-up may also depend on the dictionary itself. One example provided under a definition is succinct in dictionary's layout to learners, but it may not convey sufficient information which may result in learners' inability to differentiate definitions. Lexicographers should find a solution to this problem and they have to take special care of those entries which provide deficient definition.

8. Conclusion

This article reported on the results of a meaning determination task requiring the participants to determine the appropriate meaning of polysemous verb phrases in given sentences with the use of MEDO. Although MEDO is useful for facilitating meaning determination in general, learners still encountered several problems during their consultation. It is suggested that online dictionary skills should be taught to the Chinese EFL learners, and the entry structure and information of an online dictionary should be improved.

Acknowledgements

This study was funded by the National Social Science Foundation of China (No. 15BYY062), by the MOE Project of Key Research Institute of Humanities

and Social Sciences at Universities in P.R. China (No. 12JJD740010), and by the Research Project of Humanities and Social Sciences at Universities in Guangdong Province (No. 2012WYXM-0057). We are grateful to Dr. Alice Y.W. Chan of City University of Hong Kong for generously sending us her questionnaires, and to two anonymous adjudicators and Prof. Elsabe Taljard for their insightful comments and valuable suggestions. We would also like to thank all the students for their participation in this study.

References

Chan, A.Y.W. 2005. Tactics Employed and Problems Encountered by University English Majors in Hong Kong in Using a Dictionary. *Applied Language Learning* 15(1-2): 1-27.

Chan, A.Y.W. 2012. The Use of a Monolingual Dictionary for Meaning Determination by Advanced Cantonese ESL Learners in Hong Kong. *Applied Linguistics* 33(2): 115-140.

Chan, A.Y.W. 2013. Advanced Cantonese ESL Learners' Use of a Monolingual Dictionary for Language Production. *Lexikos* 22: 109-138.

Chen, Y. 2010. Dictionary Use and EFL Learning. A Contrastive Study of Pocket Electronic Dictionaries and Paper Dictionaries. *International Journal of Lexicography* 23(3): 275-306.

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

Dziemianko, A. 2010. Paper or Electronic? The Role of Dictionary Form in Language Reception, Production and the Retention of Meaning and Collocations. *International Journal of Lexicography* 23(3): 257-273.

Dziemianko, A. 2012. On the Use (fulness) of Paper and Electronic Dictionaries. Granger, S. and M. Paquot (Eds.). 2012: 319-341.

Granger, S. and M. Paquot (Eds.). 2012. *Electronic Lexicography*. Oxford: Oxford University Press.

Howarth, P. 1996. *Phraseology in English Academic Writing: Some Implications for Language Learning and Dictionary Making*. Tübingen: Max Niemeyer.

Hulstijn, J.H. and B. Laufer. 2001. Some Empirical Evidence for the Involvement Load Hypothesis in Vocabulary Acquisition. *Language Learning* 51(3): 539-558.

Kobayashi, C. 2007. Comparing Electronic and Printed Dictionaries: Their Effects on Lexical Processing Strategy Use, Word Retention, and Reading Comprehension. Bradford-Watts, K. (Ed.). 2007. *JALT2006 Conference Proceedings*: 657-671. Tokyo: JALT.

Komuro, Y., S. Uchida and Y. Shitara-Matsuo. 2006. An Analysis of the *Oxford Advanced Learner's Dictionary Current English*, with Special Reference to the CD-ROM. *Lexicon* 36: 55-146.

Koyama, T. and O. Takeuchi. 2007. Does Look-up Frequency Help Reading Comprehension of EFL Learners? Two Empirical Studies of Electronic Dictionaries. *Calico Journal* 25(1): 110-125.

Laufer, B. and M. Hill. 2000. What Lexical Information Do L2 Learners Select in a CALL Dictionary and How Does it Affect Word Retention. *Language Learning and Technology* 3(2): 58-76.

Lew, R. 2012. How Can We Make Electronic Dictionaries More Effective. Granger, S. and M. Paquot (Eds.). 2012: 343-361.

Lew, R. and G.-M. de Schryver. 2014. Dictionary Users in the Digital Revolution. *International Journal of Lexicography* 27(4): 341-359.

Lew, R. and K. Galas. 2008. Can Dictionary Skills Be Taught? The Effectiveness of Lexicographic

Training for Primary-School-Level Polish Learners of English. Bernal, E. and J. DeCesaris (Eds.). 2008. *Proceedings of the XIII EURALEX International Congress, Barcelona, 15–19 July 2008*: 1273-1285. Barcelona: Universitat Pompeu Fabra.

Lew, R. and P. Tokarek. 2010. Entry Menus in Bilingual Electronic Dictionaries. Granger, S. and M. Paquot (Eds.). 2010. *eLexicography in the 21st Century: New Challenges, New Applications*: 193-202. Louvain-la-Neuve: Cahiers du CENTAL.

Li, L. 2005. The Growing Prosperity of On-line Dictionaries. *English Today* 21(3): 16-21.

Müller-Spitzer, C., A. Koplenig and A. Töpel. 2012. Online Dictionary Use: Key Findings from an Empirical Research Project. Granger, S. and M. Paquot (Eds.). 2012: 425-457.

Nesi, H. 2000a. Electronic Dictionaries in Second Language Vocabulary Comprehension and Acquisition: The State of the Art. Heid, U., S. Evert, E. Lehmann and C. Rohrer (Eds.). 2000. *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000, Stuttgart, Germany, August 8th–12th, 2000. Vol.2*: 839-847. Stuttgart, Germany: Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.

Nesi, H. 2000b. On Screen or in Print? Students' Use of a Learner's Dictionary on CD-ROM and in Book Form. Howarth, P. and R. Herington (Eds.). 2000. *EAP Learning Technologies. BALEAP Conference Proceedings*: 106-114. Leeds: Leeds University Press.

Nesselhauf, N. 2005. *Collocations in a Learner Corpus.* Amsterdam/Philadelphia: John Benjamins.

Osaki, S., N. Ochiai, T. Iso, and K. Aizawa. 2003. Electronic Dictionary vs. Printed Dictionary: Accessing the Appropriate Meaning, Reading Comprehension, and Retention. Murata, M., S. Yamada and Y. Tono (Eds.). 2003. *Dictionaries and Language Learning: How can Dictionaries Help Human and Machine Learning*: *Papers submitted to the Third ASIALEX Biennial International Conference, Meikai University, Uruyasu, Chiba, Japan, August 27–29, 2003:* 205-212. Tokyo: Asian Association for Lexicography.

Wray, A. 2002. *Formulaic Sequences and the Lexicon*. Cambridge: Cambridge University Press.

Wray, A. and M.R. Perkins. 2000. The Functions of Formulaic Language: An Integrated Model. *Language and Communication* 20(1): 1-28.

Appendix 1: Meaning identification task sheet

Please determine the most appropriate meaning of the underlined verb phrases.

Name: Number: Score:

1. He has a butler who <u>waits on</u> him.

A. waits for a result before being able to make a decision
B. provides someone with everything they need or want
C. stays somewhere after other people have left
D. stays at home because someone is going to visit

2. He had to <u>break down</u> their opposition to his ideas.

A. separate into parts
B. stop working
C. remove a barrier or obstacle
D. end negotiations unsuccessfully

3. The report has <u>brought forth</u> a lot of criticism of the policy.

A. made something happen earlier than originally planned
B. removed something from where it is kept or hidden
C. produced something, make it known or visible
D. made something happen

4. He <u>called the speaker on</u> several mis-statements of fact.

A. ordered someone to stop attacking
B. asked someone to do something, especially to speak in public.
C. made something happen
D. challenged

5. The opposition party is <u>feeding on</u> the government's weaknesses.

A. consuming in an animal's diet
B. giving someone a particular food
C. growing stronger
D. using something to your advantage

6. The old lady managed to <u>fight the muggers off</u> and they didn't get her purse.

A. resist an illness or emotion
B. fight an attacker and force him or her back
C. struggle to see who wins, both by arguing or fighting
D. hit or kick someone who is attacking you

7. It'll be tricky, but we will find a way to get around the regulations.

A. become known
B. persuade, convince
C. avoid a problem
D. finally manage to do something, make the effort to do something

8. The budget cuts are starting to kick in and people are struggling.

A. pay someone illegally as part of the price
B. contribute money
C. start having an effect
D. cause trouble or pain

9. He ran up a few examples for them to look at.

A. encountered problems, often unexpected
B. moved quickly to where someone is
C. did or made something very quickly
D. explained quickly

10. They rang up the bill for the groceries.

A. entered figures into a till or cash register
B. achieved an amount or number
C. made a sudden loud sound
D. charged excessively or obtained money unfairly

http://lexikos.journals.ac.za

Multimodal Definition: The Multiplication of Meaning in Electronic Dictionaries

Xiqin Liu, *School of Foreign Languages, South China University of Technology, China* and *Guangdong University of Foreign Studies, China (flxqliu@scut.edu.cn)*

Abstract: With the popularity of electronic dictionaries, multimodality plays an increasingly important role in lexicography. To broaden the horizons of e-dictionary definitions, this article argues for the establishment of a new term, multimodal definition, as a key component of multimodal lexicography. This term integrates verbal definitions and the complementary multimodal resources from a holistic viewpoint of meaning. In a multimodal definition, a dynamic meaning ecology can be formed, with two critical variables functioning, (semiotic) mode selection and intermodal synergy. In this ecology, meaning expressed verbally can be multiplied in four dimensions: content, form, space and time. Future directions for research are discussed, including dictionary user studies and multimodal corpora. The findings of this article shed light on the construction of a theoretical model for e-lexicography which remains an urgent task in the ongoing digital revolution.

Keywords: MULTIMODALITY, MULTIMODAL DEFINITION, MULTIMODAL LEXICOGRAPHY, E-DICTIONARY, MODE SELECTION, INTERMODAL SYNERGY

Opsomming: Multimodale definisie: Die verveelvoudiging van betekenis in elektroniese woordeboeke. Met die oplewing in elektroniese woordeboeke speel multimodaliteit 'n toenemend belangrike rol in leksikografie. Om die bestek van e-woordeboekdefinisies te vergroot, word daar in hierdie artikel geargumenteer ten gunste van die vestiging van 'n nuwe term, multimodale definisie, as 'n sleutelkomponent van die multimodale leksikografie. Hierdie term word vanuit 'n holistiese betekenisoogpunt gewy aan die integrering van verbale definisies en die komplementêre multimodale bronne. In 'n multimodale definisie kan 'n dinamiese betekenisekologie geskep word waar twee kritiese veranderlikes funksioneer, (semiotiese) modusseleksie en intermodale sinergie. In hierdie ekologie kan betekenis wat verbaal uitgedruk word, in vier dimensies verveelvoudig word: inhoud, vorm, ruimte en tyd. Toekomstige navorsingsrigtings, insluitende woordeboekgebruikerstudies en multimodale korpora, word bespreek. Die bevindings van hierdie artikel werp lig op die samestelling van 'n teoretiese model vir e-leksikografie wat 'n dringende taak in die voortgesette digitale revolusie bly.

Sleutelwoorde: MULTIMODALITEIT, MULTIMODALE DEFINISIE, MULTIMODALE LEKSIKOGRAFIE, E-WOORDEBOEK, MODUS-SELEKSIE, INTERMODALE SINERGIE

1. Introduction

In the new era of digital media, electronic dictionaries (e-dictionaries) are booming as a common tool of language learning and use. In the field of e-lexicography, such core notions as multimodality and meaning representation have aroused the interest of researchers.

As "the normal state of human communication" (Kress 2010), multimodality is defined as the diverse ways in which multiple semiotic resources (1anguage, visual images or sound, etc.) are co-deployed and co-contextualized in the making of a text-specific meaning (Thibault 2001). The collection of different semiotic modes (textual, visual, audio, spatial, etc.) can determine how an audience perceives information. Therefore, multimodality is crucial for meaning representation in dictionaries, especially in e-dictionaries.

As the heart of meaning representation in a dictionary, definition plays a key role in stating or explaining the meaning of a word or phrase. This term, if used as a countable noun, can also refer to the product of the process of defining, as indicated under the entry of **definition** in *Oxford Advanced Learner's Dictionary of Current English* (the 6th edition, 2000):

> **definition**
> *noun*
> 1 [C, U] an explanation of the meaning of a word or phrase, especially in a dictionary; the act of stating the meanings of words and phrases

To deepen our understanding of meaning representation in e-dictionaries, we propose the term multimodal definition under the umbrella of multimodal lexicography proposed by Lew (2010). It is tentatively defined as "(the act of) meaning explanation of a word or phrase with multimodal devices for achieving better defining effect than language does alone, especially in an e-dictionary". Meanings explained by written language are called verbal definitions in this research. By viewing verbal definitions and the complementary multimodal resources as an organic whole and combining them into one cover term, the close relationship between them would be valued, and the horizons of e-dictionary definitions could be expanded.

So a holistic view of meaning is taken in this article. On the one hand, following the theory of "meaning as use" (Wittgenstein 1953), meaning and usage are inseparable, and "in a living language vocabulary and grammar do not have their own independent existences" (Tarp 2008: 135). On the other hand, cognitive findings tell us that linguistic information of a word or phrase is blended with encyclopedic information. In brief, meaning is of multiple facets integrated into a whole.

In the above-mentioned context, this article argues for the establishment of the notion of multimodal definition and aims to investigate the mechanism and function of multimodal definitions in terms of meaning representation. Within

a unified framework of multimodal lexicography (Lew 2010) and multimodal discourse analysis[1] from a systemic functional perspective (O'Halloran 2008), the present study was guided by the following research questions:

1) Is the concept of multimodal definition necessary in practice and theory?
2) What are the dominant variables operating in the semantic ecology of multimodal definitions?
3) In which dimensions can meaning be multiplied in this ecology?

2. Literature review

2.1 Limitations of verbal definitions

Verbal definitions are of various types, ranging from the traditional genus-differentia format to the recent full-sentence definition (Atkins and Rundell 2008: 416-441). However, many studies show that such a variety still can't ensure that verbal definitions are effective enough to meet the personal needs of dictionary users (Zhang 2002: 133-145). Problems like circularity, inaccuracy, over-defining, redundancy or lack of necessary pragmatic and cultural information can be found even in leading dictionaries (Zhang 2015: 79-82). The problem of inaccuracy seems to be intensified by the controlled defining vocabulary in learner's dictionaries and children's dictionaries (Feng 2009: 153-154). Furthermore, the lexicographic treatment of words expressing abstract relations, such as that of prepositions, is still unsatisfactory in many dictionaries (Coffey 2006; Lew 2010).

Such problems often seem unavoidable. For one thing, meaning is often subtle, multiple and even capricious so that it is hard to fully grasp it. For another, a verbal definition is, by its nature, a rather abstract construct (Atkins and Rundell 2008: 454), whether it is at the same semiotic level (periphrastic synonymy) as the defined or at the level of a constructed metalanguage. So there are many cases where the full sense of a difficult concept only becomes clear with a graph or video clip illustrating its contextual features and usage (cf. ibid: 454). Definition devices beyond words often complement verbal definitions and can even provide inspirations for lexicographers to optimize the verbal information. For example, solutions to the above problems, like using schematic graphs to illustrate the meaning of prepositions advocated by Adamska-Sałaciak (2008), often entail the use of multimodal devices, especially in e-dictionaries.

2.2 Meaning representation problems in e-dictionaries

In terms of meaning representation, many studies show there is much room for improvement in current e-dictionaries. Electronic dictionaries are often regarded as inferior in quality (Tono 2009: 48; Chen 2010), and they are still to a large extent presented in a form similar to printed formats, so the potential is still under-utilized (Prinsloo 2012). Some scholars complain about the unimagi-

native use of the new powerful technologies in specialized online dictionaries (Caruso 2014; Fuertes-Olivera and Tarp 2014). Many popular online bilingual dictionaries translate in a fairly primitive way without information or labels to indicate register differences or restrictions on use (Nesi 2012). Many local e-dictionaries in East Asia, like the widespread *Jin Shan Ci Ba* in China, are of poor quality, as evidenced by mistranslations and lack of information on usage (Zhang 2004; Nesi 2012). As shown by Kurtz (2012), there is inaccurate information in *You Dao*, a very popular e-dictionary in China.

In short, researchers find that too many lexicographic e-products were developed without any influence from innovative theoretical suggestions, and as a result current e-dictionaries too often do not live up to the expectations of their users (Gouws 2014). Therefore, a theoretical anchor point is needed to enable a breakthrough in improving e-dictionary definitions.

3. Rationale for proposing multimodal definition

3.1 Practical rationale

As illustrated above, the limitations of verbal definitions and the meaning representation problems in e-dictionaries justify the necessity of a new horizon of defining in the digital revolution. The establishment of the notion of multimodal definition is expected to help solve those problems by arousing lexicographers' attention to the multimodality of e-dictionary definitions and awareness of how definitions can be improved both verbally and non-verbally. In fact, the multimodality of the definitions in e-dictionaries is quite different from that in paper dictionaries. The former involves many more complex factors for consideration in lexicographical practice.

Furthermore, since the e-dictionary is often used as a vocabulary learning tool by language learners (Nesi 2010), they may depend even more on e-dictionaries for definitions in the new era of multiliteracies. The widespread use of e-dictionaries embedded in multimodal texts has intensified the urgent need to solve the current problems in meaning representation. A conceptual expansion in defining theories may give fresh impetus to change at this watershed of e-dictionaries. As Amsler (2012) warns, fundamentally, electronic dictionaries "think" of themselves as print dictionaries being offered via electronic access, and this is a very limiting vision.

3.2 Theoretical rationale

In essence, the poor quality of e-dictionaries can be traced to the lack of theoretical models which lexicographers could rely on. The golden opportunity to lead the way in devising models for e-dictionaries was unfortunately not utilized sufficiently by metalexicographers (Gouws 2014). Where practical lexicog-

raphers did utilize lexicographic theory, it often was a version that had been devised for printed dictionaries.

So it is necessary to devise theoretical models for e-dictionaries that focus on critical areas like the data to be included in these dictionaries, the structures to present and accommodate the data, and the way they should respond to the needs of their target users (Gouws 2014). In fact, all these areas are closely related to the meaning representation and multimodality of e-dictionaries. What's more, the recent findings in multimodal discourse analysis (O'Halloran 2011: 120), such as intermodal relations, can shed light on how to construct theoretical models, but those findings have not been fully exploited by lexicographers.

Lew (2010) proposed the framework of multimodal lexicography by illustrating various modes of meaning indication in e-dictionaries and showing how they can complement verbal definitions. With a solid base in e-lexicography, his research gained support from many researchers, such as Yang (2012) and Luo (2012). However, multimodal lexicography covers too broad a range and a clear picture of its structure and mechanism is still lacking. It is of great significance to investigate one of its key components, multimodal definition, especially from the perspective of multimodal discourse analysis.

After all, the concepts of "multimodality" and "definition" both focus on meaning and meaning-making. Theoretically they can be integrated naturally across discipline boundaries between multimodal discourse analysis and lexicographical definition.

In brief, only when the status of multimodal definition is recognized as a major component of multimodal lexicography, can its nature be fully understood and theoretical models and guiding principles be established for lexicographers.

4. The semantic ecology of multimodal definitions

4.1 Overview of the semantic ecology

From a holistic viewpoint, the meaning system of a multimodal definition can be compared to an ecology and it is characterized by complexity, diversity and dynamicity. There are interactions among its members, as well as cooperation and competition. Its components or communities are conceptually linked together as an integrated whole in a hierarchy. A close look at the components of this semantic/meaning ecology can help understand e-dictionary definitions.

In the trichotomy of multimodal devices proposed by Lew (2010), there are three categories: written language, audio presentation of the verbal elements including human voice recordings and synthesized speech, and other devices. Considering the differences between paper dictionaries and electronic ones, the first category is labeled as the verbal mode/definition, and the last two are non-verbal ones in this study.

The third category of Lew (2010) can be further divided into five types:

1) audio (recordings of non-linguistic sounds);
2) visual (static pictorial illustration, photographs, graphs, silent animations, tables, icons, typography etc.);
3) audio-visual (animations and video clips);
4) spatial (framing[2], hyperlinks, floating tips[3], sense menus, pop-up windows etc.) and
5) other affordances to be realized (e.g. tactile or gustatory modes).

All these members coexist in a semantic ecology tentatively illustrated by Figure 1.

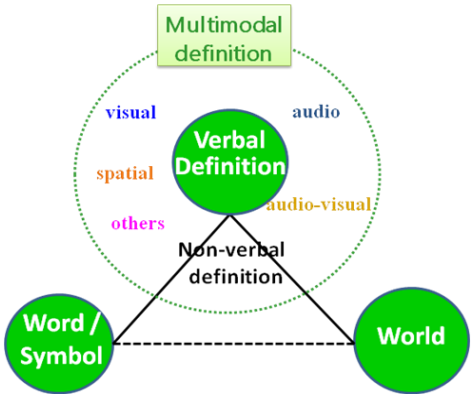


Figure 1: Semantic ecology of multimodal definitions

The top of Figure 1 shows the relationships between the two components of a multimodal definition: verbal and non-verbal, with the former at the core and latter on the periphery. In this sense, the latter can "wrap" and "enlarge" (multiply) the meaning expressed by the former.

Following the semantic triangle of Ogden and Richards (1923), there are two relating ecologies: the word/symbol and the world. From a cognitive perspective, a non-verbal definition can shorten the distance between the verbal definition and the defined word or phrase, serving as a link between the verbal definition and the outside world, hence better defining effect is likely to be achieved.

4.2 Dominant variables of the semantic ecology

Based on Multimodal Discourse Analysis (Kress and Van Leeuwen 1996; Royce 1998, 2007; O'Halloran 2008; Chan 2011: 144-165) from a Systemic Functional perspective (Halliday 1985), two dominant variables are found to be operating in a robust semantic ecology of the multimodal definitions: mode selection and intermodal synergy. The first means the choice of semiotic modes, and the latter refers to the cooperation and interaction among different modes.

4.2.1 Mode selection

Mode selection is a function of many variables, such as the purpose of compilation, number of languages involved, type of medium used, target users, items to be defined and semiotic modal resources available. The purpose of a dictionary can be language comprehension or production. It can be monolingual, bilingual or multilingual. Its medium can be a desktop, laptop, tablet or mobile phone, etc. Its target users can be children or adults, foreign language learners or native speakers. The items to be defined can be concrete or abstract, commonly-used or rare. Additionally, modal resources are different from each other in how to represent meaning, and a good multimodal corpus should give definers a wide range of options to choose from.

Most important of all, it is the correspondence between the nature of the item to be defined and the characteristics of a semiotic mode that plays an essential role in mode selection. We have some tentative suggestions about the correspondence while a systematic study of mode selection is still to be conducted.

First, there is usually an audio-audio or visual-visual match between the item to be defined and the mode for meaning explanation. For example, if the word is related to some sound, an audio recording of that sound is of course a good choice for complementing its verbal definition.

Second, there can be a concrete-concrete or abstract-abstract match between them. However, an abstract-concrete match may be more effective for young target users of the dictionary, and a concrete-abstract match may be enough for adult users if storage space is limited (as in a mobile phone dictionary). If an abstract process is to be defined, a flow chart alone may be enough for adults, but an animation or video clip is often more effective for children.

Third, typicality is one of the most important factors to be considered in mode selection. Non-verbal modes should provide the most typical context of language use for instantiation. This is crucial for learner's dictionaries, following the Theory of Family Resemblance (Wittgenstein 1953) and the Prototype Theory (Rosch 1973). For instance, a robin may be a prototype for the headword "bird", so a picture of a robin is preferable to those of other kinds of birds. When defining a verb, the most typical semantic roles/arguments are often desirable or even essential. Taking "wash" as another example, a scene of washing clothes or hands is preferred to one of washing something rare.

Fourth, the differences between semiotic modes (especially language vs. images) should be taken into account. A verbal text unfolds over time in a dynamic, sequential way and language has a rich potential for the construal of temporal deixis, sequencing, location phasing and aspect, and this is in contrast with the instantaneous holistic apprehension of an individual image and the corresponding potential of the visual semiotic for nonsequencing spatial and comparative relationships (Painter et al. 2013: 133).

Last, mode selection should be very careful as some modal resources may

impede the learning process of users if not properly designed or chosen. There is evidence that users who viewed animations for the target entries had significantly lower vocabulary retention rates than those who ignored the animations (Lew and Doroszewska 2009).

4.2.2 Intermodal synergy

In a robust semantic ecology, the synergy between different semiotic modes is critically important, especially the text-image (verbal-visual) interface which is in the heart of multiliteracies.

Identifying cohesive links and examining the logical relations in ideational meaning that extend across semiotic modes may help us gain an in-depth understanding of intermodal synergy. In this study, Chan's (2011: 144-165) model of intermodal relations has been modified to suit e-dictionary definitions. Based on her framework, intermodal synergy falls into two categories: concurrence and complementarity (see Figure 2), each of which can be further divided into four types, forming a hierarchical system.

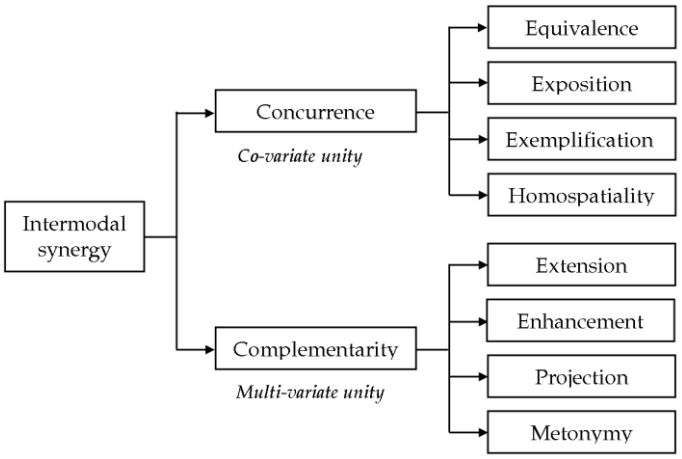


Figure 2: Toward intermodal synergy in multimodal definitions

4.2.2.1 Intermodal concurrence

The (intermodal) concurrence or agreement, where meanings across modes are similar, is one of co-variate unity (ibid: 144-165). Concurrence may entail some form of redundancy across modes, but this is not a simple inter-modal duplication of meaning (Unsworth 2006). It includes four basic types: equivalence, exposition, exemplification and homospatiality[4].

Equivalence of meaning across semiotic modes means that they are mutu-

ally identifying, following Gill (2002). Being most common in both paper and electronic dictionaries, equivalence can be partial or full. For example, the verbal definition of "napkin" is accompanied by the (partially) equivalent picture in the online Longman Dictionary of Contemporary English (http://www.ldoceonline.com/dictionary/napkin) (see Figure 3).

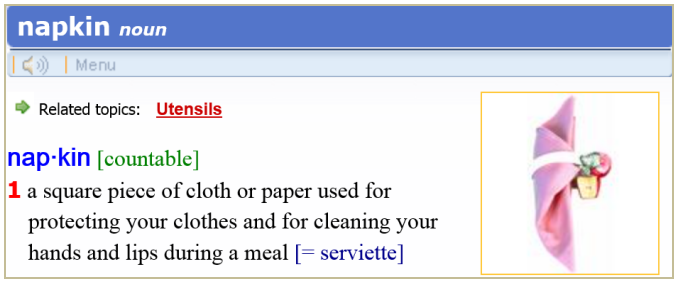


Figure 3: *Napkin* in LDOCE

Exposition means that, by expressing the same meaning in different ways, different modes reinforce each other. An example (see the entry of **lever** in Figure 4) is taken from the online Merriam-Webster Visual Dictionary (http://visual.merriam-webster.com/science/physics-mechanics/lever.php). The picture is not just an equivalence of the word "lever", but also a vivid illustration of the mechanism of the equipment stated in the verbal definition, with arrows pointing various directions.

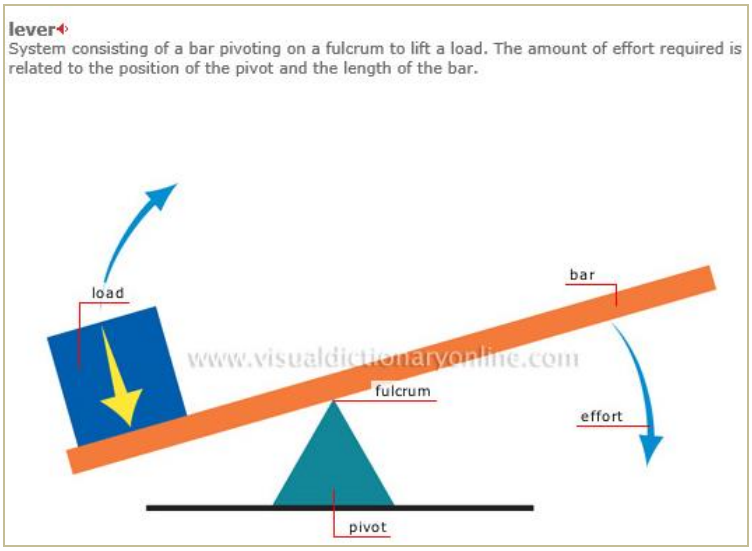


Figure 4: *Lever* in MWVD

Exemplification refers to a case in which a non-verbal mode exemplifies the verbal, giving an example to make it clear. The modes represent different levels of generality (Martinec and Salway 2005), and are of a class-member relationship. An example taken from Oxford Learner's Dictionary is the charts (bar, flow and pie) under the entry of **flow chart** (http://www.oxfordlearnersdictionaries.com/definition/english/flow-chart?q =flow+chart). Figure 6 is a pop-up window emerging when the user clicks the icon in the lower right corner of Figure 5.

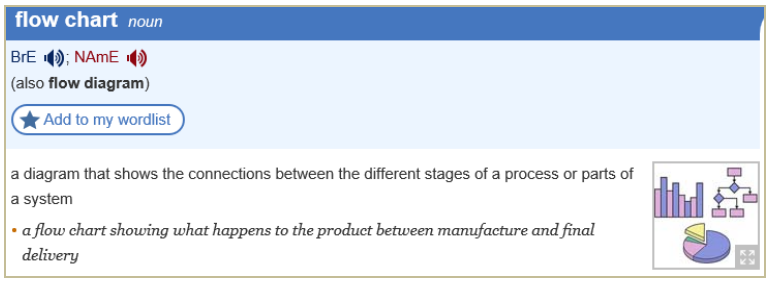


Figure 5: *Flow chart* in OLD

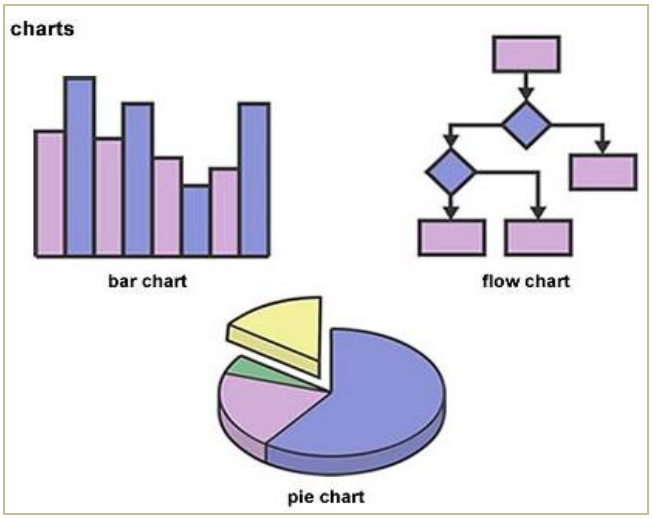


Figure 6: *Charts* in OLD

Homospatiality means that different semiotic modes co-occur in one spatially bonded homogenous entity (Lim 2004; Unsworth 2006). An example of homospatiality is the word "arc" arranged in the shape of an arc, or the word "blue" printed in the color blue. It is rare in dictionaries. The following picture accompanying the linguistic meaning of the word "onion" may be regarded as a vari-

ety of homospatiality (cited from Unsworth 2006), the picture and the word overlap each other, and the latter is in purple, a prototypical color of onions.

Figure 7: Homospatiality of *onion*

4.2.2.2 Intermodal complementarity

The (intermodal) complementarity, where meanings across modes are different but complement each other, is oriented towards multivariate unity (Chan 2011: 144-165). In Chan's (ibid: 144-165) model, it is divided into three basic types: extension, enhancement and projection. But we have found a new type, metonymy, in multimodal definitions. These distinct but interrelated types are respectively illustrated as follows.

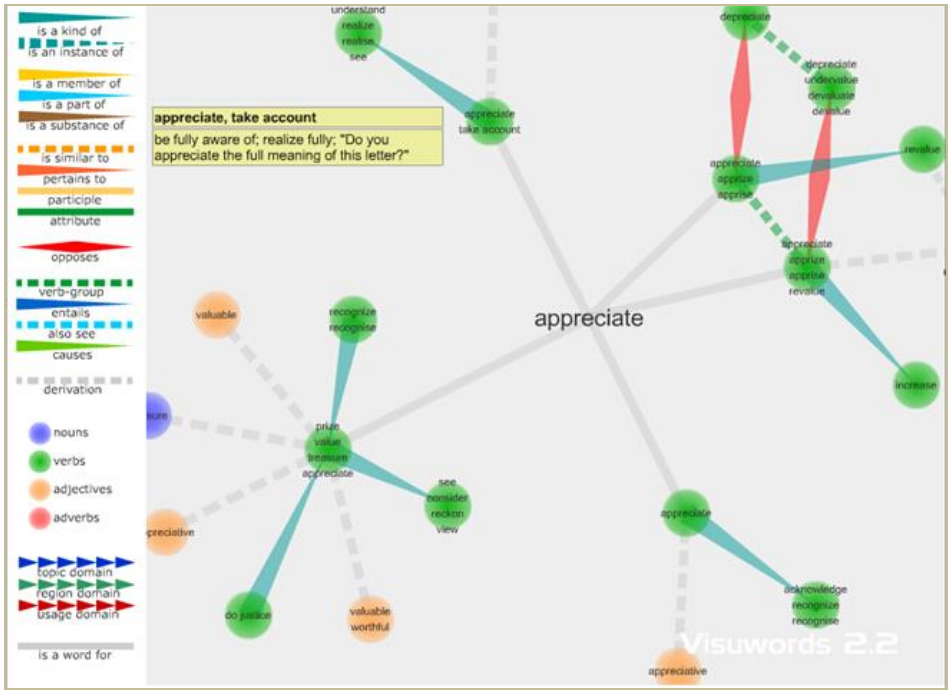


Figure 8: A floating quasi-3D graph for *appreciate* in VISUWORDS

Extension means that meanings additional to those in one mode are represented in another (Unsworth 2006, 2008). New, related information is added. For example, images might illustrate the inherent relationship between the defined item and related ones in the semantic network. A case in point is a floating quasi-3D graph (see Figure 8), demonstrating the sense relations for the entry of **appreciate** in the VISUWORDS interface (http://www.visuwords.com/, cf. Lew 2010).

Enhancement means that one mode provides meanings which expand another spatially, temporally or causally, etc. The image might depict the end result of a process described in the verbal text (Unsworth 2006). If a word has negative connotations or semantic prosody, like "funeral", the audio recording can carry an unhappy tone to achieve enhancement.

Projection originally means the use of speech or thought bubbles, but in e-dictionaries, the floating tips are the major type of projection, which facilitates fast access to meaning. In VISUWORDS (see Figure 8), the verbal definition emerges when the mouse rests on any word in the graph, and such dynamicity enables more interactions between the user and the dictionary. So is Visual Thesaurus (http://www.visualthesaurus.com).

Metonymy, if cross-modal or multimodal, refers to a contiguity relationship between different modes, especially a part-whole one. It has mostly a referential or predictive function. As an example, the entry of **wild** is taken from LDOCE (http://www.ldoceonline.com/dictionary/wild_1). In Figure 9, the picture of a tiger is used to describe the adjective "wild", since being wild is a dominant feature of a tiger and this animal is a prototype symbolizing wildness. The user's cognitive process involves four major steps: the tiger head, the whole tiger, a typical wild animal/thing and the nature of being wild. It is metonymy that facilitates every transition between consecutive steps. The image of a tiger head is metonymic for it represents the whole body of the tiger. The tiger is also metonymic since it stands for any wild animal/thing. So is the typical wild animal/thing as it symbolizes the quality of being wild.

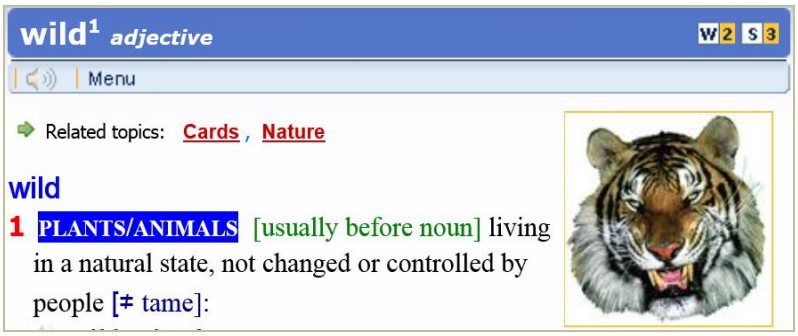


Figure 9: Metonymy of *wild* in LDOCE

The above may not be a complete list of the logical relationships between verbal and non-verbal modes in multimodal definitions. But what seems certain is the relationships are unidirectional in most cases because the verbal mode is regarded as central and non-verbal modes are intended to complement or reinforce it. Where decoding linguistic meaning was unsuccessful or only partially successful, dictionary users would rely on non-verbal modes to support their interpretation. Complementary items were found to be more difficult than those where there was concurrence of meaning across modes, and among concurrent relationships, exposition was more difficult than equivalence (Chan 2011: 144-165). So equivalence is most recommended for dictionaries designed for elementary language learners and/or dictionary users.

5. Dimensions of meaning multiplication

In the semantic ecology of a multimodal definition, meaning expressed verbally can be enriched, expanded or multiplied by the combination of verbal and non-verbal modes (cf. Royce 1998; Lemke 2002; O'Halloran 2008) in various dimensions. The concurrence or complementarity (Chan 2011: 144-165) between different modes amplifies meaning. For a detailed analysis from a systemic functional perspective, we employ the widely accepted framework of three metafunctions of the semiotic system: ideational, interpersonal and textual (Halliday 1985). They can be respectively realized by the dimensions of content, form and space in multimodal definitions. A fourth dimension, time, is added for a longitudinal view (see Figure 10).

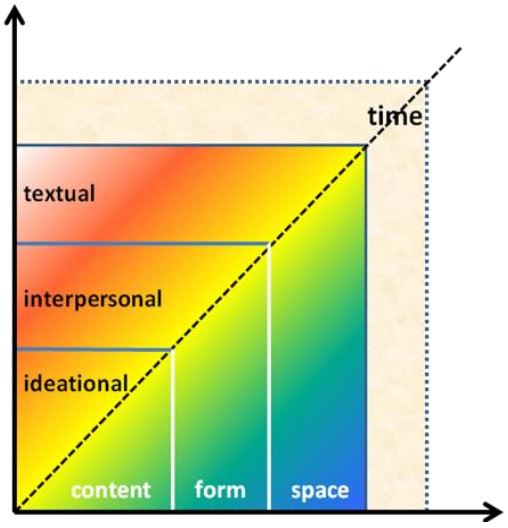


Figure 10: Meaning multiplication by multimodal definition

5.1 Content

The ideational metafunction is the most important for the defining of dictionaries because it is closely related to the content of meaning. Concerning the material reality, ideational structures construct the nature of events, the objects and participants, and the circumstances (Unsworth 2006) which can be better represented by multimodal devices than language alone does, enriching the content of meaning in a dictionary.

With multimodal devices, e-dictionaries often have the ability of showing more senses of a word and sense relations between words than their paper versions when necessary. This is an explicit and direct augmentation of meanings.

Implicitly, e-dictionaries are often more powerful in showing the characteristic co-text and context of a word/phrase which constitute part of the meaning. In such circumstances, encyclopedic and pragmatic meanings can be easily added. For instance, in bilingual dictionaries, when some culture-specific or confusing words/phrases are difficult to explain by language, proper multimodal devices are likely to vividly clarify or accurately represent the opaque meaning expressed by words, such as pictures showing the differences between "Roman nose" and "Grecian nose", and those between "in front of" and "in the front of". As elaborated above, meaning can be multiplied in the interactions between different modes, especially through intermodal concurrence and complementarity which add to the meaning of language.

Last, new media enable more interactions between the user and the e-dictionary (compiler/writer), which is part of the process of making meaning and has the potential of enriching meaning. In this sense, the ideational metafunction of multimodal definition is interwoven with its interpersonal metafunction.

5.2 Form

Interpersonally, a dictionary entry constructs the compiler/writer as a giver of information and the user as a receiver. In terms of an e-dictionary, the user often has more opportunities to probe meaning information than he does when passively receiving it from a printed one. Its multimodality can actually change the interpersonal relationship. The interpersonal metafunction is mainly realized by the forms of expressing meaning and it is often embodied by user-friendliness of dictionaries. Interpersonal resources construct the nature of relationships among addressers and addressees, and viewers and the viewed. Besides language, visual, audio and spatial modes provide diverse ways of showing meaning. The presentation of lexicographical data can take many different forms, depending on the needs of specific users and the situations with which they are confronted (Tarp 2008).

As electronic dictionaries start to exploit the opportunities of the medium more imaginatively, they could well offer users a range of defining options, like

word clouds and semantic networks. Figure 11 shows the word cloud "deduction" in *Cambridge Dictionaries Online* (http://dictionary.cambridge.org/dictionary/british/deduction), where the size, color and font etc. of each word all represent meaning.

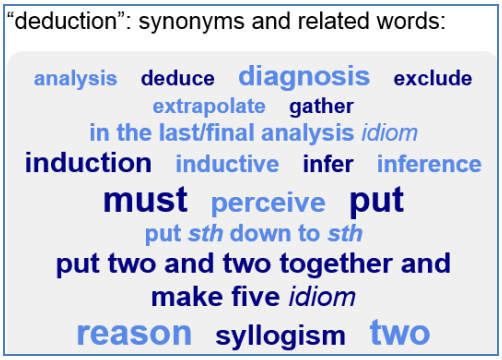


Figure 11: Word cloud of *deduction* in CDO

Other new affordances of the digital era are gaze-contingent systems utilizing modern eye tracking (Bulling and Gellersen 2010) and a system like Google Glass (Lew and De Schryver 2014). With the development of digital technology, more modes like the gustatory might be realized. For instance, the smell could be transmitted to the user via the internet someday. The diversity of the forms of such multimodal resources is likely to help improve the user-friendliness of a dictionary if specific user needs can be considered and satisfied.

5.3 Space

The textual metafunction is mainly realized by the space arrangement in multimodal definitions. Textual meanings concern the distribution of the information value or relative emphasis among elements of the multimodal text (cf. Kress and Van Leeuwen 1996: 175-214; Unsworth 2006).

Generally speaking, a dictionary entry locates the headword as the "Theme" or given information, and the "Rheme" or new information is the following verbal definition. If the verbal definition is complemented by non-verbal resources, it may become the given information in comparison with the latter, the new. Thus a cumulative given–new structure is formed. In Figure 3, the verbal definition of "napkin" (old information) is placed on the left while the picture (new information) is on the right because human beings usually read from left to right. In Figure 4, the verbal definition is old information on the top, and the picture is new information below it for we usually read from top to bottom. The information on the top is usually ideal, but that on the bottom is real, specific and detailed (Kress and Van Leeuwen 1996: 186). But a picture

may distract the reader's attention from the verbal definition, so the reading path is not necessarily from left to right, and from top to bottom.

Non-verbal resources cohere into textual compositions in different ways. E-dictionaries can offer a layered, hierarchical inner access structure (as indicated by Figure 5 and Figure 6). If the user selects a particular sense from the menu displayed at the top of the entry (Tono 2000: 855; Lew 2011), the space for a specific definition can be adapted flexibly to the user need. This type of layered presentation embodies the role of the medium in developing and altering patterns of electronic dictionary consultation as well as its effects on language reception, production, and learning (Dziemianko 2012: 321).

Furthermore, the interconnections between different pieces of information, including the ubiquitous hyperlinks and internal links to word-processing applications allow users to copy text from dictionary to document.

5.4 Time

The last dimension is time. E-dictionaries can renew its information more frequently and easily than their print counterparts. This can help guarantee the recency and accuracy of meaning which prove to be the major concerns of users (Müller-Spitzer et al. 2012). For instance, some online dictionaries let users help update and add new senses to an entry.

In addition, the users' consultation history can be recorded. The use of cookies, i.e. small text files stored on the user's hard drive, allows the server to uniquely identify a returning visitor (De Schryver and Joffe 2006: 69; Verlinde and Peeters 2012: 150). This inherently helps build the semantic network in the user's mind during the process of learning.

In brief, multimodal definitions can extend the meaning system chronically for the sustainable development of (e-)dictionaries. In essence multimodal definition is dynamic and vigorous.

6. Future directions for research

Much recent literature has shown us two future directions for research: user studies and database construction, which is evidenced by the 2014 special issue of *the International Journal of Lexicography*, *Using Online Dictionaries* (Müller-Spitzer 2014) and *Electronic Lexicography* (Granger and Paquot 2012), etc. In the field of multimodal lexicography, further explorations of such critical areas as mode selection and intermodal synergy should be made in these two directions.

6.1 User research

E-dictionary user research has attracted a lot of attention recently, but there is still much work to do (Tarp 2009; Nesi 2013; Lew and Schryver 2014). While dictionary use has moved dynamically into the digital medium, user research

on digital dictionaries has been somewhat slow in coming (Lew 2015). Lew (2010) points out factors to be considered in the design of multimodality, including dictionary culture, consultation goals and context, language proficiency level in general and in the specialized domain, and type of the lexical item to be defined.

The methods of investigating user studies are various. Welker (2010) identifies six of them: questionnaire surveys, interviews, observation, tests and experiments, protocols and log files. New laboratory-based methods have enabled researchers to observe in detail the way users interact with dictionary information on the computer screen (Nesi 2013). Representative examples are eye-tracking techniques from the cognitive science and psycholinguistics (e.g. Kaneta 2011; Simonsen 2011; Tono 2011; Lew et al. 2013) and usability testing methods from the information science (Heid and Zimmermann 2012). In the future, neuro-cognitive methods may be introduced, like Event-Related Potentials and Functional Magnetic Resonance Imaging, to explore how the mind works during dictionary consultations. Those advanced techniques can reveal microscopic aspects of dictionary use that probably cannot be discovered by behavioral observations or experiments.

In fact, previous user studies have shown different or even contradictory findings, such as those on the relationship between dictionary users' strategies and language learning (Gavriilidou 2013). Such variation of results may not be as contradictory as they seem, if we consider that the differences found could be attributed to the lack of investigations into the stratification of dictionary users. No previous research, however, has stressed the importance of such investigations. User classification, especially the one disclosing the latent or underlying classes among users, may serve as a mediating or moderating variable in related quantitative studies, including those aimed at optimizing the design of multimodal resources. A holistic analysis of these resources should consider the diversity of modes and submodes involved in meaning making while adopting a dynamic perspective on the users' cognitive processes and their taxonomy.

6.2 Multimodal corpora

Multimodal corpora in line with lexical databases, if annotated in the auditory-visual domains, can provide useful resources for dictionary making in automated processes.

Online data mining seems to be indispensable to the construction of multimodal corpora. Fujii and Ishikawa (2005) proposed a method for searching the Web, seeking images associated with a specific word sense to generate content for multimedia encyclopedias. A particularly convincing implementation of the effective visualization of semantic relations is the VISUWORDS interface (www.visuwords.com), serving as a front-end to English WordNet data (Lew 2010).

Unfortunately, annotating multimodal data is still problematic for both theoretical and technical reasons: first, there is a lack of linguistic and cognitive theories taking into consideration all the different aspects of multimodality; second, we need to specify a standardized way of representing multimodal information in order to give access to large multimodal corpora, as richly annotated as possible (Blache et al. 2009). So there is still a long way to go in this field.

7. Conclusion

In summary, a holistic view of meaning and a systemic-functional approach to the semiotic system have been adopted to broaden the horizons of e-dictionary definitions. Located at the intersection of multimodal discourse analysis and dictionary definition, multimodal definition is characterized by:

1) one key relationship (between multimodal devices and defining effect);
2) two key words (multimodality and definition) and two dominant variables (mode selection and intermodal synergy);
3) three metafunctions (ideational, interpersonal and textual);
4) four dimensions of meaning multiplication (content, form, space and time).

The three research questions raised at the beginning of the article have been answered. We think multimodal definition deserves its own place within the framework of multimodal lexicography since both practical and theoretical rationales support the necessity of having this new term. Only with proper mode selection and optimal intermodal synergy, can a robust semantic ecology be maintained in a multimodal definition. As it multiplies meaning in content, form, space and time, multimodal definition has great potential for transcending the limitations of verbal definitions and alleviating the meaning representation problems in current e-dictionaries. It would usher e-lexicography to a new era if the present theoretical and technical challenges could be overcome.

This study was inspired by and based on the pioneering work of Lew (2010). By examining online dictionaries, his trichotomy of multimodal devices was refined, with some new submodes added. Furthermore, a new type of intermodal synergy, (multimodal) metonymy, was found in e-dictionaries, complementing previous research in the field of multimodal discourse analysis with lexicographical evidence.

All these findings have cast some light on the construction of a theoretical model for e-lexicography, but the current study has its limitations for being largely dependent on retrospective analysis. A detailed plan was not worked out on systematic investigations into mode selection and intermodal synergy in multimodal definitions. This leaves much room for future work. It should also be noted that, to build a complete theoretical system of multimodal lexicography, another important subfield, multimodal exemplification, might be explored in

line with multimodal definition.

Notes

1. Multimodal discourse analysis (MDA) is an emerging paradigm in discourse studies which extends the study of language per se to the study of language in combination with other resources, such as images, scientific symbolism, gesture, action, music and sound (O'Halloran 2011: 120).
2. Framing is the use of boundaries connecting or disconnecting one space from another (Van Leeuwen 2005: 18).
3. A floating tip is a temporary window (usually small) indicating the meaning or translation etc. of a word/phrase and emerging only when the mouse rests on it.
4. A cover term "elaboration" is used for all the types by some researchers (cf. Unsworth 2006, 2008; Painter and Martin 2011).

Acknowledgements

I would like to express my deep gratitude to Prof. Yihua Zhang and Dr. Liang Li of Guangdong University of Foreign Studies (GDUFS) for their discussions about dictionary definitions. I am greatly indebted to Prof. Gunther Kress, Prof. Len Unsworth, Dr. Francis Low and Dr. Yiqiong Zhang for their talks at the forums on multiliteracies hosted by GDUFS and the Hong Kong Polytechnic University in 2014 and 2015. Special thanks are also due to Dr. Dongping Zheng of the University of Hawai'i at Mānoa and the anonymous reviewers for their comments and suggestions on earlier drafts of the article.

Bibliography

A. Dictionaries

CDO. *Cambridge Dictionaries Online.* http://dictionary.cambridge.org.

LDOCE. *Online Longman Dictionary of Contemporary English.* http://www.ldoceonline.com/dictionary.

MWVD. *Online Merriam-Webster Visual Dictionary.* http://visual.merriam-webster. com.

OALD. Wehmeier, Sally (Ed.). 2000. *Oxford Advanced Learner's Dictionary of Current English.* Sixth edition. Oxford: Oxford University Press.

OLD. *Oxford Learner's Dictionary.* http://www.oxfordlearnersdictionaries.com.

VISUWORDS. *VISUWORDS Online Graphical Dictionary.* http://visuwords.com.

VT. *Visual Thesaurus.* http://www.visualthesaurus.com.

B. Other references

Adamska-Sałaciak, Arleta. 2008. Prepositions in Dictionaries for Foreign Learners: A Cognitive Linguistic Look. Bernal, E. and J. DeCesaris (Eds.). 2008. *Proceedings of the XIII EURALEX International Congress, Barcelona, 15–19 July 2008*: 1477-1485. Sèrie Activitats 20. Barcelona:

Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada.

Amsler, Robert. 2012. *The Future of Our Field.* http://www.freelists.org/post/euralex/The-future-of- our-field-Part-1.

Atkins, B.T. Sue and Michael Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford/ New York: Oxford University Press.

Blache, Philippe, Roxane Bertrand and Gaëlle Ferré. 2009. Creating and Exploiting Multimodal Annotated Corpora: The ToMA Project. Kipp, Michael, Jean-Claude Martin, Patrizia Paggio and Dirk Heylen (Eds.). 2009. *Multimodal Corpora: From Models of Natural Interaction to Systems and Applications*: 38-53. Berlin/Heidelberg: Springer-Verlag.

Bulling, Andreas and Hans Gellersen. 2010. Toward Mobile Eye-Based Human-Computer Interaction. *IEEE Pervasive Computing* 9(4): 8-12.

Caruso, Valeria. 2014. Key Issues in Fuertes-Olivera and Tarp's *Theory and Practice of Specialised Online Dictionaries*. *Lexikos* 24: 362-377.

Chan, Eveline. 2011. Integrating Visual and Verbal Meaning in Multimodal Text Comprehension: Towards a Model of Intermodal Relations. Dreyfus, Shoshana, Susan Hood and Maree Stenglin (Eds.). 2011. *Semiotic Margins: Meaning in Multimodalities*: 144-167. London: Continuum.

Chen, Yuzhen. 2010. Dictionary Use and EFL Learning. A Contrastive Study of Pocket Electronic Dictionaries and Paper Dictionaries. *International Journal of Lexicography* 23(3): 275-306.

Coffey, Stephen. 2006. High-frequency Grammatical Lexis in Advanced-level English Learners' Dictionaries: From Language Description to Pedagogical Usefulness. *International Journal of Lexicography* 19(2): 157-173.

De Schryver, Gilles-Maurice and David Joffe. 2006. The Users and Uses of *TshwaneLex One*. De Schryver, Gilles-Maurice (Ed.). 2006. *DWS 2006: Proceedings of the Fourth International Workshop on Dictionary Writing Systems, Turin, Italy, 5 September 2006:* 41-46. Pretoria: (SF)^2 Press.

Dziemianko, Anna. 2012. On the Use(fulness) of Paper and Electronic Dictionaries. Granger, Sylviane and Magali Paquot (Eds.). 2012: 319-342.

Feng, Chunbo. 2009. *Research on Pictorial Illustrations in English Dictionaries*. Shanghai: The Press of East China University of Science and Technology.

Fuertes-Olivera, Pedro A. and Sven Tarp. 2014. *Theory and Practice of Specialised Online Dictionaries*. Berlin: De Gruyter Mouton.

Fujii, Atsushi and Tetsuya Ishikawa. 2005. Toward the Automatic Compilation of Multimedia Encyclopedias: Associating Images with Term Descriptions on the Web. *Proceedings of the 2005 IEEE/WIC/ACM International Conference on Web Intelligence, Compiegne, France, 19–22 September 2005*: 536-542. Washington, DC: IEEE Computer Society.

Gavriilidou, Zoe. 2013. Development and Validation of the Strategy Inventory for Dictionary Use (SIDU). *International Journal of Lexicography* 26(2): 135-153.

Gill, Talia. 2002. *Visual and Verbal Playmates: An Exploration of Visual and Verbal Modalities in Children's Picture Books.* Unpublished. Sydney: University of Sydney.

Gouws, Rufus H. 2014. Article Structures: Moving from Printed to e-Dictionaries. *Lexikos* 24: 155-177.

Granger, Sylviane and Magali Paquot (Eds.) 2012. *Electronic Lexicography*. Oxford: Oxford University Press.

Halliday, Michael. 1985. *An Introduction to Functional Grammar*. London: Edward Arnold.

Heid, Ulrich and Jan Timo Zimmermann. 2012. Usability Testing as a Tool for e-Dictionary Design: Collocations as a Case in Point. Fjeld, Ruth Vatvedt and Julie Matilde Torjusen (Eds.).

2012. *Proceedings of the 15th Euralex International Congress, 7–11 August 2012:* 661-671. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo.

Kaneta, Taku. 2011. Folded or Unfolded: Eye-Tracking Analysis of L2 Learners' Reference Behavior with Different Types of Dictionary. Akasu, K. and S. Uchida (Eds.). 2011. *Asialex 2011 Proceedings Lexicography: Theoretical and Practical Perspectives, 22-24 August 2011*: 219-224. Kyoto: Asian Association for Lexicography.

Kress, Gunther. 2010. *Multimodality: A Social Semiotic Approach to Contemporary Communication.* New York: Routledge.

Kress, Gunther and Theo van Leeuwen. 1996. *Reading Images: The Grammar of Visual Design*. London: Routledge.

Kurtz, Lindsey M. 2012. *Learning from Twenty-First Century Second Language Learners: A Case Study in Smartphone Use of Language Learners.* M.A. Thesis. Ames, Iowa: Iowa State University.

Lemke, Jay L. 2002. Travels in Hypermodality. *Visual Communication* 1(3): 299-325.

Lew, Robert. 2010. Multimodal Lexicography: The Representation of Meaning in Electronic Dictionaries. *Lexikos* 20: 290-306.

Lew, Robert. 2011. Online Dictionaries of English. Fuertes-Olivera, Pedro A. and Henning Bergenholtz (Eds.). 2011. *e-Lexicography: The Internet, Digital Initiatives and Lexicography*: 230-250. London/New York: Continuum.

Lew, Robert. 2015. Research into the Use of Online Dictionaries. *International Journal of Lexicography* 28(2): 232-253.

Lew, Robert and Joanna Doroszewska. 2009. Electronic Dictionary Entries with Animated Pictures: Lookup Preferences and Word Retention. *International Journal of Lexicography* 22(3): 239-257.

Lew, Robert, Marcin Grzelak and Mateusz Leszkowicz. 2013. How Dictionary Users Choose Senses in Bilingual Dictionary Entries: An Eye-Tracking Study. *Lexikos* 23: 228-254.

Lew, Robert and Gilles-Maurice de Schryver. 2014. Dictionary Users in the Digital Revolution. *International Journal of Lexicography* 27(4): 341-359.

Lim, Victor Fei. 2004. Developing an Integrative Multi-Semiotic Model. O'Halloran, Kay (Ed.). 2004. *Multimodal Discourse Analysis: Systemic Functional Perspectives*: 220-246. London/New York: Continuum.

Luo, Yongsheng. 2012. A Tentative Study on Image-Word Definition in Multimodal English Learner's Dictionary. *Journal of Guangxi University for Nationalities* 34(1): 165-170.

Müller-Spitzer, Carolin (Ed.). 2014. *Using Online Dictionaries*. Lexicographica. Series Maior 145. Berlin: Walter de Gruyter.

Müller-Spitzer, Carolin, Alexander Koplenig and Antje Töpel. 2012. Online Dictionary Use: Key Findings from an Empirical Research Project. Granger, Sylviane and Magali Paquot (Eds.). 2012: 425-457.

Martinec, Radan and Andrew Salway. 2005. A System for Image-Text Relations in New (and Old) Media. *Visual Communication* 4(3): 337-371.

Nesi, Hilary. 2010. The Virtual Vocabulary Notebook: The Electronic Dictionary as Vocabulary Learning Tool. Blue, George (Ed.). 2010. *Developing Academic Literacy*: 213-226. Oxford: Peter Lang.

Nesi, Hillary. 2012. Alternative e-Dictionaries: Uncovering Dark Practices. Granger, Sylviane and Magali Paquot (Eds.). 2012: 363-378.

Nesi, Hilary. 2013. Researching Users and Uses of Dictionaries. Jackson, Howard (Ed.). 2013. *The Bloomsbury Companion to Lexicography*: 62-74. London: Bloomsbury.

Ogden, Charles Kay and Ivor Armstrong Richards. 1923. *The Meaning of Meaning. A Study of the Influence of Language upon Thought and of the Science of Symbolism.* London: Routledge and Kegan Paul.

O'Halloran, Kay L. 2008. Systemic Functional-Multimodal Discourse Analysis (SF-MDA): Constructing Ideational Meaning Using Language and Visual Imagery. *Visual Communication* 7(4): 443-475.

O'Halloran, Kay L. 2011. Multimodal Discourse Analysis. Hyland, K. and B. Paltridge (Eds.). 2011. *Companion to Discourse:* 120-137. London/New York: Continuum.

Painter, Clare and James Martin. 2011. Intermodal Complementarity: Modelling Affordances across Verbiage and Image in Children's Picture Books. Huang, Guo Wen (Ed.). 2011. *Studies in Functional Linguistics and Discourse Analysis*: 132-158. Beijing: Higher Education Press.

Painter, Clare, James R. Martin and Len Unsworth. 2013. *Reading Visual Narratives: Image Analysis of Children's Picture Books.* London: Equinox.

Prinsloo, Danie J. 2012. Electronic Lexicography for Lesser-resourced Languages: The South African Context. Granger, Sylviane and Magali Paquot (Eds.). 2012: 119-144.

Rosch, Eleanor. 1973. Natural Categories. *Cognitive Psychology* 4(3): 328-350.

Royce, Terry. 1998. Synergy on the Page: Exploring Intersemiotic Complementarity in Page-Based Multimodal Text. *JASFL Occasional Papers* 1(1): 25-49.

Royce, Terry. 2007. Intersemiotic Complementarity: A Framework for Multimodal. Royce, Terry and Wendy Bowcher (Eds.). 2007. *New Directions in the Analysis of Multimodal Discourse*: 63-109. Mahwah, NJ: Lawrence Erlbaum.

Simonsen, Henrik Køhler. 2011. User Consultation Behaviour in Internet Dictionaries: An Eye-Tracking Study. *Hermes* 46: 75-101.

Tarp, Sven. 2008. *Lexicography in the Borderland between Knowledge and Non-knowledge: General Lexicographical Theory with Particular Focus on Learner's Lexicography.* Tübingen: Max Niemeyer.

Tarp, Sven. 2009. Reflections on Lexicographical User Research. *Lexikos* 19: 275-296.

Thibault, Paul J. 2001. Multimodality and the School Science Textbook. Torsello, Carol Taylor, Giuseppe Brunetti and Nicoletta Penello (Eds.). 2001. *Corpora Testuali per Ricerca, Traduzione e Apprendimento Linguistico*: 293-335. Padua: Unipress.

Tono, Yukio. 2000. On the Effects of Different Types of Electronic Dictionary Interfaces on L2 Learners' Reference Behaviour in Productive/Receptive Tasks. Heid, U., S. Evert, E. Lehmann and C. Rohrer (Eds.). 2000. *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000, Stuttgart, Germany, 8–12 August 2000*: 855-862. Stuttgart: Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.

Tono, Yukio. 2009. Pocket Electronic Dictionaries in Japan: User Perspectives. Bergenholtz, H., S. Nielsen and Sven Tarp (Eds.). 2009. *Lexicography at a Crossroads. Dictionaries and Encyclopedias Today, Lexicographical Tools Tomorrow*: 33-67. Bern: Peter Lang.

Tono, Yukio. 2011. Application of Eye-Tracking in EFL Learners' Dictionary Look-up Process Research. *International Journal of Lexicography* 24(1): 124-153.

Unsworth, Len. 2006. Towards a Metalanguage for Multiliteracies Education: Describing the Meaning-Making Resources of Language-Image Interaction. *English Teaching: Practice and Critique* 5(1): 55-76.

Unsworth, Len. 2008. Multiliteracies and Metalanguage: Describing Image/Text Relations as a Resource for Negotiating Multimodal Texts. Coiro, Julie, Michele Knobel, Colin Lankshear

and Donald J. Leu (Eds.). 2008. *Handbook of Research on New Literacies*: 377-406. Mahwah, NJ: Lawrence Erlbaum.

Van Leeuwen, Theo. 2005. *Introducing Social-Semiotics*. London/New York: Routledge.

Verlinde, Serge and Geert Peeters. 2012. Data Access Revisited: The Interactive Language Toolbox. Granger, Sylviane and Magali Paquot (Eds.). 2012: 147-162.

Welker, Herbert Andreas. 2010. *Dictionary Use: A General Survey of Empirical Studies. Brasilia.* Author's edition: An updated English version of Welker's earlier monograph, published in 2006 under the title *O uso de dicionários: Panorama geral das pesquisas empíricas* (reviewed by Lew, R., in 2007 in *International Journal of Lexicography* 20(4): 401-403).

Wittgenstein, Ludwig. 1953. *Philosophical Investigations*. Edited by Anscombe, G.E.M. and R. Rhees. Oxford: Blackwell.

Yang, Xinzhang. 2012. Text-image Relations in Learner's English–Chinese Dictionaries. *Journal of University of Science and Technology Beijing (Social Sciences Edition)* 28(4): 45-51.

Zhang, Ping. 2004. Is the Electronic Dictionary Your Faithful Friend? *CELEA Journal* 27(2): 23-28.

Zhang, Yihua. 2002. *Semantics and Lexicographical Definition*. Shanghai: Shanghai Lexicographical Publishing House.

Zhang, Yihua. 2015. *Second Language Acquisition and Learner's Dictionaries.* Beijing: The Commercial Press.

http://lexikos.journals.ac.za

Abbreviations and Acronyms: The Case of *Tlhalosi ya Medi ya Setswana*

Joyce T. Mathangwane, *Department of English, Faculty of Humanities, University of Botswana, Gaborone, Botswana (mathanjt@mopipi.ub.bw)*

Abstract: This paper looks at how abbreviations and acronyms are treated in African language dictionaries in general compared to selected mainstream English dictionaries. Specifically, the study looks at their treatment in T.J. Otlogetswe's (2012) *Tlhalosi ya Medi ya Setswana* dictionary. Altogether, a survey of twenty selected dictionaries was carried out examining the treatment of abbreviations and acronyms in these dictionaries. Ten of these dictionaries are mainstream English dictionaries and the remaining ten are dictionaries of varied African languages spoken in the Southern African region e.g. Shona, Ndebele, Venda, Setswana and Northern Sotho. The study addresses four questions: (a) whether African lexicographers include abbreviations and acronyms in their dictionaries as is practice in mainstream English dictionaries; (b) if so, how these have been treated; (c) what linguistic features are highlighted in these entries, if any; and, (d) what recommendation the study makes for the way forward.

The results showed that in most of the African dictionaries in the survey, unlike in mainstream English dictionaries, abbreviations and acronyms are not included despite the fact that many of them are coined and used by native speakers of these languages. An exception is Otlogetswe (2012) with a list of 25 abbreviations and acronyms. The paper recommends that African lexicographers include abbreviations and acronyms as part of their lexicon because these lexical items are coined by the communities making them part of the vocabulary of the language. Users of these dictionaries should find entries of abbreviations and acronyms in these dictionaries whenever they want to confirm the meaning, or when teaching.

Keywords: ABBREVIATIONS, ACRONYMS, AFRICAN LANGUAGES, DICTIONARY, FRENCH ACRONYMS, LEMMAS, LEXICOGRAPHERS, MONOLINGUAL, PRONUNCIATION, SEMANTIC PROPERTIES, STRESS, TONE

Opsomming: Afkortings en akronieme: Die voorbeeld van *Tlhalosi ya Medi ya Setswana*. Hierdie artikel kyk na hoe afkortings en akronieme oor die algemeen hanteer word in Afrikataalwoordeboeke in vergelyking met verkose hoofstroom Engelse woordeboeke. Spesifiek word gekyk na die hantering daarvan in T.J. Otlogetswe (2012) se woordeboek *Tlhalosi ya Medi ya Setswana*. 'n Ondersoek na die hantering van afkortings en akronieme in altesaam twintig verkose woordeboeke is uitgevoer. Tien van hierdie woordeboeke is hoofstroom Engelse woordeboeke en die res is woordeboeke van verskillende Afrikatale wat in Suider-Afrika gepraat word, bv. Shona, Ndebele, Venda, Setswana en Noord-Sotho. Die studie spreek vier kwessies aan: (a) of Afrikataalleksikograwe afkortings en akronieme in hul woordeboeke insluit soos wat by hoofstroom Engelse woordeboeke die gebruik is; (b) indien wel, hoe hulle hanteer is; (c) watter linguis-

tiese kenmerke, indien enige, in hierdie inskrywings beklemtoon word; en, (d) watter aanbevelings hierdie studie vir die pad vorentoe maak.

Die resultate het getoon dat afkortings en akronieme in die meeste Afrikataalwoordeboeke in die ondersoek, anders as in hoofstroom Engelse woordeboeke, nie opgeneem is nie, alhoewel baie van hulle deur moedertaalsprekers van hierdie tale geskep is en gebruik word. 'n Uitsondering is Otlogetswe (2012) met 'n lys van 25 afkortings en akronieme. In hierdie artikel word aanbeveel dat Afrikataalleksikograwe afkortings en akronieme as deel van hul leksikon insluit, aangesien hierdie leksikale items geskep is deur gemeenskappe wat hulle deel maak van die woordeskat van die taal. Gebruikers van hierdie woordeboeke behoort inskrywings van afkortings en akronieme in dié woordeboeke te vind wanneer hulle die betekenis daarvan naslaan of hulle tydens onderrig raadpleeg.

Sleutelwoorde: AFKORTINGS, AKRONIEME, AFRIKATALE, WOORDEBOEK, FRANSE AKRONIEME, LEMMAS, LEKSIKOGRAWE, EENTALIG, UITSPRAAK, SEMANTIESE EIENSKAPPE, KLEM, TOON

Introduction

According to Cannon (1989), the use of abbreviations and acronyms go back several millennia and was driven by the desire to economise. However, the real explosion of 'initialisms' (as they are referred to in his study) began with World War II when American English introduced so many of these initialisms, some of which were ridiculously absurd and long, e.g. USAMSMADHS 'United States Army Medical Service Meat and Dairy Hygiene School' (Cannon 1989: 101). Cannon notes that "Over the centuries, the purpose of creating these initialisms also has changed from the original medieval need for economy and efficiency …" According to Hamilton 1918 (cited in Cannon 1989: 103) "The use of abbreviations and signs is often a convenience and sometimes a temptation. … a saving of time and labour which is entirely justifiable under certain conditions." As a result, so many dictionaries of acronyms have been written on different areas of specialised vocabulary. This paper predicts another boom in the creation of new abbreviations and acronyms because of new technologies such as mobile phones and the increasing use of SMSs (Short Message Service), WhatsApp and other social media such as Facebook and Twitter. Already, the *Oxford Dictionary and Thesaurus* lists verbs derived from the noun SMS, as SMSs (third person singular verb), SMSing (progressive aspect) and SMSed (past tense) (Waite 2007: 980).

The purpose of this paper is to look at how abbreviations and acronyms are treated in dictionaries in general. Specifically, the study looks at how abbreviations and acronyms are treated in T.J. Otlogetswe's *Tlhalosi ya Medi ya Setswana* which was published in 2012. Otlogetswe's (2012) treatment of abbreviations and acronyms is compared to that of selected English dictionaries and other African languages in the Southern African region. This study carried out a comparative survey of twenty dictionaries, ten English diction-

aries and ten dictionaries of African languages, to outline the differences between these two sets of dictionaries in terms of their treatment of acronyms and abbreviations. Several questions are addressed in this study: (a) do African lexicographers include abbreviations and acronyms in their dictionaries as seems to be practice in mainstream English dictionaries; (b) if so, how have these been treated in these dictionaries; (c) what linguistic features are highlighted in these entries; and, (d) what recommendation does the study make regarding the inclusion of these forms in African dictionaries as a way forward?

Definitions of abbreviations and acronyms

According to the *Concise Oxford Companion to the English Language* edited by McArthur (1998: 1, 8), an abbreviation is the "shortening of words and phrases …" whereas an acronym "is an abbreviation from the first letters of a series of words and pronounced as one word". His definition makes a distinction between the two in that the former is pronounced as letters of the alphabet while the latter is pronounced as one word e.g. ATM (Automated Teller Machine) and WAR (Women Against Rape) respectively.

Yule (1996: 68) refines the definition of acronyms a little further and writes:

> **acronyms**, are formed from the initial letters of a set of words. These can remain essentially 'alphabetisms' such as *CD* ('compact disk') … where the pronunciation consists of a set of letters. More typically, acronyms are pronounced as single words, as in *NATO*, *NASA* or *UNESCO*.

Landau (1989: 27), on the other hand, attempts to make a clear distinction between abbreviations and acronyms when he states:

> An abbreviation is a shortened form for a word or phrase, consisting of the word or the first letter of each of the words in the phrase, or sometimes the first two letters.
>
> An acronym is a form of abbreviation composed of the first or the first two letters, or a syllable from each of the words in a compound term or phrase, so ordered that the resulting series of letters is usually pronounced as a word.

Landau (1989: 27) further notes that sometimes the distinction between abbreviations and acronyms is arbitrary as speakers often pronounce and treat them as one, in that what one considers an acronym, others may choose to spell out the individual letters, thus making it an abbreviation or alphabetism as well. He cites the acronym AWOL (absent without leave), which some speakers pronounce as a word, making it an acronym, while others pronounce it as a sequence of letters, in which case it is an abbreviation. In this paper though, we adopt Landau (1989)'s definition of the two terms, abbreviation and acronym. The terms *alphabetisms* or *initialisms* will not be used.

The treatment of acronyms in selected dictionaries

English Dictionaries

Altogether, ten selected English dictionaries were surveyed looking at whether abbreviations and acronyms are included in these dictionaries and, if so, how are they represented. The following table gives the results of the survey:

Table 1: Abbreviations and acronyms in selected English dictionaries

Dictionary	Year of Publication	Classification	Acronyms/ Abbreviations
The New American Webster Handy College Dictionary. Morehead, Albert and Loy Morehead (Eds.). (NAWHCD)	1981	Monolingual	Yes, a list of abbreviations and acronyms is given separately as a back matter text
The Concise Oxford Dictionary of Current English. Sykes, J.B. (Ed.). (CODCE)	1982	Monolingual	Yes, listed as part of the lexicon/lemmatised in the macrostructure
Collins Concise Dictionary (4th Edition). (CCD)	1999	Monolingual	Yes, listed as part of the lexicon/lemmatised in the macrostructure
Longman Dictionary of Contemporary English. New Edition. (LDCE)	1987	Monolingual	Yes, as part of the lexicon/lemmatised in the macrostructure
Longman Dictionary of English Language and Culture. (LDELC)	1998	Monolingual	Yes, listed as part of the lexicon/lemmatised in the macrostructure
Oxford Advanced Learner's Dictionary of Current English. Hornby, A.S. (7th Edition). Wehmeier, Sally (Chief ed.). (OALDCE)	2005	Monolingual	Yes, as part of the lexicon/lemmatised in the macrostructure
Cambridge International Dictionary of English. (CIDE)	1996	Monolingual	Yes, listed as part of the lexicon/lemmatised in the macrostructure

Longman Dictionary of Language Teaching and Applied Linguistics. Richards, Jack C. and Heidi Platt. (LDLTAL)	1999	Monolingual	Yes, a limited list of specialised acronyms and abbreviations/lemmatised in the macrostructure
Macmillan English Dictionary for Advanced Learners. International Student Edition. (MEDAL)	2002	Monolingual	Yes, as part of the lexicon/lemmatised in the macrostructure
Oxford Dictionary and Thesaurus. Waite, Maurice (Ed.). (ODT)	2007	Monolingual	Yes, listed as part of the lexicon/lemmatised in the macrostructure

Nine of the ten English dictionaries lemmatise abbreviations and acronyms in the macrostructure as part of the vocabulary while one includes them in a list as a back matter text.

In seven of the above dictionaries, an interesting observation can be made regarding the phonological representation of these abbreviations and acronyms in the lexicon. In six of these dictionaries, the pronunciation of the acronyms and abbreviations is given using IPA symbols (International Phonetic Alphabet), and these transcriptions are also marked for stress. Acronyms, which are pronounced as the spelling suggests, are marked for primary stress only, which is the strongest type, while abbreviations carry both primary stress represented with a high mark [ˈ] and secondary stress (less prominent) represented with a low mark [ˌ], as illustrated below:

Abbreviations:

FBI abbrev. /ˌefbiːˈaɪ/ … for the Federal Bureau of Investigation; the police department in the US which is controlled by the national government and which deals with serious crimes that involve people or places in more than one of the states in the US … (LDELC, p. 472)

PX /ˌpiːˈeks/ …abbreviation for post exchange … (LDELC, p. 1089)

BA /ˌbiːˈei/ … abbreviation for Bachelor of Arts …. (CIDE, p. 88)

UAE /ˌjuːeɪˈiː/ … abbreviation for United Arab Emirates (CIDE, p. 1577)

CBS /ˌsiːbiːˈes/ abbr (in the USA) Columbia Broadcasting System: *a CBS news broadcast: to CBS* (OALDCE, p. 235)

Acronyms:

UNICEF /ˈju:nɪsef/ n … abbreviation for United Nations International Children's Fund (CIDE, p. 1589)

ACAS /ˈeɪkæs/ abbr (in Britain) Advisory, Conciliation and Arbitration Service (an organization that helps with negotiation during industrial disputes) (OALDCE, p. 7)

NASA /ˈnæsə/ (in the USA) National Aeronautics and Space Administration (OALDCE, p. 1013)

UNESCO /ju:ˈneskəʊ/ abbr United Nations Educational, Scientific and Cultural Organization (OALDCE, p. 1670)

OPEC /ˈəʊpek/ n [the] Organization of Petroleum Exporting Countries; a group of countries that produce oil and plan together how to sell it … (OALDCE, p. 1062)

Abbreviations, on the one hand, are considered complex 'words', hence they are marked for both primary and secondary stress. It seems that acronyms, on the other hand, are phonologically considered to be simple 'words' and carry tonic or primary stress. Both monosyllabic and disyllabic acronyms carry primary stress on the first syllable while polysyllabic ones can carry stress either on the first or second syllable as indicated above.

However, some exceptions to the above treatment include The *New American Webster Handy College Dictionary* (Albert and Loy Morehead 1981), the *Oxford Dictionary and Thesaurus* (Waite 2007), and the *Concise Oxford Dictionary of Current English* (Sykes 1982). *The New American Webster Handy College Dictionary* (1981) provides a separate list of abbreviations and acronyms at the end of the dictionary as a back text, while the latter two indicate the pronunciation of acronyms by way of orthographic rewriting, and none at all for abbreviations.

Dictionaries of African Languages

A survey of dictionaries of some African languages in the Southern African region was also carried out to establish how African lexicographers have dealt with abbreviations or acronyms in their respective languages. This list, though by no means exhaustive, indicates that many of these dictionaries do not include abbreviations or acronyms in their lexicon, even those that are coined and commonly used by speakers of the respective languages. Table two below provides the results of a survey of a selection of dictionaries.

Table 2: Abbreviations and acronyms in a selection of dictionaries on African languages in Southern Africa

Dictionary	Year of Publication	Classification and Language(s)	Acronyms/ Abbreviations
Kgasa, M.L.A. *Thanodi ya Setswana ya Dikole*.	1976	Monolingual Setswana	None
Van Warmelo, N.J. *Venda Dictionary: Tshivenda–English*.	1989	Bilingual Tshivenda and English	None
Matumo, Z.I. (Compiler). *Setswana–English–Setswana Dictionary*.	1993	Bilingual Setswana and English	None
Kgasa, M.L.A. and J. Tsonope *Thanodi ya Setswana*.	1995	Monolingual Setswana	None
Chimhundu, H. *Duramazwi reChiShona*.	1996	Monolingual Shona	None
Otlogetswe, T.J. *English–Setswana Dictionary*.	2007	Bilingual English and Setswana	2 (HIV and Aids)
De Schryver, G.-M. (Ed.). *English–Northern Sotho. Oxford Bilingual School Dictionary*.	2007	Bilingual English and Northern Sotho	3 in the Northern Sotho side and 6 in the English side (HIV and Aids included)
Mojela, V. (Ed.). *Pukuntšutlhaloši ya Sesotho sa Leboa*.	2007	Monolingual Sesotho as Leboa	2 (HIV and Aids)
Hadebe, S. *Isichazamazwi SesiNdebele*.	2001	Monolingual Ndebele	None
Otlogetswe, T.J. *Tlhalosi ya Medi ya Setswana*.	2012	Monolingual Setswana	25 abbreviations and acronyms

From the above table, it is obvious that African lexicographers in general have not been paying much attention to abbreviations or acronyms when producing dictionaries in their respective languages. Of the ten dictionaries surveyed, including languages such as Setswana, IsiNdebele, ChiShona, Northern Sotho or Sesotho sa Leboa, Otlogetswe (2012) is the only one with larger list of abbreviations and acronyms in his lexicon of Setswana, altogether 25 are listed. This omission is even more noticeable in older dictionaries dating as far back as the 1980s or 1990s, for example, *Setswana English Setswana Dictionary* (Matumo 1993)

on Setswana and English, and *Venda Dictionary: Tshivenda–English* (Van Warmelo 1989) on Tshivenda. The only acronyms which are slowly making their way into some of the more current dictionaries are HIV and Aids which appears in four of the above dictionaries. These four are, *Pukuntšutlhaloši ya Sesotho sa Leboa* (Mojela 2007); *English–Northern Sotho Oxford Bilingual School Dictionary* (De Schryver 2007); *English–Setswana Dictionary* (Otlogetswe 2007) and *Tlhalosi ya Medi ya Setswana* (Otlogetswe 2012). It is interesting to note that Shona has a word for HIV/AIDS, **mukonombera**. Hence, there was no need to include the entry HIV/AIDS in the dictionary. With the exception of *Tlhalosi ya Medi ya Setswana* (Otlogetswe 2012), most of the selected dictionaries on African languages surveyed have a very brief list of abbreviations and acronyms in their lexicon ranging from zero (e.g. Kgasa 1976; Van Warmelo 1989; Matumo 1993) to nine (as in De Schryver 2007).

Abbreviations and acronyms in *Tlhalosi ya Medi ya Setswana* by T.J. Otlogetswe

As already noted above, *Tlhalosi ya Medi ya Setswana* (Otlogetswe 2012), which is a more recent dictionary, includes a short list of 14 abbreviations and 11 acronyms commonly used in the Setswana language. The list is reproduced below in its entirety.

Table 3: List of abbreviations and acronyms in *Tlhalosi ya Medi ya Setswana*, adapted from Otlogetswe (2012)

Aids **** /eìdz/ •*ln. 1a. Ø, *2a. bo-*• bolwetsi jo bo tshelwanang fa motho yo o nang le jone a tlhakanela dikobo le yo mongwe ba sa itshireletsa, kgotsa ka go arogana dieledi tsa mmele le molwetsi wa jone; bo bolayang masole a mmele mme bo paledise mmele go itshireletsa kgatlhanong le malwetsi *Acquired Immune Deficiency Syndrome* = SEGAJAJA

BCP /bì:sì:pì/ •*ln. 1a. Ø, *2a. bo-*• nngwe ya diphathi tsa sepolotiki tsa Botswana *Botswana Congress Party*

BDP /bì:dì:pì:/ •*ln. 1a. Ø, *2a. bo-*• nngwe ya diphathi tsa sepolotiki tsa Botswana *Botswana Democratic Party*

BIDPA /bídpà/ •*ln. 1a. Ø, *2a. bo-*• lekalana le le ikemetseng ka nosi la Botswana le le itebagantseng le go sekaseka mananeo le go netefatsa gore a dirisiwa ka fa go tshwanetseng *Botswana Instituition of Development and Planning Analysis*

BMD /bì:emdì:/ •*ln. 1a. Ø, *2a. bo-*• nngwe ya diphathi tsa sepolotiki tsa Botswana *Botswana Movement for Democracy*

BNF *** /bì:enef/ •*ln. 1a. Ø, *2a. bo-*• nngwe ya diphathi tsa sepolotiki tsa Botswana *Botswana National Front*

BONEPWA /bùnèpwá/ •*ln. 1a. Ø, *2a. bo-*• lekalana le le ikemetseng ka nosi mo Botswana le le thusang batho ba ba tshelang ka mogare wa HIV/Aids ka go ba fa dikgakololo tse ba di tlhokang *Botswana Network of People Living with HIV and AIDS*

BOTUSA /bʊ̀túsà/ •*ln. 1a. Ø, *2a. bo-*• kgokagano ya puso ya Botswana le lekalana la twantsho le taolo ya malwetsi la Amerika le le lwantshang malwetsi a tshwana le kgotlholo e kgolo le HIV/AIDS *Botswana USA*

BPC /bì:pì:sì:/ •*ln. 1a. Ø, *2a. bo-*• kompone e e anamisang motlakase mo lefatsheng la Botswana *Botswana Power Corporation*

BPP /bì:pì:pì:/ •*ln. 1a. Ø, *2a. bo-*• nngwe ya diphathi tsa sepolotiki tsa Botswana *Botswana People's Party* = MAGATAMMÔGÔ

BSB /bì:esbì:/ •*ln. 1a. Ø, *2a. bo-*• banka ya poloko-madi ya Botswana *Botswana Savings Bank*

BSE /bì:es ì:/ •*ln. 1a. Ø, *2a. bo-*• lekgotla le le rekisang diabe la Botswana *Botswana Stock Exchange*

BTA **** /bì:tì:eì/ •*ln. 1a. Ø, *2a. bo-*• lekgotla le le okametseng tsa ditlhaeletsanyo la Botswana *Botswana Telecommunications Authority*

BTC /bì:tì:sì:/ •*ln. 1a. Ø, *2a. bo-*• kompone e e okametseng ditlhaeletsano tsa megala mo Botswana *Botswana Telecommunications Corporation*

BTU /bì:tì:jù:/ •*ln. 1a. Ø, *2a. bo-*• lekgotla la barutabana ba Botswana *Botswana Teachers Union*

Btv /bì:tì:vì:/ •*ln. 1a. Ø, *2a. bo-*• thelebišene ya setšhaba ya lefatshe la Botswana *Botswana Television*

COSAFA /kʰɔ̀sáfà/ •*ln. 1a. Ø, *2a. bo-*• lekgotla le le okametseng kgwele ya dinao mo borwa jwa Aforika *Confederation of Southern African Football Association*

COSATU /kʰɔ̀sátù/ •*ln. 1a. Ø, *2a. bo-*• mokgatlho o o okametseng mekgatlho ya badiri ba Aforika Borwa *Congress of South African Trade Unions*

FIFA /fífà/ •*ln. 1a. Ø, *2a. bo-*• mokgatlho o o tsamaisang motshameko wa kgwele ya dinao mo lefatsheng ka bophara *Fédération Internationale de Football Association*

NDP /endì:pì:/ •*ln. 1a. Ø, *2a. bo-*• lenaneo la ditlhabololo tsa setšhaba *National Development Plan*

UCCSA /jù:sì:sì:eseì/ •*ln. 1a. Ø, *2a. bo-*• kereke ya Lontone e e mo mafatsheng a borwa jwa Aforika *United Congregational Church of Southern Africa*

UNESCO /jùnìskʰù/ •*ln. 1a. Ø, *2a. bo-*• lekgotlana la mafatshe a a kopaneng le le dirang le thuto,tsa boranyane le ngwao *United Nations Educational Scientific and Cultural Organisation*

UNIBO /jùníbù/ •*ln. 1a. Ø, *2a. bo-*• leina la yunibesithi ya Aforika Borwa e e neng e le mo kgaolong ya Bophuthatswana *University of Bophuthatswana*

UNICEF /jùnìsɛ̀f/ •*ln. 1a. Ø, *2a. bo-*• lekgotlana la mafatshe a a kopaneng le le inaakantseng le go sireletsa bana lefatshe ka bophara *United Nations Childrens Fund*

WABO /wàbù/ •*ln. 1a. Ø, *2a. bo-*• mokgatlho wa bakwadi mo Botswana *Writers Association of Botswana*

All these twenty-five abbreviations and acronyms are also lemmatised in the dictionary in alphabetical order together with all the other lexical items. And, just like all the other lexical entries in the dictionary, the following information is provided for each:

(a) Stars or asterisks (****) to indicate how frequent the lexical item (word) is used in Setswana, e.g. Aids, which is followed by four asterisks (****), indicating that it is one of the 1000 words most frequently used in Setswana

(b) Pronunciation of the abbreviation or acronym using IPA e.g. /eìdz/ (Aids)

(c) Tone marks on the transcription indicating whether High or Low tones are used, using the appropriate diacritic marks e.g. **UNIBO** /jùníbò/

(d) An indication of whether or not the lexical item is a name by way of the abbreviation *ln* (leina — 'name')

(e) Noun Class of the entry followed by its class prefix, e.g. Aids = class 1a. Ø (zero prefix)

(f) Class prefix for the plural form of the lexical item and its prefix e.g. Aids = *2a. bo-

(g) Explanation/meaning of the entry in Setswana

(h) The actual name represented by the abbreviation or acronym in English, e.g. Aids = *Acquired Immune Deficiency Syndrome,* or any other language, such as French

(i) The last information given for this entry is the Setswana name for Aids i.e. SEGAJAJA.

(*Tlhalosi ya Medi ya Setswana* (Otlogetswe 2012))

The above data entries (a–h) cover the phonology, morpho-syntactic and semantic aspects of each lexical item. This data is provided for all the other lexical items in the dictionary, abbreviations and acronyms included, an indication that these are equal lexical items in the language even if borrowed from a foreign language. This is illustrated well by a word such as FIFA, which is derived from French. The acronym FIFA is derived from the French *Fédération Internationale de Football Association*. However, this acronym is widely used as is in many languages across the globe. The word FIFA is treated as follows in this dictionary:

> **FIFA** /fífà/ •*ln. 1a. Ø, *2a. bo-*• mokgatlho o o tsamaisang motshameko wa kgwele ya dinao mo lefatsheng ka bophara *Fédération Internationale de Football Association*
>
> (*Tlhalosi ya Medi ya Setswana* (Otlogetswe 2012: 104))

An example such as FIFA is an interesting one in that it also appears in many of the conventional English dictionaries and where listed, the French name from which it is derived may also be provided. For instance, the *Longman Dictionary of English Language and Culture* (1998: 478) writes,

> **FIFA** /ˈfi:fə/ abbrev. for Fédération Internationale de Football Association; the organization that controls international football and organizes the WORLD CUP competition.

The *Macmillan English Dictionary for Advanced Learners,* International Student Edition (2002: 518) also writes,

> **FIFA** /ˈfi:fə/ the international organization that makes rules and decisions relating to the sport of football. FIFA is an abbreviation of the organization's name.

Another interesting example of an acronym derived from French and yet used as such in the English language is CERN listed in the 5th edition of the *Oxford Advanced Learner's Dictionary of Current English* (5th Edition) (1995) as follows:

> **CERN** … /ˈsɜ:n/ abbr European Organisation for Nuclear Research (French *Conseil Européen pour la Recherche Nucléaire*)

Such entries clearly indicate that acronyms may even be borrowed from another language, a fact that some users may be unaware of. The fact that they are used in any particular language of the dictionary, justifies their inclusion in the dictionary of that language as some dictionary users may consult it for acronyms.

Conclusion

From the surveyed dictionaries above, we conclude that all the English dictionaries in the survey include abbreviations and acronyms in their lemma list. They are considered part of the lexicon as they are lemmatised in alphabetical order with all the other lemmas. With the exception of *Tlhalosi ya Medi ya Setswana* (Otlogetswe 2012), dictionaries of African languages have tended to ignore these abbreviations and acronyms even though many of them are commonly used in the respective languages. Some of these abbreviations or acronyms are coined by the particular language communities as new organizations are formed within the country. In Botswana, for instance, new entries come into the language every time a new organisation is formed, e.g. BIUST (Botswana International University of Science and Technology); BITRI (Botswana Institute for Technology Research and Innovation); BOCRA (Botswana Communications Regulatory Authority), and many others.

This paper therefore argues for the inclusion of abbreviations and acronyms when compiling dictionaries of African languages, something which many African lexicographers have tended to overlook until recently. It is necessary that African lexicographers include these abbreviations and acronyms in their dictionaries as they form part of vocabulary of the various African languages, for which information may be sought in dictionaries. Many of these abbreviations and acronyms have high frequencies in these African languages justifying their inclusion in these dictionaries. For instance, Otlogetswe (2014)'s

keyword analysis of the top 100 business key words from a Setswana corpus of well over a million tokens indicate high frequencies for some of these abbreviations and acronyms as follows: BTC — (Botswana Telecommunications Corporation (69); PEEPA — Public Enterprises Evaluation and Privatisation Agency (27); BOBS — Botswana Bureau of Standards (29); SACU — Southern African Customs Union (13); BOCCIM — Botswana Confederation of Commerce, Industry and Manpower (20); and, GPH — Gaborone Private Hospital (13) (Otlogetswe 2014: 280-283).

Otlogetswe (2012) not only treated abbreviations and acronyms as part of the lexicon, he provided the phonological, morpho-syntactic and semantic properties of these lemmas, as well as the actual names of the organizations that they stand for. Even though Otlogetswe's (2012) list is very limited at twenty-five entries only, it is a good start. Hopefully more entries of abbreviations and acronyms used in Botswana and elsewhere will be captured in the second edition of this dictionary, as ideally it should. Furthermore, Otlogetswe (2012) lists the borrowed acronym FIFA (*Fédération Internationale de Football Association*) and its French name just as some of the conventional English dictionaries such as the Oxford, Longman and Macmillan dictionaries. Even if the name of the organization is borrowed from another language, it deserves a place in the dictionary to guide the users if in fact the abbreviation or acronym is known to and used by the speakers of the particular language. This paper therefore recommends that African lexicographers should include abbreviations and acronyms in the later editions of their dictionaries to assist language learners and users to know what these forms stand for and how to use them. Also worth noting is that, some of these abbreviations and acronyms are created or coined by the communities of these languages, thus making them part of the vocabulary of these languages. The practice should be to also include forms that are part and parcel of the society's language or languages. Such practice does not, in any way, tarnish the purity of that particular language, but rather adds to the communicative value of the dictionary by giving users access to commonly used abbreviations and acronyms, their meaning, how to pronounce them, what they mean, etc.

References

Dictionaries

CIDE = Procter, Paul (Ed.). 1996. *Cambridge International Dictionary of English.* Cambridge: Cambridge University Press.

Chimhundu, H. (Ed.). 1996. *Duramazwi reChiShona*. Allex Project. Harare: College Press.

CCD = *Collins Concise Dictionary*. 1999. Fourth Edition. Glasgow: HarperCollins.

Concise Oxford Companion to the English Language. 1998 = McArthur, Tom (Ed.). 1998. *Concise Oxford Companion to the English Language.* Oxford: Oxford University Press.

CODCE = Sykes, J.B. (Ed.). 1982. *The Concise Oxford Dictionary of Current English.* Oxford: Clarendon Press.

English–Northern Sotho. Oxford Bilingual School Dictionary. 2007 = De Schryver, Gilles-Maurice (Ed.). 2007. *Oxford Bilingual School Dictionary: Northern Sotho and English/Pukuntšu ya Polelopedi ya Sekolo: Sesotho sa Leboa le Seisimane. E gatišitšwe ke Oxford*. Cape Town: Oxford University Press Southern Africa.

Hadebe, S. et al. (Eds.). 2001. *Isichazamazwi SesiNdebele*. Harare: College Press.

Kgasa, M.L.A. (Ed.). 1976. *Thanodi ya Setswana ya Dikole*. Cape Town: Longman Penguin Southern Africa.

Kgasa, M.L.A. and J. Tsonope. 1995. *Thanodi ya Setswana*. Gaborone: Longman Botswana.

LDCE = Summers, D. (Ed.). 1987. *Longman Dictionary of Contemporary English.* New Edition. Harlow: Longman.

LDELC = Summers, D. (Ed.). 1998. *Longman Dictionary of English Language and Culture.* New Edition. Harlow: Longman.

LDLTAL = Richards, Jack C., John Platt and Heidi Platt (Eds.). 1999. *Longman Dictionary of Language Teaching and Applied Linguistics*. New Edition. Harlow: Longman.

MEDAL = Rundell, M. (Ed.). 2002. *Macmillan English Dictionary for Advanced Learners*. International Student Edition. Oxford: Macmillan Education.

Matumo, Z.I. (Compiler). 1993. *Setswana–English–Setswana Dictionary.* Gaborone: Botswana Book Centre and Macmillan Botswana Publishers.

Mojela, V. (Ed.). 2007. *Pukuntšutlhaloši ya Sesotho sa Leboa*. Pietermaritzburg: Nutrend Publishers.

NAWHCD = Morehead, Albert and Loy Morehead (Eds.). 1981. *The New American Webster Handy College Dictionary.* New York: Penguin Group.

Otlogetswe, T.J. 2007. *English–Setswana Dictionary*. Gaborone: Pentagon Publishers.

Otlogetswe, T.J. 2012. *Tlhalosi ya Medi ya Setswana*. Gaborone: Medi Publishing.

Oxford Advanced Learner's Dictionary of Current English. 1995 = Hornby, A.S. 1995. *Oxford Advanced Learner's Dictionary of Current English.* Fifth Edition. Oxford: Oxford University Press.

OALDCE = Hornby, A.S. 2005. *Oxford Advanced Learner's Dictionary of Current English.* Seventh Edition. Chief Editor Sally Wehmeier. Oxford: Oxford University Press.

Van Warmelo, N.J. 1989. *Venda Dictionary: Tshivenḓa–English*. Pretoria: J.L. van Schaik.

ODT = Waite, Maurice (Ed.). 2007. *Oxford Dictionary and Thesaurus*. Oxford: Oxford University Press.

Other literature

Cannon, G. 1989. Abbreviations and Acronyms in English Word-formation. *American Speech* 64(2): 99-127.

Landau, S.I. 1989. *Dictionaries: The Art and Craft of Lexicography*. Cambridge: Cambridge University Press.

Otlogetswe, T.J. 2014. Corpus Linguistics Method: A Case of Extracting Business Terms for Dictionary Subject Label. Khumalo, Langa (Ed.). 2014. *African Languages and Linguistic Theory: A Festschrift in Honour of Professor Herbert Chimhundu*. Cape Town: The Centre for Advanced Studies of African Society (CASAS).

Yule, G. 1996. *The Study of Language*. Second Edition. Cambridge: Cambridge University Press.

http://lexikos.journals.ac.za

Legal Terms in General Dictionaries of English: The Civil Procedure Mystery

Sandro Nielsen, *Centre for Lexicography, Aarhus University, Denmark* and *International Centre for Lexicography, University of Valladolid, Spain (sn@bcom.au.dk)*

Abstract: Many general language dictionaries contain specialized terms, including legal terms relating to civil lawsuits. The existing literature provides general discussions of scientific and technical terms in ordinary dictionaries but does not specifically address the inclusion of legal terms. This study examines four general dictionaries of English to see how they treat civil procedure terms used in England and Wales in the light of the change of structure of and terminology used in civil proceedings that took place in 1999. Despite being based on large, up-to-date corpora the dictionaries contain some of the old terms but fail to include the new terms that have been in use for more than 15 years. Why this is the case is a mystery. However, some clues indicate that if they pay more attention to the link between dictionary functions, corpora and the data presented in dictionaries, lexicographers may be able to work in a more focussed way that would likely ensure the inclusion of legal terms as well as terms from other subject fields in general dictionaries. This would also satisfy the needs of users.

Keywords: LEXICOGRAPHY, INFORMATION TOOLS, CORPORA, CORPUS LEXICOGRAPHY, CULTURE-DEPENDENT DOMAINS, KNOWLEDGE, LEMMATIZATION, LEGAL LANGUAGE, DICTIONARY FUNCTIONS, COMMUNICATIVE FUNCTIONS, DECODING, ENCODING

Opsomming: Regsterme in algemene Engelse woordeboeke: Die raaisel van die siviele proses. Baie algemene woordeboeke bevat gespesialiseerde terme, insluitend regsterme wat verband hou met siviele hofsake. Die bestaande literatuur verskaf algemene besprekings van wetenskaplike en tegniese terme in gewone woordeboeke, maar hanteer nie spesifiek die insluiting van regsterme nie. In hierdie artikel word vier algemene Engelse woordeboeke bestudeer om te sien hoe hulle siviele regsterme wat in Engeland en Wallis gebruik is, hanteer. Dit word gedoen teen die agtergrond van die verandering in die struktuur van siviele hofsake wat in 1999 plaasgevind het en die terminologie daarin gebruik. Alhoewel die woordeboeke op groot, bygewerkte korpusse gebaseer is, bevat hulle sommige van die ou terme, maar slaag nie daarin om die nuwe terme wat al meer as 15 jaar gebruik word, in te sluit nie. Waarom dit só is, is 'n raaisel. Sommige leidrade dui daarop dat, as leksikograwe meer aandag skenk aan die verband tussen woordeboekfunksies, korpusse en die data wat aangebied word in woordeboeke, hulle op 'n meer gefokusde manier sou kon werk, wat waarskynlik sal verseker dat regsterme sowel as terme uit ander onderwerpsvelde in algemene woordeboeke ingesluit sal word. Dit sal ook aan gebruikersbehoeftes voldoen.

Lexikos 25 (AFRILEX-reeks/series 25: 2015): 246-261

Sleutelwoorde: LEKSIKOGRAFIE, INLIGTINGSHULPMIDDELS, KORPUSSE, KORPUS-LEKSIKOGRAFIE, KULTUUR-AFHANKLIKE DOMEINE, KENNIS, LEMMATISERING, REGSTAAL, WOORDEBOEKFUNKSIES, KOMMUNIKATIEWE FUNKSIES, DEKODERING, ENKODERING

1. Introduction

Statistics show that many people come into contact with the legal system every year, in one way or another. Some are interviewed or arrested by the police, some quibble with family members over the validity of testamentary gifts, while others come into contact with the legal system in the course of their work. As members of civilized societies we are often exposed to situations in which we directly or indirectly meet legal terms, for example when we read about rules that are intended to protect consumers, and when we complain to shops because we think that the goods and services we have bought do not comply with the terms of the contract. According to the UK Ministry of Justice (2015: 8), between 1.4 million and 2.1 million cases annually were brought before the civil courts in England and Wales from 2006 to 2012 and in the fourth quarter of 2014, approximately 379,000 civil proceedings were commenced (the figures do not include family cases such as divorce). These figures suggest that we are all likely to become involved in civil proceedings at some point during our lives and it is therefore not surprising to find that general dictionaries of English contain some legal terms.

Literature discussing the treatment of terms (often called scientific and technical words) in general language dictionaries is sparse and concerns terms in general. For instance, Béjoint (1988) provides a broad discussion of the difficulties lexicographers face when they write definitions of scientific and technical terms from various domains in general language dictionaries, while Jessen (1996) gives a general analysis of the treatment of terms in selected monolingual and bilingual general language dictionaries covering English and French. Josselin-Leray and Roberts (2005) investigate the use of general dictionaries by language specialists, scientists and laypersons when they look for terms from a range of subject fields in monolingual and bilingual dictionaries of English and French, and Vrbinc and Vrbinc (2013) examine the inclusion and treatment of technical and scientific terms in three editions of *Oxford Advanced Learner's Dictionary* by looking at the use of subject field labels, short cuts and sense indicators. However, none of the contributions specifically examines the treatment of legal terms in general dictionaries of English so it is relevant to look closer at this aspect of lexicography.

Many people own medium-sized general dictionaries of English, either as native speakers or as learners of English as a foreign language. Taking the annual number of civil proceedings into consideration, many people will look for help in their dictionaries when they come across terms relating to civil proceedings that they do not know or if they are uncertain about the meanings of

those terms. Another relevant aspect is that the system of civil procedure in England and Wales was reformed in 1999 and part of this reform was to change the somewhat old-fashioned terminology. Even though legal terms are often regarded as relevant for expert-to-expert communication, the fact that so many ordinary people come into contact with the civil justice system indicates that a wide range of communicative instances are expert-to-layperson and layperson-to-layperson. This may be one of the reasons why general language dictionaries contain legal terms. However, as these dictionaries are intended for the general public the aim of this paper is to examine to which extent they lemmatize the specialized terminology of civil proceedings. This involves a description of the methods used (Section 2), and a study of how four internationally renowned general dictionaries of English treat the basic terms used in civil proceedings in England and Wales with particular focus on the selection of lemmas (Sections 3 and 4), and how the (non-)lemmatization of civil procedure terms affect communicative dictionary functions, and the use of corpora (Section 5).

2. Specifying legal terms and dictionaries

It may be argued that legal terminology does not belong in general language dictionaries. This is probably correct when we talk about "small" dictionaries but the idea of including legal terms in general dictionaries is not new. In his elementary reference work, Richard Mulcaster expresses the wish that "som one well learned and as laborious a man, wold gather all the words which we use in our English tung, whether naturall or incorporate, out of all professions, as well learned as not, into one dictionarie" (Mulcaster 1582: 166), and he explicitly refers to the field of law several times in his book. In a modern setting, this wish appears to have come true as Svensén (2009: 3) observes that "General-language dictionaries usually include a considerable number of technical terms, particularly those encountered by everyone in everyday life". As indicated above, some legal terms are part of the everyday life of modern citizens, so dictionaries should provide assistance with at least some of the most important terms. The question is then to find a way in which to select the legal terms to be lemmatized and here Magay (1984: 223) gives some sound advice: "The lexicographer has to differentiate between what is new and what is really important, i.e. what the user really needs". One interpretation of this statement is that lexicographers should lemmatize those terms that users really need and according to statistics, assistance with civil procedure terms are needed by users in their everyday lives.

It is beyond the scope of this paper to examine the (non-)lemmatization of all civil procedure terms in the selected dictionaries. For one thing, English civil procedure terms are found in many different geographical locations, for instance in the United Kingdom, the United States of America, Australia and Canada. Secondly, each of these countries has its own legal system that constitutes a jurisdiction of law with its own culture-dependent terminology, though

there may be the occasional overlap. Furthermore, it should be appreciated that the United Kingdom is divided into three jurisdictions with differing terminology — this study only concerns terms used in England and Wales, while terms used in the other two jurisdictions, i.e. Scotland and Northern Ireland, will be excluded for practical purposes.

The way in which civil proceedings are systematised and carried out is based on long-standing tradition but changes do occur, most significantly in 1999 when the civil procedure system was radically reformed. This reform included a change of terminology so that several core terms were replaced by new ones. The following analysis is based on an appendix to the Civil Procedure Rules that came into force in England and Wales in April 1999, which lists 33 new terms and the old ones they replaced (British and Irish Association of Law Librarians 2015). Since many of the terms represent very specific legal concepts that are predominantly used in expert-to-expert communication, general language dictionaries cannot be expected to contain all, so the point of departure will be those old civil procedure terms that are lemmatized in a benchmark dictionary: *Oxford Dictionary of English* (ODOE). The main reason for choosing this dictionary as benchmark is the following description:

> The foremost single volume authority on the English language, the *Oxford Dictionary of English* is at the forefront of language research, focusing on English as it is used today. It is informed by the most up-to-date evidence from the largest language research programme in the world, including the two-billion-word Oxford English Corpus. […] Ideal for anyone who needs a comprehensive and authoritative dictionary of current English; for professionals, students, academics and for use at work or at home. (*Oxford Dictionary of English*)

Oxford Dictionary of English contains 11 of the 33 old civil procedure terms and these lemmas are marked as belonging to the subject field *law*, i.e. 11 lemmas and definitions that relate to the legal domain (see Table 1). This seems to indicate that the 11 terms are so important for the description of English usage that they warrant inclusion in general dictionaries of English, i.e. they are part of the public domain. Consequently, these old terms and those that have replaced them should be included in that dictionary and other medium-sized general dictionaries of English.

Old term	**ODOE**
writ (of summons)	writ of summons
plaintiff	plaintiff
discovery	discovery
Mareva injunction	Mareva injunction
in camera	camera, in camera

interlocutory	interlocutory
statement of claim	statement of claim
pleading(s)	pleadings
interrogatory/-ies	interrogatories
Anton Pillar order	Anton Pillar order
subpoena	subpoena

Table 1: The eleven old civil procedure terms lemmatized in *Oxford Dictionary of English*

In order to be as specific as possible, it should be appreciated that the eleven terms are classified as belonging to the legal field on the basis of the concepts they refer to. This may sound obvious, but some of the lemmas have more than one meaning and only the legal meaning is relevant. For example, ODOE defines the term *plaintiff* as "a person who brings a case against another in a court of law", i.e. a legal term proper that is marked as such with the subject field label *Law*. The term *discovery* has an unmarked general sense, "the action or process of discovering or being discovered", as well as a marked legal sense: "the compulsory disclosure, by one party to an action to another, of relevant testimony or documents". ODOE treats *pleading* as a mass noun in a general sense and as a plural noun in a legal sense with the following definition: "(usu. **pleadings**) *(Law)* a formal statement of the cause of an action or defence." Finally, *in camera* is included as a cross-reference article that merely contains a cross-reference to the lemma **camera** (sense 2), which is defined as "a chamber or round building." This article contains the phrase *in camera* and the associated definition: "*chiefly (Law)* in private, in particular taking place in the private chambers of a judge, with the press and public excluded." The lemma **camera** is not a specific legal term, but the phrase *in camera* is marked as belonging to the legal domain and, therefore, qualifies as one of the eleven terms in Table 1. In their analysis of terms in three editions of OALD Vrbinc and Vrbinc (2013: 444-449) found that technical and scientific terms are not only marked directly as terms by subject field labels following lemmas, such as *law* and *physics*, but are also marked indirectly by short cuts and sense indicators in the definitions. However, the phrase *in camera* in ODOE, which has the status of a term in the legal domain, is marked directly as belonging to that domain in connection with the phrase itself and not the lemma, the sense or the definition, a practice also observed by Jessen (1996: 96).

The dictionaries selected for the analysis are all internationally well-known, medium-sized dictionaries of English. Together they occupy a considerable part of the international lexicographical landscape and are, therefore, relevant objects of study. The most recent editions of four dictionaries have

been selected for analysis and comparison with the benchmark dictionary: *Longman Dictionary of Contemporary English* (LDOCE), *Oxford Advanced Learner's Dictionary* (OALD), *Macmillan English Dictionary* (MED) and *Cambridge Advanced Learner's Dictionary* (CALD). This includes the print editions (except MED) as well as the free online editions. Furthermore, the dictionaries are all based on large electronic corpora that have been subjected to treatment by modern software and thus share important features with the benchmark dictionary. For example, MED states that its source is "a corpus, a database containing millions of examples of English as used around the world. […] using state-of-the-art software, has allowed the dictionary writers to reveal fresh information about how and when words are used" (*Macmillan English Dictionary*) and CALD is "Based on the 1.5 bn word Cambridge English Corpus" (*Cambridge Advanced Learner's Dictionary*). Finally, some of the publishers have one edition for British English and another for American English and the dictionaries studied all cover British English, because the civil procedure terms used in England and Wales differ considerably from those used in the USA. Having specified which civil procedure terms to focus on and in which dictionaries to look, I will go on to examine which civil procedure terms are included in the four dictionaries.

3. Which old terms are in the dictionaries?

The benchmark dictionary establishes a basis for comparing the lemmatization of civil procedure terms in the four selected dictionaries. LDOCE, OALD, MED and CALD all contain lemmas that are marked with a subject label referring to the legal domain, *law* and *legal* respectively. Table 2 shows how many of the 11 old civil procedure terms lemmatized in the benchmark dictionary are also lemmatised in the four dictionaries examined. CALD scores the highest as it includes 7 of the old terms, while LDOCE has the fewest with only 4 of the eleven terms; note, however, that the terms *in camera*, *interlocutory injunction*, *statement of claim* and *pleadings* are only included in the online edition of CALD. The other two dictionaries come between these two extremes, in that OALD and MED both include 6 of the old civil procedure terms.

Old term	**LDOCE**	**OALD**	**MED**	**CALD**
writ (of summons)	writ	writ	writ	writ
plaintiff	plaintiff	plaintiff	plaintiff	plaintiff
discovery	—	—	discovery	—
Mareva injunction	—	—	—	—
in camera	camera, in c.	camera, in c.	in camera	in camera
interlocutory	—	—	—	interlocutory injunction

statement of claim	—	—	—	statement of claim
pleading(s)	—	pleading	pleading	pleadings
interrogatory/-ies	—	interrogatory	—	—
Anton Pillar order	—	—	—	—
subpoena	subpoena	subpoena	subpoena	subpoena

Table 2: Old civil procedure terms lemmatized in the four medium-sized dictionaries

Two of the eleven terms are included only in the benchmark dictionary: *Anton Pillar order* and *Mareva injunction*. This means that, together, the four dictionaries cover nine of the eleven old civil procedure terms, though CALD print edition and LDOCE include less than half of the terms. The terms studied so far are those that were used before 1999, so it is relevant to examine whether the dictionaries include the new terms that have applied since 1999.

4. Which new terms are in the dictionaries?

All the dictionaries analysed are based on corpora that are intended to contain systematic collections of texts that document usage features of English, in particular British English. As indicated in Section 2, the corpus-based approach to dictionary-making is emphasised by the publishers and is a fact that should be taken into consideration. Firstly, the benchmark dictionary is examined and Table 3 shows 1) the old civil procedure terms, 2) the new terms that have replaced them, and 3) which of the new terms are lemmatized in *Oxford Dictionary of English*.

Old term	**New term**	**ODOE**
writ (of summons)	claim form	—
plaintiff	claimant	(—)
discovery	disclosure	—
Mareva injunction	freezing injunction	—
in camera	in private	—
interlocutory	interim	—
statement of claim	particulars of claim	—
pleading(s)	statement of case	—

interrogatory/-ies	request for further information	—
Anton Pillar order	search order	—
subpoena	witness summons	—

Table 3: The eleven new civil procedure terms lemmatized in *Oxford Dictionary of English*

The analysis shows that ODOE lemmatizes none of the eleven new civil procedure terms. This result is surprising. Since the eleven old terms were judged relevant for inclusion, it is reasonable to expect that the new terms would also be included and given the same treatment as the old terms, in particular because the new terms have been in use for more than 15 years and, therefore, must be expected to have replaced the old terms in the public domain. The reason for the parenthesis against the term *claimant* is that ODOE treats it in a special way. The article **plaintiff** contains the following note: "USAGE: In England and Wales the term **plaintiff** was officially replaced by **claimant** in 1999". This is a proper way in which to treat this lemma but ODOE's definition of the lemma **claimant** is not optimal: "a person making a claim, especially in a lawsuit or for a state benefit." First of all, as the term *claimant* has replaced the term *plaintiff*, it is reasonable to expect that the two are defined in the same way because they both refer to the same concept in the real world. Secondly, persons claim benefits from the state whereas they usually start legal proceedings against individuals or companies. This means that the lemma **claimant** should have been treated as polysemous and the legal sense should have been marked with an appropriate subject field label, which would have been in line with the treatment given to other lemmas such as **discovery** discussed in Section 2 above. It should also be noted that the words *in private* and *interim* are lemmatized in ODOE but are described in a general sense and not in a specific legal sense, i.e. they are not treated as terms.

The next step is to examine the lemmatization of the new civil procedure terms in the four medium-sized dictionaries. Table 4 shows how many of the 11 new civil procedure terms are lemmatized in the four dictionaries and this may be compared with the findings in Table 3. Again CALD is the highest scorer as it includes 2 of the new terms, while LDOCE, OALD and MED contain none of the eleven terms. Note that the two terms lemmatized in CALD are only included in the online edition.

New term	**LDOCE**	**OALD**	**MED**	**CALD**
claim form	—	—	—	—
claimant	—	—	—	—
disclosure	—	—	—	—

freezing injunction	—	—	—	—
in private	—	—	—	—
interim	—	—	—	—
particulars of claim	—	—	—	particulars of claim
statement of case	—	—	—	—
request for further information	—	—	—	—
search order	—	—	—	—
witness summons	—	—	—	witness summons

Table 4: New civil procedure terms lemmatized in the four medium-sized dictionaries

As Table 4 indicates, the four dictionaries do not treat the new civil procedure terms in any significant way. LDOCE and CALD are the only dictionaries to include a lemma called **claim form** but neither defines this term in the legal sense, i.e. identical to the old term *writ*. All dictionaries include the word *claimant* but none of them defines it in a legal sense, though LDOCE provides the following definition: "someone who claims something, especially money, from the government, a court etc. because they think they have a right to it". From a legal perspective, it is unlikely that anyone will claim money from a court; they are, however, likely to claim money from a person or company in a case heard by a court. All dictionaries include the word *disclosure* but only in its ordinary meaning of "revealing a secret", and none of the dictionaries include the term *freezing injunction*. The words *in private* and *interim* are included in the dictionaries but defined in their ordinary meanings with no relation to civil proceedings. CALD online includes the term *particulars of claim*, i.e. "detailed information that you have to provide when asking a court, government department, or company to give you something such as money or property that you believe you have a legal right to", and the term *witness summons*, i.e. "a legal document ordering someone to appear in a court of law to give information about a particular person or event", both with their legal definitions and subject field labels. None of the four dictionaries includes the terms *statement of case*, *request for further information*, and *search order*.

The findings in Tables 2 and 4 show that the four general language dictionaries lemmatize some old civil procedure terms but not new terms and the difference in treatment is not clear. A discussion of lemmatization principles

and the use of electronic corpora may provide clues that can help solve the mystery.

5. Discussion of findings and implications

Users consult general language dictionaries because they think these information tools can help them solve various problems. The mainly quantitative analyses in Sections 3 and 4 reveal that the four dictionaries provide limited help with civil procedure terms but a discussion including qualitative aspects may help explain the difference in treatment. Some of the findings and their implications for dictionary making, dictionary use in communicative situations and use of corpora will be discussed in the following sub-sections.

5.1 Selecting lemmas according to dictionary function

The theory of dictionary functions is relevant when assessing the data contained in dictionaries because dictionary functions determine all active decisions from the selection of lemmas, over the selection of data types, to the way in which the data are presented. Bergenholtz and Tarp (2010: 30) define a dictionary function as "the satisfaction of the specific types of lexicographically relevant needs that may arise in a specific type of potential user in a specific type of extra–lexicographical situation". This means that lexicographers make their decisions with due regard to the basic needs of the intended users identified through user profiling (see e.g. Bergenholtz and Nielsen 2006: 286) and attempt to match those needs with the dictionary function(s).

The four dictionaries examined have the same main types of communicative function. Generally speaking, dictionaries with communicative functions are designed to provide specific types of help to specific types of user in specific types of usage situations involving ongoing or planned acts of communication, for example providing help to understand words and terms users meet when reading texts, and providing help to choose the correct word and use it correctly when writing texts. According to the informative material on the dictionaries, they have two primary communicative functions: to provide help with the meaning of words and to provide help to produce oral and written texts. Table 2 shows that the four dictionaries together include nine of the old civil procedure terms analysed, which must be regarded as a good degree of coverage compared to the eleven old terms included in the benchmark dictionary. However, a closer study reveals that the matter is not that simple.

The dictionary that provides the best help to understand the meaning of old terms is the online edition of CALD: it provides the meanings of seven of the civil procedure terms analysed. The dictionaries that provide the least help to understand terms are the print edition of CALD and LDOCE in that users will only be able to find the meanings of three and four civil procedure terms,

respectively. The implication of the study of old terms reported in Table 2 is that users will have to consult CALD online, OALD as well as MED in order to find the meanings of all nine terms included in the four dictionaries. Consequently, the four dictionaries provide varied help to understand those civil procedure terms that were in use in England and Wales before 1999.

Table 4 shows that help to understand civil procedure terms in use after 1999 is almost non-existent. CALD online is the only dictionary that lemmatizes new terms so that users will be able to find the meanings of two of the eleven terms; LDOCE, OALD, MED and CALD print edition provide no help to understand any of the eleven new civil procedure terms analysed. Nevertheless, the print editions of the four medium-sized dictionaries treat the new terms in the same way as the benchmark dictionary (CALD online scores higher). This means that LDOCE, OALD, MED and CALD print edition give users no help with the meaning of the new terms and that CALD online explains the meanings of two of the eleven new civil procedure terms that replaced the old terms in 1999 and have been in the public domain since then.

It is somewhat more complicated to assess the coverage of civil procedure terms related to the communicative function called text production. The dictionary that provides the best help to produce oral and written communication is also the online edition of CALD as it contains seven of the eleven old terms, see Table 2. Again CALD print edition and LDOCE provide the least help to produce oral and written communication because they lemmatize only three and four of the eleven old terms, respectively. That being said, the four dictionaries provide some help to produce texts which contain terms that were used pre-1999 but it is reasonable to expect that users of the dictionaries rarely need help to produce texts that contain those terms. Users are, however, more likely to look for assistance with producing texts that contain the new civil procedure terms.

Table 4 shows that the four dictionaries provide almost no help to produce oral and written communication containing civil procedure terms in current use after 1999. CALD online is the only dictionary that lemmatizes new terms so that users will be able to find help with two of the eleven terms; LDOCE, OALD, MED and CALD print edition provide no help to produce texts with any of the eleven new civil procedure terms analysed. Nonetheless, the print editions of the four medium-sized dictionaries treat the new terms in the same way as the benchmark dictionary (CALD online is the exception). This means that LDOCE, OALD, MED and CALD print edition give users no assistance in producing oral and written texts in which the new terms occur and that CALD online provides help with two of the eleven new civil procedure terms that have been part of the public domain since 1999. Since the fact that they include old civil procedure terms indicates that the dictionaries are designed to provide help to understand and use civil procedure terms (communicative functions), the lack of help with the new terms is puzzling, in particular since the dictionaries are based on large text corpora.

5.2 Corpora: what seems to be the problem?

Lexicographers have benefited greatly from the introduction of electronic corpora in dictionary making. Corpora can help lexicographers to identify such features as the meanings of words and their grammatical properties with relative ease, and finding the meanings of words is particularly relevant for this study. The four dictionaries examined are all informed by large corpora and the informative material on CALD online explains that this dictionary "includes up-to-date vocabulary" and that it is based on the Cambridge English Corpus, which is "a multi-billion word collection of written and spoken English." Similarly, OALD claims that it is "based on the authoritative Oxford English Corpus", which was also used for compiling the benchmark dictionary (see Section 2 above). So why do these dictionaries score so low?

The meanings of legal terms found in corpora can be discovered from the context in which they occur, in particular from external text-type indicators. As explained by Atkins and Rundell (2008: 299) "For lexicographers, text-type features such as domain, time and regional dialect often provide valuable evidence to support the process of identifying dictionary senses" and these features can be used to discuss the lemmatization of legal terms. As indicated in section 2, the field of law is a culture-dependent domain because each jurisdiction has its own internal structure that is reflected by the terms used. This means that texts from this domain included in corpora should be clearly marked (tagged) as belonging to the field of law. Secondly, the fact that the domain is culture-dependent means that lexicographers should carefully study the "regional dialect" used in the texts, so that texts are marked as belonging to the legal domain in the UK, the USA, Australia etc. based on the language variety in which they are written and the geographical location of the publisher. Finally, lexicographers should also pay attention to the date of the texts included in corpora because the field of law is a dynamic field in which structural as well as terminological changes occur. As far as civil procedure terms in England and Wales are concerned, the important year is 1999; for the text-reception function, lexicographers should examine corpora for both old and new civil procedure terms, while they should look for new terms for the text production function.

A look at what is going on in the world with specific focus on civil proceedings and how this is reflected in text types that are likely to be included in corpora is relevant. Texts from UK newspapers with nationwide distribution are likely to be included in lexicographical corpora for general language dictionaries so the homepages of The Times and The Guardian were visited in order to see how often the two civil procedure terms *plaintiff* and *claimant* occur in these papers. The two homepages were visited on the same day and in order to make the searches as focussed as possible, the strings "plaintiff AND court" and "claimant AND court" were typed in the search boxes; the search results are presented in Table 5.

Legal term	The Times	The Guardian
plaintiff	923	5,160
claimant	2,970	42,400

Table 5: The number of texts containing legal terms in two UK newspaper databases

Table 5 shows that the new term *claimant* occurs in far more texts than the term *plaintiff*. If the term *plaintiff* is included in the dictionaries studied because of frequency of occurrence then the term *claimant* should also be lemmatized because it is much more frequent than *plaintiff*. A brief examination of the texts in which the term *plaintiff* occurs in the electronic archives of both newspapers reveals that most articles either appeared before the year 2000 (the time feature) or refer to civil proceedings outside England and Wales, particularly in the USA where the term *plaintiff* is in current use. One implication of this is that corpus builders and lexicographers should not only tag legal texts as British texts (the regional dialect feature) because they were published in the UK but they should also tag them with respect to the factual and culture-dependent contents (the domain feature), so that those texts dealing with civil proceedings in England and Wales are marked as such while texts describing civil proceedings in, for instance, the USA should be marked as such. This is one way in which to avoid the type of situation described by Magay (1984: 224): "Whether traditional human methods or computer-aided procedures are used, help must be given to the practicing lexicographer because he is often lost in a sea of words and terms, having to make haphazard choices". Another way in which to avoid haphazard choices is to have practicing lexicographers with factual and domain-specific knowledge about the subject fields and topics that are represented in dictionaries, because they possess the knowledge required to resolve issues regarding domain, time and regional dialect features.

In connection with legal terms in general dictionaries, it should be appreciated that there is a direct link between dictionary functions and underlying corpora. As mentioned in sub-section 5.1 above, dictionary functions affect all decisions for selecting lemmas and this includes decisions for compiling the corpus that is to form the basis of the selection. An assessment of the lemmatization of civil procedure terms reported in Tables 2 and 4 in the light of the extensive use of corpora, begs the following questions: 1) Do the corpora reasonably represent what is going on in the real word? 2) Are the corpora used in a reasonable way? The above discussion only answers these questions tentatively but at the same time highlights some important facts to be taken into consideration when including civil procedure terms — and other legal terms — in general language dictionaries. First of all, the corpora should be compiled so that they include texts that can help lexicographers find data that enable them to make dictionaries with, in this case, communicative functions, i.e. contain

data that help users understand civil procedure terms and use those terms in contemporary settings. This implies that lexicographers should possess the relevant factual and domain-specific knowledge in order to identify and lemmatize those civil procedure terms users really need in everyday usage situations so users can understand and talk about what is going on in the real world. It seems reasonable to suggest that the use of corpora for general language dictionaries treating legal terms, and perhaps terms from other culture-dependent subject fields, should be examined more extensively.

Secondly, the lemmatization of primarily old terms assists users who need help to understand old and new texts dealing with pre-1999 subject matter, while the lemmatization of primarily old terms does not help users talk and write about contemporary situations involving civil proceedings in England and Wales. Finally, one consequence of the practice followed by lexicographers of general language dictionaries when they do not lemmatize terms, when they remove subject field labels, and when they attempt to write very generalized definitions that (attempt to) cover both general and domain-specific meanings of words and terms is that the dictionaries become less authoritative. There are significant lemma lacunae and definitions become vague because the dictionaries contain definitions written for everyone and no-one with users ending up being uncertain or confused (see Nielsen 2012 for a similar discussion of the treatment of legal terms in *Collins COBUILD Advanced Dictionary*).

6. Concluding remarks

One of the problems facing lexicographers is that dictionaries can never be completely up to date due to various time lags involved in the dictionary-making process; though many time lags in online dictionaries are considerably shorter than those for print dictionaries. This means that when they consult general language dictionaries, users look at the present (the dictionary in front of them) but see the past. This is often no problem if the past still applies to the present but in connection with civil procedure terms some of the past does not apply to the present, as discussed above. The fact that the general language dictionaries examined do not lemmatize the new civil procedure terms that have applied since 1999 seems to indicate that a time lag of 15 years is too long. Furthermore, the fact that the four dictionaries include several of the old terms but not the new terms is a mystery, in particular because the dictionaries claim to contain up-to-date vocabulary.

The above findings and discussion indicate that lexicographers should pay more attention to the link between dictionary functions and underlying corpora. One way in which to ensure the lemmatization of new legal terms, in this case civil procedure terms, is to compile electronic corpora that contain texts that include relevant data, i.e. data that can help lexicographers make dictionaries that have the two communicative functions: provide help to understand terms and provide help to produce oral and written communication using

those terms. In addition, lexicographers should have the necessary factual and culture-dependent knowledge in order to identify and select the terms to be lemmatized. The discussion shows that the inclusion of old civil procedure terms is appropriate since the dictionaries provide help to understand those terms and to write about the concepts referred to by the terms. However, the failure to include the new civil procedure terms does not satisfy the needs of users because it prevents users from understanding and communicating about things that go on in the real civil procedure world today even though many people are likely to be involved in civil lawsuits at one time or another. A more stringent adherence to the principle of lexicographical functions would be one way in which to make dictionaries that contain up-to-date legal terms that provide help in communicative usage situations.

References

Dictionaries

Deuter, M., J. Bradbery and J. Turnbull (Managing Eds.). 2015. *Oxford Advanced Learner's Dictionary*. Ninth edition. Oxford: Oxford University Press.

Deuter, M., J. Bradbery and J. Turnbull (Managing Eds.). 2015. *Oxford Advanced Learner's Dictionary*. Ninth edition. [Online]. http://www.oxfordlearnersdictionaries.com/ [24 April 2015].

Fox, C. and R. Combley (Managing Eds.). 2014. *Longman Dictionary of Contemporary English.* Sixth edition. Harlow: Pearson.

Fox, C. and R. Combley (Managing Eds.). 2015. *Longman Dictionary of Contemporary English.* Updated edition. [Online]. http://www.ldoceonline.com/ [25 April 2015].

Hands, P. (Senior Ed.). 2009. *Collins COBUILD Advanced Dictionary of English*. Glasgow/ Boston: Harper Collins Publishers/Heinle Cengage Learning.

MacIintosh, C. (Editor-in-chief). 2013. *Cambridge Advanced Learner's Dictionary.* Fourth edition. Cambridge: Cambridge University Press.

MacIntosh, C. (Editor-in-chief). 2013. *Cambridge Advanced Learner's Dictionary*. [Online]. http://dictionary.cambridge.org/dictionary/british [24 April 2015].

Rundell, M. (Editor-in-chief). 2015. *Macmillan English Dictionary*. [Online]. http://www.macmillandictionary.com/ [25 April 2015].

Stevenson, A. (Ed.). 2015. *Oxford Dictionary of English. Third edition. 2015 Online Version*. [Oxford Reference Online]. DOI: 10.1093/acref/9780199571123.001.0001 [25 April 2015].

Other literature

Atkins, B.T.S. and M. Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.

Béjoint, H. 1988. Scientific and Technical Words in General Dictionaries. *International Journal of Lexicography* 1(4): 354-368.

Bergenholtz, H. and S. Nielsen. 2006. Subject–field Components as Integrated Parts of LSP Dictionaries. *Terminology* 12(2): 281-303.

Bergenholtz, H. and S. Tarp. 2010. LSP Lexicography or Terminography? The Lexicographer's Point of View. Pedro A. Fuertes-Olivera (Ed.). 2010. *Specialised Dictionaries for Learners*: 27-38. Lexicographica Series Maior 136. Berlin/New York: Walter de Gruyter.

British and Irish Association of Law Librarians. 2015. *Appendix Two. Civil Procedure Rules Terminology*. [Online]. Available at: http://www.biall.org.uk/data/files/Publications/Moys%20Classification, %20App%202,%20extractpages.pdf [25 April 2015].

Jessen, A. 1996. *The Presence and Treatment of Terms in General Dictionaries*. M.A. thesis, University of Ottawa, School of Translation and Interpretation, Ottawa. [Online]. Available at: https://www.ruor.uottawa.ca/handle/10393/4427 [29 April 2015].

Josselin-Leray, A. and R.P. Roberts. 2005. In Search of Terms: An Empirical Approach to Lexicography. *Meta* 50(4): 256-265. doi: 10.7202/019920ar.

Magay, T. 1984. Technical or General: Problems of Vocabulary Selection in a Medium-size Bilingual Dictionary. Hartmann, R.R.K. (Ed). 1984. *LEXeter '83 Proceedings: Papers from the International Conference on Lexicography at Exeter, 9–12 September 1983*: 221-225. Lexicographica. Series Maior. Tübingen: Max Niemeyer.

Ministry of Justice. 2015. *Civil Justice Statistics Quarterly October to December 2014*. Available at: https://www.gov.uk/government/statistics/civil-justice-statistics-quarterly-october-to-december-2014 [20 April 2015].

Mulcaster, R. (1582). The First Part of the Elementarie vvhich Entreateth Chefelie of the Right Writing of our English Tung, Set Furth by Richard Mulcaster. London: Thomas Vautroullier dwelling in the blak-friers by Lud-gate. Available at: http://quod.lib.umich.edu/cgi/t/text/text-idx?c=eebo;idno=A07881.0001.001 [26 April 2015].

Nielsen, S. 2012. Domain-specific Knowledge in Lexicography: How it Helps Lexicographers and Users of Accounting Dictionaries Intended for Communicative Usage Situations. *Hermes - Journal of Language and Communication in Business* 50: 51-60.

Svensén, B. 2009. *A Handbook of Lexicography. The Theory and Practice of Dictionary-making.* Cambridge: Cambridge University Press.

The Guardian [online]. http://www.theguardian.com [29 April 2015].

The Times [online]. http://www.thetimes.co.uk [29 April 2015].

Vrbinc, M. and A. Vrbinc. 2013. Differences in the Inclusion and Treatment of Terminology in OALD3, OALD4 and OALD8. *Lexikos* 23: 440-455.

http://lexikos.journals.ac.za

Treatment of Spelling Variants in Setswana Monolingual Dictionaries

Thapelo J. Otlogetswe, *Faculty of Humanities, University of Botswana, Gaborone, Botswana (otlogets@mopipi.ub.bw)*

Abstract: This paper argues that the Setswana language is characterised by spelling variants which are a consequence of multiple factors. It considers spelling variants found amongst individual words as well as those found in multi-word expressions (MWEs). It argues that spelling variation may be a result of historical fissions and amalgamations of the Batswana groups as well as borrowings from adjacent languages such as Afrikaans and English. The paper considers how three monolingual Setswana dictionaries of the past twenty years, *Thanodi ya Setswana* (Kgasa and Tsonope 1995), *Thanodi ya Setswana* (Mareme 2007) and *Tlhalosi ya Medi ya Setswana* (Otlogetswe 2012) have lemmatised spelling variants. The paper argues that spelling variants must be included in a general monolingual dictionary and that how such variants are handled must be informed by frequency information from corpus data. The paper concludes by proposing three strategies for addressing variation in MWEs where a difference between the two or more MWEs occurs because of a single word in the MWE or where variation is caused by the presence or absence of a word in a MWE. The third solution applies to cases where the variants differ in too many places such that it would be much more elegant to treat them as separate entries.

Keywords: SPELLING VARIATION, DIALECT, SETSWANA CORPUS, MULTI-WORD EXPRESSION, BORROWING, HISTORY, MONOLINGUAL DICTIONARY, LEMMATISATION, CROSS-REFERENCING

Opsomming: Die hantering van spellingvariante in eentalige Setswana woordeboeke. Hierdie artikel argumenteer dat die Setswanataal gekenmerk word deur spellingvariante wat die gevolg is van verskeie faktore. Oorweging word geskenk aan spellingvariante wat gevind word by individuele woorde sowel as dié wat in meerwoordige uitdrukkings gevind word (MWU's). Daar word geargumenteer dat spellingvariasie die gevolg kan wees van historiese verdelings en vermengings van die Batswanagroepe, sowel as ontlenings aan aanliggende tale soos Afrikaans en Engels. Daar word gekyk na die lemmatisering van spellingvariante in drie eentalige Setswanawoordeboeke van die afgelope twintig jaar, *Thanodi ya Setswana* (Kgasa en Tsonope 1995), *Thanodi ya Setswana* (Mareme 2007) en *Tlhalosi ya Medi ya Setswana* (Otlogetswe 2012). Hierdie artikel argumenteer dat spellingvariante in 'n algemene, eentalige woordeboek ingesluit moet word en hoe hierdie variante hanteer moet word deur gebruik te maak van frekwensie-inligting uit korpusdata. As slot word drie strategieë aangebied wat variasie in MWU's aanspreek, waar daar 'n verskil tussen die twee of meer MWU's bestaan as gevolg van 'n enkelwoord in die MWU of waar variasie veroorsaak word deur die teenwoordigheid of afwesigheid van 'n woord in 'n MWU. Die derde strategie geld vir gevalle waar die variante in soveel opsigte verskil dat dit 'n beter opsie sal wees om hulle as aparte inskrywings te hanteer.

Lexikos 25 (AFRILEX-reeks/series 25: 2015): 262-284

Sleutelwoorde: SPELLINGVARIASIE, DIALEK, SETSWANA KORPUS, MEERWOORDIGE UITDRUKKING, ONTLENING, GESKIEDENIS, EENTALIGE WOORDEBOEK, LEMMATISERING, KRUISVERWYSING

Introduction

Natural languages are characterized by numerous variants. The variants may be pronunciation variants, as those found, for instance, in words such as *data* (deɪtə or dɑ:tə), *potato* (pəteɪtəʊ or pəteɪtoʊ) and *either* (aɪðə or i:ðˀr) (Wells 2000). There are also spelling variants such as distinctions between American and British spelling peculiarities; found in terms such as *color* and *colour* and *behavior* and *behaviour*. The Setswana language like all natural languages is characterised by variants of various kinds. Some of the variants are dialectal (Southerland and Katamba 1996: 565) while others are spelling variants. By dialects we refer to features in an individual's speech that are associated with pronunciation, vocabulary, grammar and idiom (Honey 1997: 3; Crystal and Davy 1969: 67). A dialect may also be defined by other factors, such as social class or education. This paper restricts itself to the investigation of Setswana spelling variants most of which are a consequence of phonological features of different dialects. The paper argues that Setswana is characterized by multiple variants that pose a challenge to Setswana lexicography. The challenge that confronts a lexicographer is whether all variants of a form should be entered in a dictionary and if they were to be entered, how they would be treated in a monolingual Setswana dictionary.

First, we outline historical events which have had a bearing on spelling variation. We then demonstrate, with much evidence, that spelling variation is common in Setswana texts. The paper then considers how three monolingual Setswana dictionaries of the past twenty years have dealt with spelling variation. The dictionaries used are *Thanodi ya Setswana* (Kgasa and Tsonope 1995), *Thanodi ya Setswana* (Mareme 2007) and *Tlhalosi ya Medi ya Setswana* (Otlogetswe 2012). We then measure the frequency of variants in a twenty million-word Setswana corpus to determine which variant is to be the primary form in the dictionary. Variants of multi-word expressions are discussed and three strategies of how such variants could be lemmatized proposed. This paper therefore answers a specific question: How should variants be treated in Setswana dictionaries? In answering this specific question the paper proposes different strategies of how variants should be handled in a general Setswana dictionary.

Sources of variation in Setswana

Spelling variation in Setswana words and expressions is pervasive and is a consequence of numerous causes. Some of the spelling variation in Setswana has historical roots. It may be traced to the historical splits that occurred hundreds of years ago amongst the various Batswana groups. The civil wars, inter-

nal dissension and drought amongst the Batswana resulted in a permanent splitting of Batswana tribes which over the years led to dialectal variation (Otlogetswe 2014).

Below we discuss briefly how the separations between different Batswana groups gave way to dialectal variation amongst them.

Historical conditions for dialects

Dialects are caused largely by isolation and distance. Speakers of the same language who are separated from one another by a border, geographical distance, a hill, or a political border fence will end up developing dialects of the same language. If through separation, speakers come into contact with speakers of other languages, their language will continue to change over time. For Setswana, this is evident in the development of the Sekhurutshe dialect (which has developed from Sehurutshe), in northern Botswana which has been influenced extensively by Kalanga. Setawana has also been influenced by Seyei and languages in the Maun/Chobe areas of Botswana. Setswana in general has borrowed words extensively from English and Afrikaans.

Amongst the Batswana, civil wars, internal dissension and drought resulted in a permanent splitting of tribes. There is even a widely-held theory that the name Batswana is derived from the reciprocal verb stem *-tšwana*, (come or go out from one another, to separate) suggesting that its meaning is "(the offshoots or separatists), reference being made either to the separation of the Batswana from the main Bantu (or Sotho) stock to which they originally belonged, or to the separation from one another of the various tribes which we know today" (Cole 1955: xxi). While there is no compelling linguistic evidence to support this argument, there is however sufficient evidence that secession was common amongst the Batswana (Schapera 1963: 164). It is no wonder that Tlou and Campbell (1997: 96) characterise the history of the Batswana thus: "This is the history of the Batswana: groups of people splitting up and then other groups joining together." Such splits were a consequence of many factors, amongst these droughts which made subsistence difficult. Family and tribal feuds also contributed to more splits and separations. Since separation creates boundaries between people and boundaries between people create dialects of the same language, we now discuss the separation that occurred amongst the Batswana that has given way to dialectal variation.

The splits amongst the Batswana will be expounded through a discussion of the historical splits amongst the Bahurutshe. A comprehensive discussion of Batswana splits and their linguistic implications is found in Otlogetswe (2014). Much of the discussion that follows relies heavily on Ngcongco (1979) and Tlou and Campbell (1997).

Historically the Bahurutshe and Bakwena used to be a single group termed the Baphofu Confederacy (Tlou and Campbell 1997: 97). This group included the ancestors of the Bahurutshe, Bakwena (including Bangwaketse

and Bangwato), Batlharo, Bakgatla and some Bapedi. Towards the end of the 15th century the Confederacy began to disintegrate. First, the Batlharo separated and later a group led by Mokgatla seceded moving northwards to live with the Bapedi and later established an independent settlement.

Around 1500, the remaining Baphofu under Malope were living in upper Limpopo near Mabjanamatshwana (near current Pretoria). It is believed that Malope had heirs in two houses: the first born child in Malope's senior house was a daughter, Mohurutshe, while the first born child in the second house was a son, Kwena. A dispute arose as to whether the chiefdom should be ceded into the hands of the eldest child in the senior house despite her being female, or whether the leadership should be kept male by appointing the senior son of the second house as chief. Following Malope's death, the dispute caused a split around 1475 to 1505. The followers of Mohurutshe were forced to leave Mabjanamatshwana and moved south as a separate group, with a separate totem (the baboon).

Between five and seven generations after the separation of the Bahurutshe and the Bakwena (around 1625–1655), while Mogopa was still ruling the Bakwena, a terrible famine which was famously termed '*tlala e e boitshegang*' scattered and dispersed the Bakwena clans far and wide. As a result of this famine, many Bakwena clans migrated south of the Lekwa or Vaal River into the modern Free State.

Mogopa and the remaining Kwena groups, which included the Modimesana clans and those that later formed the Botswana group of the Bakwena migrated to Mabjanamatshwana along the Odi River to its confluence with Madikwe and there built a settlement named Rathatheng. After a period of very strenuous or difficult existence owing to scarcity of food and water, Mogopa migrated back to Mabjanamatshwana in the modern Brits district of the former Transvaal.

Thus, partly as a result of the droughts and famines that occurred during the generation c.1625–c.1655, two Kwena kingdoms in the former western Transvaal emerged. These were the Bakwena-Mogopa based at Mabjanamatshwana and the Bakwena-Kgabo at Rathatheng.

Splits caused by droughts and famines also applied to the Bahurutshe state which gave birth to the Manyana and Gopane chiefdoms, the Bakaa and the Phuduhutswana-Tlhaping hived off from the Barolong-Tshidi, and the Bakwena-Modimosana split up and formed the four chiefdoms known as Ramanela, Maaka, Mmatau and Matlhaku. It is believed that it was at Rathatheng that Kgabo II was succeeded by his son Motshodi although according to some traditions, he (Kgabo II) led the migration of his followers across the Madikwe into present-day Botswana.

Towards the end of the long reign of Motshodi, the Kwena-Kgabo kingdom broke up. From this fission two new independent states came into being; the Ngwato and Ngwaketse kingdoms. Parsons (1973: 84) estimates the Ngwato secession to have occurred around 1790. Towards the end of the 18th century a group of Bangwato led by Tawana seceded to form an independent tribe on the shores of Lake Nghabe.

The above discussion gives a broad picture of some of the Batswana splits and fissions which subsequently gave rise to distinct dialectal features some of which are discussed in the development of this paper.

Some phonological differences between Setswana dialects

Because of the various separations amongst the Batswana, there are a number of linguistic variations that developed. Many of these are lexical while others are phonological with lexical implications. We therefore consider a few phonological differences between the various Batswana groups. By phonology we refer to the characteristic pronunciation patterns of a speech community. We only discuss those which have had impact on written Setswana. Phonological differences are important markers of linguistic variation and have been previously used to differentiate dialects (Batibo 1999; Cole 1955). Phonological features discussed here are only those that are reflected in certain spelling peculiarities.

a. The [tl/t] and [tl^h/t^h] distinction

One of the obvious differences between the northern Setswana dialects (Sengwato and Setawana found around the areas of Serowe and Maun) and central dialects (Sengwaketse, Serolong and Sehurutshe found around Kanye and Mahikeng areas) is the distinction between [t^h/t] and [tl^h/tl] as in *batla/bata* (want) and *kgotlha/khotha* (poke). Northern Setswana uses [t^h/t] where all other Setswana dialects use [tl^h/tl] as well as where central dialects use [tl/t] (Batibo 1999). The northern Setswana dialect is sometimes mistakenly labelled by some South Africans Setswana speakers as *Setswana sa Botswana* (Setswana spoken in Botswana). This is because the northern Setswana dialect speakers, Bangwato and Batawana, are found only in Botswana and not in South Africa compared to Bakgatla ba ga Kgafela, Barolong and Bahurutshe who are found both in Botswana and South Africa. Table 1 presents comparative data of how Northern Setswana words are pronounced compared to central Setswana dialect speech.

Table 1: A comparison [tl] and [t] in Sengwato and other dialects

Sengwato dialect	Southern dialects	English
thaga	*tlhaga*	grass
othaya	*otlhaya*	discipline
setha	*setlha*	bladder
thaba	*tlhaba*	pierce; kill
ta	*tla*	come
bata	*batla*	search for
tota	*tlotla*	respect

Phonologically, this manner of pronunciation has led to lexical ambiguity between certain words which does not exist elsewhere in other Tswana dialects. Table 2 demonstrates the challenges that may result as a consequence of pronouncing [tl] as [t] and [tl^h] as [t^h].

Table 2: Ambiguity as a result of the conflation of [tl] and [t]

Southern dialect	**Sengwato**	**Ambiguous with**
batla "want"	*bata*	*bata* (bath)
tladi "thunderbolt"	*tadi*	*tadi* (striped cat)
tlala (fill)	*tala*	*tala* (green)
tlhaba (pierce)	*thaba*	*thaba* (hill)
tlhaka (letter)	*thaka*	*thaka* (pupil of eye)
tlhapa (bath/wash)	*thapa*	*thapa* (employ)

Setswana orthographies over the years have standardized the Setswana writing system so that though northern Setswana speakers can speak using [t^h/t] instead of [tl^h/tl], in writing they are taught to write [tl^h/tl] instead of [t^h/t] (Ministry of Education 1981; Chebanne 2008 and Chebanne et al. 2008). Although this is the case, it is common that many of the writings of northern Setswana speakers reflect their speech peculiarities resulting in lexical ambiguity. The problem outlined in this section is much more common in Botswana than in South Africa since the northern Setswana dialect is spoken only in Botswana.

The challenge for a Setswana lexicographer in this instance is not severe since the problem is limited to a northern Setswana dialect and it is rare in written text though it exists phonologically and in some students' essays and in the social media such as Facebook text.

b. The difference between [lo-] and [le-]

One conspicuous dialectal difference between northern and central Setswana varieties is the distinction between the noun class singular prefix [le-] of class 5 nouns and [lo-] singular prefix of class 11. In the central dialects such as Sengwaketse and Serolong the distinctions between the two are still maintained while in the northern dialects such as Sengwato and Setawana the distinctions have been collapsed into singular prefix [le-] of class 5. This difference has been identified by Cole (1955). All the words which in other dialects are class 11 nouns are class 5 nouns in northern dialects. This is illustrated in Table 3 below.

Table 3: The [lo-] and [le-] in Setswana dialects

Southern dialects	Northen dialects	Gloss
logong	*legong*	wood
lokotswana	*lekotswana*	wall
lobadi	*lebadi*	door
lofeelo	*lefeelo*	broom

According to general linguists Setswana books (e.g. Mogapi 1984; Cole 1955) and Setswana orthography, the determination of whether a word takes a [le-] or [lo-] prefix is based on the plural prefix that the word takes (Chebanne et al. 2008). Class 11 nouns when pluralised take a class 8 prefix [di-] while class 5 nouns when pluralised take a class 6 prefix [ma-] (Cole 1955; Chebanne et al. 2008). This can be summarised as follows:

Singular noun class prefix	→	**Plural noun class prefix**
noun class 5 [le-]	→	noun class 6 [ma-]
noun class 11 [lo-]	→	noun class 8 [di-]

The rule above can be applied to the following linguistic data as follows:

Table 4: Formation of class 5 and 11 plurals

Noun	English meaning	Plural form	NOT
logong	wood	*dikgong*	*legong > magong*
lobone	light/bulb	*dipone*	*lebone > mabone*
lobota	wall	*dipota*	*lebota > mabota*
lebante	belt	*mabante*	*lobante > dipante*
lerato	love	*marato*	*lorato > dithato*
leeto	travel/visit	*maeto*	*loeto > dieto*

Although these linguistic rules exist, they have been applied inconsistently by Setswana writers. A lexicographer is therefore bound to find in a Setswana corpus spelling variation as a consequence of the inconsistent use of [le-] and [lo-] in the language. For instance in a twenty million-token Setswana corpus (Otlogetswe 2008), *logong* occurs 511 times while *legong* occurs 180 times. *Loeto* occurs 925 times while *leeto* occurs 558 times. To resolve these spelling variations, a lexicographer has two possible approaches to adopt. He can enter both variants in a dictionary and at the *wrong* spelling offer a cross-reference to the properly spelt word and indicate that the offered spelling is unacceptable. Alternatively, a lexicographer could only lemmatise those headwords which are consistent with Setswana spelling and treat grammatical information, including spelling rules, in the front matter of the dictionary.

c. Borrowings and variants

Setswana has borrowed extensively from both English and Afrikaans. Many borrowed words result with spelling variation since words are usually borrowed as they are heard. Since the Setswana language, especially in Botswana, Zimbabwe and Namibia lacks a language board which fixes spellings of borrowed terms, the variants have remained in Setswana texts. A good example is the word *computer* which has been borrowed into Setswana using any of the following spelling variants: *khompiutara, khomphiutara, khomputara, komputara* or *khompiuta*. The 1988 *Terminology and Orthography No.4* (Setshedi et al. 1988: 129) gives *khomphutara* as the Setswana equivalent for *computer*. The word *bus* has been borrowed into Setswana as *base*, whose pronunciation is closer to the English pronunciation of *bus*, and *bese*, whose pronunciation is significantly distinct from English. Table 5 shows borrowings from English and their variants in Setswana. The data is derived from a twenty million-word corpus.

Table 5: Variants of borrowed terms

Borrowed variants	**Original borrowed term**
Baebele, Baebela, Bibele	Bible
boronse, boronso	bronze
diabolo, diabolose	devil
enfolopo, enfelopo, enfolopo	envelope
Keresemese, Keresemose, Khirisimore	Christmas
khompiutara, khomphiutara, khomputara, komputara or *khompiuta*	computer

The challenge that confronts a lexicographer is how a dictionary should capture the spelling variations that are found in a corpus. For a general dictionary variants are interesting and important to include in a dictionary so that users may find what they are looking for. This is particularly important if the variation occurs at the beginning of a word and therefore necessitates distinct lemmatization of variants in different parts of a dictionary. Variants are, however, a challenge since they take space in a dictionary and contribute no distinct meaning to their variant form. In some dictionaries (e.g. Otlogetswe 2012) spelling variants are treated as synonyms. However, treating them as synonyms implies that they constitute separate lexical items (lexemes) when in fact all variants constitute a single lexeme only having alternate forms. Such forms are therefore better marked as *variants*, that is, identical terms that only happen to have an idiosyncratic spelling. Writing about the English language Svensén observes that:

> Many words have spelling variants. Since we are dealing here with well-established languages with a long lexicographic tradition, it is seldom difficult to ascertain which spelling is to be preferred and, consequently, is to appear as a

> lemma. Frequent spelling variants should be included, although as a rule only in dictionaries entirely or partly intended for reception; in production dictionaries they are mostly unnecessary (Svensén 2009: 110).

Svensén's observation holds true for English as well as for Setswana. As has been demonstrated before, there are various reasons which may lead to variation in Setswana and as we will show in the development of this paper, sometimes it is not clear which variant form should be considered the preferred lemma to which other variants should be cross-referenced.

The data presented in the tables that follow demonstrates how widespread the variant phenomenon is across different Setswana word classes. In our study we have identified one thousand variants in *Tlhalosi ya Medi ya Setswana* dictionary (Otlogetswe 2012) which constitutes 6.5% of all of the dictionary headwords. We start with evidence of spelling variation amongst nouns.

Table 6: Variants amongst nouns

Variants	**Word equivalent**
bokoso, lebokisi, lebokose, lebokoso, bokose, lepokisi	box
bolakaboroto, boroto	blackboard, board
kgonagalo, kgonafalo	a possibility, a chance, a likelihood
poresidente, moporesitente, moporesidente, poresitente	president
boperesiti, boperesita	priesthood
dithutopuisano, dithutapuisano, dithutopuisanyo	workshops
dithutopuisano, dithutapuisano, dithutopuisanyo	blouse
lekwaladikgwang, lekwalodikgang	a newspaper
letlebekwane, letemekwane, letemelekwane	a dumbling
bopelokhutshwane, bobelokhutshwane	impatience
sekarete, sekarese, sekerete, sakerete	cigarette
fatuku, faatuku, fatukwi	a table cloth
base, bese	bus
sepalamo, sepagamo, sepalangwa	a transport
moitseanape, moitsaanape, maitseanape	an expert
ntšhe, ntšhwe, mmantšhe, mmantšhwe,	ostrich
khompiutara, khomphiutara, khomputara, komputara	computer
leje, lejwe	a stone
lerapo, lesapo	a bone
Bopelotshetlha, bobelotshetlha	greed
Tlhotlhwa, tlhwatlhwa	a price

It is not only nouns that are characterized by variants. Verbs also have numerous variants as shown in Table 7.

Table 7: Variants amongst verbs

Verb spelling variants	**Meaning/equivalent**
atamêlana, atamalana, atumêlana	come closer to each other
abêlana, abalana	share
atolosa, katolosa	widden
Tšhemola, tšhamola, tšhwemola, tšhwamola	slip away
batola, bitola	beat severely
Fologa, gologa	step down
Gagoga, kgagoga	tear apart
Katogana, katologana	spread out
Kgorometsa, kgarametsa	push
Palama, pagama	climb onto
Pepetla, papetla	quash
Reetsa, theetsa	listen
Rutlolola, rutlomolola	unroof
Thela, tshela	pour
Thologa, tshologa	pour out
Tsisa, tlisa	bring
Thathologa, thanthologa	unwind

The spelling variations are also found not just in common nouns but also in proper nouns as shown in Table 8.

Table 8: Variants of Names of persons and places

Setswana variants	**English**
Aforika, Aferika	Africa
Baebele, Babele, Beibele	Bible
Ennyelane, Engelane, Enngelane, Enyelane	England
Faro, Farwe	Pharaoh
Balete, Bamalete	Balete
Baphoting, Baphuting	Baphoting
Borithani, Borithane	Britain
Diphalane, Phalane	September
Goleate, Goliata, Goliate, Goliathe	Goliath
Gouteng Gauteng	The city of gold (Johannesburg)
Hirikgong, Herikgong, Ferikgong, Firikgong	January
Keresemose, Keresemese, Keresemose, Khirisemose	Christmas
Lenyelesemane, lenyesemane, leesemane, moesemane	An Englishman man
Mmantaga, Mantaga	Monday
Ngwanaitseele, Ngwanatsele, Ngwanaatsele	November
Jorotane, Joretane	Jordan

Spelling variants are also to be found amongst interjectives in Setswana. We present these in Table 9.

Table 9: Variants on interjectives

Interjective variants	**Interjective meanings**
uši, utšhi, kuši, kutšhi, iši, itšhi,	an expression of pain
haa! haaa! haahaaa!	used to represent laughing
haleluja, haleluya	an expression of praise to god
ijaa! ija!	an expression of shock
še! šeng!	an expression of shock or surprise
thakaa! thakanaa!	an expression of shock or surprise
tshi! tshikhi!	an expression used to capture a cold environment
tshuu! tshutshu! tšhuutšhuu!	an expression used to show that one is burning
tswee, tsweetswee, tswedintswerere	a plea

There are also lexical variations which are as a consequence of the variation of [f] and [h] in Setswana words

Table 10: Variants caused by f/h alternation

Words	**English**
fora, hora	deceive
funolola, hunolola	untie
futswela, hutswela	add milk to porridge
fuduga, huduga	move to another place
mohaladi, mofaladi	a stranger from elsewhere
sehatlha, sefatlha	a naughty person
lefuto, lehuto	a knot
bofafalele, bohahalele	carelessness
mahura, mafura	fat/oil
firisa, hirisa	to rent

The tables above have demonstrated how widespread variants are in Setswana across different word classes. There are variants of different types. Some variants are caused by a difference in a single vowel or consonant, for instance *fata/hata* (dig), *Gouteng/Gauteng* (Johannesburg), *Aforika/Aferika* (Africa), *reetsa/theetsa* (listen), *Jorotane/Joretane* (Jordan) etc. Other variants are caused by the presence of an additional syllable(s) in one of the variants, for instance, *Phalane/Diphalane* (October), *Ngwanatsele/Ngwanaitseele* (November), *Mmantaga/Mantaga* (Monday), *katogana/katologana* (part/move away from each other), and *bokoso/lebokoso* (a box).

Treatment of variation in Setswana monolingual dictionaries

In this section we test how three Setswana monolingual dictionaries have lemmatized a randomly selected group of words with spelling variants. We consider three dictionaries of the past twenty years: *Thanodi ya Setswana* (Kgasa and Tsonope 1995), *Thanodi ya Setswana* (Mareme 2007) and *Tlhalosi ya Medi ya Setswana* (Otlogetswe 2012). Kgasa and Tsonope's (1995) dictionary was for a long time the definitive monolingual Setswana dictionary which was widely used in Botswana and South Africa. For years it was the prescribed Setswana dictionary in Botswana schools. Mareme (2007) is the largest Setswana dictionary to come out of South Africa. It has much in common with Kgasa and Tsonope (1995). It was produced by the Setswana Lexicographic Unit based at the North West University (the Mahikeng campus). Otlogetswe (2012) is the most recent of the three dictionaries compiled with a great dependency on a large Setswana corpus. The three dictionaries are therefore contemporary monolingual dictionaries which are in general circulation and use.

Table 11: Measure of variation in Setswana dictionaries

Word	Kgasa and Tsonope (1995)	Mareme (2007)	Otlogetswe (2012)
lebele (breast)	✓	✓	✓
letsele (breast)	✓	✓	✓
iši (ouch!)	✓	x	✓
uši (ouch!)	x	x	✓
itšhi (ouch!)	x	x	✓
utšhi (ouch!)	✓	x	✓
gofele (the youngest child)	x	x	✓
gofejane (the youngest child)	✓	✓	✓
kgonotšwe (thumb)	x	x	✓
kgonotswe (thumb)	x	✓	x
kgonope (thumb)	✓	✓	✓
kgonojwe (thumb)	x	✓	✓
khomphiutara (computer)	✓	x	✓
khompiutara (computer)	x	x	✓
khomputara (computer)	x	✓	✓
khomphutara (computer)	x	✓	x
khomphuta (computer)	x	x	x
jala (plant)	x	✓	✓
jwala (plant)	✓	x	✓
TOTAL	**8**	**9**	**16**

The results as shown in Table 11 indicate that different Setswana dictionaries adopt different strategies in their inclusion of variants. Of the twenty variants tested in Table 11 Kgasa and Tsonope (1995) have eight variants. Mareme

(2007) has nine while Otlogetswe (2012) has sixteen. The dominant approach with Kgasa and Tsonope (1995) seems to be a choice of one of the variants and including it as a headword while other variants are excluded though this approach is abandoned in the treatment of other variants such as *lebele* and *letsele*.

It isn't clear on what basis one of the variants was chosen amongst the many to refer to the other variants.

With Mareme (2007), there isn't much of a discernible pattern in the treatment of spelling variants since in some cases the variants are lemmatized as in the case of *letsele/lebele* and *kgonope/kgonojwe*, while in other cases one of the variants is lemmatized to the exclusion of one of the variants. For instance *jala* is lemmatized while *jwala* is not. *Khomputara* and *khomphutara* are lemmatized while *khomphiutara,* and *khompiutara* are not lemmatized. Otlogetswe (2012) tends to lemmatize many of the variants with the exception of only a few. The Setswana monolingual dictionaries therefore adopt different strategies in their lemmatization of spelling variants in the dictionary.

Below we give examples of how the variants have been treated in the three dictionaries. We only consider *lebele/letsele* (breast) from Kgasa and Tsonope (1995), *kgonojwe/kgonotswe* (thumb) from Mareme (2007) and *khomphiutara*, *khompiutara* and *khomputara* (computer) in Otlogetswe (2012) for brevity.

> **lê·bêlê** TGG *ln/5*. ma-. nama e e mo sehubeng sa motho, mo basading e ipotokile, mo diphologolong e fa gare ga maoto kana mo mpeng; e kgona go ntsha maši mo ditsheding tse di namagadi
> **lê·tsêlê** TGG *ln/5*. ma-. nama e e mo sehubeng e e amusang mo diphologolong tse di namagadi
> (Kgasa and Tsonope 1995: 115, 134)

Kgasa and Tsonope (1995) do not link the two variants through cross-referencing; instead they offer the full treatment of tone, part of speech and definition to both *lebele* and *letsele*, though *lebele*'s definition is more elaborate than that of *letsele*. We now consider *kgonotswe/kgonojwe* variants in Mareme (2007).

> **kgonojwe** *(ln) (**9**/10)* di monwana o motona wa leoto kgotsa seatla
> **kgonotswe** *(ln) (**9**/10)* **di- BONA kgonojwe**
> (Mareme 2007: 230)

Mareme (2007) links *kgonotswe* to *kgonojwe* through cross-referencing and does not offer any definition at *kgonotswe* only offering it at *kgonojwe*. He however does not link *kgonojwe* to *kgonotswe*. This means that when users consult *kgonojwe* they are unlikely to know that *kgonojwe* has a variant *kgonotswe*. We now move to consider the *khompiutara* variants.

> **khomphiutara** /kʰʊ̀mp̀ʰíútà:rá/ •*ln. 9. n-, *10. din-*• = KHOMPIUTARA
> ⇐ SeE: computer

khompiutara /kʰʊ̀m̀píútà̀rá/ •*ln. 9. n-, *10. din-*• motšhine o o dirisiwang go boloka ditlhaka, ditshwantsho, le tse dingwe, o dirisiwa go kwala, dipalo le dipuisano tsa imeile jalojalo ⇐ SeE: computer

khomputara /kʰʊ̀m̀pútà̀rá/ •*ln. 9. n-, *10. din-*• = KHOMPIUTARA ⇐ SeE: computer

(Otlogetswe 2012: 210)

Otlogetswe (2012) only offers a definition at *khompiutara* and then links *khomphiutara* and *khomputara* to *khompiutara* through cross-referencing. However at *khompiutara* he does not show the other variants to which *khompiutara* is linked. This means that when one consults a dictionary for *khompiutara* one is unlikely to know its variants.

It is however not clear if any of the Setswana dictionaries have handled the matter of cross-referencing of variants appropriately.

It is the argument of this paper that the decision of what to lemmatize and what constitutes a primary form to which other variants refer, should be based on corpus frequency. We also argue that to preserve space only one of the variants should have full lexicographic treatment and carry a definition or definitions in the case of polysemous entries. Such a variant, we call a primary form. Frequency list information can assist, not just in determining which variants to include, but also in deciding which of the variant forms is the primary form on which a full lexical treatment can be made. Below we test the treatment of *khompiutara*, *lebele* and *kgonotswe* variants against corpus evidence to determine the appropriate primary form of each. The table below presents the corpus frequencies of the variants

Table 12: Frequency of selected variants in a corpus

Variant	**Frequency**
khomputara	181
khompiutara	38
khomphiutara	6
khompiotara	6
khomphuitara	4
khomphutara	2
lebele	412
letsele	136
kgonotswe	60
kgonope	29
kgonojwe	17
kgonotšwe	11

The corpus data demonstrates that amongst the *khomputara, khompiutara, khomphiutara, khompiotara, khomphuitara and khomphutara,* variants, *khomputara* is

the most frequent variant with a frequency of 181 that qualifies to be entered in the dictionary as a primary variant form. It should therefore have the full meaning treatment with all the other variant forms linked to it through cross-referencing. At the *khomputara* entry all the other variants should be listed at the end of the entry as variant forms. This is so that a user who consults the dictionary meaning at *khomputara*, may also be aware of other variant forms.

Lebele is the most frequent variant with a frequency of 412 in the corpus. It would have the full meaning treatment in the dictionary as a primary variant with *letsele* linked to it through cross-referencing. The same approach will be applied to *kgonotswe* and its variants as above. *Kgonotswe* would receive a full lexicographic treatment with *kgonope, kgonojwe* and *kgonotšwe* cross-referenced to it since *kgonotswe* has the highest frequency comparatively. The treatment of *lebele/letsele* variants by Kgasa and Tsonope in which both variants are defined is therefore not recommended since it fails to show the link between the two variants. A more preferred approach is the one similar to that of Otlogetswe (2012) in the treatment of *khomputara* as the primary form and having all the other variants cross-referenced to it. Corpus evidence however shows that instead of *khompiutara* receiving full lexicographic treatment, as Otlogetswe (2012) has done, it is *khomputara* that must receive full treatment and have all the other variants cross-referenced to it since it is the most frequent of all the variants. The inclusion of variants cross-referenced to a primary form is attractive since it is user friendly and assists the learner (Svensén 2009). Moon prefers this approach also arguing that:

> Some, particularly learners' dictionaries, help users by routinely setting cross-references to the place of treatment. This solves the problem of handling variations, and at least means that users are supported during their search for information. (Moon 1994: 342)

Linking variants by cross-referencing brings together related dictionary material which may be spread throughout the dictionary. It saves space by giving one of the variants full lexicographic treatment and linking related headwords.

Multi-word expression variants

Variants do not only exist at the spelling level of the word, but also occur at the level of multi word expressions. By multi word expression is meant:

> … any word combination (adjacent or otherwise) that has some feature (syntactic, semantic or purely statistical) that cannot be predicted on the basis of its component words and/or the combinatorial processes of the language. Such units need to be included in any language description that hopes to account for actual usage (Bannard 2007: 1).

MWEs therefore include idioms, phrasal verbs, proverbs, compound words, etc. (Otlogetswe 2009). English examples include *ad hoc, by and large, kick the*

bucket, in step, take up, take off, shake up, telephone booth, pull strings, fresh air, fish and chips, salt and pepper etc. Setswana examples include *solegela molemo* (benefit), *kukega maikutlo* (be upset), *iphaga dikoro* (involve yourself in other people's business), *tsholetsa maoto/dinao* (walk faster), *opisa tlhogo* (cause trouble), *tsaya karolo* (participate), *tsaya tsia* (pay attention), *nna le seabe* (take part), and *ja monate* (enjoy).

Moon has observed that:

> idioms are always difficult to treat lexicographically. This is not just because of the problems of variation and lexical form. There are other problems presented by idioms: how to convey the meaning and usages of what are essentially context-bound items, with vague or plastic meanings and heavy connotations (Moon 1994).

MWEs just like single word forms also display variation. For instance English has the following variants:

Get a raw deal
Have a raw deal
Have cold feet
Develop cold feet
Have one's feet on the ground
Keep one's feet on the ground
With one's feet on the ground
With one's nose in the air
Stick one's nose in the air
Drag one's feet
Drag one's heels
Hit the roof
Hit the ceiling
At least
At the least
At the very least
Shake in one's shoes
Shake in one's boots
Quake in one's shoes
Quake in one's boots
Quiver in one's boots
Quake in one's Doc Marten's
(Moon 1994: 90-100)

Just like English, Setswana also has MWEs that display variation. Examples include the following:

Table 13: Variation amongst MWEs

MWE VARIANTS	**MEANING**
a. *bolwetse jwa pelo* (Noun Phrase) b. *bolwetsi jwa pelo*	Heart disease
a. *koma ka tlhogo* (Idiom) b. *koma tlhogo*	To nod one's head
a. *kgomo e e maši ga e itsale* (Proverb) b. *e e maši ga e itsale*	Just because a parent is good, it doesn't mean that they will have good children

a. *kgwele ya dinao* (Noun Phrase) b. *kgwele ya maoto*	Football
a. *matlho a ngwana a roga mogolo ga a gake* (Proverb) b. *matlho a roga mogolo ga a gake*	A person with a secret is revealed by how they behave
a. *bana ba motho ba kgaogana tlhogo ya ntsi* (Proverb) b. *bana ba motho ba kgaogana tlhogo ya tsie*	People who associate in some form share the little that they have
a. *mogama kgomo tsa mafisa o di gama a lebile tsela* (Proverb) b. *mogama kgomo tsa mafisa o gama a gadimile tsela* c. *mogama kgomo ya mafisa o e gama a gadimile kwa morago*	A person who is using something that is borrowed, uses it knowing that it may be wanted back any time
a. *monate o tswa ka ditsebe* (Idiom) b. *monate o tswa ka phogwana*	This idiom is used to mean something (e.g. food) is very nice
a. *monna o bolawa ke se a se jeleng* (Proverb) b. *monna o jewa ke se a se jeleng*	It is the actions of an individual which bring them harm
a. *moswela Tebele ga a jewe o tshwana le moswela gagabo* (Proverb) b. *moswela gae ga a jewe o tshwana fela le moswela nageng*	People can live where they can best make a living even though it is not where they were born
a. *bodiba bo jeleng ngwana'a mmago sika ka bone o bo kakologe* (Proverb) b. *bodiba ba go ja ngwana wa ga mmago, e re o feta ka jone o bo dikologe* c. *bodiba ba go ja ngwana wa ga mmago, e re o feta ka jone o bo tlarologe*	An individual must learn from other people's mistakes and avoid danger

Treatment of MWEs in Setswana dictionaries

Table 13 has demonstrated the extent as well as types of variation in Setswana MWEs. As in variation amongst individual words, the MWEs such as idioms and proverbs have variations which pose unique challenges to their lemmatisation in a monolingual general dictionary.

Svensén observes that:

> There are two options as regards the positioning of an idiom in dictionaries. Either it can be entered as an indication in the entry for one of the component words, or it can be entered as a lemma, and in the latter either as an independent lemma or sublemma. (Svensén 2009: 194)

Kgasa and Tsonope (1995) do not lemmatize the MWEs as headwords, they include them as subentries or in Svensén terminology, as sublemmata. For instance the idiom *go bona ka bobi jwa segokgo* (to see a little bit) is entered as a subentry of *bona* (see):

> **bôna** GT tpt. fa o lebile sengwe ka matlho o a se bôna *go bona ka bobi jwa segokgo = go bona go le go nnye
> (Kgasa and Tsonope 1995: 25)

Mareme (2007) doesn't deviate from Kgasa and Tsonope's (1995) approach. For instance:

> **kakabolola** *(ldr)* **1** go utlwisa botlhoko ka go itaya kgotsa go tsietsa thata *Go kakabolola ditsebe: go betsa thata* **2** go thibolola **3** go bulela
> (Mareme 2007: 189)

In this example *go kakabolola ditsebe* (to beat severely) is a subentry of the headword *kakabolola* (unblock).

However Svensén argues that:

> Presenting an idiom as an indication under one of its components, which is the traditional and probably still prevailing method, is actually contrary to the idiom's nature of an independent lexical item. (Svensén 2009: 194)

This is because the idiom or proverb is semantically non-compositional and usually semantically unrelated to the headword under which it is subsumed. For instance *kabolola ditsebe* is not semantically related to *kabolola* or *ditsebe*. It is therefore flawed to associate it with any of the words that constitute it. Otlogetswe (2012) adopts a different approach from Kgasa and Tsonope (1995) and Mareme (2007). He enters MWEs such as idioms and proverbs as independent lemmas in the dictionary without subsuming them under a headword. For instance:

> **botlhale jwa phala bo tswa phalaneng, phala e se nang phalana lesilo** •*seane*• bagolo ba ka thusiwa ke bana ka megopolo
> (Otlogetswe 2012: 63)

This treatment of the MWE as a separate lemma is the one we support since an idiom is "an independent lexical item having an opaque meaning" (Svensén 2009: 194).

Treatment of MWE variants in Setswana dictionaries

Returning to the matter of entering variants in dictionaries, an examination of Kgasa and Tsonope (1995) and Mareme (2007) dictionaries reveal that they both exclude variants whilst Otlogetswe (2012) includes them. For instance the

case of *go itaya kgomo lenaka*:

> **itaya** TTT |itaa tpt. > itaya –ile. utlwisa botlhoko ka go betsa ka letswele kana go ngata. * *go itaya kgomo lenaka = go bua selo fela jaaka se ntse kana nnete*
> (Kgasa and Tsonope 1995: 72)

The idiom *go itaya kgomo lenaka* (to get something right) has multiple variants that are not captured in either Kgasa and Tsonope (1995) or Mareme (2007). These are *go opa kgomo lonaka* and *go otla kgomo lonaka* which are included in Otlogetswe (2012) though they are included as synonyms:

> **itaya kgomo lonaka** •*leele*• go bua nnete jaaka e ntse = ÔPA MAGÔGWÊ MO NKONG, ÔPA KGOMO LONAKA.
> (Otlogetswe 2012: 148)

> **ôtla kgomo lonaka** •*leele*• = ÔPA MAGÔGWÊ MO NKONG, ÔPA KGOMO LONAKA.
> (Otlogetswe 2012: 452)

> **ôpa kgomo lonaka** •*leele*• go bua se e leng buammaaruri *Le fa o ne o bua fela o se na bosupi, o ne wa opa kgomo lonaka ka go diragetse fela jaaka o buile*
> (Otlogetswe 2012: 450)

The evidence therefore demonstrates that the three Setswana dictionaries adopt different strategies to deal with multi-word expression variants. Kgasa and Tsonope (1995) and Mareme (2007) do not include variants while Otlogetswe (2012) does, although some of them are given a full treatment with meaning. In the following section we propose three different ways of handling multi-word expressions in Setswana.

Three strategies of handling MWE variants in a Setswana dictionary

We propose that multi-word expressions be handled in any one of the three different ways in a general dictionary.

1. Where a difference between the two or more multi-word expressions is just because of a single word in the MWE in the same part of a phrase then a forward slash [/] may be used in the entry of a MWE or fixed expression to save space to show alternates. This approach is the one used by the *Macmillan English Dictionary for Advanced Learners* (Rundell 2007) in the treatment of multi-word expressions, for instance, in the treatment of the word *framework*.

framework […] **1.** …

◆ **provide/establish/create a framework** *Einstein's research provided much of the theoretical framework for particle physics*
(Rundell 2007: 595)

For Setswana, this means that the following multi-word expressions will be treated in the following manner:

The variants *monate o tswa ka ditsebe* and *monate o tswa ka phogwana* only differ in terms of the use of the words *ditsebe* and *phogwana* in the same spot which engender variation between the two structures. We therefore propose that these be treated in the following manner:

monate o tswa ka ditsêbê/phogwana •*leele*• monate o o kwa godimo *Re ne re itumetse kwa moketeng, monate o tswa ka ditsebe phogwana*

This approach therefore avoids the approach that follows which uses much space in the dictionary.

monate o tswa ka ditsêbê •*leele*• monate o o kwa godimo *Re ne re itumetse kwa moketeng, monate o tswa ka ditsebe*

monate o tswa ka phogwana •*leele*• monate o mogolo = MONATE O TSWA KA DITSÊBÊ
(Otlogetswe 2012: 376)

Additionally *monna o bolawa ke se a se jeleng* and *monna o jewa ke se a se jeleng* only differ in the use of the words *bolawa* and *jewa*. The proverb would therefore be represented as follows in the dictionary:

monna o bolawa/jewa ke se a se jeleng •*seane*• ditlamorago tsa sengwe di bakwa ke se motho a se dirileng *O ne a tshwanetse go solofela sone se ka gore monna o bolawa ke se a se jeleng*

The attractive matter about this approach is that it saves space since it doesn't unnecessarily repeat words.

2. The second solution to dealing with MWEs is that in instances where variation is caused by the presence or absence of a word in a MWE, then brackets may be used around a word that may be left out. For instance the variants *kgomo e e maši ga e itsale* and *e e maši ga e itsale* may be treated as

(kgomo) e e maši ga e itsale •*seane*• motsadi ga se gantsi a tshola bana ba ba ditiro dintle jaaka ene = E E MAŠI GA E ITSALE

This approach differs from the Otlogetswe (2012) approach below that consumes space.

kgomo e e maši ga e itsale •*seane*• motsadi ga se gantsi a tshola bana ba ba ditiro dintle jaaka ene = E E MAŠI GA E ITSALE (Otlogetswe 2012: 197)

and

e e maši ga e itsale •*seane*• ga se gore motho yo o ditiro dintle o tlaa nna le bana ba ba ditiro dintle jaaka ene *Ngwana yo ga a na botho mo go tlhabisang ditlhong, mme mmaagwe ke motho yo o maitseo tota; e le ruri e e maši ga e itsale* (Otlogetswe 2012: 91).

3. The third solution applies to cases where the variants differ in too many places such that it would be much more elegant to treat them as separate entries. For instance:

— moswela *Tebele* ga a jewe o tshwana le moswela *gagabo*
— moswela *gae* ga a jewe o tshwana *fela* le moswela *nageng*

The two variants differ in *Tebele/gae* and *gagabo/nageng* and one of the variants has *fela* that is absent in the other. The argument here is that it is inelegant to combine solution one and two above to come up with a single entry as below:

moswela Tebele/gae ga a jewe o tshwana (fela) le moswela gagabo/nageng •*seane*• motho o gololesegile go nna gongwe le gongwe kwa a ka tshelang sentle teng mo lefatsheng, le fa e se gagabo ka motsing a swang ga a ka ke a itemogela pharologano epe ka a tlaabo a setse a sule

The above entry is confusing since it ends up generating variants that do not exist in Setswana. *Moswela Tebele/gae ga a jewe o tshwana (fela) le moswela gagabo/nageng* generates the following variants:

1. moswela Tebele ga a jewe o tshwana le moswela nageng
2. moswela Tebele ga a jewe o tshwana fela le moswela nageng
3. moswela Tebele ga a jewe o tshwana le moswela gagabo
4. moswela Tebele ga a jewe o tshwana fela le moswela gagabo
5. moswela gae ga a jewe o tshwana le moswela nageng
6. moswela gae ga a jewe o tshwana fela le moswela nageng
7. moswela gae ga a jewe o tshwana le moswela gagabo
8. moswela gae ga a jewe o tshwana fela le moswela gagabo

In this instance, the use of brackets and forward slashes complicates matters and results with unintended variants that do not exist in the language. This paper therefore proposes that in this instance both variants

should be entered as separate headwords and one of the variant forms be cross referenced to the primary form as demonstrated below:

moswêla gae ga jewe, o tshwana fêla le moswêla nageng •*seane*• motho o gololesegile go nna gongwe le gongwe kwa a ka tshelang sentle teng mo lefatsheng, le fa e se gagabo ka motsing a swang ga a ka ke a itemogela pharologano epe ka a tlaabo a setse a sule

and

moswêla Tebele ga a jewe o tshwana le moswêla gagabô •*seane*• = MOSWÊLA GAE GA JEWE, O TSHWANA FÊLA LE MOSWÊLA NAGENG

Conclusion

Variants in the Setswana language are important since many words and expressions are written in a variety of ways. This may be because of dialectal variation or as a consequence of pronunciation pattern amongst speakers. Some variation may be because of the way words are borrowed into the language. The variants occur even though Setswana orthographies have been developed over the years both in Botswana (Chebanne et al. 2008) and South Africa (PanSALB 2008) as well as by independent research centres such as Centre for Advanced Studies of African Society (Chebanne et al. 2003). This paper has shown that variants are treated differently in Setswana dictionaries – mostly in a non-consistent manner. It has argued that a well-designed Setswana corpus is an important source in exposing variants that could be lemmatised in a dictionary and it is useful in showing the variant to which a cross-reference should be made. This paper has also proposed a variety of ways of treating MWEs in a monolingual dictionary.

References

Bannard, C. 2007. A Measure of Syntactic Flexibility for Automatically Identifying Multi-word Expressions in Corpora. *Proceedings of the ACL Workshop on a Broader Perspective on Multi-word Expressions, Prague, Czech Republic, June 2007*: 1-8.

Batibo, H.M. 1999. A Lexicostatistical Survey of the Setswana Dialects Spoken in Botswana. *South African Journal of African Languages* 19(1): 2-11.

Chebanne, A. 2008.A Sociolinguistic Perspective of the Indigenous Communities of Botswana. *African Study Monographs.* 29(3): 93-118.

Chebanne, A., M. Mokitimi, L. Matlosa, R. Nakin, M.W. Nkolola, S. Mokgoatsana and M. Machobane. 2003. The Standard Unified Orthography for Sotho/Tswana Languages: Botswana, Lesotho, Namibia, South Africa, Zambia and Zimbabwe. Cape Town: CASAS.

Chebanne A., T.J. Otlogetswe, P. Seloma, A. Molosiwa, B. Moatlhodi, N. Kgolo and M. David. 2008. *Mokwalo o o lolameng wa Setswana.* Gaborone: Pentagon Publishers.

Cole, D.T. 1955. *An Introduction to Tswana Grammar*. Cape Town: Longman.

Crystal, D. and D. Davy. 1969. *Investigating English Style*. London, Longman.

Honey, J. 1997. *Language is Power: The Story of Standard English and its Enemies.* London: Faber and Faber.

Kgasa, M.L.A. and J. Tsonope. 1995. *Thanodi ya Setswana*. Gaborone: Longman Botswana.

Mareme, G.B. 2008. *Thanodi ya Setswana*. Pietermaritzburg: Nutrend Publishers.

Ministry of Education. 1981. *Setswana Standard Orthography of 1981*. Gaborone: Ministry of Education.

Mogapi, K. 1984. *Thutapuo ya Setswana.* Gaborone: Longman.

Moon, R. 1994. *Fixed Expressions and Text: A Study of the Distribution and Textual Behaviour of Fixed Expressions in English*. Unpublished Ph.D. Thesis, University of Birmingham.

Ngcongco, L.D. 1979. Origins of the Tswana. *Pula: Botswana Journal of African Studies* 1(2): 21-46.

Otlogetswe, T.J. 2008. *Corpus Design for Setswana Lexicography*. Unpublished Ph.D. Thesis. Pretoria: University of Pretoria.

Otlogetswe, T.J. 2009. Populating Sub-entries in Dictionaries with Multi-word Units from Concordance Lines. *Lexikos* 19: 446-457.

Otlogetswe, T.J. 2012. *Tlhalosi ya Medi ya Setswana.* Gaborone: Medi Publishing.

Otlogetswe, T.J. 2014. Dialectical Variations among Tswana Peoples of South Africa and Botswana: An Historical Linguistic Perspective. *Boleswa Journal of Theology, Religion and Philosophy* 4(2): 326-346.

PanSALB. 2008. *Melawana ya mokwalo le mopeleto Setswana.* Pretoria: PanSALB.

Parsons, Q.N. 1973. On the Origins of the bamaNgwato. *Botswana Notes and Records* 5: 82-103.

Rundell, M. (Ed.). 2007. *Macmillan English Dictionary for Advanced Learners*. Second edition. Oxford: Macmillan.

Schapera, I. 1963. Kinship and Politics in Tswana History. *Journal of the Royal Anthropological Institute* 93(2): 159-173.

Setshedi, J.E., J.H.K. Malao, O.S. Mathule, C.D. Ditsele, P.M. Rakgokong, W.M. Legotlo, G.R. Letsholo, J.W.P. Mashike, R.O. Komane, A.M.I. Rabaji, T.L. Tsambo and K.D. Rahlaga. 1988. *Setswana Terminology and Orthography No 4*. Pretoria: Department of Education and Training and Department of Education.

Southerland, R.H. and F. Katamba. 1996. Language in Social Contexts. O'Grady, W., M. Dobrovolsky and F. Katamba (Eds.). 1996. *Contemporary Linguistics: An Introduction*: 540-590. Third edition. London/New York: Longman.

Svensén, B. 2009. *A Handbook of Lexicography: The Theory and Practice of Dictionary-making.* Cambridge: Cambridge University Press.

Tlou, T. and A. Campbell. 1997. *History of Botswana.* Gaborone: Macmillan Botswana Publishing.

Wells, J.C. 2000. *Longman Pronunciation Dictionary*. Second Edition. Harlow: Pearson Education.

Corpus-based Lexicography for Lesser-resourced Languages — Maximizing the Limited Corpus

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

Abstract: This article focuses on lesser-resourced languages for which only very limited corpora are available and how such relatively small and often unbalanced, raw corpora could be maximally utilized for lexicographic purposes to obtain similar results as for bigger corpora. Sepedi and Afrikaans will be studied in this regard. The aim is to determine to what extent enlarging a corpus from e.g. one to 10 million, and from 10 million to 100 million words enhances its potential for (a) macrostructure compilation, (b) sourcing information on the most important microstructural aspects and (c) the creation of lexicographic tools. It will be argued that valuable and even sufficient data for the compilation of a specific dictionary can be extracted from a relatively small corpus of approximately one million words but that "bigger" in some instances indeed means "better".

Keywords: CORPUS-BASED LEXICOGRAPHY, LESSER-RESOURCED LANGUAGES, LIMITED CORPORA, CORPUS TOOLS, LEXICOGRAPHIC TOOLS

Opsomming: Korpusgebaseerde leksikografie vir hulpbronbeperkte tale — die maksimalisering van die beperkte korpus. Die fokus in hierdie artikel is op hulpbronbeperkte tale waarvoor slegs baie beperkte korpusse beskikbaar is en hoe sodanige relatief klein en dikwels ongebalanseerde, rou korpusse maksimaal benut kan word vir leksikografiese doeleindes om soortgelyke resultate as van groter korpusse te verkry. Sepedi en Afrikaans, word in hierdie verband bestudeer. Die doel is om te bepaal tot watter mate die vergroting van 'n korpus van byvoorbeeld een na 10 miljoen, en van 10 miljoen na 100 miljoen woorde die potensiaal sal verhoog vir (a) makrostruktuur samestelling, (b) die inwin van inligting omtrent die belangrikste mikrostrukturele aspekte en (c) die ontwerp van leksikografiese hulpmiddels. Daar sal aangevoer word dat waardevolle en selfs voldoende data vir die samestelling van 'n spesifieke woordeboek onttrek kan word uit 'n relatief klein korpus van ongeveer een miljoen woorde maar dat "groter" wel in sekere omstandighede "beter" is.

Sleutelwoorde: KORPUSGEBASEERDE LEKSIKOGRAFIE, HULPBRONBEPERKTE TALE, BEPERKTE KORPUSSE, KORPUSGEREEDSKAP, LEKSIKOGRAFIESE HULPMIDDELS

Introduction

The days of a default corpus size of one million words such as the groundbreaking first computer-readable general text corpus, the *Brown Corpus of Stan-*

dard American English being regarded as an acceptable norm, are long gone. Currently corpora for major languages typically run into hundreds of millions and even billions of words, for example *Google Books* with 155 billion for American English, 45 billion for Spanish and 34 billion for British English, and are typically referred to as "big corpora".

In many cases sincere attempts at corpus designs and the compilation of balanced and representative corpora reflecting stratified speaker groups have been made, e.g. in the compilation of the *Brown* corpus. Different levels of corpus annotation and sophisticated corpus manipulation tools e.g. *Sketch Engine*, *Dante*, *Interactive language Toolbox*, *WordSmith Tools* and *AntConc* became the norm as an international standard and represent the typical scenario for major languages of the world.

This article, however, focuses on lesser-resourced languages for which only very limited corpora are available and how such relatively small and often unbalanced, raw corpora could be maximally utilized for lexicographic purposes to obtain similar results in the absence of large corpora. It presents empirical research for Sepedi. English and Afrikaans corpora are used as measurement instruments to determine the power of limited corpora for lexicographic purposes.

"Big corpus" is a relative term. For lesser-resourced languages with a limited number of printed material such as many of the African languages, a corpus of 10 million words can be regarded as a "big corpus". The aim is to determine to what extent enlarging a corpus from e.g. one to 10 million, and from 10 million to 100 million words enhances its potential for (a) macrostructure compilation, (b) sourcing information on the most important microstructural aspects and (c) the creation of lexicographic tools. It will be argued that valuable and even sufficient data for the compilation of a specific dictionary can be extracted from a relatively small corpus of approximately one million words. The question is how much energy should be invested for lexicographic purposes in the maximum utilization of a limited corpus for macrostructural and microstructural compilation versus increasing the corpus size. Macrostructural compilation mainly concerns the compilation of the lemmalist and microstructural aspects include sense distinction, collocations, idioms and examples of usage.

English, Afrikaans and Sepedi corpora

For the purpose of this study corpora for English, Afrikaans and Sepedi were used. For English the *Pretoria English Internet Corpus* (PEIC) consisting of 12 million words and a subsection of approximately one million words were used. These corpora will be referred to as the 10m PEIC and 1m PEIC respectively. For Afrikaans a small section of the *Media 24* archive for the newspaper *Beeld* consisting of 119 million words as well as two subsections consisting of approximately 10 million and one million words respectively were used and will be referred to as 100m MED 24, 10m MED 24 and 1m MED 24 respectively.

For Sepedi a 10 million-word corpus and a one million subsection thereof were used and will be referred to as 10m PSC and 1m PSC respectively. The corpora and subsections of the corpora are schematically indicated and their exact sizes are given in figure 1:

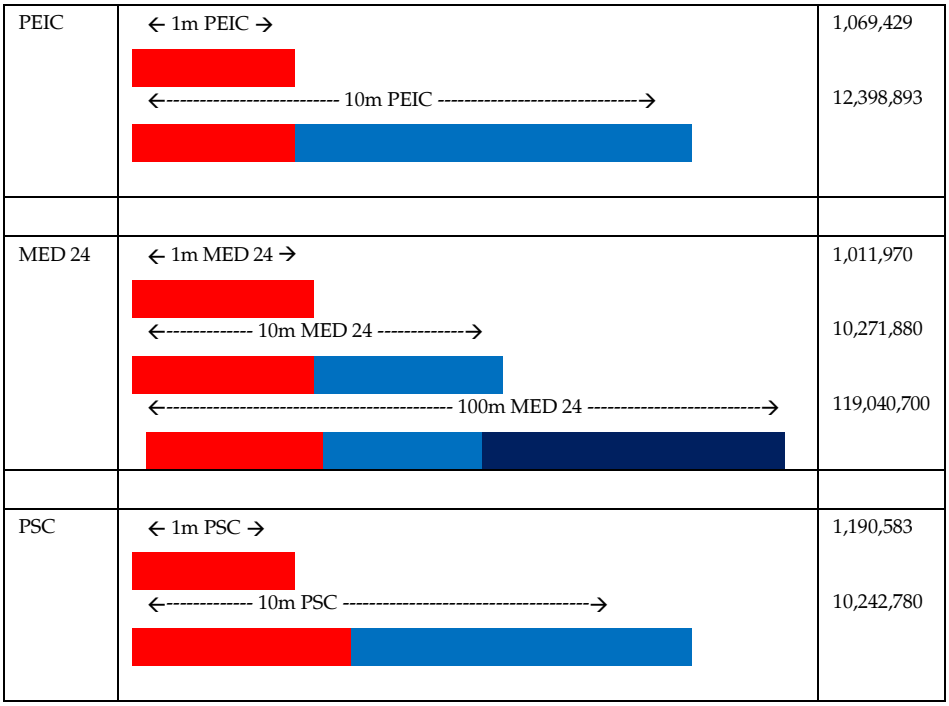


Figure 1: Corpora and sub-corpora used for English, Afrikaans and Sepedi

Macrostructure

In Africa publishers normally restrict dictionaries to a very limited number of pages. 5000 articles are often the norm and by necessity put the focus on commonly used words for inclusion in the dictionary. This study thus assumes that the basic/common words of a language are most likely to be looked for especially by learners of the language in such a small dictionary. These are the frequently used words typically marked by means of e.g. a star-rated system, filled diamonds, and/or by a different colour in dictionaries such as the *Macmillan English Dictionary* (MED), and *Collins COBUILD English Dictionary* (COBUILD), e.g. **car** … *** (MED) and **car** … **cars** ♦♦♦♦♦ (COBUILD). MED states that a word marked with three stars is one of the most basic words in English. COBUILD, as indicated in table 1, states that the 1,900 most frequently used words in the language, marked with four or five filled diamonds represent 75% of all written

and spoken words in English and that the top 14,700 words account for 95% of English words.

Number of filled diamonds	Lemmas per category	Totals	% of all written and spoken English
5	700		
4	1200		
(Total 5 + 4)		1900	75
3	1500		
2	3200		
1	8100		
(Total 3 + 2 + 1)		12800	20
(Total 5 + 4 + 3 + 2 + 1)		14700	95

Table 1: Summary of frequency band values in COBUILD (p. xiii)

On the macrostructural level an evaluation was made of frequency lists compiled from the 1m PEIC and 10m PEIC for English, the 1m MED 24, the 10m MED 24 and the 100m MED 24 for Afrikaans, and the 1m PSC and 10m PSC for Sepedi. The most basic words in English indicated with three stars (***) in MED were used as a benchmark against the 1m PEIC and 10m PEIC English corpora. There are 2,275 three-starred words in MED. Of these words 2,203 occur in the 31,982-word frequency list culled from the 1m PEIC; thus an overlap of 96.8%. Since it is hardly feasible for a lexicographer to work through a frequency list of this size when compiling a lemmalist, a more realistic number of words were considered, i.e. 11,559 which occurred five times or more in the corpus. 2,061 three-starred words in MED remained, i.e. an overlap of 90.6%. This means that the lexicographer who only had a one million English corpus at his/her disposal, and willing to read through a list of 11,000 words would be in a position to capture 90.6% of the most basic English words. A 90%+ figure can surely be regarded as quite a significant achievement on such a small corpus.

This experiment was repeated for the entire 10m PEIC. Of the 2,275 three-starred words in MED, 2,272 (only three not: e-mail, long-term and no-one), and with the exception of *metre* with a frequency of 1, appear in the 10m PEIC. All of these 3-starred words have a frequency count higher than 10 and occur in the 118,202-word frequency list of the 10m PEIC; thus an overlap of 99.9%. Once again, a more realistic number of words were considered, i.e. 11,161, which occurred 65 times or more in the corpus. 2,191 three-starred words in MED remained. This means that the lexicographer who only had a 10 million English corpus at his/her disposal, and willing to read through a list of 11,000

words would be in a position to capture 96.3% of the most basic English words. Once again, a relatively small corpus of 10 million words enabled the lexicographer to capture the most basic words. It is also significant that a tenfold increase in the corpus size from one million to 10 million only resulted in a 5.7% increase in the three-starred words retained.

Consider table 2 as summary:

MED	**1m PEIC**	**10m PEIC**
2,275 (three-starred words)	2,203 MED *** in 1mPIC (overlap with MED ***): 2,061 = 90.6% (Lexicographer considers freq. >4) (11,559 words to consider)	2,272 MED *** in PEIC (overlap with MED ***): 2,191 = 96.3% (Lexicographer considers freq. >64) (11,161 words to consider)

Table 2: MED 3-starred words versus the 1m PEIC and the 10m PEIC

For the Afrikaans experiment the aim was to see to what extent increasing a one-million word corpus to 10 million and again to a 100-million word corpus would enhance the quality of the lemmalist in terms of the most basic words of Afrikaans.

In the absence of a benchmark for basic words such as the three-starred words for English, an alternative approach and criterion for comparison had to be found. This was done through comparison of top frequencies in the 1m MED 24 with those in the 10m MED 24 with 100m MED 24 in order to determine internal stability in terms of top frequencies, or formulated differently, to what extent the top frequencies differ when a corpus is enlarged from one to 10 to 100 million words. The ideal situation would be if the top frequencies were identical as schematically illustrated by the single centre dot in figure 2a. Figure 2b represents a situation where there is great overlap in terms of this top frequency core and figure 2c a possible situation where the top frequencies do not overlap.

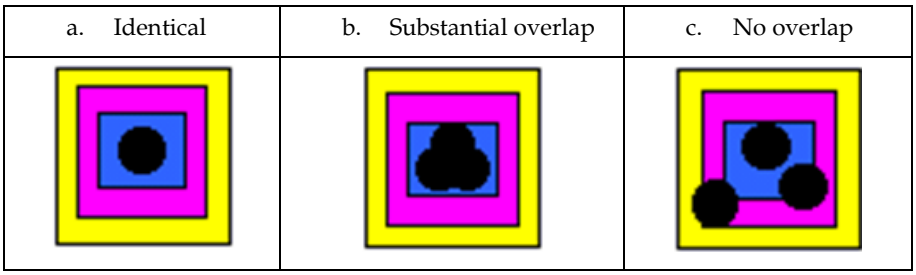


Figure 2: Possible scenarios of overlap in top frequencies

Consider table 3 where the top-ranking 100 words in terms of frequency in the 100n MED 24 are compared to the 1m MED 24 in columns 2 and 3. Columns 4 and 5 indicate the difference in ranks and the extent of the deviation respectively.

	100m	1m	Comp.	Diff.
DIE	1	2	-1	1
#	2	1	1	1
VAN	3	3	0	0
HET	4	4	0	0
IN	5	5	0	0
EN	6	7	-1	1
N	7	6	1	1
IS	8	8	0	0
NIE	9	9	0	0
WAT	10	10	0	0
TE	11	11	0	0
OP	12	14	-2	2
OM	13	12	1	1
MET	14	15	-1	1
SE	15	13	2	2
WORD	16	17	-1	1
SY	17	18	-1	1
VIR	18	16	2	2
HY	19	21	-2	2
DIT	20	19	1	1
DAT	21	20	1	1
SAL	22	24	-2	2
AS	23	23	0	0
AAN	24	22	2	2
WAS	25	25	0	0
MNR	26	50	-24	24
KAN	27	26	1	1
BY	28	27	1	1
DEUR	29	31	-2	2
DIÉ	30	30	0	0
OOR	31	28	3	3
OOK	32	33	-1	1
GESÊ	33	34	-1	1
HULLE	34	29	5	5
TOT	35	32	3	3
SUID	36	38	-2	2
NA	37	36	1	1
MAAR	38	37	1	1
OF	39	35	4	4
DAAR	40	41	-1	1
ONS	41	39	2	2
MOET	42	43	-1	1
JAAR	43	40	3	3
EK	44	44	0	0
HUL	45	42	3	3
TEEN	46	47	-1	1
AFRIKA	47	46	1	1
GAAN	48	45	3	3
UIT	49	52	-3	3
MEER	50	48	2	2
SÊ	51	49	2	2
NET	52	51	1	1
TWEE	53	53	0	0
NOG	54	54	0	0
TOE	55	56	-1	1
WEES	56	61	-5	5
GISTER	57	57	0	0
EEN	58	55	3	3
HAAR	59	58	1	1
MENSE	60	63	-3	3
HOM	61	59	2	2
ANDER	62	65	-3	3
BAIE	63	64	-1	1
NUWE	64	67	-3	3
SOOS	65	60	5	5
EERSTE	66	68	-2	2
AL	67	62	5	5
NOU	68	72	-4	4
ONDER	69	69	0	0
GROOT	70	70	0	0
VOLGENS	71	71	0	0
NÁ	72	75	-3	3
VERLEDE	73	66	7	7
BEGIN	74	73	1	1
DRIE	75	88	-13	13
MY	76	78	-2	2
WIL	77	82	-5	5
MAAK	78	79	-1	1
SO	79	80	-1	1
EGTER	80	92	-12	12
WEER	81	85	-4	4
SOWAT	82	81	1	1
AFRIKAANSE	83	90	-7	7
VOOR	84	94	-10	10
NADAT	85	74	11	11
REEDS	86	84	2	2
TUSSEN	87	83	4	4
OMDAT	88	86	2	2
LAAT	89	89	0	0
WAAR	90	76	14	14
MILJOEN	91	87	4	4
DE	92	105	-13	13
THE	93	77	16	16
GEEN	94	102	-8	8
PRETORIA	95	106	-11	11
KRY	96	96	0	0
KOM	97	97	0	0
VANJAAR	98	120	-22	22
LAND	99	100	-1	1
DOEN	100	91	-1	1
				Average 3.1 positions different

Table 3: Top 100 ranks in 100m MED 24 versus 1m MED 24

From this table the stability in terms of the top 100 frequencies in the one million corpus versus the 100 million corpus is illustrated. Only 4 items, e.g. 92. *de*, 94. *geen*, 95. *Pretoria* and 98. *vanjaar* in the top 100 ranks of the 100 million corpus do not appear in the top 100 ranks of the one million corpus. Furthermore the actual difference in the rank numbers is very small. So, for example, are the rank numbers for rank 3, i.e. *van*, 4 *het*, 5 *in*, 8 *is*, 9 *nie* and 10 *wat* identical in both corpora. For the top 100 ranks the average variation in rank positions is only 3.1%. For the compilation of a dictionary with approximately 5,000 lemmas in mind, a random cut-off point of the top ranks at approximately 7,700 ranks were made in all three corpora. The aim is to determine which words likely to be looked for by the target user will be missed if only a one million corpus was available instead of a 10 million corpus and only a one million corpus versus a 100 million corpus. 7,737 words occur in the one million Afrikaans corpus with a frequency of 11 and more. Compared with the closest match in terms of frequency, 7,734 words occur in the 10 million corpus with a fre-

quency of 100 and more and 7,733 in the 100 million corpus with a frequency of 1081 and more. The overlap between these selected sections of the 1m MED 24 corpus' frequency list and the 10m MED 24 corpus is 6,449, i.e. 83.4%. The overlap between these selected sections of the 1m MED 24 and the 100m MED 24 is 5,991, i.e. 77.5%. This means that 1,742 words, i.e. 22.5% of the selected top section of the 100 million corpus would not have been available for consideration if the lexicographer only had the one million corpus available and 1,285 words or 16,6% if a 10 million corpus was available.

1m MED 24	**10m MED 24**	**100m MED 24**
Top 7,737 ranks considered Frequency of 11 and more	Top 7,734 ranks considered Frequency of 100 and more	Top 7,733 ranks considered Frequency of 1081 and more
Overlap 1m MED 24 versus 10m MED 24: 6,449 = 83,4%		
Overlap 1m MED 24 versus 100m MED 24Million: 5,991 = 77,5%		

Table 4: Comparison of top frequencies in the 1m MED 24, 10m MED 24 and 100m MED 24

The question is how significant this presumed 22.5% "loss" is for the compilation of the lemmalist. Among the words occurring with a high frequency are *Kersfees* 'Christmas', *koningin* 'queen', *toesig* 'supervision', *eksamen* 'exam', *koor* 'choir', *volk* 'nation', *aardbewing* 'earthquake', *skandaal* 'scandal', *digter* 'poet', *opskrif* 'heading', *strook* 'strip', *tjek* 'cheque' and *gogga* 'bug'. The Afrikaans lexicographer would probably regard these words as likely to be looked for and that they deserve a place in the dictionary.

For Sepedi the same procedure was followed in order to determine to what extent increasing a one-million word Sepedi corpus to a 10-million word corpus would enhance the quality of the lemmalist, i.e. to see which words likely to be looked for by the target user will be missed if only the 1m PSC was available instead of the 10m PSC. Consequently, the top 7,646 ranks occurring 8 times or more in the 1m PSC were compared to the top 7622 ranks occurring 62 times or more in the 10m PSC. The overlap was 5,553 words, i.e. 72.8%. This means that 2,069 high frequency words in 10m PSC were missed by the 1m PSC.

1m PSC	**10m PSC**
Top 7,646	Top 7,622
With frequency 8 times or more	With frequency 62 times or more
Overlap 5,553 words = 72.8%	

Table 5: Comparison of the top frequencies in 1m PSC and 10m PSC

As for Afrikaans, words occurring with high frequency in 10m PSC but not in the top 7,646 of 1m PSC were considered. These words include *bjalobjalo* 'et cetera', *diteng* 'contents', *seyalemoya* 'radio', *metara* 'metre', *semolao* 'legal', *kamano* 'relationship', *Bathobaso* 'Black people' and *komiti* 'committee'. Once again it is likely that the Sepedi lexicographer would regard them as common words likely to be looked for and that they should be included in the dictionary.

Microstructure

On the microstructural level the evaluation focused on the value of information drawn from limited corpora in terms of meaning, sense distinction, examples of usage, collocations and proverbs/idioms.

Consider as a first example the randomly selected adjective *great* in *Sketch Engine* in figure 3.

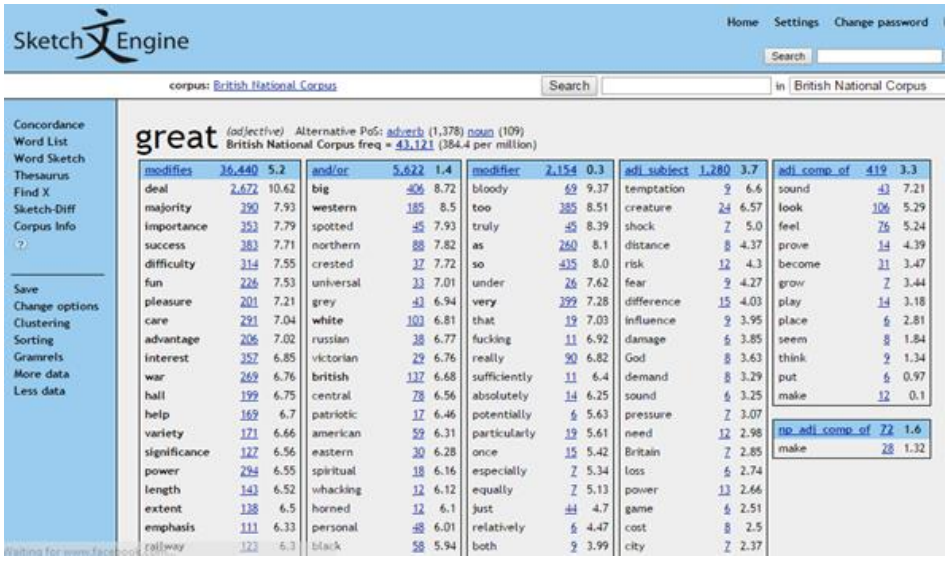


Figure 3: Collocations: *great* in Sketch Engine

The top 20 combinations of *great* + a noun in column 1 were then compared to the collocations for *great* given in MED, the 1m PEIC and the 10m PEIC as given in table 6. There were in total 1,709 occurrences of *great* in the 1m PEIC and 15,887 in the 10m PEIC.

	GREAT …	Sketch Engine	MED	1mPEIC (1,709)	10m PEIC (15,887)
1	great deal	✓	✓	22	58
2	great majority	✓	✓	12	51
3	great importance	✓		13	72
4	great success	✓		5	25
5	great difficulty	✓	✓	8	70
6	great fun	✓		0	12
7	great pleasure	✓	✓	9	51
8	great care	✓		1	52
9	great advantage	✓	✓	10	53
10	great interest	✓		2	33
11	great war	✓		3	35
12	great hall	✓	✓	4	56
13	great help	✓		0	8
14	great variety	✓		3	33
15	great significance	✓		1	2
16	great power	✓		3	67
17	great length	✓		3	20
18	great extent	✓		10	36
19	great emphasis	✓		0	3
20	great railway	✓		0	0

Table 6: Sketch Engine's *great* as modifier vs. MED, 1m PEIC and 10m PEIC

From table 6 column 4 it is clear that MED accounts for six of the 20 collocations, i.e. 30%. The 1m PEIC has examples of 16 (80%) and the 10m PEIC of 19 (95%). 80% for the 1m PEIC is significant for such a small corpus but a corpus should provide more evidence to the English lexicographer for common combinations such as *great fun*, *great care*, *great help* and *great significance*, etc. which are under-represented or missing in the 1m PEIC.

As a second example the senses of the verb *count* were studied in the 1m PEIC and the 10m PEIC. The senses distinguished in MED given in table 7 were used as a benchmark. As in the case of the frequency lists, it is not feasible for the lexicographer to read through thousands of concordance lines generated for a specific keyword in context – 100-300 lines could be regarded as a reasonable number to consider for detecting senses and to find typical collocations and authentic examples of use. The first deficiency encountered in the 1m PEIC was

an insufficient number of concordance lines. For *count* only 66 concordance lines were found in the 1m PEIC in contrast to 813 in the 10m PEIC. In the 10m PEIC a sufficient number of concordance lines were found for at least four out of five of the senses listed in table 7 but no or insufficient information for all senses, with the possible exception of the first sense *to calculate* in the 1m PEIC.

	Sense description	**1m PEIC**	**10m PEIC**
1	To calculate how many people or things there are in a group e.g. *all the votes have been counted*	3	27
2	Say numbers one after another in order e.g. *I can count up to ten in German*	1	5
3	To include someone or something in a calculation e.g. *sick pay is counted as income*		7
4	To think of someone or something as a particular thing e.g. *that counts as a lie*	1	11
5	To be important, or to have influence e.g. *what really counts is …*		1

Table 7: Verbal senses of *count* in MED compared to their occurrence in 1m PEIC and 10m PEIC

As a third example, consider three randomly selected Sepedi idioms in table 8: *monna ke nku (o llela) teng* 'a man is a sheep (he cries inside)', *bana ba tau (ga re jane)* 'children of a lion (we do not eat each other)' and *go sepela ke go bona* 'to travel is to see (become experienced)'.

Idiom	**1m PSC**	**10m PSC**
Monna ke nku …	11	127
Bana ba tau …	9	25
Go sepela ke go bona …	4	35

Table 8: Occurrence of idioms in 1m PSC versus 10m PSC

From table 8 it is clear that although in a limited number, these idioms do occur in a one million corpus but the lexicographer is more likely to detect them in a bigger corpus such as the 10m PEIC.

As for finding authentic examples of use, a one-million corpus proved to be quite significant for commonly used words of the language and as such could go a long way in supplementing the lexicographer's intuition when compiling a relatively small dictionary. Consider, for example, the potential for good examples even for the limited number of collocations *great success*, *great*

care and *great interest* in table 6 that can be found in the concordance lines from the 1m PEIC given in table 9.

troops that day was about twelve miles. This I regarded as a	**great**	success, and it removed from my mind the most serious
of his making his escape, that the Southern troops had had	**great**	success all day. Johnston forwarded the dispatch to Ri
opportunities should present themselves which would insure	**great**	success. General Meade was left in command of the few
destroy the railroad between Petersburg and Richmond, but no	**great**	success attended these latter efforts. He made no grea
entry into politics, a career he followed ever after with	**great**	success, and in which he died enjoying the friendship,
uniform and in prescribed order. Orders were prepared with	**great**	care and evidently with the view that they should be a
back to his grandfather. On the other side, my father took a	**great**	interest in the subject, and in his researches, he fou
change his position. While at Cairo I had watched with very	**great**	interest the operations of the Army of the Potomac, lo

Table 9: Concordance lines for *great success*, *great care* and *great interest* in 1m PEIC

Lexicographic tools

As for the creation of lexicographic tools, the aim was to determine whether a relatively small corpus of one million words can be utilized to create useful tools such as rulers, block systems, indicators of spreading-across-sources, etc. So, for example, the aim was to see whether, in the absence of larger corpora, a one-million word corpus would be sufficient to build a sensible guide for the lexicographer for balancing alphabetical stretches in the dictionary or whether larger corpora would contribute substantially to the refinement of such tools. Prinsloo and De Schryver (2002) introduced the concept of a measurement instrument for the relative length of alphabetical stretches in dictionaries and referred to it as a *lexicographic ruler*. Such a ruler guides the compiler of a dictionary to appropriately balanced alphabetical stretches in terms of overall length and the number of lemmas treated, i.e. not to over/under treat a specific alphabetic stretch in relation to the other alphabetic stretches. They indicate how, for example, a compiler could enthusiastically treat the first few alphabetic categories exhaustively but 'gets tired' towards the end of the alphabet. Formulated differently, a lexicographic ruler tells the compiler when alphabetic stretch 'A' has been sufficiently treated, i.e. when it is time to move on to 'B'. So, for example, Prinsloo and De Schryver (2003: 110) give a schematic illustration of a ruler for Afrikaans in figure 4.

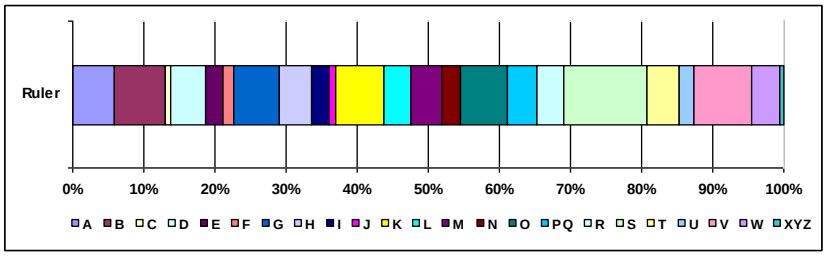


Figure 4: A lexicographic ruler for Afrikaans

This ruler indicates at a glance that e.g. B, K, O, S and V are relatively big stretches in Afrikaans whilst C, F, J, X, Y and Z are small. Figure 4 also gives a basic indication in terms of percentage of progress through the alphabetic stretches moving from A to Z. For example that M roughly represents the middle of the dictionary and that concluding S means reaching the 80% stage of compilation. They performed a formal breakdown of the ruler into percentages to guide dictionary compilation referred to as a block system. Consider, for example, the block system for Setswana in figure 5.

1	**ALAF**	21	**FELE**	41	**KOUS**	61	**MOTL**	81	**SELE**
2	**AROG**	22	**FOLO**	42	**LAEL**	62	**MPHE**	82	**SERA**
3	**BADI**	23	**GAGW**	43	**LEBO**	63	**NATE**	83	**SETO**
4	**BANN**	24	**GATS**	44	**LEKI**	64	**NGWA**	84	**SIMO**
5	**BATW**	25	**GOLO**	45	**LERI**	65	**NKUK**	85	**SUAS**
6	**BIRO**	26	**GWET**	46	**LETS**	66	**NTEM**	86	**TALE**
7	**BOGA**	27	**HUBE**	47	**LOKO**	67	**NTSH**	87	**THAA**
8	**BOLA**	28	**IJES**	48	**MAAD**	68	**NYOR**	88	**THIB**
9	**BONK**	29	**IKGO**	49	**MAHA**	69	**OOMA**	89	**THWE**
10	**BORU**	30	**INOL**	50	**MALE**	70	**PANT**	90	**TLAM**
11	**BOUT**	31	**IPUS**	51	**MARA**	71	**PHAK**	91	**TLHA**
12	**DAAM**	32	**ITIS**	52	**MATL**	72	**PHIM**	92	**TLHO**
13	**DIFA**	33	**ITSH**	53	**MEFA**	73	**PITL**	93	**TLWA**
14	**DIKG**	34	**JOKO**	54	**MESU**	74	**PUDU**	94	**TSAP**
15	**DINK**	35	**KANY**	55	**MMAL**	75	**RAMO**	95	**TSHE**
16	**DIRA**	36	**KERO**	56	**MMOL**	76	**RENG**	96	**TSHW**
17	**DITH**	37	**KGAR**	57	**MOFI**	77	**ROKG**	97	**TSUN**
18	**DITU**	38	**KGOM**	58	**MOKG**	78	**RURU**	98	**UBAU**
19	**EGEP**	39	**KHAN**	59	**MONG**	79	**SEBA**	99	**WABO**
20	**ETLH**	40	**KODU**	60	**MORW**	80	**SEHI**	100	**ZIMB**

Figure 5: A block system for Setswana

A useful practical application of a block system is to pace dictionary compilation in terms of time and resources. It suggests that the compiler should be at IN when 30% of time and resources have been spent, that MA roughly repre-

sents 50% of completion but that 15% of time and resources should be spent on M, and that SE is the 80% mark.

Rulers are calculated by determining the percentage of words in each alphabetic category from an alphabetic list of words culled from a corpus. This simply means how many words start with a, b, c, … z. The same data is used for calculating a block system but instead of the 26 letters of the alphabet, the list is broken down into 100 sections to each represent 1%.

The question here is whether a ruler compiled from a one-million word corpus could provide a reliable ruler when compared to a 10 million corpus. In table 10 the breakdown of words into alphabetical stretches of both the 1m PSC and the 10m PSC is given. Columns 3 and 5 reflect the percentage breakdown per alphabetical stretch in the 1m PSC versus the 10m PSC and the difference between these percentages is given in column 6.

	1m PSC	% 1m PSC	10m PSC	% 10m PSC	Difference
A	1164	2.13	6521	2.55	-0.41
B	5045	9.25	23123	9.02	0.22
C	98	0.18	1853	0.72	-0.54
D	3486	6.39	17241	6.73	-0.34
E	753	1.38	4271	1.67	-0.29
F	1475	2.70	5703	2.23	0.48
G	1945	3.57	8697	3.39	0.17
H	2275	4.17	9147	3.57	0.60
I	2475	4.54	10668	4.16	0.37
J	206	0.38	1311	0.51	-0.13
K	3519	6.45	16433	6.41	0.04
L	3657	6.70	15466	6.04	0.67
M	9005	16.51	40687	15.88	0.63
N	3357	6.15	14010	5.47	0.69
O	715	1.31	4032	1.57	-0.26
P	2484	4.55	12123	4.73	-0.18
Q	0	0.00	386	0.15	-0.15
R	1581	2.90	9663	3.77	-0.87
S	4629	8.49	22433	8.76	-0.27
T	5872	10.77	26155	10.21	0.56
U	270	0.50	1521	0.59	-0.10
V	68	0.12	1601	0.62	-0.50
W	247	0.45	1742	0.68	-0.23
X	45	0.08	324	0.13	-0.04
Y	154	0.28	901	0.35	-0.07
Z	20	0.04	215	0.08	-0.05

Table 10: Alphabetical stretches in 1m PSC compared to 10m PSC

The final column indicates that the difference between the rulers is very small with the difference in all stretches less than 1%. The similarity is visually illustrated in figure 6 where the two lines of the graph are very close to each other.

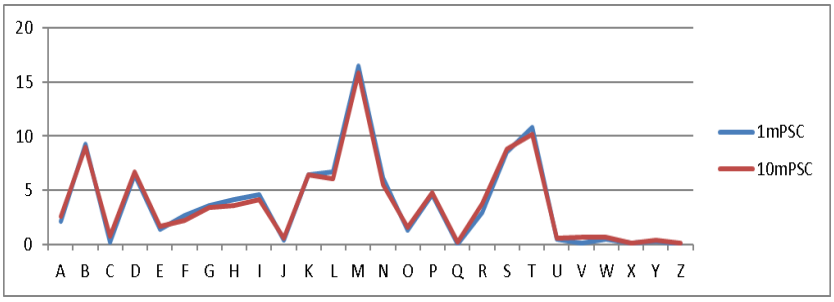


Figure 6: A ruler graph for 1m PSC versus 10m PSC

The same similarity is observed in the breakdown in the block systems calculated from the 1m PSC versus the 10m PSC in table 11.

%	1m PSC	10m PSC
1	ALO	AKO
2	ATH	ARE
3	BAH	BAF
4	BAR	BAP
5	BEL	BEF
6	BLO	BJE
7	**BOH**	**BOH**
8	BOL	BOL
9	BON	BOM
10	BOS	BOR
11	BUA	BOT
12	DIB	COM
13	DIK	DIB
14	DIM	DIK
15	DIP	DIN
16	DIT	DIP
17	**DIT**	**DIT**
18	EDI	DIU
19	ERI	DUT
20	FEE	ENK
21	FIH	FAR
22	FUL	FIH
23	GAM	GAB
24	GIL	GAM
25	GON	GOB
26	HLA	GRA
27	HLA	HLA
28	HLO	HLA
29	HOM	HOL
30	**IDI**	**IDI**
31	IKI	IKG
32	IPH	IPA
33	ITH	ITH
34	ITS	IWE
35	KAM	KAN
36	KGA	KGA
37	KGE	KGO
38	KGO	KGO
39	KHU	KIL
40	KON	KOT
41	KWE	LAB
42	LEA	LEB
43	LEF	LEH
44	LEK	LEN
45	LEP	LET
46	LET	LLA
47	LOG	MAA
48	MAB	MAF
49	MAG	MAJ
50	MAI	MAM
51	MAL	MAR
52	MAR	MAT
53	MAS	MED
54	MAZ	MEP
55	MEL	MMA
56	MET	MMO
57	MME	MOG
58	MOD	MOK
59	MOH	MOM
60	MOL	MOR
61	MON	MOT
62	MOS	MPH
63	MOT	NAG
64	MPO	NGW
65	NEE	NKG
66	NIK	NTA
67	NKU	NTS
68	NTE	NYA
69	NTS	OLO
70	NYS	PAF
71	OKS	PET
72	PAL	PHE
73	PHA	PHU
74	PHE	POT
75	PHU	RAG
76	PŠH	REI
77	RAP	ROB
78	RIP	ROT
79	RUR	ŠAR
80	SEB	SEE
81	SEG	SEJ
82	SEK	SEL
83	SEN	SER
84	SER	SET
85	SET	SIS
86	SOB	SOU
87	ŠUT	SWA
88	TAL	TAU
89	**THA**	**THA**
90	THE	THI
91	THU	TIA
92	TIT	TLE
93	TLH	TLW
94	TOM	TSE
95	TSE	TŠH
96	TŠH	TSI
97	TŠI	TSW
98	TŠW	UTI
99	UTS	WEB
100	**ZUL**	**ZUL**

Table 11: Sepedi block systems: 1m PSC versus 10m PSC

So, for example, both block systems indicate that the compiler should be at the sub-stretch ID after 30% of the available time and resources for the project, at MA after 50%, SE after 80%, etc. All of the other comparative blocks are alphabetically very close to each other.

Conclusion

In this article it has been argued that raw corpora built only from written data, although not reflecting an ideal situation, can substantially assist the lexicographer in the compilation of especially small bilingual and monolingual dictionaries.

On the macrostructural level a corpus of one million words is useful to pinpoint the most commonly used words in the language and would be a useful tool for the lexicographer tasked with the compilation of a relatively small dictionary of approximately 5,000 lemmas. Additional common words will however have to be found. Consider in this regard high-ranking words in the 100m MED 24 mentioned which were not found in the 1m MED 24. The lexicographer will have to find such words through other means, e.g. introspection, field work and reading and marking. If a one million corpus is extended to 10 million words the offering of commonly used words in the top frequency ranks becomes more reliable and represents a gradual enhancement. If the corpus is further extended to a 100 million words, the frequently used words provide a reliable account of the commonly used words in the language and little additional collection is required from the lexicographer for a small dictionary.

As far as microstructural elements are concerned, it is clear that a one million corpus is useful in determining the basic senses of a word as well as typical examples of usage of these basic senses. Such a corpus would typically include a limited number of idioms. Increasing the corpus to 10 million words gradually improves the situation in the sense that more senses are detected, more idioms can be found and more evidence on the use and meaning of such words and idioms is available.

As for lexicographic tools, the results clearly indicate that reliable lexicographic rulers and block systems could be compiled from a corpus as small as one million words. In this case enlarging the corpus to 10 million did not substantially enhance the quality/accuracy of the tool.

In conclusion it could be recommended that the lexicographer should carefully analyse the situation for each specific language. If no written sources are available (s)he should attempt to compile, say, a one-million token corpus of the spoken language. If a limited number of written sources are available, (s)he should try to compile a 10 million corpus and if sources are available in abundance, especially in electronic format, a 100 million corpus will be extremely valuable.

Acknowledgement

This research is (a) conducted within the SeLA project (Scientific e-Lexicography for Africa), supported by a grant from the German Ministry for Education and Research, administered by the DAAD and (b) supported in part by the National Research Foundation of South Africa (Grant specific unique reference number (UID) 85763). The Grantholder acknowledges that opinions, findings and conclusions or recommendations expressed in any publication generated by the NRF supported research are those of the author, and that the NRF accepts no liability whatsoever in this regard.

References

AntConc: http://www.laurenceanthony.net/software/antconc/ (Consulted 25 June 2015).

Brown Corpus of Standard American English: http://www.essex.ac.uk/linguistics/clmt/w3c/corpus_ling/content/corpora/list/private/brown/brown.html (Consulted 25 June 2015).

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

Dante: http://www.webdante.com/ (Consulted 25 June 2015).

Google Books: http://googlebooks.byu.edu/ (Consulted 25 June 2015).

Interactive Language Toolbox: https://ilt.kuleuven.be/inlato/ (Consulted 25 June 2015).

MED: Rundell, M. 2007. *Macmillan English Dictionary for Advanced Learners*. Second Edition 2007. Oxford: Macmillan.

MEDIA 24: Subsection of the archive for the newspaper *Beeld* http://argief.beeld.com/cgi-bin/beeld.cgi (Extract made available by Pharos/Media 24).

PSC: Pretoria Sepedi Corpus compiled at the University of Pretoria.

PEIC: Gauton, Rachélle: The University of Pretoria English Internet Corpus.

Prinsloo, D.J. and G.-M. de Schryver. 2002. Designing a Measurement Instrument for the Relative Length of Alphabetical Stretches in Dictionaries, with Special Reference to Afrikaans and English. Braasch, A. and A. and C. Povlsen (Eds.). 2002. *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002, Copenhagen, Denmark, August 13–17, 2002*: 483-494. Copenhagen: Center for Sprogteknologi, University of Copenhagen.

Prinsloo, D.J. and G.-M. de Schryver. 2003. Effektiewe vordering met die *Woordeboek van die Afrikaanse Taal* soos gemeet in terme van 'n multidimensionele Liniaal [Effective Progress with the *Woordeboek van die Afrikaanse Taal* as Measured in Terms of a Multidimensional Ruler]. Botha, W. (Ed.). 2003. *'n Man wat beur. Huldigingsbundel vir Dirk van Schalkwyk*: 106-126. Stellenbosch: Buro van die WAT.

Sketch Engine: http://www.sketchengine.co.uk/ (Consulted 10 January 2015).

WordSmith Tools: http://www.lexically.net/wordsmith/index.html (Consulted 10 January 2015).

http://lexikos.journals.ac.za

From Print to Digital: Implications for Dictionary Policy and Lexicographic Conventions

Michael Rundell, *Lexicography MasterClass (www.lexmasterclass.com)*
and *Macmillan Dictionary, London, United Kingdom*
(michael.rundell@lexmasterclass.com)

Abstract: Editorial policies and lexicographic conventions have evolved over hundreds of years. They developed at a time when dictionaries were printed books of finite dimensions — as they have been for almost the whole of their history. In many cases, styles which we take for granted as "natural" features of dictionaries are in reality expedients designed to compress maximum information into the limited space available. A simple example is the kind of "recursive" definition found in many English dictionaries where a nominalization (such as *assimilation*) is defined in terms of the related verb ("the act of assimilating or state of being assimilated"), and the user is required to make a second look-up (to the base word). Is this an ideal solution, or was it favoured simply as a less space-intensive alternative to a self-sufficient explanation?

As dictionaries gradually migrate from print to digital media, space constraints disappear. Some problems simply evaporate. To give a trivial example, the need for abbreviations, tildes and the like no longer exists (though a surprising number of dictionaries maintain these conventions even in their digital versions). So the question arises whether we need to revisit, and re-evaluate, the entire range of editorial policies and conventions in the light of changed circumstances. This paper looks at some familiar editorial and presentational conventions, and considers which are no longer appropriate in the digital medium — and what new policies might replace them.

Keywords: DEFINITIONS, EXAMPLE SENTENCES, DIGITAL MEDIA, EXCLUSION CRITERIA, GATEKEEPER, LEXICOGRAPHIC CONVENTIONS, ONLINE DICTIONARY, USER PROFILE

Opsomming: Van druk na digitaal: Implikasies vir woordeboekbeleid en leksikografiese norme. Redigeringsbeleide en leksikografiese norme ontwikkel al oor honderde jare. Dit het ontstaan in die tyd toe 'n woordeboek 'n gedrukte band was met vasgestelde dimensies — soos dit was vir die grootste deel van die geskiedenis van die woordeboek. In baie gevalle is die stylelemente wat as "natuurlike" eienskappe van woordeboeke beskou word, in der waarheid hulpmiddels wat ontwerp is om die maksimum hoeveelheid inligting in 'n beperkte beskikbare ruimte saam te pers. 'n Eenvoudige voorbeeld is 'n rekursiewe definisie wat in 'n aantal Engelse woordeboeke verskyn, waarby 'n nominalisering (soos bv. *assimilasie*) in terme van die verwante werkwoord gedefineer word ("die daad om te assimileer of die toestand van geassimileer wees"), en die gebruiker word genoodsaak om 'n tweede keer (die basiswoord) na te slaan. Is hierdie 'n ideale oplossing of word dit verkies bloot omdat dit minder ruimte in beslag neem as 'n onafhanklike verduideliking?

Soos woordeboeke geleidelik van druk- na digitale medium beweeg, verdwyn hierdie ruim-

tebeperkings. Sekere probleme verdamp eenvoudig. Om 'n nietige voorbeeld te gee, die behoefte aan afkortings, tildes, en dies meer bestaan nie meer nie (alhoewel 'n verbasende aantal woordeboeke hierdie norme selfs in hul digitale weergawes behou). Die vraag ontstaan dus of ons die volledige reeks redigeringsbeleide en norms in die lig van die veranderde omstandighede behoort te beskou en te herevalueer. Hierdie studie neem 'n paar bekende redigeringsbeleide en aanbiedingsnorme in oënskou, en oorweeg dan watter daarvan nie meer toepaslik is in die digitale medium nie en met watter nuwe beleide hulle vervang kan word.

Sleutelwoorde: DEFINISIES, VOORBELDSINNE, DIGITALE MEDIA, UITSLUITINGSKRITERIA, HEKWAGTER, LEKSIKOGRAFIESE KONVENSIES, AANLYN WOORDEBOEK, GEBRUIKERSPROFIEL

1. Setting the scene: from print to digital

This paper revisits a number of familiar and well-established editorial policies and lexicographic conventions. The aim is to discover whether policies which developed during the long period when dictionaries existed only as printed books remain appropriate in the 21st century, when many — if not most — dictionaries are now published in digital media.[1]

In early 2015, in one of the regular updates to its dictionary, Macmillan added entries for 64 chemical elements. This completed the dictionary's coverage of all 118 elements. But it is legitimate to ask why they were not all included in the first place. It is usual practice for dictionaries to cover all members of any clearly-defined set (days of the week, signs of the zodiac, and so on) but in this case it was decided to omit the rarer elements in the interests of including other, more frequent vocabulary items. When dictionaries are published in the form of printed books, editors make decisions like this all the time: a book of finite dimensions sets up a "zero-sum" game, in which the addition of one category of information entails the omission of something else.

The problem can become acute when major new editions are created (typically every four or five years). Newly-emerging words, phrases and meanings need to be added in order to ensure that the dictionary remains current. At this point we have to decide whether to remove an equivalent amount of material in order to accommodate the newcomers (and if so, using what criteria?); whether to increase the size of the book (a popular option, but unsustainable in the long term); or whether to create more space by making typographical adjustments and increasing the amount of text on the page (which may alienate users). Each strategy carries its own risks, which we generally try to minimise through a carefully calibrated combination of all three expedients. For editors of printed dictionaries, the optimal use of limited space is a major preoccupation.

Macmillan's dilemma regarding chemical elements is just one of countless similar decisions forced upon editors working in print media. This is one of the reasons why digital media are so much better adapted as a platform for reference materials of all types (encyclopedias and maps, as well as dictionaries).

Lexicography is going through a turbulent phase and, as dictionaries

gradually migrate from old to new media, "lexicographers … currently live in a sort of interregnum" (Hanks 2015: 87). As always happens when changes are driven by technology, the global picture is uneven. In many parts of the world, paper dictionaries still have a healthy future ahead of them. Furthermore, certain *types* of dictionary — such as those designed for schools, or special-subject dictionaries, or dictionaries of "smaller" languages — may show a preference for print for some time to come. But for three major categories of dictionary — which collectively account for a large chunk of the global dictionary market, and have also had the greatest impact in terms of lexicographic innovation — the long-term decline in sales of printed editions is irreversible and has led publishers to focus increasingly on digital versions. These are: general monolingual dictionaries aimed at adult mother-tongue speakers; bilingual dictionaries for "big" language pairs; and monolingual learner's dictionaries. Progress, for now, is somewhat uneven, but the direction of travel is clear. Nor should this be seen as a cause for nostalgic regret: with unlimited space and digital functions such as multimedia and hyperlinking, new media provide exciting opportunities for innovation and improved coverage, and open up endless possibilities for reference resources which will serve users' needs more effectively than their print-bound predecessors. This paper looks at the implications of this change for the way dictionaries present information and for the type and range of information they include, and asks how well current online dictionaries have responded to the new reality.

2. Background: changes in the publishing model

Though dictionaries in some form pre-date the invention of print (see e.g. Hanks 2010), the dictionaries we are familiar with today largely evolved in the medium of the printed book. For English, this means over 400 years in which editorial policies and lexicographic conventions have developed and become settled. People know what to expect — and what not to expect — of their dictionary: numbered word senses, concise definitions employing familiar (if sometimes incomprehensible) formulae, devices for conveying the sounds of words in written form, and so on. But much of what we take for granted as "natural" features of dictionaries are in reality expedients. They evolved not because they are the best possible way of conveying information to users, but because they satisfy the imperative of shoehorning large amounts of information into a limited space.

Users of Merriam-Webster's dictionaries, for example, will be familiar with their idiosyncratic defining style. This was introduced in the 1950s by chief editor Philip Gove who, according to Kory Stamper (Stamper 2015), was tasked with "saving" 300 pages from the Second International to create the Third. Another source notes that "Every editorial decision Gove made was dictated by space: the need to create as much of it as possible so he could cram new words into the finite boundaries of the printed book. ... Gove claimed he

saved 80 pages in the Third by using fewer commas" (Fatsis 2015). In a typical set of Merriam entries, we learn that *expectant* means "characterized by expectation", and *expectation*, in turn, is defined as "the act or state of expecting". (*Expectancy*, meanwhile, is the subtly different "act, action, or state of expecting".) So a user who genuinely doesn't know the meaning of *expect* (unlikely, but the same "recursive" approach is used for less familiar sets of words too) can only resolve the meaning of *expectant* by making two further look-ups: tiresome for the user, no doubt, but undeniably economical.

The use of telegraphic definitions is not the only space-saving strategy. Until at least the 1990s, the *Concise Oxford Dictionary*, living up to its name, deployed a range of techniques geared to cramming an impressive amount of data into limited space, as this entry for the word *bag* (from the 7th (1982) edition) illustrates:

> **băg**[1] *n.* **1.** receptacle of flexible material with closable opening at top (esp. w. prefixed word showing contents or purpose; DIPLOMATIC *bag*, GAME[1] *bag*, HAND[1]*bag*, KIT[1]*bag*, *mailbag*, *travelling-bag*, VANITY *bag*); (w. such prefix understood) particular kind of this; hence ~FUL. 2 *n.* **2.** contents of bag; MIXED *bag*; amount of game a sportsman has shot or caught (also fig.) **3. ~and baggage**, with all belongings; **~of bones** lean creature**; (whole) ~of tricks** every… [etc]

Figure 1: Partial entry for *bag*, *Concise Oxford Dictionary*, 7th edition, 1982

In the UK, at least, dictionaries had been gradually moving away from these extreme forms of lexicographese even before the move to digital media. But there is always a trade-off, and improved user-friendliness generally meant reduced coverage of the lexicon.

The Macmillan Dictionary, when originally developed for print, had an explicit policy of favouring the most central vocabulary of English. The goal was to provide detailed information (on syntax, collocation, phraseology, register, and so on), backed up by abundant example sentences, for a core set of 7500 high-frequency words. The unavoidable downside of this approach was that words outside this set received more perfunctory treatment, and often lacked examples altogether. (Steps are now being taken to remedy this.) The policy is far from ideal, but it is perfectly defensible in the context of print publishing: adding an example sentence at a word like *parsimonious* could mean that an important pattern at a verb like *instruct* would be left without an example — and for the student who needs to use *instruct* productively, this could be problematic.

Information about morphology is another area where difficult choices have to be made. Among the well-known English monolingual learner's dic-

tionaries, most (in their print editions) provide inflections only for words with irregular morphology: it is assumed, rightly or wrongly, that target users know how the verb *walk* conjugates but may have problems with a verb such as *strive*. The exception is the COBUILD family of dictionaries which have always supplied full inflectional information for every headword, regular or otherwise. Either policy is defensible, but the COBUILD approach carries a space penalty: the three inflected forms shown at *psychoanalyse* take up a full line, and the systematic application of this policy is one of the reasons that COBUILD dictionaries covered a significantly smaller part of the lexicon than their competitors.

If traditional editorial policies and dictionary conventions are — as these cases illustrate — at least partly driven by the space constraints of the printed medium, what happens when those constraints no longer exist?

3. The response so far

The digital revolution has already led to a redefining of what we mean by "dictionary". Contemporary general-purpose monolingual dictionaries now routinely include some or all of the following: a thesaurus, multilingual content, a blog, language-related games or puzzles, "Ask the Editor" features, videos, infographics, and user-generated content of various kinds (Rundell forthcoming). These are supported by almost constant activity on social media. But the focus of this paper is not on the novel features which complement and enhance what is there already, but on the dictionary's central function: describing the meanings and usage of the words in a language. The questions here are how well publishers have adapted to the new medium, how this has affected dictionary macrostructures and microstructures, and what more needs to be done.

In keeping with the uneven way in which innovation is distributed during this transitional period, we currently find several dictionary models co-existing. Broadly, these are: dictionaries published in print form only; those appearing in both print and digital media; and digital-only dictionaries. The second category is probably the most common (at the time of writing), but even digital-only dictionaries are — in most cases — derived from print products. The Macmillan Dictionary is an example of the last type, having started as a printed book in 2002 and moving to a digital-only model in 2013. But the same applies to the so-called "aggregators", online resources such as dictionary.com and thefreedictionary.com. Though apparently "new" products for the digital age, they recycle dictionary data from traditional sources. The smallest category consists of dictionaries conceived and compiled from scratch as digital products. Examples include *Elexiko* ("an online dictionary of contemporary German") published by the *Institut für Deutsche Sprache*; the *Diccionario de Aprendizaje del Español como lengua Extranjera* (DAELE), a Spanish learner's dictionary being developed at Pompeu Fabra University; and the *Algemeen Nederlands Woordenboek* (ANW) being compiled at the *Instituut voor Nederlandse Lexicologie* in Leiden. All are works in progress rather than complete dictionaries,

and all are from non-commercial institutions.

In its most primitive form, a digital dictionary simply makes the text of a printed dictionary available on a website. A notable example is the *Diccionario de la lengua española* (published by the Real Academia española), whose online dictionary is virtually identical to the print product it is derived from. In this entry for *traducción*, no attempt has been made to exploit the possibilities of the new medium: the entry retains the abbreviations, tildes, and recursive definitions ("the act or result of translating") of the original, with no hyperlinks to words referred to in the definitions:

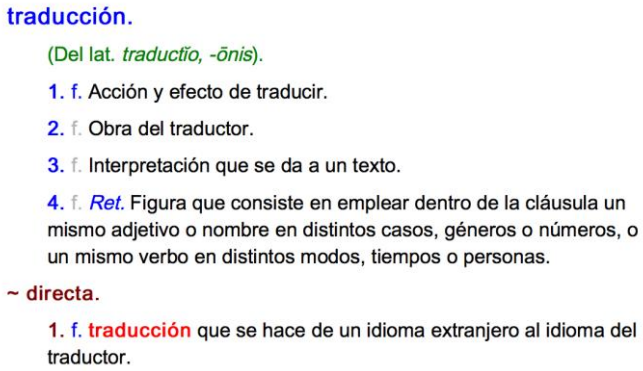


Figure 2: Entry for *traducción*, *Diccionario de la lengua española*

This is an extreme case, but in many online dictionaries, old and new features sit uneasily together. Though never existing in print, *Wordnik* — with its two-column presentation — shows contemporary web-derived example sentences on the right side of the screen, supported, on the left, by definitions from a range of traditional dictionaries. Thus at its entry for *tweet*, we find up-to-the-minute examples of the social-media sense, while the corresponding definitions (derived from an old edition of the *American Heritage Dictionary*) fail to record this recently-coined meaning.

Wiktionary is an especially interesting case. On the face of it, this is a very "modern" dictionary: an entirely web-based resource, its entries created from user-generated content, and with no roots in traditional print lexicography. But things are not quite so simple. Though most entries for subject-specific terminology are newly created, usually by people with specialist knowledge, many of the definitions for more "everyday" vocabulary are simply copied from other dictionaries. Worse, *Wiktionary*'s contributors — rightly concerned about intellectual property issues — tend to borrow material from a safely-out-of-copyright edition of *Webster's Revised Unabridged Dictionar*y published in 1913. Thus many of Wiktionary's entries exhibit long-outdated defining styles and an analysis of word senses which reflects old-fashioned ideas about meaning dating from the

pre-corpus age. As Robert Lew has commented: "It seems that the web community, while enthusiastically embracing the novelty of online collaboration, propagates the traditional model of lexicographic description" (Lew 2014: 17).

In the best "hybrid" dictionaries (where a dictionary created for print publication is also available online), there are still remnants of older ways of doing things. But conscientious efforts are being made to adapt to the new medium. As well as making obvious changes (spelling out abbreviated forms and grammar codes, more "open" design where different information types start on a new line and often in a new colour, and so on), dictionary-makers are rethinking the role of alphabetical order. In a traditional macrostructure, alpha order is the mechanism through which users find what they are looking for. It is so fundamental to the way print dictionaries are organised that early digital dictionaries clung on to this model: they continued to display dictionary entries in alphabetically ordered lists, seemingly reluctant to recognise the irrelevance of this approach in an online resource. But alphabetical order is an arbitrary system which brings together completely unrelated words in sequences like:

redneck, redness, redo, redolent, redoubtable

After some delay, this model is giving way to one more suited to the new medium. The most usual method now is that a search for a specific word brings up the entry for that word and that word only, typically with links to "related words" (as opposed to alphabetically-similar words) shown in a sidebar, as in this entry for *area* from the online version of the *Oxford Advanced Learner's Dictionary*:

Figure 3: Entry for *area*, http://www.oxfordlearnersdictionaries.com

There are residual issues with cross-referencing policy: this entry for *area* still ends with a "see also" list inherited from the print edition, even though the "Other Results" column makes this redundant. But these are teething troubles. A further development relates to phrasal verbs and idioms. These would traditionally be "nested" under main entries, so that *set off*, *set up*, and *set someone's mind at rest* could all be found at the end of the entry for *set*. A newer model — now adopted in many English dictionaries — is to make these items standalone entries. This makes sense if we see phrasal verbs and idioms as distinct lexemes (and many of the former and some of the latter have more than one sense). Why should a user who wants to understand the expressions *put up with* or *set the cat among the pigeons* be obliged to scroll through a long entry before eventually locating their search item somewhere near the bottom?

Some interesting alternatives to conventional macrostructure can be found in dictionaries with no print legacy. This entry from DAELE gives a flavour:

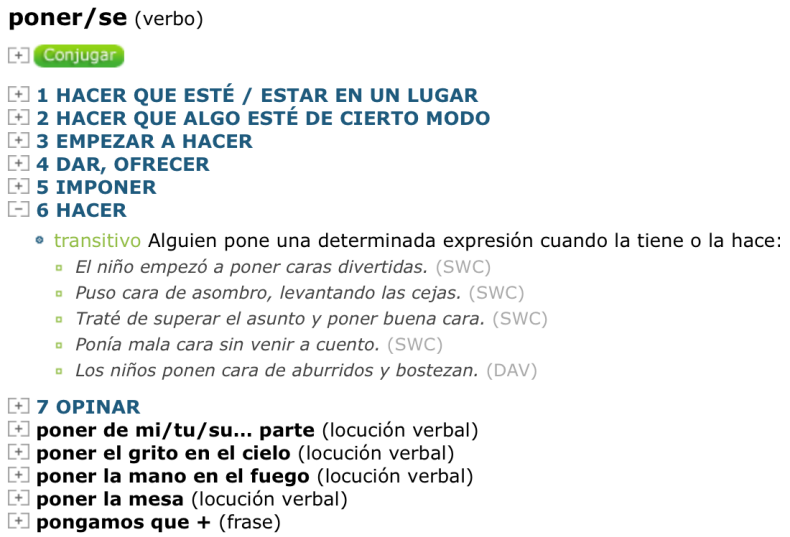


Figure 4: Entry for poner/se, DAELE

Here the "Conjugar" button gives users the option of seeing morphological information, while each of the example sentences comes with information showing the corpus it derives from. But the most notable feature is that each main part of the entry can be opened up or collapsed using the + and - buttons. Starting from a bare menu giving signposts to each sense or usage, the user can pick a specific meaning to see a fuller definition supported by several corpus examples. Some of these features can also be seen in this entry for *Beratung* (counselling or guidance) in *Elexiko*:

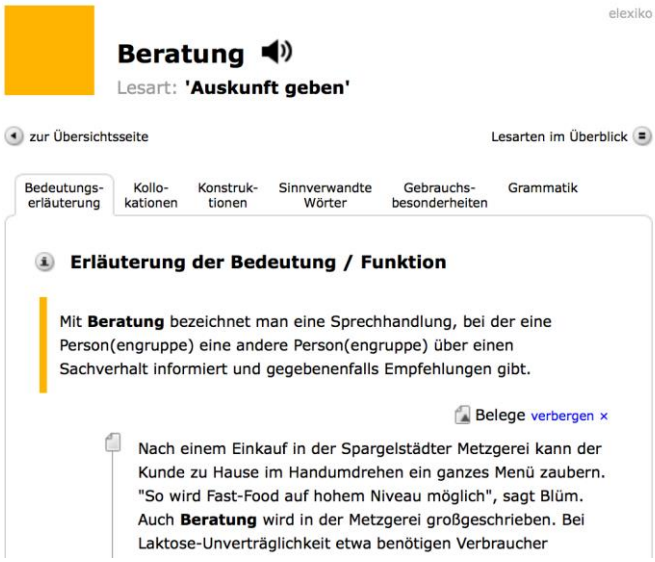


Figure 5: Entry for *Beratung*, *Elexiko*

As in DAELE, users can show or hide examples. And as in DAELE, we find a less traditional approach to defining — in this case a full-sentence "explanation of the meaning or function" of *Beratung*. But what is most interesting here is the use of tabs: these provide access to other categories of information (about collocation, syntactic behaviour, related words, and so on) but at the same time they give users the option of ignoring any information type which they are not (currently) interested in.

Before we conclude this section on current and emerging practice, a few observations are in order. Firstly, there are as yet no "standard" models for the macrostructure of a digital dictionary. What we are seeing at present is a great deal of trial-and-error, as publishers experiment with different approaches. One encouraging sign is the use of generic conventions which can now be assumed to be familiar to anyone using the Web. The + icon in DAELE and the tabs used in *Elexiko* are devices used in many (non-dictionary) websites for managing information, and for suppressing or making available different information-types. The goal in all cases is to avoid overwhelming the user with data, while at the same time making a large quantity of information easily accessible.

The risk of information overload was a challenge for publishers even before dictionaries migrated from print to digital, as the corpus revolution allowed us to provide more complete descriptions of a word's contextual features. To a degree, digital media supply the tools to meet these challenges (even if publishers are still trying to work out the most effective solutions). Part of the problem relates to what Robert Lew has called "presentation space". As Lew

points out, "*storage space* in electronic dictionaries is relatively unlimited", but *presentation space*, which "refers to how much can be presented (displayed, visualized) at a given time to the dictionary user", is self-evidently not (Lew forthcoming). Regardless of how much information the dictionary database contains, the amount that can be presented effectively on a single screen is limited. And with the growing trend for consulting dictionaries on mobile devices, the problem becomes more acute. But dictionary-makers also recognise that users consult dictionaries for different purposes in different situations — broadly, in receptive or productive modes, but with various subdivisions of these types. All these factors intersect, and the challenge for publishers is to design macrostructures which minimise the problem of "too much information", and take advantage of search techniques already familiar to web-savvy users, while facilitating access to different layers of information which will meet different situations of use.

We are, then, in a transitional phase. The challenges and opportunities created by the move to digital media are fairly well understood, and most publishers have grasped the point that the removal of limits on "storage space" is not a licence to abandon the traditional virtues of conciseness. As John Simpson has observed, "if editors were to allow the extent of individual entries to range out of proportion to utility this would result in making the user's task of interpreting an entry much more difficult" (Simpson 2014: 21). There is plenty of experimentation, but as yet little consensus on the way forward.

This is therefore a good moment for a fundamental re-appraisal of policies and conventions which have become so familiar that we may mistake them for being an essential part of any dictionary's DNA. What is needed now is "models for e-dictionaries that focus on critical areas like the data to be included … the structures to present and accommodate the data, the functions of these dictionaries and the way they should respond to the needs of their target users" (Gouws 2014: 157). In the sections that follow, we will look at three specific areas where traditional policies may need rethinking: inclusion criteria, definitions, and example sentences.

Some specific issues: (1) inclusion policies

One of the first questions any dictionary publisher has to consider is "which words get into the dictionary". The theoretical background to this is the observation that the lexicon is an unbounded set. As Hanks points out, "the lexicon is dynamic: new words are being added all the time" (Hanks 2013: 29). When even the mighty OED does not claim to include every English word, it follows that all dictionary publishers need to have robust criteria governing decisions about what to include. But do these criteria need rethinking for the digital age?

Inclusion criteria typically take account of the corpus evidence for a word's frequency, currency, and dispersion across text-types and regions. Some of these criteria are already being modified for digital dictionaries. For exam-

ple, the Oxford Dictionaries site addresses the question of longevity, acknowledging that traditional, stricter criteria regarding how long a word had been current may no longer be appropriate: "It used to be the case that a new term had to be used over a period of two or three years before we could consider adding it to a print dictionary. In today's digital age, the situation has changed" (http://www.oxforddictionaries.com/words/how-do-new-words-enter-oxford-dictionaries).

But two key factors are the "user-profile" of a particular dictionary, and the availability of space. Space limitations require dictionary-makers to be selective about what they include (this has contributed to the public perception of dictionaries as "gatekeepers", only admitting to "the dictionary" words of which they approve), and a good user-profile is the most reliable way of ensuring that the resulting headword list is fit for purpose. A user profile "seeks to characterize the typical user of the dictionary, and the uses to which the dictionary is likely to be put" (Atkins and Rundell 2008: 28). A clear idea of the target user's receptive and productive needs, pre-existing knowledge, language proficiency, and reference skills, is an indispensable aid to inclusion decisions when space is limited.

But neither factor has the same weight when the dictionary is online. It is obvious that unlimited space means inclusion policies can be relaxed, but in an online setting it also becomes much harder to predict who the user will be. In the case of the familiar English monolingual learner's dictionaries, well over 50% of people consulting the site have arrived there through what is known as "organic search": they have submitted to their search engine a search-string (such as "definition of X") which does *not* specify a particular dictionary, then clicked on one of the links in the output. So-called "direct search", where the searcher specifies a particular source (such as Oxford or Macmillan) accounts for a smaller segment of total traffic to most dictionary sites. Consequently, the potential user group is harder to pin down, and this makes it more difficult to feel confident about inclusion decisions.

Samuel Johnson noted rather gloomily that "they that take a dictionary into their hands have been accustomed to expect from it a solution of almost every difficulty" (Johnson 1747: 6). In the digital age, users' expectations are higher than ever. The former "gatekeeper" notion is giving way to a situation where dictionary users (especially younger users) no longer consider that a word is somehow invalid if it is not in "the dictionary". They are more likely to think that if a given dictionary doesn't include a word which they have heard, then the fault lies with the dictionary rather than with the word — and they will simply try a different source. So when there are no space constraints, it may make sense to turn the question around and — rather than asking "does this word pass my inclusion tests?" — we should ask instead "are there good reasons for *not* including this word"? Some traditional principles still apply: candidate words have to be "real" — not invented, or used by only a small group (co-workers, family, or the like) — and they must be supported by independent evidence.

With a general approach based on "exclusion criteria" as our starting point, we need to look at some specific categories. The most difficult of these is so-called "named entities" — broadly speaking, names of people, places, institutions, companies, and so on. Should dictionaries include them? Traditionally, most dictionaries do not include encyclopedic information, but the boundary between encyclopedic and lexical has never been clear-cut, and there is a long list of exceptions. For example, dictionaries generally include the names of institutions which have well-established metonymic uses: thus t*he Kremlin, the Pentagon*, and *Buckingham Palace* will usually feature in any headword list, because corpus data frequently includes sentences like these:

> ***The Kremlin*** *wants the presidential term extended from four to seven years*
> *The Bush White House and* ***the Pentagon*** *seem not to have planned for such contingencies.*
> *How can we be sure that* ***Buckingham Palace*** *has behaved properly in this case?*

The same applies to places real and imaginary which have extended meanings, such as *Mecca* (*At its tip lies Sharm-el-Sheikh, a* ***Mecca*** *for divers and sun-worshippers*) and *Shangri-la* (*a weekend in New York's gay* ***Shangri-la***). In cases like these, dictionaries typically define only the extended uses. Similarly, names like *Google* and *Facebook* only enter most dictionaries as verbs, with no definition for the proper nouns they derive from. There is a host of other quasi-encyclopedic information in dictionaries (such as names of religions, or of trademarked products such as *Band-Aid* and *Memory Stick* which are often used generically). Making these lexical/encyclopedic distinctions is difficult enough for lexicographers, but to the average user they will look arbitrary. But there is clearly a thin-end-of-the-wedge aspect to this. If we decide, for example, to provide definitions for countries (as well as for languages and nationalities, as is the usual convention currently), then why not also for cities, and what is the cut-off point here? And if countries and cities, why not people too — and if so, which ones? The whole issue of which named entities a dictionary should include needs to be revisited in the light of changed circumstances (and changed expectations among users) — though in resources like Babelnet, the lines between encyclopedic and lexical data are already breaking down.

A number of other categories need to be considered. At the end of 2014, the American Dialect Society named as its Word of the Year (WOTY) the social media hashtag #blacklivesmatter. This is a new departure. Dictionary publishers and others routinely nominate Words of the Year, and up to now they have been recognisable as words and have found their way into dictionaries: Oxford's WOTY for 2013, for example, was the now ubiquitous *selfie*. But the American Dialect Society is not alone in extending the scope of what counts as a word. In a readers' poll hosted by dictionary publisher Collins in 2014, the hashtag #nomakeupselfie was a popular choice as the word people most wanted to see in the Collins dictionary, attracting enough votes to come a creditable fourth. Do hashtags belong in dictionaries? Almost certainly not — most are trans-

parent in meaning, and few of them last more than a few weeks — but the question needs to be addressed.

So too does the issue of lexical creativity. This is a pervasive feature of language in use, and lexicographers routinely face inclusion decisions when confronted by examples of it. A newspaper article by the author Margaret Atwood, discussing the concept of freedom in the age of the Internet, provides two interesting examples:

(1) *We human beings have been exploring the border between freedom and* ***unfree-dom*** *for a very long time.*
(2) *Minus our freedom, we may find ourselves no safer; indeed we may be* ***double-plus unfree****, having handed the keys to those who promised to be our defenders.* (Atwood 2015)

Unfreedom is a legitimate formation, but corpus data shows that it is extremely rare (with a hit rate of less than 0.02 per million words) and, being paired here with *freedom*, its meaning is completely clear. Understanding *double-plus unfree* requires a little more background knowledge: the *double-plus* prefix is a feature of Newspeak, the fictional language used by the government of Oceania in Orwell's *Nineteen Eighty-Four*, where it functions as an intensifier. Atwood's choice of words is interesting in the context of her argument, but is any of this lexicographically relevant? The key, as Hanks has argued over many years, is to distinguish between "norms" and "exploitations" (e.g. Hanks 2013: 10-15). In some cases, exploitations can become norms, as one individual's creative coinage gets picked up by others and settles into the language. But such instances are hugely outnumbered by one-off examples of creativity which barely register in corpus data — and which have no place in a dictionary. (See also the discussion here: http://www.macmillandictionaryblog.com/what-goes-in-the-dictionary-when-the-dictionary-is-online.)

The cases discussed above all contribute to fleshing out what we mean by exclusion criteria, and we can now attempt a work-in-progress summary of what these might include:

— user-profile: although (as noted above) this is harder to pin down when a majority of users arrive directly from a search engine, it remains relevant. A general-purpose dictionary is a different animal from a more specialised resource (such as a dictionary of engineering or economics, or a comprehensive historical dictionary), so some filtering is still needed (see also below on technical terms)

— named entities: some broadening of what is acceptable for inclusion seems reasonable, but the question needs more discussion so that robust criteria can emerge

— hashtags: this looks an unlikely category. There may well be a case for a (separate) online resource which lists and explains the most commonly

used hashtags, but there is no real case for these to be included in a general-purpose dictionary

— exploitations: as noted, the traditional position on excluding these is well-founded, and evidence of frequency and dispersion will usually resolve difficult cases

— anything ephemeral: it is always difficult to make reliable predictions about a word's longevity. When dictionaries existed only in print, the problem was less acute. With new editions appearing only every four or five years, editors could often track a word's currency over a longer period. Editors of online dictionaries do not have this luxury, but in principle it is not the function of dictionaries (even those specifically devoted to neologisms) to record the many coinages which appear and disappear in a short space of time

— anything parochial: this is a vague category and not easily defined. But (unlike specialised dialect dictionaries) most dictionaries do not include usages whose range is very limited (whether geographically or socially).

— anything highly technical: the range of specialist "sublanguages" is vast, and few dictionaries (as opposed to specialised glossaries) even scratch the surface in recording their vocabulary. Terminology of the type found in scientific journals like *Nature*, where subject-specialists are addressing other subject-specialists, was rarely included in print dictionaries, and there is no reason for this principle to be relaxed for general-purpose dictionaries in a digital environment.

Individual items can sometimes move unexpectedly into the mainstream. Seemingly ephemeral coinages or parochial usages will in some cases confound our expectations and become part of general vocabulary. Similarly, events in the real world may propel a specialist word towards wider currency: following the global financial crisis of 2008, numerous longstanding technical terms from that sector (*credit default swap, quantitative easing*, and *LIBOR*, among many others) suddenly became part of general discourse — and so merited inclusion in general-purpose dictionaries. But none of this invalidates the broad principles.

Some specific issues: (2) definitions

In section 2 we looked briefly at some of the characteristics of traditional methods of defining. We saw how a focus on economy can lead to definitions which achieve conciseness (and aspire to precision) through the use of standard formulae ("the act of X-ing", "characterised by Y", and so on) and through a "recursive" strategy, where (for example) the entries for *expectant* and *expectancy* feature definitions which are not self-sufficient but depend on the definition at *expect*. And somewhat disappointingly, the user-generated definitions in Wik-

tionary often perpetuate these styles (as does the Spanish dictionary of the Real Academia). In all cases, the goal of saving space is achieved, but the costs are loaded onto the user, who has to learn these conventions in order to fully understand what the dictionary is saying.

In the last 30 years, publishers — especially in the UK — have addressed this issue by developing more open defining styles which approximate to "normal" prose. Even at Merriam-Webster, the digital medium has brought changes in the way words are defined. Its online dictionary (effectively the Merriam-Webster *Collegiate*) provides two layers of definition, as the entry below illustrates: the traditional style — as enjoined by editor Philip Gove — is still there, lower down the entry. But the first thing we see is two new explanations of *expectant* — and these (unlike the so-called FULL DEFINITIONS which follow) require no familiarity with lexicographic conventions and can be fully understood without the need to consult other entries.

Figure 6: Entry for *expectant*, http://www.merriam-webster.com

A related issue is the question of defining vocabularies. Definitions in a learner's dictionary have to be accessible to users with relatively low language proficiency. Most English learner's dictionaries address this issue by identifying a small list (typically of 2000–3000 words) of high-frequency words, and using these, and only these, when writing definitions — the contention being that even quite low-level users will successfully decode any definition. User-research broadly supports this claim, and the use of a defining vocabulary (DV) has been a salient feature of publishers' marketing since these lists were first introduced (see Atkins and Rundell 2008: 449-450). But, for lexicographers and users alike, defining vocabularies are not without their problems, and the cost of clarity can sometimes be a loss of precision. Does the digital medium offer opportunities for improvement? In the digital editions of most (if not all) of the British learner's dictionaries, every word in an entry is hyperlinked. So if a user

is unsure about any word in a definition, they can rapidly find the entry for *that* word. But it is never ideal to have to look from one entry to another in order to get the full picture. A more promising approach may be to create somewhat larger DVs with two or three bands based on frequency. In this model, any word could (and ideally, should) be defined using words from Band 1. But lexicographers would have the option, when necessary, of using words from a higher band — for example, when defining a technical term of low frequency.

As far as the *content* of definitions is concerned (as opposed to the way they are worded), online dictionaries — unconstrained by the need for economy — need to "find the balance between telling the fullest story and deciding what's useful to or necessary for the average reader" (Fatsis 2015). Much has been written about the inadequacies — for many areas of the lexicon — of "classical" approaches to defining, with their insistence on "substitutability" and a model based on *genus* and *differentiae* (e.g. Atkins and Rundell 2008: 416-17). The problem is not merely a practical one — does the definition enable the user to grasp what a word means? — but a theoretical one too. As Hanks has observed "The very word *definition* implies identifying boundaries" (Hanks 2013: 85), and this reflects a traditional view of word meaning which is now being challenged. The assumption that meanings are fixed entities, which "can be attributed to the word in isolation, rather than in context" (Hanks 2015: 87) and which can be described in terms of "necessary and sufficient conditions", is undermined by research in lexical semantics and prototype theory, backed up by the findings of corpus linguistics.

This has led to new approaches to defining (or better, explaining) word meanings. A common thread is a greater focus on context and co-text, and some of these experiments pre-date the digital era.

A notable early example is the "full-sentence definitions" (FSDs) pioneered by the first COBUILD dictionary in 1987, and subsequently adopted (though not systematically) by many other pedagogical dictionaries. This format allows us to include in the definition itself significant (and helpful) information about the definiendum's colligational and collocational preferences, and sometimes also its illocutionary features. I have argued elsewhere (Rundell 2006) that FSDs are not well-adapted to explaining *every* category of word in the lexicon, but they work well in many cases, and they represent an important addition to the definer's repertoire.

On a smaller scale, the Macmillan Dictionary introduced a model for conveying connotative (or pragmatic) information by means of a second sentence. Here a conventional definition explains a word's denotative meaning, then a second sentence adds information about the attitude or motivation of a speaker who chooses to use this word. For example:

> **bureaucrat** someone who is employed to help run an office or government department. This word can suggest that you do not like people like this because you think they have too much power and care too much about rules and systems

Similar examples can be found at the Macmillan entries for *blue-eyed boy*, *nerd*, *just good friends*, *bourgeois*, and many others. Although this style was used in the first (paper) edition of the dictionary (2002), it is clearly well-adapted to a digital structure in which different information types are made available to the user when needed.

In the online *Elexiko* dictionary, traditional definitions are replaced by a statement providing an "explanation of the meaning or function" of the headword. In the case of *Beratung* (see section 3 above), this explanation tells us that it is a speech act, and describes a kind of "frame" in which one person provides another with information about an issue and, where appropriate, makes recommendations.

Yet another innovation can be found in the *Algemeen Nederlands Woordenboek* (ANW) — like *Elexiko*, a genuinely "from scratch" digital resource. Here, conventional definitions are supplemented by "semagrams". A semagram is "the representation of knowledge associated with a word" (Schoonheim and Tempelaars 2010: 721). Thus at the entry for *koe* (cow) we are told about the sound cows make, and we learn that (among other things) they provide milk and meat, have to be milked daily, have udders and four stomachs, and are thought of as being friendly but lazy. All of which "leads to a much richer semantic description, in which the implicit knowledge of the definitions has been made explicit" (ibid.).

A radically different approach is found in Hanks' *Pattern Dictionary of English Verbs* (PDEV), where conventional word senses are replaced by patterns. Here a syntactic pattern (such as *allow someone to do something*) is described in terms of the semantic types (such as Human, Institution, Eventuality) which instantiate the pattern. This description is supported not by anything we would recognise as a definition, but by an "implicature" which maps a meaning onto the specific pattern and its participants. As the site explains, "No attempt is made … to identify the meaning of a verb or noun directly, as a word in isolation. Instead, meanings are associated with prototypical sentence contexts" (Hanks, PDEV).

As all these instances show, dictionary-makers are beginning to explore the possibilities of the new medium. For the time being, there is not much convergence around any new standards, but there are encouraging signs. At one end of the scale, many of the aggregators reproduce material from other sources, with *Wordnik*, for example, featuring definitions from (among others) an ageing edition of the *American Heritage Dictionary*, *Wiktionary* (many of whose definitions come from much older dictionaries), and the truly ancient *Century Dictionary and Cyclopedia*, whose sole definition of *computer* is "One who computes; a reckoner; a calculator". This does not look like a constructive way of exploiting the availability of limitless space. But some of the innovations described here look a great deal more promising.

Some specific issues: (3) example sentences

Dictionary users appreciate example sentences. They help to elucidate meanings, they illustrate contextual preferences, and (especially useful in pedagogical dictionaries) they provide models for language production (e.g. Atkins and Rundell 2008: 452-455). Well before dictionaries went online, the older model of invented, often truncated examples was giving way to the use of authentic examples in the form of complete sentences taken from a corpus. From the 1990s, the provision of *additional* examples became a common feature of dictionaries published on optical disks (CD-ROM and DVD-ROM). Typically these would be taken from a corpus, but in most cases there was little or no filtering (for quality, appropriacy etc) and — critically — examples for polysemous words were not mapped to specific senses.

Now, without the space constraints imposed by the printed medium, publishers of online dictionaries are experimenting with new ways of providing larger numbers of examples. One approach is to give users direct access to the corpora that underpin the dictionary. The *Digitale Wörterbuch der deutschen Sprache* (DWDS), for example, allows users to see concordances in several different corpora, as in this entry for the lemma *Hausarrest*:

Figure 7: Concordances for *Hausarrest*, *Digitale Wörterbuch der deutschen Sprache*

Several other European dictionaries (including the Dutch ANW and the Danish *Den Danske Ordbog*) have a similar concordance feature. This is likely to be a useful resource for linguists and other researchers, but whether non-specialist users want this kind of data (or know what to do with it) is a question which needs to be investigated through user-research.

A different model, which may be better-adapted to the needs and skills of the general user, can be found in *Oxford Dictionaries Online* (ODO). Here, the user will, by default, find one or two examples at most words or senses, but now has the option of clicking on a "MORE EXAMPLE SENTENCES" link to bring up (typically) three further corpus-derived examples. What is especially impressive in the way this is implemented in ODO is that, when the word in question is polysemous, the link appears at individual senses and (as this partial entry for *party* shows), the extra examples are mapped to the meaning which they instantiate:

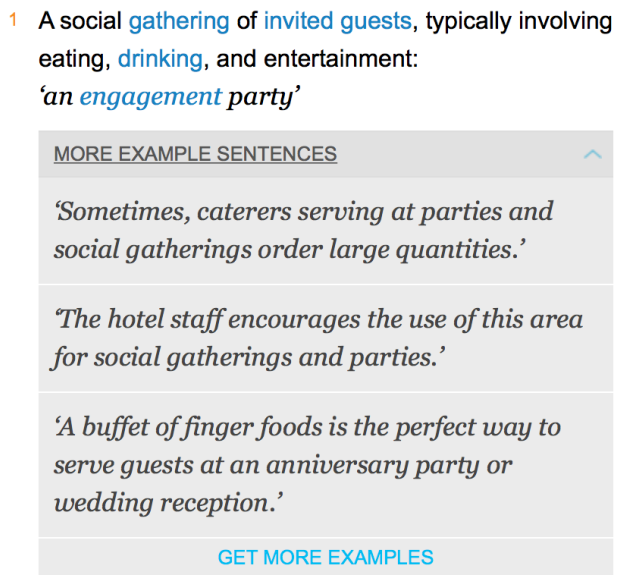


Figure 8: Extra examples feature (*party*) in oxforddictionaries.com

As with the other features, the picture is uneven and not every venture in this area has been entirely successful. As noted earlier (section 3), *Wordnik*'s two-column display (with definitions on one side, and web-sourced examples on the other) runs into problems when the headword has acquired newer senses. In the entry for *toxic*, for instance, almost all of the ten example sentences illustrate more recent uses (toxic assets, toxic relationships etc), but all the definitions on the left side, though taken from five different sources, fail to account for these meanings.

Conclusions

At the time of writing, the online edition of the Spanish Academy's *Diccionario de la lengua española* still defines *diccionario* as "a book" — with no mention in the definition of the medium in which the dictionary appears:

> Libro en el que se recogen y explican de forma ordenada voces de una o más lenguas…

At the other end of the scale, we have a resource like Babelnet. describing itself as "both a multilingual encyclopedic dictionary … and a semantic network which connects concepts and named entities", Babelnet takes full advantage of the digital medium, and provides one model of how "the dictionary" may develop as people experiment with new ways of presenting and linking reference information of various kinds.

This range of responses illustrates how well (and how badly) some dictionary-makers are adapting to the new paradigm. A dictionary is a work-in-progress at the best of times, and as dictionaries steadily migrate to digital media there is a growing flexibility in our ideas about what a dictionary should look like and what information it should contain.

As Gouws has suggested, "The dynamic nature of e-dictionaries enables lexicographers to move away from a static to a dynamic data display that includes the use of a multi-layered structure of dictionary articles" (Gouws 2014: 164). In some of the innovations described here, there is evidence of sensible moves in this direction. One tendency is the increasing use of generic (as opposed to dictionary-specific) conventions for displaying and linking information: hyperlinks, icons for collapsing and expanding a specific category of information, the use of tabs, and so on. A degree of standardisation is emerging in the Web as a whole, and there is a certain "vocabulary" of search strategies which users can now be assumed to be familiar with. So it makes obvious sense for these to be used in dictionary sites, too, since the data on how people arrive at dictionary sites shows that — for many users — the destination is simply an outcome of search, rather than an instance of "looking it up in the dictionary". More generally, dictionary-designers need input from new Web-oriented dictionary-user research and from the field of information science.

But we have also seen that many of these structural innovations are applied to outdated content. Most aggregators recycle entries from dictionaries which pre-date the transformations in lexicography that followed the corpus revolution and the influence of cognitive linguistics. Even a resource as groundbreaking as Babelnet depends, for most of its dictionary content, on Wiktionary — whose definitions of everyday words are in many cases taken from 100-year-old sources. As the scope of the dictionary expands and its structures develop to fully exploit the possibilities of digital media, the lexical data it delivers should also reflect the most up-to-date linguistic thinking about how humans create and understand meanings. This calls for the use of high-

quality corpus-based content, as well as resources such as Hanks' PDEV.

We have looked at three specific areas (inclusion, definitions, and examples) where traditional lexicographic policies are being adjusted to take account of the change in publication medium. This is no more than a first step towards the wholesale re-evaluation of editorial policies and lexicographic conventions which is now needed.

Endnote

1. This paper is based on a talk I gave at the 20th International Conference of Afrilex, held at the University of KwaZulu-Natal, Durban, in July 2015. I am grateful to the Afrilex Board for inviting me as a keynote speaker, and to the conference hosts in Durban for their warm hospitality.

References

Atkins, B.T.S. and M. Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.

Atwood, M. 2015. We are Double-plus Unfree. *The Guardian*, September 2015. http://www.theguardian.com/books/2015/sep/18/margaret-atwood-we-are-double-plus-unfree.

Fatsis, S. 2015. The Definition of a Dictionary. *Slate Magazine*. January 2015. http://www.slate.com.

Gouws, R.H. 2014. Article Structures: Moving from Printed to e-Dictionaries. *Lexikos* 24: 155-177.

Hanks, P. 2010. Lexicography, Printing Technology, and the Spread of Renaissance Culture. Dykstra, A. and T. Schoonheim (Eds.). 2010. *Proceedings of the XIV EURALEX International Congress, Leeuwarden, 6–10 July, 2010*: 988-1006. Leeuwarden/Ljouwert: Fryske Akademy.

Hanks, P. 2013. *Lexical Analysis: Norms and Exploitations.* Cambridge, Mass.: MIT Press.

Hanks, P. 2015. Cognitive Semantics and the Lexicon (Review Article). *International Journal of Lexicography* 28(1): 86-106.

Johnson, S. 1747. *The Plan of an English Dictionary.* Available at https://andromeda.rutgers.edu/~jlynch/Texts/plan.html.

Lew, R. 2014. User-generated Content (UGC) in English Online Dictionaries. *OPAL: Online publizierte Arbeiten zur Linguistik* 2014(4): 8-26. Institut für Deutsche Sprache.

Lew, R. Forthcoming. Space Restrictions in Paper and Electronic Dictionaries and their Implications for the Design of Production Dictionaries. Bański, P. and B. Wójtowicz (Eds.). Forthcoming. *Issues in Modern Lexicography*. München: Lincom Europa.

Rundell, M. 2006. More than One Way to Skin a Cat: Why Full-sentence Definitions have not been Universally Adopted. Corino E., C. Marello and C. Onesti (Eds.). 2006. *Proceedings of the 12th EURALEX International Congress, Torino, Italia, September 6–9, 2006*: 323-337. Alessandria: Edizioni Dell'Orso.

Rundell, M. Forthcoming. Crowdsourcing, Wikis, and User-generated Content, and their Potential Value for Dictionaries. Hanks, P.W. and G.-M. de Schryver (Eds.). Forthcoming. *International Handbook of Modern Lexis and Lexicography*. Berlin: Springer.

Schoonheim, T. and R. Tempelaars. 2010. Dutch Lexicography in Progress: *The Algemeen Nederlands Woordenboek* (ANW). Dykstra, A.and T. Schoonheim (Eds.). 2010. *Proceedings of the XIV EURALEX International Congress, Leeuwarden, 6–10 July, 2010*: 718-725. Leeuwarden/Ljouwert: Fryske Akademy.

Simpson, J. 2014. *What Would Dr Murray Have Made of the OED Online Today?* Slovenščina 2.0, 2 (2): 15-36. Available at http://www.trojina.org/slovenscina2.0/arhiv/2014/2/Slo2.0_2014_2_03.pdf.

Stamper, K. 2015. *This Wild and Barbarous Jargon, Reduced: Practical Lexicography in an Age of Digital Abundance.* Video of talk given at eLex 2015: Electronic Lexicography in the 21st Century. https://elex.link/elex2015/videos/.

Dictionaries and other online resources

Babelnet: http://babelnet.org.
Digitale Wörterbuch der deutschen Sprache (DWDS): http://www.dwds.de.
Macmillan Dictionary: http: //www.macmillandictionary.com.
Merriam-Webster online: http: //www.merriam-webster.com.
Oxford Dictionaries Online (ODO): http: //www.oxforddictionaries.com.
Oxford Advanced Learner's Dictionary Online: www.oxfordlearnersdictionaries.com.
Pattern Dictionary of English Verbs (PDEV): http: //pdev.org.uk.
Wiktionary: https: //en.wiktionary.org/.
Wordnik: https: //www.wordnik.com.

http://lexikos.journals.ac.za

Methods in Lexicography and Dictionary Research*

Stefan J. Schierholz, *Department Germanistik und Komparatistik, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany (Stefan.Schierholz@fau.de)*

Abstract: Methods are used in every stage of dictionary-making and in every scientific analysis which is carried out in the field of dictionary research. This article presents some general considerations on methods in philosophy of science, gives an overview of many methods used in linguistics, in lexicography, dictionary research as well as of the areas these methods are applied in.

Keywords: SCIENTIFIC METHODS, LEXICOGRAPHICAL METHODS, THEORY, METALEXICOGRAPHY, DICTIONARY RESEARCH, PRACTICAL LEXICOGRAPHY, LEXICOGRAPHICAL PROCESS, SYSTEMATIC DICTIONARY RESEARCH, CRITICAL DICTIONARY RESEARCH, HISTORICAL DICTIONARY RESEARCH, RESEARCH ON DICTIONARY USE

Opsomming: Metodes in leksikografie en woordeboeknavorsing. Metodes word gebruik in elke fase van woordeboekmaak en in elke wetenskaplike analise wat in die woordeboeknavorsingsveld uitgevoer word. In hierdie artikel word algemene oorwegings vir metodes in wetenskapfilosofie voorgelê, 'n oorsig word gegee van baie metodes wat in die taalkunde, leksikografie en woordeboeknavorsing gebruik word asook van die areas waarin hierdie metodes toegepas word.

Sleutelwoorde: WETENSKAPLIKE METODES, LEKSIKOGRAFIESE METODES, TEORIE, METALEKSIKOGRAFIE, WOORDEBOEKNAVORSING, PRAKTIESE LEKSIKOGRAFIE, LEKSIKOGRAFIESE PROSES, SISTEMATIESE WOORDEBOEKNAVORSING, KRITIESE WOORDEBOEKNAVORSING, HISTORIESE WOORDEBOEKNAVORSING, NAVORSING OP WOORDEBOEKGEBRUIK

1. Introduction

In dictionary production and in scientific work which is carried out in the field of dictionary research, methods are used to reach certain results. Currently there is no comprehensive and up-to-date documentation of these particular methods in English. The article of Mann and Schierholz published in Lexico-

* This article is based on the article from Mann and Schierholz published in *Lexicographica* 30 (2014). The English version was presented on the StellenLex conference 2015 in Stellenbosch organized by Rufus Gouws. Thanks to Laura Schierholz who was an excellent and reliable help in proofreading in order to get the final English version of this text.

graphica 30 gives an overview in German and is a first attempt to structure the existing methods in our discipline. The other articles which are published in the Lexicographica yearbook (2014) treat methods in specific areas: the functional perspective (Tarp 2014), dictionary planning of online-dictionaries (Geyken 2104), research in dictionary use (Müller-Spitzer 2014b), the saving of lexical data (Romary and Witt 2014), translation science (Bielińska 2014), specialized lexicography (Bocorny Finatto 2014), pronunciation dictionaries (Hirschfeld and Stock 2014), items giving the paraphrase of meaning (Töpel 2014a).

Even this article cannot give a complete overview of all the methods but it provides a general overview of methods in lexicography and dictionary research, as well as in contiguous disciplines. Selected methods will be described in detail to show how to apply the methods.

The basic aim of this article is to foster the work on methods and to encourage young researchers to show more interest in the methods of this research field.

An overview of all methods should be theory independent or should be orientated to the different lexicographical theories, i.e. the "Lexicography and Dictionary Research" (Wiegand 1998a) or the "Function Theory" (Bergenholtz and Tarp 2002; Tarp 2008; Tarp 2013). Above this, all methods which are used in lexicographical practice should be enumerated and ordered. This is rather complicated because one has to keep in consideration that some methods will be applied only with a special dictionary type or are only specific to one dictionary project.

Besides this aspect regarding methods, it has to be noted that the individual theories are not critically reviewed with respect to existing methods, in "Lexicography and Dictionary Research", in the "Function Theory"[1], and in "Lexicography as an Art and Craft" (Landau 1989). Exceptions are (a) the *Historical Dictionary Research* and (b) the *Research on Dictionary Use* where in the past the state of the art of the empirical methods was summarized by Ripfel and Wiegand (1988), in Wiegand (1998a: 568ff.), in Welker (2006, 2013a, 2013b). For the period from 1993 to 2012 a very good overview of the research projects concerning the use of electronic dictionaries including a lot of critical comments on the used methods is given by Töpel (2014b).

Apart from these observations regarding the documentation of methods, one can find methods reflected more or less in the publications of lexicography. These are the publications where (a) the subject matter itself is methodology, (b) single methods are discussed critically, (c) specific methods are used aligned to the subject matter, (d) methods are applied without any reflection on their use but also without any consequences for the quality of the research results.

2. What a method is

A method is a procedure or a technique used to reach a certain aim. In all sciences, scientific methods are applied. They are part of a general theory of sci-

ence on its own or of theories of an individual science. In general, what all methods have in common is that they are considered as single scientific procedures which serve to gain secured and verifiable insights into scientific objects and entities.[2] Accordingly, this counts also for methods in metalexicography and also for all the methods in practical lexicography. Wiegand (2010: 251) writes: "A method is an ordered set of instructions to act which have to be followed in at least one orderly sequence and in compliance with the conditions of correctness which are part of the method properties in order to receive the desired result."[3]

Starting from the basic elements in science theories you can transfer principles — e.g. observational adequacy, explanatory adequacy, hypothesizing, theory construction (and sub theory construction), methods, application — to theories which are used in a single discipline. For metalexicography, different theories are necessary but a dominant role is played by an *action theory* and a linguistic theory.

The *action theory* plays a decisive role for the subject "dictionary" because dictionaries are produced for the purpose of being used. Therefore, the actions of a potential user of a dictionary are in the center of the considerations (cf. Wiegand 1987). An extensive explanation especially in context with the research on dictionary use is found in Wiegand (1998a: 268ff.) but also in the function theory of the Aarhus group (cf. Tarp 2008: 33ff.; Tarp 2014: 58ff.).

The *linguistic theory* plays a decisive role since, not only in language dictionaries but also in subject dictionaries, linguistic expressions are described in/ with linguistic terms.

In both theories one will find methods that can be considered as inductive (from single propositions, observations or data leading to general insights) or deductive (starting from a general regularity to a statement about a single case). This can be seen from the perspective of the philosophy of science but also from the single sciences (research on dictionary use, linguistics etc.). In German linguistics, this was discussed extensively by weighing up the different positions (cf. among others Schaeder 1981: 28ff.), and it is convincingly demonstrated that the strict separation between "inductive" and "deductive" cannot be kept if the specific features of the subject matter *language* and the special requests for lexicography are included.

From a single science perspective a differentiated way of looking at things is necessary. Thus, it is better to speak of a *primarily inductive method* when a systematically collected huge amount of data and the formulation of the hypotheses are used in combination during the research process (cf. Schaeder 1981: 45ff.; Schierholz 2001: 76ff.).

The following methods shall be considered to be used in practical lexicography and in metalexicography. All of them should be guided by the enumerated principles which are universal in general philosophy science. The following text will not contain:

— the field of teaching methods in dictionary use where you can learn the

skills in dictionary use because these methods belong to the area of didactics.

— the methods which belong to terminology and terminography because this would go beyond the scope of this contribution.

— scientific methods in practical lexicography.

Firstly, a short overview to the methods in linguistics will be given because these methods play a dominant role in the production of linguistic lexicographical items.

3. Methods in Linguistics

Within philosophy of science the theories and methods in linguistics can be distinguished under the aspects of the differentiations described above. Very often, a classification in the qualitative and the quantitative approach can be found in linguistics, which takes into account more the specific properties of the subject matter. In this case the mutual dependence of both approaches is absolutely necessary (cf. Schlobinski 1996: 15ff.).

Qualitative methods are procedures in which data measurable in numbers are primarily not collected (e.g. linguistic descriptions and explanations), and in which new insights are obtained from critical hermeneutical work, e.g. by the processes of comprehension and interpretation. Methods of data collection include interviews, written questionnaires, and observations. For the recording and interpretation of these data, special tools were developed for example, content analysis, text analysis, conversation analysis or discourse analysis (cf. Deppermann 2008; Warnke and Spitzmüller 2008). Qualitative methods are also the classical ones in traditional grammar, such as elimination tests, permutation tests or commutation tests.

Quantitative methods are always aligned with data which are measurable in numbers and which are collected with the aim to assess a judgment which is based on results presented in numbers. The data collection must underlie a systematic method of gathering the data. Koplenig (2014: 65f.) distinguishes between "obtrusive" and "unobtrusive methods". Another distinction can be made between experimental and non-experimental procedures. Many of the tools of data survey and data evaluation used are adapted from empirical social sciences and are taken over into linguistics (e.g. oral and written questionnaires, observations, tests, experiments, probabilistic statistics), other tools have on a larger scale genuine linguistic properties (e.g. the query to and the evaluation of text corpora) (cf. the summary in Albert and Marx 2010: 12ff.; Meindl 2011: 25ff.; with a perspective from social sciences e.g. Bortz and Döring 2006 and fundamental Diekmann 2007).

All linguistic procedures play a dominant role in lexicography and dictionary research (cf. Rundell 2012: 64ff.). If such linguistic data are collected it

shall be the basis for the writing of lexicographical items. This means that one has to consider;

(a) the procedures of phonetic transcription, procedures of segmentation and classification, many procedures in phonology and morphology, procedures to categorize the parts of speech, syntactical procedures for the categorization of valence partners, procedures of meaning analysis, content analysis, prototype theory and many more (cf. Rundell 2012: 67ff.);

(b) the methods of observation and questioning, statistical procedures for the collection of linguistic data, for the evaluation of huge and small data sets, for samples as well as for the selection from huge data sets (e.g. text corpus data) (cf. amongst others Schlobinski 1996; Albert and Koster 2002; Albert and Marx 2010);

(c) the knowledge the lexicographer has from experience related to their profession.

The methods of corpus linguistics must be used with special attention because linguistic data is collected on the basis of huge text corpora.[4] The required size of the corpus, the balance of the corpus, the state of the corpus (original text or scanned text; with annotation or current text), the text varieties which exist in the corpus; but first the query procedures and the procedures of the citation evaluation must be taken into consideration and must be used with a critical reflection and perspective to the investigation aims (cf. Heid 2008, Lemnitzer and Zinsmeister 2006, Schierholz 2008a, Schierholz 2008b, Schierholz 2013). It must also be distinguished in the lexicographical process between corpus controlled, corpus supported or corpus validated (cf. amongst others Engelberg and Lemnitzer 2009: 238).

Attention must be paid to the use of linguistic methods in that they serve very often only for the material processing and material preparation for the following lexicographical process. In each case it must be decided in which way data extracted from natural language sources have to be further worked on in the lexicographical process with linguistic methods.

4. Methods in Lexicography and Dictionary Research

Lexicography and dictionary research is marked by pluralism of methods: The lexicographical work is determined by the type of the dictionary but reverts to the results, the methods and the theories of different sciences.[5] (Wiegand 1984: 559; cf. also Wiegand 1998a: 100).

The methods are (a) linguistic methods which are used for the collection of the linguistic data, (b) methods which source from social sciences and especially psychology (e.g. quantitative data analysis, systematic empirical methods, case studies), (c) furthermore philological methods, in which the introspection

together with the personal language competence is important, (d) the methods of comprehension and interpretation which are adapted to the text material but also to all empirical data. Even the philological methods have to fulfill the condition that the intersubjective examination of the results must be possible in each part of the workflow.

In the following section, typical methods and their application in either practical lexicography or in the sub theories of dictionary research will be presented with the essential features.

4.1 Methods in Practical Lexicography

During dictionary-making, different phases of the lexicographical process can be distinguished (cf. Gouws and Prinsloo 2005: 9ff., Svensén 1993). Wiegand (1998a: 134ff.) calls (a) the phase of preparation, (b) the acquisition of the material and the data, (c) the treatment of the material and the data, (d) the evaluation of the material and the data, (e) the preparation of the print process; Engelberg and Lemnitzer (2009: 228) complete with (f) the phase of further development, as well as the upkeeping and cultivation of the data material. In each of the phases different decisions must be taken, actions must be done and different methods must be used. The phases are valid for print dictionaries and for online dictionaries; in some areas in the same way, in others not (cf. Klosa 2013 about online-dictionaries). Furthermore, it must be determined whether the production of a dictionary is (a) a complete new project, (b) a dictionary derived from one or more existing dictionaries[6], (c) a translation of another dictionary, (d) a revision and/or actualization of an existing edition or a retro-digitalization. The latter one can be done by copying (e.g. double-keying-method) or scanning with text recognition (cf. Piotrowski 2012: 25-52; Burch et al. 2000 to the example of *DWB*), or by parsing the tape of type setting instructions (cf. Hauser and Storrer 1996), where the digitalized data (automatically or semi-automatically) can be transferred into single data (e.g. in XML) or into a data base system (cf. also Engelberg and Lemnitzer 2009: 223ff.).

In the following section the methods and the application area where methods must be used will be enumerated in connection with the phases of the lexicographical process.[7] All cases will be led from a complete new production of a dictionary. Only some methods will be dealt with in detail.

4.1.1 Methods during the preparation phase

In the preparation and planning phase of the lexicographical process a needs analysis must be done to calculate the costs, the work flow, the agreed time period of the making and the size of the dictionary. To this belongs an evaluation of the market opportunities, a calculation of production costs, the personal planning, the time management, the development of a payment system and the

profit prospects (cf. Beltrami 2013). To be able to do this one needs methods from business management, which have to be selected and applied with consideration of the dictionary type, the users of the dictionary, as well as the size and the structure of the dictionary (cf. Engelberg and Lemnitzer 2009: 227ff.).

If a dictionary is made by an editorial staff, it is necessary to have a work schedule of the lexicographical work, in which the distribution of the tasks is arranged, e.g. the conditions of the compilation of the articles and the distribution of the articles to the lexicographers: (a) It is possible that each article is written completely only by one person. The sequence of the article writing can start with the articles under *A*, continuing onto the initials of the word list until *Z*. (b) It is possible that a lexicographer will write only one item type or a set of item types in every article so that the production of item types is done thematically. The approach under (a) — from *A* to *Z* — as it was practiced in many dictionary projects in the last centuries, has very often the consequence that the lemma list under *A* is excessively long and that also the dictionary articles are more extensive i.e. produced more carefully and with a bigger expenditure of time than the lemma list under *Z* (cf. Schierholz and Windisch 1991). The more specialist field knowledge that is necessary for specific items of the dictionary articles the better it is when single item types are worked on by a lexicographer who is an expert in the respective special field. The methods which should be used here to achieve an efficient work flow are very different and their availability may depend on the financial situation and on the economic behavior of the dictionary producer. Furthermore the methods belong to the procedures which are used in the section of work organization and they are based on experience and knowledge but they have nearly nothing to do with any linguistic method. Instead of that, all these processes should be carried out in cooperation with an experienced lexicographer because the experience from the lexicographical workplace is helpful and necessary to organize the work efficiently.

Another part of the preparation phase is the development of a concrete dictionary conception where the dictionary type and the dictionary functions are laid down; these decisions also influence other decisions on methods (cf. part 4.1.8). Besides, a text compound structure and a data distribution programme in which the distribution of the lexicographical data and the components of the text compound are arranged must be developed. Apart from the data distribution in the wider sense the following aspects concerning methods must be arranged, although some corrections are possible in some cases later on (cf. also Schierholz and Wiegand 2004):

— the way of the article arrangement in the word list and/or in registers or indices, where alphabetical arrangement and thematic arrangement have to be distinguished fundamentally: in the case of an alphabetical order one has to choose a method of alphabetization, where especially the filling of alphabet external signs must be taken into account (cf. Wiegand 1989: 376ff.; Wiegand and Beer 2013; Wiegand and Gouws 2013). In the

case of a thematic arrangement, not only a system of concepts or of subject groups must be determined, but also a fitting method how the lemma signs or rather dictionary entry can be categorized systematically in this system (cf. to this complex of problems amongst others Wiegand 2004: 62*ff.; Quasthoff 2004: 199*ff.).

— the microstructural programme and the microstructural method of lexicographical treatment (e.g. integrated, not-integrated, semi-integrated) (cf. amongst others Wiegand and Smit 2013a)

— the use of textual condensation procedures (cf. amongst others Wolski 1989), of abbreviations, symbols etc.

— the utilization of transcription methods and transliteration methods with corresponding tables etc.

— guidelines for the drawing up of items giving the paraphrase of meaning (e.g. Wahrig 1973)

— the reference prerequisites and the mediostructural programme which describes under which circumstances and in how a reference carrying out can be undertaken (cf. amongst others Wiegand 2002a; Wiegand and Smit 2013b)

— the layout and the typographical marking up of all dictionary components, especially for the dictionary articles with their items and structural indicators.

Also the working tools (software) to be used must be committed (cf. part 4.1.8) because homogeneous methods must be applied for the sake of uniform results concerning the form. All the methods relevant for the lexicographer to make a dictionary in a systematic procedure should be written down in an instruction book[8].

On the basis of these decisions, example articles should be produced so that the practical suitability of the instructions can be tested and the instruction book can be corrected and possibly changed. Consequently, example articles can function in the following working phases as best example articles for the dictionary work.

4.1.2 Methods during the phase of material collection

In the phase of the material collection the lexicographical work is based on the experience and knowledge of the staff members and on general search routines, which are used to collect data from the internet, from libraries or from existing dictionaries. These processes will help to provide the sources for building up the dictionary basis, which is distinguished into the *primary sources* (data which are used for the purpose of the dictionary making, e.g. citation

collections, special text corpora), the *secondary sources* (other dictionaries), and the *tertiary sources* (all other linguistic material concerning the dictionary subject matter (e.g. literature, grammars) (cf. amongst others Wiegand 1998a: 140; Schierholz and Wiegand 2004: 208f.). Reichmann (1990: 1589) takes the secondary sources and the tertiary sources together. The collection of these data must be done systematically in order to make the data stock of the dictionary basis reliable.

In the case of a new compilation of a corpus for the linguistic analysis and even in the case of the takeover of an existing corpus, the method of compiling (representative, balanced, opportunistic, see above, section 3) must be applied in a very critical manner. Subsequent to this the data can be tagged or annotated so that methods (programs, algorithms) of natural language processing will be used (cf. Klosa 2013: 521).

If in the planned dictionary images and diagrams or in the case of electronic dictionaries, audio and video examples should be presented, it is necessary to compile a data collection which is determined by clear criteria and by methods, which can ensure the suitability of the multimedia material for the envisaged dictionary type and for the dictionary functions.

4.1.3 Methods in the phase of material processing

The phase of the material processing is in the traditional lexicographical process the use of the note catalogue and the excerption of recently obtained material (to the procedure cf. Wiegand 1998a: 145ff.). In case of the — nowadays usual — use of big electronic corpora, other methods are necessary. In the case of the data volume being oversized, the data must be analyzed automatically and by random principle. However, if the data volume is a manageable size each piece of data should be analyzed individually.

For this phase Klosa (2013: 521) mentions the process of corpus analysis for the sake of the extraction of lemma candidates, word frequencies and collocation candidates (to the procedures of statistical determining of collocations cf. Evert 2005). The methodical backgrounds of the lemma selection will be demonstrated more precisely in the next step.

The method of lemma selection

Before one starts writing dictionary articles, a lemma candidate list must be drawn up (cf. amongst others Bergenholtz (1989; 1995); Bergenholtz and Meder (1998); Drosdowski (1977); Scholze-Stubenrecht (2002)). This list should be an open list so that during the process of dictionary production new lemmata can be entered. The lemma selection must ensure a well-balanced distribution of the lemmata over the individual character stretches.

The distribution of the lemmata in accordance with the character stretches

should correspond to the distribution of the token and the distribution of the types in a certain natural language. The quantitative proportion which the word forms have per initial in one language can be recorded by the frequency distributions of word forms in text corpora (cf. amongst others Schierholz and Windisch 1991). In the production of a specialized dictionary a corpus with texts of the special field in question should be the basis because specialized vocabulary can have another distribution of the lemmata per initial than the distribution of the general vocabulary. In dictionaries the page volume which is at hand for a single character stretch should be distributed analogous to the lemma quantities because only then can a homogeneous dealing with the dictionary articles concerning the quantitative perspective be assumed (cf. also Engelberg and Lemnitzer 2009: 246ff.).

These quantitative analyses must be completed by examinations with regard to the contents in case of a general qualitative judgment. In the end, the lexicographer must decide which words or which lexicographical items should be estimated as important for a dictionary. During this process one has to take into account other criteria such as the dictionary type, characteristics of special field vocabularies or user-specific partial vocabularies. The selection of the lemmata can be based on corpus evaluations (frequency data), but can also be orientated at the lemma lists of other dictionaries.

According to the dictionary type the lemma selection underlies different criteria: (a) In dictionaries where the subject matter is the general vocabulary you need corpora containing the vocabulary of the standard language (e.g. newspaper texts; to the reasons cf. Schierholz 2001: 97f.), but also the general special field vocabulary from those fields which play an important role in the language community. In addition, corpora containing the oral language use must be examined. (b) For a learner's dictionary it is essential that it contains the basic vocabulary. This can be taken over from old lists but should be confirmed by actual frequency investigations of corpora and the existing word lists of linguistic didactics should also be taken into account. (c) In the selection of neologisms the lexicographer must take into account the sources of a neologism. This can be the regional spread and must be differentiated between occasionalism and neologism. (d) In dialect dictionaries the lemma selection must distinguish between the vocabulary in standard language, regional language and a dialect in the broader sense of the region of the survey. (e) In a variety dictionary on a group specific vocabulary a group external lexicographer has the methodological problem to get enough and reliable material for the selection of the lemma candidates. (f) In specialized dictionaries the lemma selection must be done on the basis of specialized texts and expert knowledge. For example in the WSK volume 1.2 (*Wörterbücher zur Sprach- und Kommunikationswissenschaft*) with the subject matter *syntax* (Dürscheid and Schierholz 2013ff.) the lemma selection was processed by (first) scanning, evaluating and comparing the indices and registers from the most important grammars and special field literature (handbooks etc.) to the subject *syntax;* (second) deciding with an

expert judgment which of the terms can be graded as dictionary worthy.

This list is not complete and can be extended in detail for every dictionary type and can also be extended for numerous further dictionary types.

4.1.4 Methods during the phase of material evaluation

The phase of the material evaluation is a phase where the dictionary articles are drawn up on the basis of the collected and prepared material (cf. Wiegand 1998a: 148ff.). The main task during the making of dictionary articles is the writing of single concrete specifications of different item types.

Methods for the formulation of lexicographical items

During the dictionary-making the lexicographer applies a philological method when lexicographical items are selected and formulated. The application of methods is based on the knowledge the lexicographer has from experience (theoretical and practical knowledge of lexicography) and on the data which were found out for the purpose of the current step of the lexicographical work. Further, methodological advice which supports the guarantee of a uniform dealing with the items have to be given in the instruction book with item specific orders. In the microstructural programme of a dictionary, which should have been drawn up during the preparation phase, it is determined which items must be written.

For the selection of the example items it must be decided in principle (a) if the text citations should be taken as citation example items in the original form; (b) if non relevant parts of a text citation can be deleted (which is left at the lexicographer's discretion), (c) if an example item should be formed with the help of the lexicographer's competence, (d) if from case to case a combination of the both procedures can be used — for example in the way that original citations should be preferred, but competence example items are permitted if the number of text citations is too small or if no text example is adequate to be used as an example item.

For the syntactical items of verbs — e.g. it is necessary to choose if a facultative complement should be inserted or left out — the lexicographer has to decide this for each concrete article or each version again.

The writing of the items follows certain methods which are partly borrowed from linguistics, especially if linguistic items are concerned. The number of different lexicographical item classes which occur in dictionaries is estimated to be ca. 1000.[9] The number of methods which is needed during the writing process is smaller because the same method can be applied repeatedly for different items (e.g. to morphology). But this "repeatedly" is not easily and precisely defined and is not described in metalexicography. (a) It can be a method of introspection with a recourse to somebody's own linguistic competence, (b)

it can be a method of copying from other lexicographical (or not lexicographical) reference works (secondary or tertiary sources such as dictionaries, grammar handbooks), (c) it can be a method of systematic corpus query where variant forms of a lemma are collected in order to decide afterwards depending on the citation situation which item giving the form should be taken into the dictionary.

This decision can be done on the basis of somebody's own competence or with statistical procedures which can be applied to the set of the collected citations. Metalexicographers have to discuss in theory and in methodology which method is the best for a certain dictionary type considering the users and the dictionary functions.

It is clear that it is impossible to have the discussion for each item and it is clear that in concrete cases the decisions must be done with practical orientation. For selected item types it was shown in Lexicographica 30 (2014) how the theoretical preconsiderations and the practical work can be combined with regard to single item types.

4.1.5 Methods during the preparation of the publication

In this phase the lexicographer has to do the editorial control (the proofreading) of all written articles (cf. Wiegand 1998a: 149f.), and in an electronic dictionary also the testing of the implemented links, of the multimedia elements etc. and of the whole electronic system (cf. Klosa 2013: 521). To be able to carry this out reliably it would help to have a list of examination methods or a catalogue of proof procedures written down in the instruction book of the dictionary in question.

In addition, the dictionary must be brought into a form which is suitable for the publication, e.g. into a pdf-file where one has to ensure that the result is sufficient for a high quality print.

4.1.6 Methods of publication

A dictionary can be published in one medium or in different forms. Since more than 200 years ago, the print version of a dictionary was the one and only relevant publication form. The production of the concrete product was the task of the printing business so that the lexicographer did not have to pay attention to the methods used in the branch of printing.

Nowadays, especially in online lexicography, the lexicographer can do the publication and can organize the processes independently so that the methods which are necessary for successful electronic publishing are in the present more relevant than in the past. This means one has to consider at least the following things:

— The publication can be a closed version dictionary (a static or completed

dictionary) or a dynamic dictionary or rather, an extension dictionary (cf. Lemberg 2001; Storrer and Freese 1996). (a) In a closed version dictionary all dictionary articles and outer texts must be finished before the dictionary can be published. (b) In an extension dictionary the dictionary articles or parts of the articles which are written can be published, the other articles can follow later step by step. This way of publication has been used in WSK since 2012 when the publisher De Gruyter started with 400 finished articles taken from different WSK-volumes followed by 3000 articles in 2013 and again in 2014. This process will be continued until 2018 and the volume editors (Schierholz and Wiegand 2013ff.) hope that in 2019 the first print version of a WSK volume can be published. (c) If only parts of dictionary articles are published it means that a lemma must not be worked out exhaustively so that frequent and unproblematic readings of a lemma can be written and published first. In this procedure it is also possible to publish at first articles which contain only automatically produced items. The other items can be added later by the lexicographer (cf. Klosa 2013: 522). Whatever opportunity is chosen, different methods will be necessary for this part of the lexicographical work.

— Updated versions can be published in regular or irregular intervals. A regular quarterly revision is published by the online edition of *OED*,[10] and *Duden online*[11] is revised in irregular intervals — which means: every day if necessary. The latter procedure has the advantage that updates can be published immediately so that the users have current access to the newest items in the dictionary and the items can be a recent reflection of reality.

4.1.7 Methods of data maintenance, reprocessing, post-production

While in a closed version dictionary maintenance does not play an important role in the concrete lexicographical process, maintenance is relevant if new editions are planned or updates are done regularly or irregularly. In online dictionaries an update is relatively simple and should be foreseen in the work flow. In such cases the following aspects concerning the methods have to be taken into account:

— Which are the update criteria for the regular inspection of the published articles?

— Will the dictionary basis, the based corpus, be updated too? If the answer is "yes", what are the criteria to start with this and when?

— Do the dictionary users have the opportunity to give feedback? What kind of feedback shall it be (direct, indirect or additional feedback cf. Abel and Meyer 2013; Mann 2010: 34, 41)? How will the feedback be integrated in the dictionary update by the lexicographer?

— Will the storing of the data be done sustainably, e.g. in a format which is not bound to special software?

4.1.8 The influence of superordinated decisions on used methods

In section 4.1.1–4.1.7 the mentioned methods were assigned to the respective phases of the lexicographical process. The choice of the methods which can be used also depends on the superordinated facts and decisions which have an effect to the work in all phases. Here the following important aspects can be enumerated: the use of the computer (cf. also Storrer 1996), the dictionary type and the dictionary function.

The *use of the computer* is nowadays a standard in the lexicographical work; in the dictionary project planning it will be unlikely that a decision between note catalogue and database will be done in favor of the first. But even today, projects exist in which the lexicographer has to work with a handwritten note catalogue, which is the heritage of several generations of lexicographers and in which these notes will be transferred into an electronic database step by step. This was the situation with the Duden publisher 15 to 20 years ago and a little bit later the situation in German academia-projects such as the Grimm dictionary and the Goethe dictionary. But even in dictionary projects where the intensive support of computers is available from the beginning the applied procedures can differ clearly. One reason can be the software used. In dictionary projects with good funding and with an extensive technical know-how of the employees, the software solutions could fit perfectly and could be adapted to project specific needs so that the software can support the methodic procedures very well. In projects with less funding or staff with less training in the technical area one may have to fall back upon simple software products which can support the lexicographical and methodological work in a limited way.

During the entire lexicographical process the *dictionary type* influences the lexicographical procedures and the used methods. With the help of the needs analysis of a certain dictionary, the determination of the individual respective group of users and — following from that — the planned size of a dictionary, the lexicographer can determine guidelines for text compound structure, for the macrostructure (lemma selection, principles of lexicographical order) and for the microstructure (number and sequence of items) of the dictionary articles. The dictionary type determines the microstructure concerning the elements of the whole set of items. It influences also the composition of the items: e.g. the pronunciation rules in a descriptive pronunciation dictionary are presented differently and more precisely than in a general explanatory dictionary; items giving the meaning in a learner's dictionary will be different from those in specialized dictionary etc. These rules influence the possibilities of the use of many lexicographical methods which are necessary for a successful handling with the data.

Similarly, the dictionary type and partly overlapping with it the *dictionary*

functions, influence the lexicographical work and the used methods in all phases. The components of the text compound structure, the lemma selection, the choice and formulation of the items etc. are directed by the dictionary functions (support of text production, text reception, translation and information function) which are orientated to the needs of the envisaged user group. A procedure that is explicitly orientated to dictionary functions must keep an eye on the relevance of the data material in every phase of the lexicographical process. Also the preparation and use of the following results — especially of the lexicographical items — must be done for the support of the dictionary functions. The relevance of all measures should be methodically verified and well-founded.

4.2 Methods in Dictionary Research

Metalexicography, as the theory of practical lexicography, has the task of investigating all methods of practical lexicography and their theoretical reflection. This is based on the principle of interest of insights in the area of metalexicography itself, but also to find possibilities of improvement which should influence the lexicographical practice positively. Metalexicography has the task of drawing up a methodology of lexicography and dictionary research altogether in order to deliver a fundamental contribution to the theory of dictionary research. But the study of methods is in some way depending on the theory from which a method comes from. Metalexicography itself has also some specific methods which play no or only an unimportant role in practical lexicography.

In the following section the methods will be considered in connection with the theory elements of the dictionary research as it was worked out by Herbert Ernst Wiegand.

4.2.1 Systematic dictionary research

All the methods which can be used for the description and investigation of the dictionary structures belong to the *systematic dictionary research*. Wiegand (2010) distinguishes between (a) the methods concerning the investigation of the dictionary form and (b) the methods of the presentation. To (a) belong the methods of text segmentation and of structure constructing; to (b) belong the methods of the presentation of the textual structures, the methods of constructing the article structure schemata a well as the methods of the presentation of typologies.

A central metalexicographical text segmentation method is the *functional-positional segmentation* (cf. to this and to other text segmentation methods Wiegand 2010: 256ff.). This procedure is applied to dictionary articles to find out the number of items and their types. For each element of a dictionary article (phrases, words, parts of words, abbreviations, punctuation marks, blanks etc. with consideration of the typographical marking up of the elements) it can be

detected step by step, if a positional determined element has an independent function or if it is part of a larger element which has its own function. With this procedure one can find out which parts of a dictionary article belong together and which tasks are given to these elements by the lexicographer. It is an intersubjective comprehensible method which is applied in dictionary research but it can also be used to explain uncertainty in the dictionary use. This can occur if the lexicographical items in a dictionary article are condensed in a way that an average user who does not know how to apply this method can not decide which interpretation of a lexicographical item is the right one.

The structure constructing methods are built on the text segmentation methods and are used to describe and investigate the structure and hierarchy which exist under the segmented elements. Wiegand's works are based on the concept of the set theory so that the dictionary form is recorded in terms like *set, set of carriers, relation, element* etc. The reproduction of the construction and the relationship of dictionary components by sets and relations is a method which gives a well-structured presentation of the data which are the constituents of a reference work. Thus many structures can be identified, and the most important ones are the text compound structures, the access structures, the macrostructures, the microstructures, the article constituent structures, the text architectures, the search area structures, the addressing structures, the mediostructures, and the data distribution structures, but there are more. The literature about the single structures is numerous and extensive, a good summary can be found in Engelberg and Lemnitzer (2009: 134ff.), in Wiegand and Fuentes Morán (2010), in the WLWF-1 (2010), and in Gouws, Heid, Schweickard and Wiegand (2013) — the latter one in English. Methods which concern reference structures are worked out in detail in the reference theory (amongst others Wiegand 2011). These procedures were applied in WLWF-1 (2010) and in the WSK so that a successful application in practical lexicography still exists.

Beyond that, even the methods which are used to build up typologies and classifications belong to *systematic dictionary research*. These procedures are borrowed from other sciences, especially from the philosophy of science. Details of the recent methods in *systematic dictionary research* can be found particularly in Wiegand (2010); Wiegand and Fuentes Morán (2010); Wiegand, Beer and Gouws (2013: 39ff.).

4.2.2 Research on Dictionary Use

In *research on dictionary use* it shall be investigated in which way a user uses a dictionary, so that the results can help to improve the quality of a dictionary. This affects print and online dictionaries in the same way.[12] In this research area, methods of empirical social sciences are used, such as observation, written questionnaire, content analysis, test, experiment, protocol on dictionary use or oral comment on dictionary use (cf. Tarp 2009, Lew 2011). Frequently used is the *questionnaire in a written form* (to advantages and disadvantages cf. e.g.

Ripfel and Wiegand 1988: 493; Ripfel 1990: 1632; Tono 2001; Lew 2002). The questionnaire can be distributed personally which was the practice in the past and is nowadays used by single persons or by students when they undertake a pilot study during their studies. But in huge projects a questionnaire is organized online, as it was done with the German online dictionary *elexiko* (cf. Klosa, Koplenig and Töpel 2014). The whole process of planning, realization, evaluation, and critical reflection is described in Koplenig and Müller-Spitzer 2014a: 79ff., Müller-Spitzer 2014c: 85ff., Koplenig and Müller-Spitzer 2014b: 127ff., Müller-Spitzer and Koplenig 2014: 143ff., Koplenig and Müller-Spitzer 2014c: 189ff. The *observation* of a certain person contains the danger that the personal presence of the researcher influences the act of usage (known as *observer's paradox*). When choosing the *eye-tracking* procedure (e.g. Tono 2011; Müller-Spitzer, Michaelis and Koplenig 2014: 209ff.: they call it "Eye-tracking technology") the influence of the researcher is avoided but it must be mentioned that this method allows only to observe the search behavior of the user.[13] After it, the lexicographer has to interpret the eye movements of the test subject and the definite intention the user/subject had cannot be interpreted unambiguously. Besides this, the expenditure of the performance of only a single experiment is enormous (cf. Runte 2015 and briefly also Kemmer 2014: 275).

A specific lexicographical method is the *protocol on dictionary usage* (cf. Ripfel and Wiegand 1988: 494f.; Wiegand 1998a: 974ff.), in which the described procedures can be combined. With the help of a protocol it can be shown in which way a subject uses a dictionary, which acts of usage are successful and which problems arise during the dictionary use in a certain dictionary to certain questions and search actions. The protocol process of the single working steps can be done (a) if the subject gets tasks which make the use of a dictionary necessary (cf. Wiegand 1985); (b) if the subject writes a content analysis, where they document the act of usage. On the one hand, this method requires a certain ability of the test subject, on the other hand it is not unproblematic to make the evaluation objectively; (c) if a *think-aloud-protocol* is made, by articulating during the usage, which action is carried out, why this action is carried out, which unexpected complications occur and how the problems were approached (method of loud thinking). The utterances of the test subject are recorded and analyzed (cf. amongst others Wiegand 1998a: 1010ff.; Hartmann 2001: 118f.).

These procedures can show ways of improving a dictionary, but additionally they give the chance to formulate a hypothesis, which can be the starting point of systematically gathering information (e.g. in experiments or tests).

Online dictionaries can also be investigated by the evaluation of the log files which contain automatically produced protocol data to record the activities of the users.[14] This method can give a relatively good overview of how often and at what time which expressions were looked up, and the log files can also give information about which expressions were never looked up in a reference work. It reveals how users go about using the online dictionary, e.g. which

search expression is entered into the search field. So the lexicographer has the possibility to react to this (cf. for example Rautmann 2014 for *Duden online*). But one has to be careful with the interpretation of the data because the following considerations should be included, for instance: (a) Does a long break between user activities mean that the user needed much time to read the dictionary article? (b) Was the user interrupted by other things, e.g. coffee break? Having (a) it can mean that the article should be designed more comprehensible and more clearly structured, having (b) it means nothing regarding the dictionary. (c) Does the looking up of words from vulgar and sexual vocabulary in a high frequency mean (cf. Docherty 2000: 73; De Schryver and Joffe 2004: 190; Bergenholtz and Johnsen 2005: 126) that this vocabulary part needs special explanation and clarification? (d) Is this kind of search part of reading dictionaries for entertainment or to kill time (which is called "lexicotainment"; cf. Bergenholtz 2011: 16)? It is relatively unlikely to give a reliable interpretation of this situation of usage because beneath the log file data no contact with the dictionary users exists.

4.2.3 Historical Dictionary Research

This section deals with all lexicographical processes in a historical context and pursues the goal to write the history of lexicography and to develop a theory of the history of lexicography (cf. Wiegand 1998a: 10). In this process the conditions of culture and society in different ages and time epochs are involved because they had a major influence on lexicographical activities. Very often it can only be adequately judged retrospectively in which way lexicographical processes were influenced by the social situation. So the methods which are used in sciences of history are important for the historical dictionary research. Beyond that, the construction of typologies of historical dictionaries (cf. Reichmann 2012: 91ff.) and the research on dictionary use in the past play a dominant role.

In the area of practical lexicography all the methods which play a role in the making of historical dictionaries must be taken into consideration.[15] These methods include the use of text corpora, the special conditions in using historical corpora, specific procedures of the lemma selection, the peculiarities of writing and — following from that — the problems with the macrostructure of the lemmata, the special conditions of citation excerption, as well as the characteristics in historical morphology and word formation. During this work, procedures of historical linguistics should be used but also the methods of grapholinguistics and of investigation of language change. Also the mentioned methods of the systematic dictionary research will be used in historical practical lexicography.

The procedures to find out the meaning of words and how to express the paraphrase of meaning are described precisely and extensively by Reichmann (2012: 218ff.) as are historical excerpts which underlie different principles because of a completely different empirical basis than the data basis which is at hand

for the contemporary languages. (Reichmann 2012: 124ff., 472ff.).

4.2.4 Critical Dictionary Research

Critical dictionary research has as the subject matter the entire amount of scientific and non scientific texts which exist in the area of lexicography and dictionary research, as well as all dictionaries because in many cases these are the subject of that entire amount of texts.[16]

This means that the critical dictionary research in a wider sense comprises the texts from every research area (sensu Wiegand 1998a), i.e. works on research of dictionary use, on the production of a dictionary, on the historical dictionary research, on the systematic dictionary research, and texts concerning the total of all dictionaries.

Critical dictionary research in a narrower sense concentrates primarily on the analysis of dictionaries themselves with regard to all properties a dictionary has. This can be done in evaluations of single items or of the entire dictionary and it will be published in dictionary criticisms, monographies, essays or in reviews.

The outcome of this is that in the framework of critical dictionary research, in a wider sense, all methods which occur in the single research areas of lexicography and dictionary research, belong to the subject matter. Even for critical research on dictionaries in a narrower sense, many methods from the mentioned research areas of dictionary research play an important role. The following can be included:

(a) The text segmentation methods and structure constructing methods (section 4.2.1), which are used by dictionary critics in order to find out the existing structures in an available dictionary. These structures can be analyzed and criticized relating to the appropriateness of the dictionary type, dictionary functions, groups of users etc.

(b) Knowledge about the methods of research on dictionary use (section 4.2.2) is relevant because a dictionary researcher must take notice of the research results on dictionary use in a critical way. This is the basis in order to formulate fundamental statements on the benefit of a dictionary for the envisaged users. Besides, investigations on dictionary use which aim to improve existing dictionaries can themselves be counted as critical or self-critical measurements and can also be counted to the critical dictionary research.

(c) Knowledge of methods and insights of historical dictionary research (section 4.2.3) are necessary to categorize the historical dictionaries in case of critics or of a comparison with other contemporary dictionaries in a reliable way.

(d) Philological (qualitative) methods are necessary for the analysis with regard to contents of dictionary texts, e.g. to find out actual or, in cul-

tural retrospective, trendy and ideological influences in meaning paraphrases, outer texts etc. (cf. Haß-Zumkehr 2000 on lexicographical methodics in national socialism).

(e) Quantitative analysis and statistical methods are necessary, for instance, if one wants to give a reasonable estimation of the number of the items in a dictionary. This includes methods on a projection of the lemma number because it is impossible to read the whole dictionary or to count all lemmata. Instead of this a sample should be taken from the dictionary, which takes into consideration the whole word list from A to Z equally. The process to do this in a reliable way is worked out among others in Wiegand (1990: 2127) and in Mann (2013: 745ff.).

(f) Another method is to count a certain lemma stretch (e.g. the stretch *I)* and to extrapolate by the well-known relations of the individual lemma stretches to each other how many lemmata the entire dictionary contains. On the basis of the results, a judgment about the macrostructural coverage of the vocabulary of the dictionary can be given. This can be set in relation to the population of a language, a variety or a corpus.

If a certain item type shall be checked critically, similar prerequisites and difficulties as for the lemma counting exist concerning the sample. It is not possible to make a meaningful analysis about a dictionary by reading or examining a few pages of the dictionary in an unsystematic way and extrapolating to the properties in the complete work. It is not possible to take the first 10 or 20 pages of a dictionary as a basis for an overall assessment of this dictionary. A possible and systematical evaluation method is for instance to start on the second page of the word list, to take from that page the first dictionary article into the sample and to repeat this procedure on every 20th page. By doing this, one collects data systematically to create a sample that is equally distributed over the entire word list. Afterwards this procedure should be repeated starting with the 12th page. Through this you will get two samples of the same size so that all analyses which will be done can be compared and can be checked with regard to their reliability. This method was chosen by Schierholz (1998; 2002) when morphological items in two learner's dictionaries of German were analyzed and compared with the data of text corpora of standard German.

The *dictionary evaluation* is an essential part of the critical dictionary research.[17] In principle, different text types must be distinguished (cf. also Engelberg and Lemnitzer 2009: 186ff.), which can be used to write such an evaluation: (a) the expert discussion (in conferences, in education, in panels etc.); (b) the dictionary review in a journal or in the press; (c) the online discussion in a newsgroup; (d) the systematic and comparing inspection of detailed questions; (e) the critical and constructive investigation of an entire dictionary carried out from different perspectives and with different formulations of a question. This was successfully performed in the area of learner's lexicography

for the German dictionaries of Langenscheidt (LGwDaF) and De Gruyter (dGwDaF) in the collection of essays on pedagogical lexicography by Wiegand (1998b; 2002b). In the contributions the correctness, the reliability, the completeness of lexicographical items, the appropriateness of the lemma selection, the textual structures of the dictionary articles and of the dictionary, the text compound structure, the reference practice, as well as the user friendliness were analyzed. Because in a learner's dictionary the group of users is defined very clearly it is easy to check if the intended functions were fulfilled and to make pedagogically motivated suggestions for the improvement of a new dictionary edition.

In the framework of an evaluation in any case the viewpoint of the researcher must be reflected critically because it there is a difference between a dictionary being evaluated from the perspective of an expert or from the perspective of a user who is the intended target of a dictionary. This can be demonstrated by two examples: (a) Many users of the Langenscheidt dictionaries think positively about the coloured typeface used for the lemma sign, but many metalexicographers do not agree on this.[18] (b) No metalexicographer has a problem identifying in the dictionary article to "singen" (cf. Fig. 1) the abbreviation "e-e" which was used in the first editions of LGwDaF (1994, 2003). The abbreviation "e-e" is "eine", but foreign-language students and also German students asked more than one time in seminars what the "e-e" would mean.

> **sin•gen;** *sang, hat gesungen;* Vt/i **1 (etw.) s.** e-e Melodie od. ein Lied mit der Stimme produzieren <ein Lied s.; falsch, richtig, laut, leise. schön, gut s.; nach Noten/vorn Blatt s.; solo. Sopran s.>: *Weihnachtslieder s.* ‖ K-: ***Sing-, -stimme, -weise;*** Vt **2 j-n in den Schlaf s.** leise s. (1), bis *bes* ein Kind einschläft; Vi **3** beruflich od. als Hobby regelmäßig s. (1): im Kirchenchor, am Theater s. **4** *gespr; mst* vor der Polizei ein Verbrechen gestehen […]

Figure 1: Dictionary article for the lemma "singen" in LGwDaF (2003).

To check these questions, corpus investigations, analysis of the literature and reference works, especially dictionaries, must be done. When assessing dictionaries the critical interpretation of the data is always placed at the end. This is in any case a philological method which is based on the individual subjective judgment ability.

5. Conclusion

This overview should have made clear that methods must be used regularly in lexicographical practice and in theory if recent and reliable dictionaries shall be produced, analyzed or improved. For this, the scientific, economic, technologi-

cal, structural and especially thematic aspects have to be considered.

The overview does not claim completeness referring to the number of methods or the description of methods or the publications on methods.[19] But it should be the starting point for further investigation into methods and methodology in lexicography and dictionary research because the reliable application of a reliable method will support the advancement of our activities in future.

6. Notes

1. Critical aspects to the Function Theory are found amongst others in Bogaards (2010), Rundell (2012: 63), and Swanepoel (2015). A general application to lexicography of language for special purposes can be found in Mihindou (2013: 112ff.).
2. Schaeder 1981: 29. In German: "[…] als einzelne wissenschaftliche Verfahren bezeichnet werden, die dazu dienen, gesicherte und nachprüfbare Erkenntnisse über wissenschaftliche Objekte bzw. Gegenstände zu gewinnen".
3. Original in German: "Eine Methode ist eine geordnete Menge von Handlungsanweisungen, deren Befolgung in mindestens einer geordneten Reihenfolge und unter Beachtung aller methodenzugehöriger Korrektheitsbedingungen erfolgen muss, damit das gewünschte Ergebnis erhältlich ist."
4. To the utilization of corpora in lexicography cf. Gouws, Heid, Schweickard and Wiegand 2013: 1336ff.
5. Wiegand 1984: 559: "Bei der lexikographischen Tätigkeit wird, und zwar determiniert durch den jeweiligen Typ des zu erarbeitenden Nachschlagewerkes, auf Ergebnisse, Methoden und Theorien aus verschiedenen Wissenschaften zurückgegriffen." Cf. also Wiegand 1998a, 100.
6. The method *dictionaries from dictionaries* ("Wörterbücher-aus-Wörterbüchern-Methode") is shown by Wiegand (1998b: 649) within dictionaries from the 16th century. Meyer and Gurevych (2014) demonstrate that in 2013 even the sources of *Wiktionary* which is a collaborative produced dictionary can be found in other dictionaries.
7. The allocation of the phases follows basically Wiegand (1998a), which is adopted by Engelberg and Lemnitzer (2009), Klosa (2013) and can be found in the WLWF-I (201: 8-21).
8. This term is a synonym to "manual" or "textbook on dictionary making" (cf. Hartmann 2013: 600) and it is a translation of the German term "Instruktionsbuch" which Schierholz and Wiegand use in their conception of the WSK project (Schierholz and Wiegand 2004: 205ff.)
9. Wiegand (2005: 344ff.) enumerates more than 1.500 item classes, including many synonyms.
10. Cf. http://public.oed.com/the-oed-today/recent-updates-to-the-oed/ [2015-06-21].
11. Cf. http://www.duden.de/woerterbuch [2015-06-21].
12. Cf. the critical perspective on user research in Bergenholtz and Bergenholtz 2011.
13. Kemmer (2014) who investigates the reception of illustrations and the items giving the meaning in online dictionaries combines the user questionnaire and the eye-tracking method.
14. Cf. Bergenholtz and Johnson (2013). Another useful application of log file data, not concerning the research on dictionary use, is shown by Koplenig, Meyer and Müller-Spitzer (2014).
15. Methods used in different languages can be found in Gouws et al. 2013: 612ff.
16. Cf. Ripfel (1989).
17. Cf. Kemmer (2010); Swanepoel (2008; 2013).

18. The normal users mentioned here were students in different seminars hold at the Friedrich-Alexander University of Erlangen-Nürnberg, the metalexicographers are colleagues who mentioned this in special field conversation in some informal situations.
19. Cf. the selected bibliography in Mann and Schierholz 2014: 33-57. The bibliography of this article contains supplements especially of publications in English.

7. Bibliography

Abel, Andrea and Christian M. Meyer. 2013. The Dynamics outside the Paper: User Contributions to Online Dictionaries. Kosem, Iztok et al. 2013. *Electronic Lexicography in the 21st Century: Thinking Outside the Paper. Proceedings of the eLex 2013 Conference, 17–19 October 2013, Tallinn, Estonia:* 179-194. Ljubljana/Tallinn: Trojina, Institute for Applied Slovene Studies/Eesti Keele Instituut. Online: http://eki.ee/elex2013/proceedings/eLex2013_13_Abel+Meyer.pdf [2015-06-21].

Albert, Ruth and Cor J. Koster. 2002. *Empirie in Linguistik und Sprachlehrforschung. Ein methodologisches Arbeitsbuch.* Tübingen: Gunther Narr.

Albert, Ruth and Nicole Marx. 2010. *Empirisches Arbeiten in Linguistik und Sprachlehrforschung. Anleitung zu quantitativen Studien von der Planungsphase bis zum Forschungsbericht.* Tübingen: Gunther Narr.

Beltrami, Pietro G. 2013. Theory of Dictionary Management. Gouws, Rufus H., Ulrich Heid, Wolfgang Schweickard and Herbert Ernst Wiegand (Eds.). 2013: 524-530.

Bergenholtz, Henning. 1989. Probleme der Selektion im allgemeinen einsprachigen Wörterbuch. Hausmann, Franz Josef, Oskar Reichmann, Herbert Ernst Wiegand and Ladislav Zgusta (Eds.). 1989: 772-779.

Bergenholtz, Henning. 1995. Lemma Selection. Bergenholtz, Henning and Sven Tarp (Eds.). 1995. *Manual of Specialised Lexicography. The Preparation of Specialised Dictionaries*: 98-104. Amsterdam/Philadelphia: John Benjamins.

Bergenholtz, Henning. 2011. Access to and Presentation of Needs-adapted Data in Monofuntional Internet Dictionaries. Fuertes-Olivera, Pedro A. and Henning Bergenholtz (Eds.). 2011: 30-53.

Bergenholtz, Henning and Inger Bergenholtz. 2011. A Dictionary is a Tool, a Good Dictionary is a Monofunctional Tool. Fuertes-Olivera, Pedro A. and Henning Bergenholtz (Eds.). 2011: 187-207.

Bergenholtz, Henning and Mia Johnsen. 2005. Log Files as a Tool for Improving Internet Dictionaries. *Hermes. Journal of Linguistics* 34: 117-141.

Bergenholtz, Henning and Mia Johnsen. 2013. User Research in the Field of Electronic Dictionaries: Methods, First Results, Proposals. Gouws, Rufus H., Ulrich Heid, Wolfgang Schweickard and Herbert Ernst Wiegand (Eds.). 2013: 556-568.

Bergenholtz, Henning and Gregor Meder. 1998. Die äußere Selektion in *Langenscheidts Großwörterbuch Deutsch als Fremdsprache*. Wiegand, Herbert Ernst (Ed.). 1998b: 285-296.

Bergenholtz, Henning and Joachim Mugdan. 1990. Formen und Probleme der Datenerhebung II: Gegenwartsbezogene synchronische Wörterbücher. Hausmann, Franz Josef, Oskar Reichmann, Herbert E. Wiegand and Ladislav Zgusta (Eds.). 1990: 1611-1625.

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

Bielińska, Monika. 2014. Methoden der Übersetzungswissenschaft und der zweisprachigen Lexikographie. *Lexicographica* 30: 213-246.

Bocorny Finatto, Maria José. 2014. New Methods for Specialized Lexicography: Brazilian Approach Examples. *Lexicographica* 30: 247-261.

Bogaards, Paul. 2010. Lexicography: Science without Theory? De Schryver, Gilles-Maurice (Ed.). 2010. *A Way with Words: Recent Advances in Lexical Theory and Analysis. A Festschrift for Patrick Hanks*: 313-322. Kampala: Menha Publishers.

Bortz, Jürgen and Nicola Döring. 2006. *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler.* 4th revised edition. With 156 Figures and 87 Tables. Heidelberg: Springer.

Burch, Thomas, Ruth Christmann, Vera Hildenbrandt and Thomas Schares. 2000. Ein "Hausbuch" für alle? — Das Deutsche Wörterbuch von Jacob und Wilhelm Grimm auf CD-ROM und im Internet. Braungart, Georg, Karl Eibl and Fotis Jannidis (Eds.). 2000. *Jahrbuch für Computerphilologie 2*: 11-34. Paderborn: mentis. Also online: http://computerphilologie.digital-humanities.de/jg00/christma/christma.html [2014-04-08].

Deppermann, Arnulf. 2008. *Gespräche analysieren. Eine Einführung in konversationsanalytische Methoden.* Qualitative Sozialforschung 3. 4. Aufl. Wiesbaden: VS Verlag für Sozialwissenschaften.

De Schryver, Gilles-Maurice and David Joffe. 2004. On How Electronic Dictionaries are Really Used. Williams, Geoffrey and Sandra Vessier (Eds.). 2004. *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004, Lorient, France, July 6–10, 2004*: 187-196. Lorient: Faculté des Lettres et des Sciences Humaines, Université de Bretagne Sud. Also online: http://tshwanedje.com/publications/euralex2004-LOGS.pdf [2014-04-10].

dGWDaF (2000) = Kempcke, Günter (Ed.). 2000. *De Gruyter Wörterbuch Deutsch als Fremdsprache*. Berlin/New York: Walter de Gruyter.

Diekmann, Andreas. 2007. *Empirische Sozialforschung: Grundlagen, Methoden, Anwendungen*. Vollständig überarb. und erw. Neuausgabe 2007, 7. Aufl., (24. Aufl. der Gesamtausgabe). Reinbek bei Hamburg 2013.

Dornseiff, Franz. 2004. *Der deutsche Wortschatz nach Sachgruppen. 8., völlig neu bearb. und mit einem vollständigen alphabetischen Zugriffsregister versehene Auflage von Uwe Quasthoff. Mit einer lexikographisch-historischen Einleitung und einer ausgewählten Bibliographie zur Lexikographie und Onomasiologie von Herbert Ernst Wiegand.* Berlin/New York: Walter de Gruyter.

Drosdowski, Günther. 1977. Nachdenken über Wörterbücher: Theorie und Praxis. Drosdowski, Günther, Helmut Henne and Herbert E. Wiegand (Eds.). 1977. *Nachdenken über Wörterbücher*: 103-143. Mannheim/Wien/Zürich: Bibliographisches Institut, Dudenverlag.

Dürscheid, Christa and Stefan J. Schierholz (Eds.). 2013ff. *Grammatik. Ein Konsultationswörterbuch mit englischen Übersetzungen (WSK-Online-Version).* Berlin/New York: De Gruyter. [http://www.degruyter.com/view/db/wsk] [2015-06-21].

Engelberg, Stefan and Lothar Lemnitzer. 2009. *Lexikographie und Wörterbuchbenutzung*. 4., überarb. und erw. Aufl. Tübingen: Stauffenburg.

Evert, Stefan. 2005. *The Statistics of Word Cooccurrences. Word Pairs and Collocations.* Dissertation. Stuttgart: Institut für maschinelle Sprachverarbeitung, University of Stuttgart. Also online: http://elib.uni-stuttgart.de/opus/volltexte/2005/2371/ [2014-04-08].

Fuertes-Olivera, Pedro A. and Henning Bergenholtz (Eds.). 2011. *e-Lexicography. The Internet, Digital Initiatives and Lexicography.* London/New York: Continuum.

Geyken, Alexander. 2014. Methoden bei der Wörterbuchplanung in Zeiten der Internetlexikographie. *Lexicographica* 30: 77-111.

Gouws, Rufus H., Ulrich Heid, Wolfgang Schweickard and Herbert Ernst Wiegand (Eds.). 2013. *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography.* Handbücher zur Sprach- und Kommunikationswissenschaft 5.4. Berlin/Boston: Mouton de Gruyter.

Gouws, Rufus H. and Daniël J. Prinsloo. 2005. *Principles and Practice of South African Lexicography.* Stellenbosch: SUN PReSS.

Hartmann, Reinhard Rudolf Karl. 2001. *Teaching and Researching Lexicography.* Applied Linguistics in Action Series. Harlow: Longman/Essex: Pearson Education.

Hartmann, Reinhard R.K. 2013. Aids in Metalexicographic Research. Gouws, Rufus H., Ulrich Heid, Wolfgang Schweickard and Herbert Ernst Wiegand (Eds.). 2013: 596-611.

Hauser, Ralf and Angelika Storrer. 1996. Probleme und Lösungen beim Parsen von Wörterbüchern. Feldweg, Helmut and Erhard W. Hinrichs (Eds). 1996. *Lexikon und Text. Wiederverwendbare Methoden und Ressourcen zur linguistischen Erschließung des Deutschen.* Lexicographica. Series Maior 73: 53-67. Tübingen: Niemeyer.

Hausmann, Franz Josef, Oskar Reichmann, Herbert E. Wiegand and Ladislav Zgusta (Eds.). 1989. *Wörterbücher. Ein internationales Handbuch zur Lexikographie/Dictionaries. An International Encyclopedia of Lexicography/Dictionnaires. Encyclopédie internationale de lexicographie. Handbücher zur Sprach- und Kommunikationswissenschaft 5.1.* Berlin/New York: Walter de Gruyter.

Hausmann, Franz Josef, Oskar Reichmann, Herbert E. Wiegand and Ladislav Zgusta (Eds.). 1990. *Wörterbücher. Ein internationales Handbuch zur Lexikographie/Dictionaries. An International Encyclopedia of Lexicography/Dictionnaires. Encyclopédie internationale de lexicographie. Handbücher zur Sprach- und Kommunikationswissenschaft 5.2.* Berlin/New York: Walter de Gruyter.

Heid, Ulrich. 2008. Corpus Linguistics and Lexicography. Lüdeling, Anke and Merja Kytö (Eds.). 2008. *Corpus Linguistics. An International Handbook. Vol. 1.* Handbücher zur Sprach- und Kommunikationswissenschaft 29.1: 131-153. Berlin/New York: Walter de Gruyter.

Hirschfeld, Ursula and Eberhard Stock. 2014. Wie kommt die Aussprache ins (Aussprache-) Wörterbuch? Methoden, Probleme und Ergebnisse normphonetischer Untersuchungen zur deutschen Standardaussprache. *Lexicographica* 30: 262-290.

Kemmer, Katharina. 2010. Onlinewörterbücher in der Wörterbuchkritik. Ein Evaluationsraster mit 39 Beurteilungskriterien. *Online publizierte Arbeiten zur Linguistik* 2: 1-33.

Kemmer, Katharina. 2014. Rezeption der Illustration, jedoch Vernachlässigung der Paraphrase? Müller-Spitzer, Carolin (Ed.). 2014a: 251-278.

Klosa, Annette. 2013. The Lexicographical Process (with Special Focus on Online Dictionaries). Gouws, Rufus H., Ulrich Heid, Wolfgang Schweickard and Herbert Ernst Wiegand (Eds.). 2013: 517-524.

Klosa, Annette, Alexander Koplenig and Antje Töpel. 2014. Benutzerwünsche und -meinungen zu dem monolingualen deutschen Onlinewörterbuch *elexiko*. Müller-Spitzer, Carolin (Ed.). 2014a: 281-384.

Koplenig, Alexander. 2014. *Empirical Research into Dictionary Use.* Müller-Spitzer, Carolin (Ed.). 2014a: 55-76.

Koplenig, Alexander, Peter Meyer and Carolin Müller-Spitzer. 2014. Dictionary Users Do Look Up Frequent Words. A Log File Analysis. Müller-Spitzer, Carolin (Ed.). 2014a: 229-249.

Koplenig, Alexander and Carolin Müller-Spitzer. 2014a. The First Two International Studies on Online Dictionaries — Background Information. Müller-Spitzer, Carolin (Ed.). 2014a: 79-84.

Koplenig, Alexander and Carolin Müller-Spitzer. 2014b. General Issues of Online Dictionary Use. Müller-Spitzer, Carolin (Ed.). 2014a: 127-141.

Koplenig, Alexander and Carolin Müller-Spitzer. 2014c. Question of Design. Müller-Spitzer, Carolin (Ed.). 2014a: 189-204.

Landau, Sidney I. 1989. *Dictionaries. The Art and Craft of Lexicography.* Cambridge: Cambridge University Press. [First print New York: Scribner, 1984.]

Lemnitzer, Lothar and Heike Zinsmeister. 2006. *Korpuslinguistik. Eine Einführung*. Tübingen: Narr.

Lew, Robert. 2002. Questionnaires in Dictionary Use Research: A Reexamination. Braasch, Anna and Claus Povlsen (Eds.). 2002. *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002, Copenhagen, Denmark, August 13–17, 2002. Volume 1*: 267-271. Copenhagen: Center for Sprogteknologi, University of Copenhagen.

Lew, Robert. 2011. User Studies: Opportunities and Limitations. In: Akasu, Kaoru and Satoru Uchida (Eds.). 2011. *ASIALEX 2011 Proceedings. LEXICOGRAPHY: Theoretical and Practical Perspectives. Papers Submitted to the Seventh ASIALEX Biennial International Conference. Kyoto Terrsa, Kyoto, Japan, August 22–24, 2011*: 7-16. Kyoto: Asian Association for Lexicography.

LGwDaF (1994) = Götz, Dieter, Günther Haensch and Hans Wellmann (Eds.). 1994. *Langenscheidts Großwörterbuch Deutsch als Fremdsprache. Das neue einsprachige Wörterbuch für Deutschlernende.* Third Edition. Berlin/Munich: Langenscheidt.

LGwDaF (2003) = Götz, Dieter, Günther Haensch and Hans Wellmann (Eds.). 2003. *Langenscheidts Großwörterbuch Deutsch als Fremdsprache.* Revised edition. Berlin [etc.]: Langenscheidt.

Mann, Michael. 2010. Internet-Wörterbücher am Ende der "Nullerjahre": Der Stand der Dinge. Eine vergleichende Untersuchung beliebter Angebote hinsichtlich formaler Kriterien unter besonderer Berücksichtigung der Fachlexikographie. *Lexicographica* 26: 19-45.

Mann, Michael. 2013. German II: Synchronic Lexicography. Gouws, Rufus H., Ulrich Heid, Wolfgang Schweickard and Herbert Ernst Wiegand (Eds.). 2013: 742-816.

Mann, Michael and Stefan J. Schierholz. 2014. Methoden in der Lexikographie und Wörterbuchforschung. Ein Überblick mit einer Auswahlbibliographie. *Lexicographica* 30: 3-57.

Meindl, Claudia. 2011. *Methodik für Linguisten. Eine Einführung in Statistik und Versuchsplanung.* Tübingen: Narr.

Meyer, Christian M. and Iryna Gurevych. 2014. Methoden bei kollaborativen Wörterbüchern [Methods in Collaborative Dictionaries/Méthodes dans le domaine des dictionnaires collaboratifs]. *Lexicographica* 30: 187-212.

Mihindou, Guy-Roger. 2013. LSP Lexicography in the Academic Discourse. Botha, Willem, Paul Mavoungou and Dion Nkomo (Eds.). 2013. *Festschrift Rufus H. Gouws*: 111-126. Stellenbosch: SUN PReSS.

Müller-Spitzer, Carolin (Ed.). 2014a. *Using Online Dictionaries.* Lexicographica. Series Maior 145. Berlin/Boston: De Gruyter.

Müller-Spitzer, Carolin. 2014b. Methoden der Wörterbuchbenutzungsforschung. *Lexicographica* 30: 112-151.

Müller-Spitzer, Carolin. 2014c. Empirical Data on Contexts of Dictionary Use. Müller-Spitzer, Carolin (Ed.). 2014a: 85-126.

Müller-Spitzer, Carolin and Alexander Koplenig. 2014. Online Dictionaries: Expectations and Demands. Müller-Spitzer, Carolin (Ed.). 2014a: 143-188.

Müller-Spitzer, Carolin, Frank Michaelis and Alexander Koplenig. 2014. Evaluation of a New Web Design For the Dictionary Portal OWID. Müller-Spitzer, Carolin (Ed.). 2014a. 207-228.

Piotrowski, Michael. 2012. *Natural Language Processing for Historical Texts.* Synthesis Lectures on Human Language Technologies 17. San Rafael, CA: Morgan & Claypool.

Quasthoff, Uwe. 2004. Methodologische Einführung. Dornseiff, Franz. 2004: 193*-206*.

Rautmann, Karin. 2014. *Duden online* und seine Nutzer. Abel, Andrea and Annette Klosa (Eds.). 2014. *Der Nutzerbeitrag im Wörterbuchprozess. 3. Arbeitsbericht des wissenschaftlichen Netzwerks "Internetlexikografie"*: 49-62. Mannheim: Institut für Deutsche Sprache. (OPAL — Online publizierte Arbeiten zur Linguistik 4/2014), Online: http://pub.ids-mannheim.de/laufend/opal/opal14-4.html [2014-05-28].

Reichmann, Oskar. 1990. Formen und Probleme der Datenerhebung I: Synchronische und diachronische historische Wörterbücher. Hausmann, Franz Josef, Oskar Reichmann, Herbert Ernst Wiegand and Ladislav Zgusta (Eds.). 1990: 1588-1611.

Reichmann, Oskar. 2012. *Historische Lexikographie. Ideen, Verwirklichungen, Reflexionen an Beispielen des Deutschen, Niederländischen und Englischen.* Studia Linguistica Germanica 111. Berlin/Boston: Walter de Gruyter.

Ripfel, Martha. 1989. *Wörterbuchkritik. Eine empirische Analyse von Wörterbuchrezensionen.* Lexicographica Series Maior 29. Tübingen: Max Niemeyer.

Ripfel, Martha. 1990. Probleme der Erhebung metalexikographischer Daten. Hausmann, Franz Josef, Oskar Reichmann, Herbert E. Wiegand and Ladislav Zgusta (Eds.). 1990: 1631-1638.

Ripfel, Martha and Herbert Ernst Wiegand. 1988. Empirische Wörterbuchbenutzungsforschung. Wiegand, Herbert Ernst (Ed.): 1986. *Studien zur neuhochdeutschen Lexikographie VI. Volume 2.* Germanistische Linguistik 87–90: 491-520. Hildesheim/New York/Zürich: Georg Olms.

Romary, Laurent and Andreas Witt. 2014. Méthodes pour la représentation informatisée de données lexicales/Methoden der Speicherung lexikalischer Daten. *Lexicographica* 30: 152-186.

Rundell, Michael. 2012. It Works in Practice but Will it Work in Theory? The Uneasy Relationship between Lexicography and Matters Theoretical. Fjeld, R.V. and J.M. Torjusen (Eds.). 2012. *Proceedings of the 15th Euralex International Congress, 7–11 August 2012, Oslo*: 47-92. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo.

Runte, Maren. 2015. *Lernerlexikographie und Wortschatzerwerb.* Lexicographica. Series Maior 150. Berlin/Boston: De Gruyter.

Schaeder, Burkhard. 1981. *Lexikographie als Praxis und Theorie.* Reihe Germanistische Linguistik 34. Tübingen: Niemeyer.

Schierholz, Stefan J. 1998. Die Grammatik der Substantive in *Langenscheidts Großwörterbuch Deutsch als Fremdsprache*. Wiegand, Herbert Ernst (Ed.). 1998. *Perspektiven der pädagogischen Lexikographie des Deutschen. Untersuchungen anhand von* Langenscheidts Großwörterbuch Deutsch als Fremdsprache. Lexicographica. Series Maior 86: 88-103.Tübingen: Max Niemeyer.

Schierholz, Stefan J. 2001. *Präpositionalattribute. Syntaktische und semantische Analysen.* Linguistische Arbeiten 447. Tübingen: Max Niemeyer.

Schierholz, Stefan J. 2002. Die Grammatik der Substantive im *de Gruyter Wörterbuch Deutsch als Fremdsprache*. Wiegand, Herbert Ernst (Ed.). 2002b: 75-90.

Schierholz, Stefan J. 2008a. Einige grundlegende Überlegungen zur Corpuslinguistik. Lenz, Friedrich and Stefan J. Schierholz (Eds.). 2008. *Corpuslinguistik in Lexik und Grammatik*: 1-14. 2. Aufl. Tübingen: Stauffenburg.

Schierholz, Stefan J. 2008b. Corpusbasierte Operationalisierungsstrategien zur Bestimmung von Valenzpartnern. Valentin, Jean-Marie (Ed.). 2008. *Akten des XI. Internationalen Germanistenkongres-*

ses Paris 2005 "Germanistik im Konflikt der Kulturen": 37-48. Bern: Peter Lang. (Jahrbuch für Internationale Germanistik. Series A: Kongressberichte. Bd. 80/4. Empirische Grundlagen moderner Grammatikforschung — Integrative Zugriffe auf Phänomene des Sprachwandels — Lexik und Lexikologie: sprachpolitische Einstellungen und Konflikte — Sprache und Diskurs in den neuen Medien).

Schierholz, Stefan J. 2013. Welches Corpus für welche Untersuchung? Grucza, Franciszek et al. (Eds.). 2013. *Akten des XII. Internationalen Germanistenkongresses Warschau 2010*: 313-321. Jahrbuch für Internationale Germanistik. Series A: Kongressberichte. Bern: Peter Lang.

Schierholz, Stefan J. and Herbert Ernst Wiegand. 2004. Die Wörterbücher zur Sprach- und Kommunikationswissenschaft (WSK). Eine neue Konzeption der linguistischen Fachlexikographie und ihre computergestützte Praxis. *Lexicographica* 20. 2004 [2005]: 164-264.

Schierholz, Stefan J. and Herbert Ernst Wiegand (Eds.). 2013ff. *Wörterbücher zur Sprach- und Kommunikationswissenschaft/Dictionaries of Linguistics and Communication Science. Online-Version.* Berlin/New York: De Gruyter. [http://www.degruyter.com/view/db/wsk] [2015-06-21].

Schierholz, Stefan J. and Eric Windisch. 1991. Der Duden unter der Lupe. Quantitative Analysen zur Makrostruktur des Duden-Universalwörterbuchs. *Lexicographica* 7: 125-168.

Schlobinski, Peter. 1996. *Empirische Sprachwissenschaft.* Wiesbaden: Westdeutscher Verlag.

Scholze-Stubenrecht, Werner. 2002. "Die Auswahl der Einträge ist äußerst beliebig". Warum Jagdherr und Pokémon nicht im Duden stehen. *Sprachwissenschaft* 27: 225-248.

Storrer, Angelika. 1996. Metalexikographische Methoden in der Computerlexikographie. Wiegand, Herbert Ernst (Ed.). 1996. *Wörterbücher in der Diskussion II.* Lexicographica. Series Maior 70: 239-255. Tübingen: Max Niemeyer.

Storrer, Angelika and Katrin Freese. 1996. Wörterbücher im Internet. *Deutsche Sprache* 24(2): 97-153.

Svensén, Bo. 1993. *Practical Lexicography. Principles and Methods of Dictionary-Making.* Oxford/New York: Oxford University Press.

Swanepoel, Piet. 2008. Towards a Framework for the Description and Evaluation of Dictionary Evaluation Criteria. *Lexikos* 18: 207-231.

Swanepoel, Piet. 2013. Evaluation of Dictionaries. Gouws, Rufus H., Ulrich Heid, Wolfgang Schweickard and Herbert Ernst Wiegand (Eds.). 2013: 587-596.

Swanepoel, Piet. 2015. The Design of Morphological/Linguistic data in L1 and L2 Monolingual Explanatory Dictionaries: A Functional and/or a Linguistic Approach? *Lexikos* 25.

Tarp, Sven. 2008. *Lexicography in the Borderland between Knowledge and Non-Knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography.* Lexicographica. Series Maior 134. Tübingen: Max Niemeyer.

Tarp, Sven. 2009. Reflections on Lexicographical User Research. *Lexikos* 19: 275-296.

Tarp, Sven. 2013. Lexicographic Functions. Gouws, Rufus H., Ulrich Heid, Wolfgang Schweickard and Herbert Ernst Wiegand (Eds.). 2013: 460-468.

Tarp, Sven. 2014. Theory-Based Lexicographical Methods in a Functional Perspective: An Overview. *Lexicographica* 30: 58-76.

Tono, Yukio. 2001. *Research on Dictionary Use in the Context of Foreign Language Learning. Focus on Reading Comprehension.* Lexicographica. Series Maior 106. Tübingen: Max Niemeyer.

Tono, Yukio. 2011. Application of Eye-Tracking in EFL Learners' Dictionary Look-up Process Research. *International Journal of Lexicography* 24(1): 124-153.

Töpel, Antje. 2014a. Methoden zur Erstellung von Bedeutungsparaphrasenangaben. *Lexicographica* 30: 291-319.

Töpel, Antje. 2014b. Review of Research into the Use of Electronic Dictionaries. Müller-Spitzer, Carolin (Ed.). 2014a: 13-54.

Wahrig, Gerhard. 1973. *Anleitung zur grammatisch-semantischen Beschreibung lexikalischer Einheiten. Versuch eines Modells.* Linguistische Arbeiten 8. Tübingen: Max Niemeyer.

Warnke, Ingo H. and Jürgen Spitzmüller. 2008. *Methoden der Diskurslinguistik. Sprachwissenschaftliche Zugänge zur transtextuellen Ebene.* Berlin/New York: De Gruyter.

Welker, Herbert Andreas. 2006. *O uso de dicionários: Panorama geral das pesquisas empíricas.* Brasilia: Thesaurus.

Welker, Herbert Andreas. 2013a. Empirical Research into Dictionary Use since 1990. Gouws, Rufus H., Ulrich Heid, Wolfgang Schweickard and Herbert Ernst Wiegand (Eds.). 2013: 531-540.

Welker, Herbert Andreas. 2013b. Methods in the Research of Dictionary Use. Gouws, Rufus H., Ulrich Heid, Wolfgang Schweickard and Herbert Ernst Wiegand (Eds.). 2013: 540-547.

Wiegand, Herbert Ernst. 1984. Prinzipien und Methoden historischer Lexikographie. Besch, Werner, Oskar Reichmann and Stefan Sonderegger (Eds.). 1984. *Sprachgeschichte. Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung*: 557-620. Handbücher zur Sprach- und Kommunikationswissenschaft 2.1. Berlin/New York: De Gruyter.

Wiegand, Herbert Ernst. 1985. Fragen zur Grammatik in Wörterbuchbenutzungsprotokollen. Ein Beitrag zur empirischen Erforschung der Benutzung einsprachiger Wörterbücher. Bergenholtz, Henning and Joachim Mugdan (Eds.). 1985. *Lexikographie und Grammatik. Akten des Essener Kolloquiums zur Grammatik im Wörterbuch, 28–30.6.1984*: 20-98. Lexicographica Series Maior 3. Tübingen: Max Niemeyer.

Wiegand, Herbert Ernst. 1987. Zur handlungstheorethischen Grundlegung der Wörterbuchbenutzungsforschung. Kučera, A., A. Rey, H.E. Wiegand and L. Zgusta (Eds.). 1987. *Lexicographica* 3: 178-227. Tübingen: Max Niemeyer.

Wiegand, Herbert Ernst. 1989. Aspekte der Makrostruktur im allgemeinen einsprachigen Wörterbuch: Alphabetische Anordnungsformen und ihre Probleme. Hausmann, Franz Josef, Oskar Reichmann, Herbert Ernst Wiegand en Ladislav Zgusta (Eds.). 1989: 371-409

Wiegand, Herbert Ernst. 1998a. *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie. Volume 1.* Berlin/New York: Walter de Gruyter.

Wiegand, Herbert Ernst (Ed.). 1998b. *Perspektiven der pädagogischen Lexikographie des Deutschen. Untersuchungen anhand von* Langenscheidts Großwörterbuch Deutsch als Fremdsprache. Lexicographica. Series Maior 86. Tübingen: Max Niemeyer.

Wiegand, Herbert Ernst. 2002a. Altes und Neues zur Mediostruktur in Printwörterbüchern. *Lexicographica* 18: 168-252.

Wiegand, Herbert Ernst (Ed.). 2002b. *Perspektiven der pädagogischen Lexikographie des Deutschen II. Untersuchungen anhand des* de Gruyter Wörterbuchs Deutsch als Fremdsprache. Lexicographica. Series Maior 110. Tübingen: Max Niemeyer.

Wiegand, Herbert Ernst. 2004. Lexikographisch-historische Einführung. Dornseiff, Franz. 2004: 9*-91*. Berlin/New York: Walter de Gruyter.

Wiegand, Herbert Ernst. 2005. Angaben, funktionale Angabezusätze, Angabetexte, Angabestrukturen, Strukturanzeiger, Kommentare und mehr. Ein Beitrag zur Theorie der Wörterbuchform. *Lexicographica* 21: 202-379.

Wiegand, Herbert Ernst. 2010. Zur Methodologie der Systematischen Wörterbuchforschung: Ausgewählte Untersuchungs- und Darstellungsmethoden für die Wörterbuchform. *Lexicographica* 26: 249-330.

Wiegand, Herbert Ernst. 2011. Funktionale Verweisartikelstrecken und Verweisartikelteilstrecken in Fachwörterbüchern vom Typ des Lern- und Konsultationswörterbuchs. *Lexicographica* 27: 299-311.

Wiegand, Herbert Ernst, Sandra Beer and Rufus H. Gouws. 2013. Textual Structures in Printed Dictionaries: An Overview. Gouws, Rufus H., Ulrich Heid, Wolfgang Schweickard and Herbert Ernst Wiegand (Eds.). 2013: 31-73.

Wiegand, Herbert Ernst and Sandra Beer. 2013. Access Structures in Printed Dictionaries. Gouws, Rufus H., Ulrich Heid, Wolfgang Schweickard and Herbert Ernst Wiegand (Eds.). 2013: 110-149.

Wiegand, Herbert Ernst and Mª Teresa Fuentes Morán. 2010. *Estructuras lexicográficas. Aspectos centrales de una teoría de la forma del diccionario.* Colección Lexicografía 2. Granada: Ediciones Tragacanto.

Wiegand, Herbert Ernst and Rufus H. Gouws. 2013. Macrostructures in Printed Dictionaries. Gouws, Rufus H., Ulrich Heid, Wolfgang Schweickard and Herbert Ernst Wiegand (Eds.). 2013: 73-110.

Wiegand, Herbert Ernst and Maria Smit. 2013a. Microstructures in Printed Dictionaries. Gouws, Rufus H., Ulrich Heid, Wolfgang Schweickard and Herbert Ernst Wiegand (Eds.). 2013: 149-214.

Wiegand, Herbert Ernst and Maria Smit. 2013b. Mediostructures in Printed Dictionaries. Gouws, Rufus H., Ulrich Heid, Wolfgang Schweickard and Herbert Ernst Wiegand (Eds.). 2013: 214-253.

WLWF-I (2010) = Wiegand, Herbert Ernst, Michael Beißwenger, Rufus H. Gouws, Matthias Kammerer, Angelika Storrer and Werner Wolski (Eds.). 2010. *Wörterbuch zur Lexikographie und Wörterbuchforschung/Dictionary of Lexicography and Dictionary Research*. *Volume 1, A–C.* Berlin/ New York: Walter de Gruyter.

Wolski, Werner. 1989. Formen der Textverdichtung im allgemeinen einsprachigen Wörterbuch. Hausmann, Franz Josef, Oskar Reichmann, Herbert Ernst Wiegand and Ladislav Zgusta (Eds.). 1989: 956-967.

http://lexikos.journals.ac.za

The Design of Morphological/ Linguistic Data in L1 and L2 Monolingual, Explanatory Dictionaries: A Functional and/or Linguistic Approach?

P.H. Swanepoel, *Departement Afrikaans en Literatuurwetenskap, Unisa, Pretoria, Suid-Afrika (swanepoel1950@gmail.com)*

Abstract: The contributions in Klosa (2013a) provide information on various aspects of the design of morphological data for (German) language dictionaries. Bergenholtz and Gouws (2013), however, reject most of these contributions as sources lexicographers could use for the design of language dictionaries because they are guided by linguistic theory (cf. their critique of Klosa 2013b), instead of Function Theory, and by a misguided application of methods for user research, (cf. their critique of Töpel's 2013 use of the questionnaire as method for user research). The first goal of this article is to provide a critical discussion of Bergenholtz and Gouws's (2013) views on the approach of Function Theory to the theoretical, methodological and practical aspects of the design of morphological/linguistic data in L1/L2 (language) dictionaries. It is argued that the approach of Function Theory provides lexicographers only with a usable overview of the design trends in printed and electronic dictionaries for the selection and presentation of morphological/ linguistic data in dictionaries, but that the theoretical, methodological and practical approach of Function Theory is fraught with difficulties. The strategy Bergenholtz and Gouws (2013) use to debunk the linguistic approach to lexicography is not new; it is a well-known strategy used in the rhetoric of Function Theory. The second goal of this article is therefore to point out some of the general problems with a Function Theory of lexicography for the planning, production and publishing of language dictionaries. Given the problems with the theoretical, methodological and practical aspects of Function Theory, it does not offer a viable alternative to the linguistic approach to the design of morphological/linguistic data in L1/L2 (language) dictionaries. It is argued that linguistic lexicography provides lexicographers with access to a vast body of theoretical, methodological and practical research to support the design of morphological data in L1/L2 (language) dictionaries. Furthermore, it is argued that Function Theory has established the importance of functional variables in the design of (language) dictionaries, but that what lexicography needs now is a truly multidisciplinary approach to lexicography, and not an approach that reduces lexicography to the status of a hand-maiden of another discipline, for example, information science, or of a reductionist Function Theory of lexicography.

Keywords: FUNCTION THEORY OF LEXICOGRAPHY, LINGUISTIC THEORY OF LEXICOGRAPHY, THEORETICAL, METHODOLOGICAL AND PRACTICAL ASPECTS OF

LEXICOGRAPHY, THE DESIGN OF MORPHOLOGICAL DATA, THE DESIGN OF LINGUISTIC DATA

Opsomming: Die ontwerp van morfologiese/linguistiese data in eentalige, verklarende L1- en L2-woordeboeke: 'n Funksionele en/of linguistiese benadering? Die artikels in Klosa (2013a) bied inligting oor verskeie aspekte van die ontwerp van morfologiese data in (Duitse) taalwoordeboeke. Bergenholtz and Gouws (2013) verwerp egter die meeste van die bydraes vir leksikografiese doeleindes omdat hulle steun op linguistiese teorie(ë) (sien hulle kritiek op Klosa 2013b), in plaas van die Funksieteorie, en 'n foutiewe toepassing van metodes vir gebruikersnavorsing (sien hulle kritiek van Töpel 2013 se gebruik van die vraelys as metode vir gebruikersnavorsing). Die eerste doelstelling van hierdie artikel is om 'n kritiese ontleding te gee van Bergenholtz en Gouws (2013) se siening van die Funksieteoretiese benadering tot die teoretiese, metodologiese en praktiese aspekte van die ontwerp van morfologiese/linguistiese data in L1/L2 (taal-) woordeboeke. Daar word aangevoer dat die Funksieteoretiese benadering net vir leksikograwe 'n bruikbare oorsig oor die ontwerprigtings in gedrukte en elektroniese woordeboeke vir die keuse en aanbieding van morfologiese/linguistiese data bied. Verder, dat die gebruik van die Funksieteorie as teoretiese, metodologiese en praktiese basis vir die leksikografie teen allerlei probleme stuit. Die strategie wat Bergenholtz en Gouws (2013) gebruik in hulle kritiek op die linguistiese benadering tot die leksikografie is nie nuut nie, maar 'n bekende strategie van die retorika van Funksieteorie. Die tweede doel van hierdie artikel is dus om sommige van die algemene probleme van die Funksieteorie vir die beplanning, produksie en publikasie van woordeboeke aan te toon. Gegee hierdie probleem met die teoretiese, metodologiese en praktiese aspekte van die Funksieteorie, bied dit nie 'n lewensvatbare alternatief vir die linguistiese benadering tot die ontwerp van morfologiese/linguistiese data in L1/L2 (taal-) woordeboeke nie. Daar word aangevoer dat leksikograwe in die linguistiese leksikografie toegang het tot 'n omvattende bron van teoretiese, metodologiese en praktiese navorsing om die ontwerp van morfologiese/linguistiese data in L1/L2 (taal-) woordeboeke te ondersteun. Verder word aangevoer dat die Funksieteorie die belang van die funksionele veranderings in die ontwerp van (taal-) woordeboeke gevestig het, maar wat nou nodig is, is 'n waarlik multidisiplinêre benadering en nie een waarin die leksikografie gereduseer word tot die dienskneg van 'n ander dissipline, soos die inligtingskunde, of tot 'n reduksionistiese Funksieteorie van leksikografie nie.

Sleutelwoorde: FUNKSIETEORIE VAN LEKSIKOGRAFIE, LINGUISTIESE TEORIE VAN LEKSIKOGRAFIE, TEORETIESE, METODOLOGIESE EN PRAKTIESE ASPEKTE VAN LEKSIKOGRAFIE, DIE ONTWERP VAN MORFOLOGIESE DATA, DIE ONTWERP VAN LINGUISTIESE DATA

1. Introduction

Lexicographers tasked with the design (selection and presentation) of the morphological data (abbreviated to: *MD*) for an explanatory, monolingual first language (abbreviated to: *L1*) or a second/foreign language (abbreviated to: *L2*) dictionary usually have a number of recent resources at their disposal for this endeavour: existing dictionaries, grammars, lexicographic manuals, research

material, corpora etc. The same applies to designers of the MD for a L1/L2 German language dictionary; most articles in Klosa (2013a), for example, focus on various aspects of the lexicographic selection and representation of MD in the planning, compilation and production of such dictionaries.

Bergenholtz and Gouws (2013: 60), however, criticise a number of the contributions on MD in Klosa (2013a) against the tenets and practices of Functional Theory as being

— theoretically misguided, by using a linguistic theory of MD and linguistic definitions of morphological terms as point of departure for the design of MD in dictionaries

— methodologically misguided, by using a typical method of empirical research but without adhering to basic criteria for such methods, viz. that a sample of representative respondents is to be selected randomly to be able to generalise the findings of such research and that the questions posed to respondents should be linguistically interesting and not those dictated by the tenets of FT (cf. Bergenholtz and Gouws 2013: 60-61; and, specifically, their critique of Töpel's 2013 questionnaire methodology)

— practically misguided, as current designs of the MD in dictionaries, especially in the case of printed language dictionaries, do not optimally support the access to and comprehension of MD in these dictionaries (cf. Bergenholtz and Gouws's 2013: 61-70 lengthy discussion of the selection and presentation of MD in printed and electronic dictionaries and where MD can be presented in dictionaries), and because they produce polyfunctional dictionaries.

The FT alternative that Bergenholtz and Gouws (2013: 60-61) propose for the theoretical, methodological and practical aspects of what will be called a linguistic approach to lexicography (abbreviated to: *LL*) can be summarised in the following well-known tenets of the Function Theory of Lexicography (again a term coined for that article, and which is abbreviated to: *FTL*):

— that the design of the MD of L1/ L2 dictionaries should theoretically be determined by the (morphological) information needs which specific types of users may have in specific types of situations or contexts of language use (prime attention being devoted in FT to the communicative functions/ situations of text production, text reception, and translation, and the cognitive function of language learning/acquisition) (cf. Bergenholtz and Gouws 2013: 61; also Tarp 2014: 62);

— (albeit implicitly) that the methodology of both theoretical and practical lexicography should be determined by the FTL approach to issues of theory and methodology; and

— that any dictionary should ideally be monofunctional, that is, support only one function, and provide users with only the data (no more and no less;

> cf. Tarp 2014: 63) they might need for the solution of only one type of communicative or cognitive problem (cf. Bergenholtz and Gouws 2013: 60-61; 70-74) so that MD should be easily accessible for users and that they should be able to comprehend the MD presented.

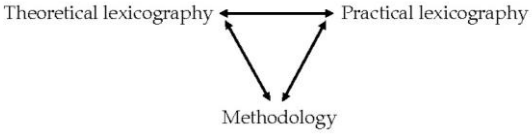
Bergenholtz and Gouws's (2013) critique of LL is not new. It is argued that the FT campaign against the theory, methodology and practical design of MD in LL is in fact only symptomatic of the general effort of FT to discredit LL. The FTL campaign against LL, especially as practised by British lexicographers, has a long, well-known history (cf., for example, Bergenholtz 2011, Gouws 2011 and Leroyer 2011 for overviews). This campaign, however, is also part and parcel of the rhetorical strategies used to promote FTL (cf., for example, various contributions to Fuertes-Olivera and Bergenholtz 2011) — a general strategy that has been criticised in, for example, Lew (2008), Piotrowski (2009), Rundell (2012) and Tono (2010). The main goal of this article, however, is not to repeat these points of critique, but to approach some of the problems of FTL via its approach to the selection and presentation of MD in L1/L2 (language) dictionaries.

The first goal of this article is to provide a critical discussion of Bergenholtz and Gouws's (2013) critique of LL and to present FTL as alternative approach to the theoretical, methodological and practical aspects of MD design in L1/L2 (language) dictionaries and, in a broader perspective, of the design of linguistic data in these dictionaries. It is argued that the FTL approach to the design of MD provides lexicographers (of whatever lexicographic persuasion, and irrespective of the planning, production and publication of printed/paper dictionaries (abbreviated to *p*-dictionaries) or electronic dictionaries (abbreviated to: *e*-dictionaries)) with a usable overview of the design trends in *p*- and *e*-dictionaries. However, it is argued that the theoretical tenets of FTL, the methods used by Bergenholtz and Gouws (2013) in the evaluation of these design trends and in determining users' needs of data on MD in text production, text perception and the acquisition and use of L1/L2, leave much to be desired.

The second goal of this article is to link the theoretical, practical and methodological aspects highlighted in Section 2 to those of FTL in general, and to point out the general problems of FTL for the planning and production of L1/L2 language dictionaries.

In the conclusion of this article it is argued that lexicographers have, in LL, access to a vast body of theoretical, methodological and practical research to support the design of morphological/linguistic data in L1/L2 dictionaries. Furthermore, it is argued that Function Theory has established the importance of functional variables in the design of (language) dictionaries, but that what lexicography needs now is a truly multidisciplinary approach to lexicography, and not an approach that reduces lexicography to the status of a hand-maiden of another discipline, for example, information science, or of a reductionist Function Theory of lexicography.

The discussion of the topics mentioned above can be organised according to three of the pillars of lexicography:



As discussed below, Bergenholtz and Gouws (2013), and Tarp (2004a,b; 2009a,b; 2014), for example, either organise both their presentation and evaluation of LL and its alternative, FTL, according to these three aspects of lexicography, or their discussion can be organised into these three topics.

2. The case against LL

2.1 Bergenholtz and Gouws (2014) on the LL approach to the design of MD in general monolingual dictionaries and FT as alternative

Bergenholtz and Gouws's (2014) exposition of the LL approach to the design of MD can be summarised as follows:

Theoretical approach

— Theoretically, LL uses linguistic terms like *derivative*, *compound*, *prefix*, *affixoid*, etc. as point of departure and focuses on how these MD are currently presented in dictionaries and how their presentation can be improved.

— Dictionary users and functions are occasionally mentioned, but the approach is from linguistics, and the interest in LL is on general information regarding MD.

— Many of the contributions in Klosa (2013a), but also in Barz, Schröder and Fix (2000), regard MD in dictionaries from the perspective of the linguist, a view which is "motivated by the fact that many lexicographers regard lexicography as a subdiscipline of linguistics and therefore want to bring as much from the field of linguistics into the dictionary" (Bergenholtz and Gouws 2013: 61).

According to Bergenholtz and Gouws (2013) the linguistic approach of many of the contributions in Klosa (2013a) is evidenced by the following quotation from Klosa (2013b):

> The presentation of word formation in a dictionary is primarily motivated by the fact that it enables a display of relations and interconnections of words. By means of the inclusion of compounds and derivatives, by means of the lemmatization of affixes and by means of the description of word formation rules in the

> dictionary grammar the interconnectivity of the vocabulary can be successfully indicated although the alphabetical ordering of head words in the dictionary can only display these relations in an inadequate way. In general the usability and effectiveness of dictionaries should be increased by the inclusion of word formation. The reception of items giving word formation can e.g. lead to an enrichment of the vocabulary especially for learners.

As has become customary in critique of LL, Bergenholtz and Gouws (2013) evaluate the theoretical, methodological and practical approach of LL to the design of MD in dictionaries from the tenets of FTL. They (cf. Bergenholtz and Gouws 2013: 59-61) offer the FTL view that the guiding theoretical principles should be the information needs experienced by dictionary users in different types of user situations. This should be followed by a consideration of the ways MD should (practically) be designed to support the envisaged target users of a given dictionary (cf. Bergenholtz and Gouws 2013: 58-59).

This theoretical approach is set within a broader consideration when it comes to the evaluation of competing theories (such as LL and FTL): "The question should be which theory, linguistic or otherwise, is appropriate to ensure that the selection, method and lexicographic presentation for a given dictionary with a specific genuine purpose can be achieved in an optimal way" (Bergenholtz and Gouws 2013: 61).

This theoretical assumption feeds into the questions used methodologically to ascertain users' views of the information needs they experience in different types of user situations.

Methodological approach

The problems Bergenholtz and Gouws (2013) have with the methods used in LL are illustrated with Töpel's (2013) use of the questionnaire methodology. The main problems are a choice of an unrepresentative sample of respondents by way of self-selection (instead of choosing a representative sample by way of random selection), and the wrong kind of questions being asked of respondents (typically the kind of questions one would put to linguists or students of linguistics: Which of the following data categories do you regard as the most important? (Cf. Bergenholtz and Gouws 2013: 61 for a list.))

As alternative to Töpel's (2013) general question on what linguistic data users view as important, they proffer ones that probe the information needs of and their ranking by users in specific user situations: "Which items are extremely important when you use a dictionary as an aid to solve a reception problem? Or: Which items are extremely important when you are learning a language?" (Bergenholtz and Gouws 2013: 61.)

Practical approach

Bergenholtz and Gouws's (2013) presentation and evaluation of the practical design of MD in *p*- and *e*-dictionaries are spread over two large sections of the

article under the heading *2. The presentation of word formation products in current dictionaries* (cf. Bergenholtz and Gouws 2013: 61-70). Their conclusion is that most of these design options (most of those in *e*-dictionaries excluded) do not contribute to, or do not optimise, access to and comprehension of MD.

The evaluation criteria Bergenholtz and Gouws (2013) apply in this case are also dictated by FTL, viz. that the selection and presentation of MD should optimise access to and the comprehensibility of MD. Bergenholtz and Gouws (2013) also accept the assumption of FT that any dictionary should ideally be monofunctional, that is, support only one function, and provide users with only the data (no more and no less; cf. Tarp 2014: 63) they might need for one type of communicative or cognitive problem.

From these two statements of FT, Bergenholtz and Gouws (2013: 61) raise a further problem with the linguistic approach, viz. that it designs and produces dictionaries as "polyfunctional" tools, i.e. to provide help for a number of functions in the same dictionary. This practice leads to information overload for users (cf. also Tarp 2014). As they note: "one can also make, especially in the case of electronic dictionaries, monofunctional information tools, i.e. dictionaries, available to the users" (Bergenholtz and Gouws 2013: 61).

2.2 An evaluation of Bergenholtz and Gouws's (2013) perspective on MD in dictionaries

Despite its simplicity of presentation, there are a number of problems with Bergenholtz and Gouws's (2013) presentation and evaluation of LL and the theoretical, methodological and practical aspects of FTL they present as alternative to LL.

Theoretical aspects

First of all, there is the reductionist presentation of LL. Bergenholtz and Gouws (2013: 61) claim that most lexicographers (of an LL persuasion) regard lexicography as a subdiscipline of linguistics and want to include in the dictionary as much from the wider field of linguistics as possible. However, they do not provide any reference to LL literature by way of justifying this opinion.

Secondly, no (FTL) alternative subtheory of MD is provided in Bergenholtz and Gouws (2013). It is left to the reader to try and build a general FT theory (from rather scant references) and to deduce a subtheory of MD from it.

Thirdly, the same reductionist approach to FT is followed in Bergenholtz and Gouws's (2013) account of the theoretical perspective in Klosa (2013). What is not mentioned is the fact that Klosa's (2013) reference here is to a complex theory of the content, structure and function of the mental lexicon and about its acquisition and use in such tasks as text reception, production and language learning. The theory of the content and structure of the mental lexicon as a network of interconnections between linguistic data associated with the lem-

mas/words of a L1/L2 is well-known, and is also the theoretical foundation of a project such as WordNet (cf. Felbaum 1998). Given that FTL is offered as an alternative to this linguistic-theoretical approach, one surely should expand on this linguistic theory to be able to evaluate the tenets of FTL as a theoretical alternative — a necessary step Bergenholtz and Gouws (2013) do not take.

A brief outline of the psycholinguistic theory of the mental lexicon is ample proof of the fact that it consists of a systematic collection of hypotheses that cover not only the theory of morphological knowledge, but also of its acquisition and teaching.

In the field of psycholinguistics, for example, there is strong, though not unqualified, theoretical and empirical support for the following, closely aligned, hypotheses and assumptions:

— *Morphological awareness* (abbreviated to: *MA*) plays a critical role in users' and learners' acquisition of the vocabulary of a language and word reading, reading comprehension and spelling development (Apel 2014: 197). Apel (2014: 200) proposes the following definition of MA (keeping in mind the proviso that MA is not static but a changing concept which develops over time, and differs according to the age and linguistic proficiency of L1/L2 users):

> Morphological awareness includes: (a) awareness of spoken and written forms of morphemes; (b) the meaning of affixes and the alterations in meaning and grammatical class they bring to base words/roots…; (c) the manner in which written affixes connect to base words/roots, including changes to those words/roots (e.g., doubling or dropping of consonants — PHS)…; and (d) the relation between base words/roots and their inflected or derived forms (e.g., knowing that a variety of words are related because they share the same base word/root … (Apel 2014: 200).

— Via the acquisition of the vocabulary of a language, users' MA plays a determining role in their acquisition and use of L1 and L2 literacy skills (i.e., narrowly defined, their reading, writing, speaking and listening skills in a language (L1 or L2)) (cf. Apel and Werfel 2014).

— MA instruction (with a control group) has some significant effects on one or more measures of MA (e.g. word reading, spelling, morphological analysis of unfamiliar words, etc.) (Carlisle 2010: 480).

— If adequately defined, and given a set of appropriate measures, users' and learners' MA of an L1 or L2 can be measured (and, as an extension to this, that users' and learners' vocabulary acquisition and use can be measured); for example, it is well known and documented with appropriate test items that EFL students have a low accuracy rate in the use of inflectional morphemes (cf. Jiang 2004: 603-604) (Apel 2014 and Apel, Diehm and Apel 2013 discuss a number of the tests tailored to the definition of MA provided by Apel 2014.)

— If necessary, users' and learners' MA can be rectified and improved by both implicit learning and explicit teaching (and acquisition) of the morphology of an L1 or/and an L2 (cf. Bowers and Kirby 2010), for which dictionaries can be used. Templeton (2011/2012) lists methods such as the following for the support of implicit and explicit learning: exercises to build awareness of the morphophonology link (morphology based spelling or the spelling-phonemic-meaning-morpheme link in, for example, *sign/signature*, *bomb/bombard*;) paradigm building (e.g., *courage*, *courageous*, *courageously*; *encourage*, *discourage*, *discouragingly*, *discourageable*); Take-Apart-Build-a-Word exercises (i.e. remove prefixes from words and then add suffixes to the remaining stem/base); word categorization activities to build awareness of the link between category and meaning in a paradigm (such as *bene-* in *benefactor*, *beneficial*, *benefit*, *benevolent*, *benediction*); unravelling the "semantic biography" (etymology) of a word to anchor its meaning.

— Given a set of appropriate instruments, the effects of various instruction techniques and strategies on users' and learners' acquisition and use of L1 and L2 literacy skills can be measured.

Carlisle (2010), for example, spells out the relationship between morphological awareness, phonology, orthography and meaning (or vocabulary) and reading ability, and lists a number of outcomes for the instruction of MA and to measure improvements in reading ability. These outcomes are listed here as they give some idea of the kind of morphological information dictionaries would have to contain to fulfil their cognitive function, i.e. to be a source for the acquisition of MA: learners' and users' recognition of spelling characters, morphemic segmentation, interpreting the meaning of novel words, derivation in sentence context, adding inflections, forming compounds and spelling polymorphemic words, forming and understanding morphological analogies, identifying and subtracting morphemes, spelling derivational prefixes and suffixes, spelling stems in pseudo words, reading pseudo words and reading words with morphological rules (cf. further, for example, Apel 2014, Apel and Werfel 2014, Bowers and Kirby 2010, Carlisle 2010, Carlisle and Stone 2005, Jiang 2004 and Templeton 2011/2012).

Needless to say, FT is presented as a theoretical alternative to LL, and specifically of MD in L1/L2 dictionaries, while it has nothing like the sort of complexity illustrated by the theoretical assumptions of MA in psycholinguistics.

Bergenholtz and Gouws (2013) also do not mention or discuss the fact that in the field of theoretical and practical lexicography, lexicography can rely on a number of resources, of which some are linguistic subtheories of morphology (cf., for example, Bauer, Lieber and Plag 2013) and others, for example, come from the field of computational (morphological) analysis of corpora (cf. for example De Pauw and De Schryver 2008). Schmid, Fitschen and Heid (2004) discuss how a computer program analyses German corpora morphologically,

and Evert and Lüdeling (2001) indicate how the productivity of morphological word forms can be measured automatically in corpus analysis. In all three of these cases, usable and appropriate guidelines for the design of dictionaries are presented or can be deduced.

Bergenholtz and Gouws (2013) uncritically accept and do not discuss the fact that the FLT approach is presented as a *lexicography theory*. There are various strategies one could follow if one does not want to problematize this fact: simply present the tenets of FT as a well-known, accepted and unproblematic concept; do not define the central concepts on which FT rests, such as "theory", "method", "guideline" (for practical lexicography); define the central concepts as vaguely as possible or define them in a reductionist way so that only FT as theory, FT methods and guidelines for dictionary design are acceptable (cf. the discussion below). Bergenholtz and Gouws (2013) opt for the first two options. However, simply presenting FT as an unproblematic concept does not make it so. Anyone could still question its status as a theory. Legitimate questions would be, for example: Is FT a theory at all or simply a set of guidelines for the practical design of dictionaries? Precisely of what is FT a theory, if a theory? and so on. This point is taken up again in Section 3.

Methodological problems

Bergenholtz and Gouws's (2013) critique of the LL/Töpel's (2013) application of the questionnaire as a method focuses, firstly, on the fact that it employs a self-selected, unrepresentative sample of respondents. This, however, is common practice in user research, and manuals of methods warn that researchers must account for this in their interpretation of the research results they generate. This is also a practice that researchers (such as Töpel 2013) adhere to in their discussion of the status of their research findings. To generalize one's research findings to a whole population if they are in fact based on a unrepresentative self-selected sample of respondents is simply bad research practice and as such unacceptable. However, Bergenholtz and Gouws (2013) do not discuss these well-known methodological hedges, but simply reject this kind of research methodology.

The second point is that Bergenholtz and Gouws (2013), in providing a set of alternative FTL inspired questions for user research, are highly misleading. Simply because FTL, by design (cf. Tarp 2014), does not embark on or support such empirical research on users, user needs, usage situations, etc., as proponents of FTL rely on the subjective opinions of lexicographers or trained teachers for this purpose (cf. Tarp 2014 and the discussion in Section 3 below).

Practical problems

Two alternative criteria on which Bergenholtz and Gouws (2013) rely in their critique of LL practice, are that in the design of dictionaries (a) ease of access to data and (b) the comprehensibility of the data, once found, are of prime importance. This methodology is evident in the way Bergenholtz and Gouws

(2013) evaluate the current design features (of MD and linguistic data) of printed and electronic dictionaries.

The first fact to be mentioned is that concern about the access to and comprehensibility of MD/linguistic data in monolingual dictionaries is not confined to FT. Various designs for the content and structure of entries, and their evaluation to improve access and comprehensibility, were among the focus areas in the development of the so-called "Big Five" monolingual dictionaries (cf., for example, Swanepoel 2001).

The second point one must make is that Bergenholtz and Gouws (2013), in their evaluation of the access to and comprehensibility of MD in printed and electronic dictionaries, revert to the FTL practice of substituting users and their experiences with the subjective opinions of lexicographers/linguists/researchers on these issues. This methodology is evident in the way Bergenholtz and Gouws (2013) act on behalf of the user in their analysis of the problems users (could) experience with the selection and presentation of MD in, especially, printed dictionaries. In their analysis of the techniques used to present MD in current monolingual dictionaries, Bergenholtz and Gouws (2013: 61-69) present possible or hypothetical problems users may experience when accessing information on MD, comprehending both the dictionary's selection and presentation policy or comprehending the presented MD itself as if they were the actual experiences of users. A few examples (cf. Bergenholtz and Gouws 2013: 61-69):

— on presenting derivatives and compounds as unexplained run-on lists in articles without indicating which of the polysemous meanings of a lemma apply to them: "There is no way the user will know which sense applies in the different complex forms, i.e. derivatives and compounds"

— on presenting complex forms as (unexplained) run-ons in articles and not also as (fully treated) lemmas in the main word list: "The user is at a loss when having to decide where to find a given compound or derivative: attached to the article of the lemma presenting the first stem (of derivations or compounds — PHS) or included in its own alphabetic position as guiding element of a separate article"

— on presenting linking morphemes in run-on lists with a tilde (to mark the place the lemma would occupy) and the linking morpheme as part of the next morpheme: "This form of textual condensation … makes it difficult for the average user to quickly interpret the compound correctly. If the linking morpheme does not occur in all complex forms the presentation confuses the user even more".

Obviously their commitment to research on the functional variables that should determine the design of dictionaries is nothing but lip service as it does not guide anything in FT.

What is noticeable is the FT requirement that all dictionaries should only provide the data users need for specific functions, not more nor less, as more

information inevitably leads to information overload for users and the abortion of consultations. This view is striking for its static image of the user of dictionaries. No allowances are made in this approach for the fact that users could progress in their acquisition of lexicographic skills, or that users could be strategic users of texts: they could have different (information) goals in approaching texts, different reading and interpretation techniques which they adjust to their goals, etc. They could acquire these strategies very early in their confrontation with various kinds of printed and electronic and verbal and visual texts, and they could activate these strategies and skills when confronted with new (kinds of) texts. There is a significant body of research (too much to list comprehensively) on the acquisition and use of literacy skills to inform research on how users in fact process lexicographic texts.

This view of users as dynamic participants in their confrontation with lexicographic texts also questions the assumption of FTL that if one designs lexicographic texts according to the functional variables discerned in FTL, this ensures that users will receive the data they need and deduce from the data the correct information. Heid (2011), however, illustrates with his research on the usability of lexicographic texts that such assumptions do not align with the findings of empirical research. One could also add that most probably — and this is an assumption for empirical research — there are numerous other variables (other than those distinguished in FTL) that intervene in the complex process in the users' heads from their interpretation of the data presented in dictionaries to what they deduce as information from this data and finally apply to satisfy their information needs in text reception, production and the acquisition and use of L1/L2. One thing that the uses and gratification framework — referred to by Leroyer 2011 as a foundation of FTL — teaches (cf., for example, Pitout 2009: 391-398), is that one should not only approach the question of dictionary design from its complex effects on users, but also from the perspective of what the user, as a dynamic participant, does with lexicographic texts (cf., for example, Pitout 2009: 398-410 on the theoretical assumptions and methodology of reception studies).

If one does, however, overlook this confusion of fact and hypotheses, Bergenholtz and Gouws (2013) do provide a useful list of all the possible ways in which MD can be presented in *p*- and *e*-dictionaries. If one simply lists all the possibilities, lexicographers have a useful tool at their disposal for the selection and presentation of MD, which, however, still has to be tested empirically for effectiveness, or, as Heid (2011) illustrates, for usability. Bergenholtz and Gouws (2013: 61-67), for example, present a list of design options within a dictionary article:

— each complex form as a lemma, with a full treatment of its morphological properties

— as an unexplained run-on list, after the treatment of a lemma, with a tilde as place marker for the lemma, with alphabetic ordering of complex forms,

but with no differentiation of types of complex forms or treatment of the complex forms in the list (if complex forms are taken to be self-explanatory with regard to their meaning)

— as a run-on list in an article with indication of morpho-syntactic/morpho-semantic properties and a combination of an alphabetic and a syntactic/semantic ordering of complex forms' properties

— as a run-on list after the senses of a polysemous lemma have been treated (thus with no indication of what sense of the lemma applies to what complex forms; usually the main sense applies to all complex forms listed, but this is not necessarily so)

— as a run-on list directly after treatment of each sense of the lemma so that it is clear what sense applies to the meaning of a list of complex forms, or by explicitly linking a complex form and the sense of the lemma that applies

— placing of linking morphemes in a run-on list after the tilde (as place holder for the lemma) as following element before the last stem of compounds

— use within an article of a morphological comment

— indication of the variants of a complex lemma either in one article or in separate articles with an indication of the preferred variant

Bergenholtz and Gouws (2013: 67-70) continue the listing of design options utilised in electronic dictionaries and include a section on where else MD can be found in dictionaries (e.g. in outer texts such as the user guide and dictionary grammar).

Much the same methodology is used in Bergenholtz and Gouws (2013: 70-74) in analysing users' needs for MD/linguistic data in text production and reception and L1/L2 acquisition. No empirical research is done to establish users' information needs in these usage situations; instead, they are the result of the subjective opinion of the lexicographer/linguist/researcher.

A further problem is that in Bergenholtz and Gouws (2013: 70-74) the length of the discussion of users' needs is no indication of the depth of analysis of users' need for MD. If one eliminates the long-winded explanations of potential problems users may have with information on MD and analyses of examples, their discussion can be summarized as follows:

— text reception: the meaning of a morphologically complex form; apart from an explanation of the meaning of a complex form (restricted to the products of derivation and compounding), a Google-like function is needed to correct incorrect spelling of complex forms or to suggest alternative spellings;

— text production: information that a complex word exists, how it is spelled, whether or not it has a linking morpheme, what morphological variants a complex form has, what the preferred form is if it has variants, and what irregular word formation processes are involved in the production of a specific complex form. Ideally a dictionary should contain all the complex forms of language as lemmas with their own articles, as including complex forms only as run-ons in articles has a very limited supporting function;

— language learning: users often consult grammars instead of dictionaries for this purpose; dictionaries must contain dictionary grammars which spell out the general, predictable word formation rules used in a language; articles should contain information on the use of word-specific irregular word formation processes; articles and the entries in the dictionary grammar must be cross-referenced; comments with complementary info on word formation can be added to articles; linguistic terms pertaining to morphological phenomena could be included as explained lemmas in the main word list.

The first problem with this list of MD that users may need, is its reductionist nature, especially if compared to Tarp's (2004a,b) discussion of the MD that users of dictionaries need in the case of text production, reception and language acquisition or, for example, Bauer, Lieber and Plag's (2013) in-depth discussion of a wide range of morphological phenomena in English. One does not even get an idea of the selection and presentation of MD in a printed, monolingual LL dictionary such as *The Cambridge Advanced Learner's Dictionary*. This dictionary includes, for example, outer texts with information about prefixes and suffixes, plus regular word formation rules, the formation of irregular verbs, word families and an extensive list of geographical names plus (ir)regular derivation of adjectives from place names and person names from place names. Affixes are also treated as lemmas of the main list with their morphological properties. These design options support Tarp's (2004a,b) contention that polyfunctional printed L2 dictionaries can in fact provide users of the MD with information they may need in text production, reception and L2 language learning.

Bergenholtz and Gouws (2013) only work with the generic labels *derivation* and *compounding*, and working through their article, one has no idea of the complexity of MD that these two generic labels are intended to cover in the design of language dictionaries.

When evaluating Bergenholtz and Gouws's (2013) presentation and evaluation of LL and FTL as an alternative to it, it is evident that their analyses, the alternatives they provide, and their evaluation of theories, methods and practical guidelines are problematic. Their presentation of LL, for example, is utterly reductionist and their evaluations on the basis of FTL criteria are not convincing. The same holds for their exposition of the tenets of FTL: as is argued extensively in the foregoing, the theoretical, methodological and practi-

cal aspects of FTL, as presented in Bergenholtz and Gouws (2013), is fraught with difficulties and as such FTL is not a viable alternative for LL.

In the next section a wider perspective of the problems of FTL is provided.

3. The tenets and rhetorical strategies of FTL: a wider perspective

As was argued in Sections 1 and 2, debunking LL and promoting FTL is not a rhetorical strategy confined to FTL views on the theoretical, methodological and practical aspects of the selection and presentation of MD in general monolingual dictionaries, but a general strategy used in the growing corpus of studies on FTL. In Section 3 this general strategy is outlined and evaluated in more detail. Given the length restrictions on articles, a choice of resources on FTL had to be made. The resources were selected to support the view that FTL underwent a large degree of expansion and precision in its formulation of the theoretical, methodological and practical aspects of FTL, but that these aspects of FTL are also highly problematic.

Reservations

Before outlining the FTL strategy, two important reservations have to be made. The first is that FTL literature in which proponents of FTL debunk LL solely on the basis of the intuitions of lexicographers/linguists/researchers, i.e. without references to relevant literature and empirical analyses to support their views, will be left out in further discussion. A typical example of this approach/method is illustrated in Bergenholtz (2011). A few examples will suffice (cf. Bergenholtz 2011 for an in-depth discussion).

— In LL, linguists "masquerade" as lexicographers (Bergenholtz 2011: 188).

— In LL, linguists regard the polyfunctional dictionary (which contains "as many items as possible in order to satisfy the largest number of different user needs") as *the* dictionary (Bergenholtz 2011: 188).

— In LL, almost "all of those who call themselves lexicographers … are of the view that lexicography is a linguistic discipline" (Bergenholtz 2011: 189).

— In LL, linguists have robbed lexicography of its status as a scientific discipline (Bergenholtz 2011: 189-190).

— In LL, "The structural description of existing dictionaries (which I believe is a reference to Wiegandian scientific studies of dictionary structures — PHS) — became partly a mixture of hard-to-grasp theories and contemplative analyses, which, at best, do not get in the way of future dictionary concepts" (Bergenholtz 2011: 190).

— In LL, "most of the studies of dictionary usage were carried out in the most unscientific way imaginable, as they were conducted without any knowl-

> edge and without use of the methods of the social sciences"; they do not satisfy the two fundamental requirements for scientific surveys, viz. that they must sample a section of the population that "can be statistically viewed as representative of the entire population" and that respondents must be selected on the basis of this principle; surveys do not probe "the real needs of users, only … linguistic phenomena" (Bergenholtz 2011: 190).

To summarise, Bergenholtz (2011: 190) makes the following statement: "… many of the lexicographic proposals put forward during the past 40 years for changing or improving lexicographic practice were often more harmful than useful" (Bergenholtz 2011: 190). Such an approach makes any discussion of its merits or problems impossible.

The second point to be made, is that despite the differences between FTL and LL, both overlap to a large extent in the production of *e*-dictionaries. Of most English L1/L2 language dictionaries, for example, electronic versions already exist, or are in production, and given the fierce competition in the market, it would be suicide if commercial producers of language dictionaries did not optimize the use of computer technologies for this purpose and employ the necessary teams of lexicographers, subject specialists (linguists), and computer experts to do so. Research (within FTL, especially) on the architecture of electronic databases to optimize the production of various dictionary types, user customization, access and comprehensibility of data and macro- and micro-architecture, etc. is therefore of equal importance for modern FTL and for LL (cf., for example, Bothma 2011, Nielsen and Almind 2011, and Spohr 2011).

How FT and LL differ — and this is crucial — is the theory and theory-driven methods used in each of the sub-phases of dictionary planning, production and publication. FT prioritizes the intended functions of a dictionary (cf. Tarp 2004a,b; 2009a,b; 2014) as the major theoretical determinant and methodology of lexicographers'/linguists'/researchers' opinions on all aspects of users that affect the design of FTL e-dictionaries. LL, in contrast, draws on a larger theoretical pool of options for its theoretic foundation and prioritizes (where applicable) the methods of empirical research commonly used in the social sciences. These aspects of FTL are discussed at length in, for example, Tarp (2004a,b; 2009a,b; and 2014), and these sources are used here to explicate and evaluate FTL.

As will be argued below, both FTL and LL are based on the assumption that they have a theoretical component (itself consisting of one or more general and various subtheories) that informs the practical planning, production and publication of dictionaries or lexicographic tools. The methodological component refers to both the methods used to build relevant theories and concepts (that inform lexicography as practice) and the methods, determined or motivated by the theoretical assumptions one uses in the various phases, generally followed in the planning, production and publication of dictionaries/lexicographic works.

The general tenets of FTL

Tarp (2009b: 157-158) distinguishes between general lexicographic theories ("containing general summarizing statements about lexicography", such as FT — PHS) and specific theories (containing statements about the sub-areas of lexicography, e.g. a theory of morphology); between a specific sub-theory which is integrated in a general theory of lexicography and "which helps to enrich the general theory and strengthen its status" (p. 158), and a sub-theory which is not integrated in a general theory of lexicography.

Typical examples of the latter, according to Tarp (2009b: 158), are linguistic theories "from which linguistic concepts, theories, and methods are transferred uncritically to lexicography", i.e. not transformed and not integrated into a general theory of lexicography.

Although Tarp (2009a) rejects non-integrated linguistic theories as candidates for (sub-)theories of lexicography, Tarp (2014) is rather generous with regard to the candidate(s) for the choice of a general lexicographical theory. Tarp (2014) defines the field broadly, in correspondence with what he defines as the research object/subject field of lexicography: dictionaries and all other kinds of reference works. As Tarp (2014: 61) notes, these reference works have traditionally covered "almost every area of human activity and every discipline related to these areas". All concepts, theories and methods of these disciplines, and, one could assume, also the way in which they determined the design of these reference works, are thus potential candidates for the choice of a general theory of lexicography and relevant methods.

However, Tarp (2014) limits the choice of theory-embedded/theory-derived methods by the first general principle of methodology (of FT), which determines that the choice of concepts, theories and methods must have been "subjected to critical analysis with a view to determining what should be rejected, what can be used, and how useful factors can be adjusted and adapted to suit the particular nature of lexicography" (Tarp 2014: 62, quoted from Tarp 2008: 12). Based on this principle, a number of methods of linguistics for the analyses, description and formalization/standardization of language (i.e. linguistic theories and methods) have been evaluated and either rejected on the grounds of or adapted to the tenets of FT (cf. Tarp 2014: 62 for a discussion).

However, the only alternative theory to existing linguistic theories and methods that Tarp (2014: 62) proposes as the candidate for a general theory of lexicography is the well-known, central tenet of FTL:

> The function theory is based upon the axiomatic statement that dictionaries and other lexicographic reference tools are utility tools which are, or should be, designed to attend specific types of punctual information need which specific type of user may have in specific types of situation of context.

As a theory, FT is taken as the yardstick for the choice of concepts, theories and methods because it addresses the "question of solving the concrete types of information needs which different types of users may experience in different

communicative situations" (Tarp 2014: 62). Evaluated from this perspective, linguistic concepts, theories and methods are rejected or adapted because "They are, so to say, not sufficiently focused on the specific tasks to be solved by lexicography (as specified in FT — PHS)" (Tarp 2014: 62).

From FT it also follows that the focus of empirical research for the design of dictionaries would be the types of users, their information/morphological needs and the usage situations in which these needs arise.

According to Tarp (2014: 63), lexicographical functions are the data a dictionary provides for the information needs specific users may have in specific situations of use. These functions, communicative (i.e. text production and reception, translation) and cognitive (i.e., knowledge in general, L1 and L2 acquisition and use in particular), are, according to a second general principle of methodology, the cornerstone of any decision regarding the selection and presentation of (linguistic/morphological) data and users' access to it.

Tarp (2009b: 156) formalizes as a general principle of methodology for the conception and production of dictionaries that it is necessary:

1. to determine the types of categories of data to be included in the dictionary;
2. to present these data in such a way that it is easy for the user to retrieve the needed information;
3. to guarantee quick and easy access to these data; and
4. to ensure that the data are correct.

1-3 require lexicographic principles which allow quick and easy data access for consultation purposes (Tarp 2009b: 160); and 1-3 must be the task of a person with specialized knowledge of lexicography; 4 requires the contribution of an expert in the specific field in question (cf. Tarp 2009a: 156-157).

To the requirement of easy access to data and the requirement that such data be comprehensible, Tarp (2014: 63) introduces by a third general rule of methodology that methods should be selected, designed and adapted "with a view to selecting, preparing and presenting as *little data as possible* to the users in each consultation", i.e. to present users with enough data to solve their information problems, but to avoid an overload of information, or, that dictionaries should provide users with the exact amount of data (a quantitative criterion) and data types (the qualitative criterion).

This restriction arises from the "well-known" fact that data overload in both printed and electronic dictionaries "tends to obstruct both access to the relevant data and retrieval of the needed information" (Tarp 2014: 63). According to Tarp (2014: 63), however, the information overload is unavoidable in polyfunctional printed dictionaries. In contrast, though, it is possible to avoid this problem by producing monofunctional electronic dictionaries (from polyfunctional databases, one must add) by utilising available computer technologies and techniques.

Tarp (2014: 60) explicitly notes that methods "should not only be com-

pared and evaluated in the light of their final result, i.e. whether or not they do lead to the desired goal" but also in terms of "the time it takes to reach this goal and the resources employed". Given, amongst others, these evaluation criteria, Tarp (2014: 61) proffers FT as a theory "capable of guiding the development of goal-focused methods which, with the employment of lesser time and resources than other methods, can lead to the desired results in the form of high-quality online lexicographical works".

In essence, FT is summarized in its central tenet (methodological principle 1), to the exclusion of all other alternative theories; a very restricted approach to research (only on the functional variables determined in the central tenet of FT); and a number of sub-statements (read: guidelines) for the planning and production of dictionaries according to the FTL theory and method. One must also note that these tenets/statements of FT are not only explicated, but also presented and used as criteria for the evaluation of concepts, theories, methods and design guidelines of any alternatives to the FT.

Of importance here is also the fact that, although not as clearly systematized and elaborated, the central tenet of FT outlined above is essentially the concept of FT as presented in Bergenholtz and Gouws (2013) and used as evaluation (and rejection) of LL theory, method and practical design and production of dictionaries.

A critical assessment of the theoretical, methodological and practical assumptions of FT

As mentioned in Section 2, Bergenholtz and Gouws (2013) uncritically accept and do not discuss the fact that FT is presented as a *lexicography theory*. Tarp (2014), on the other hand, discusses and tries to refute in this regard two claims made against FT: (1) the assumption that lexicography is a craft and as such does not have a theoretical foundation (often ascribed in FTL literature to Atkins and Rundell 2008), and (2) the claim that FT is not a theory because it is not formulated as a set of falsifiable claims/hypotheses (cf. in this regard especially Piotrowski 2009 and Tono 2010).

Given both the (linguistic-)theoretical foundation of LL and the claims of FTL about its theoretical foundation, the first claim — amended, in Rundell 2012 — will not be discussed further.

Tarp's (2014) refutation of the second claim, viz. that FT is not a theory because it is not presented as a set of falsifiable statements (as should be evident from its tenets provided above) is to reject it as an inapplicable criterion for disciplines such as lexicography, and to present an alternative set of criteria from Fuertes-Olivera and Tarp (2014) for the *evaluation* and *validation* of a theory such as FT (cf. Tarp 2014: 73).

The main assumption of his argument is that a functional theory, such as FT, "where a lot of empirically grounded theories — i.e. theories built on practice-based evidence" (Tarp 2014: 72) serve as basis for theory deduction, should not be evaluated by criteria used for (other) scientific theories. As an alternative, Tarp (2014) proposes a set of practice-based criteria for the evaluation of

FT as a theory. According to Tarp (2014: 73) practice has three dimensions which can be validated (as opposed to the criterion of falsifiability): (1) the product (a dictionary/dictionaries/lexicographic works), (2) the production process (i.e. the steps suggested for the process of the practical planning, production and publication of a dictionary) and (3) the quality of the final product (dictionary/lexicographic work).

To validate each of these dimensions, Tarp (2014) presents a number of research/validation questions plus — at times — the method(s) to be used in answering each of them:

— Can lexicographic works be produced when guided by the tenets of FT or another theory?

— Can lexicographic works be produced in less time and employing fewer human and material resources when guided by the tenets of FT or another theory? (Measure these variables (time, resources) and compare various theories in terms of their productivity (output of lexicographical works).)

— Is it possible to produce lexicographic works of higher quality when guided by the tenets of FT or another theory, i.e. works satisfying users' (information) needs in all respects? (This requires the use of time-consuming quantitative methods.)

Obviously the first question can be answered with a simple "yes" or "no", or, even, "perhaps", by simply checking if a dictionary has been produced by following the tenets of some theory. A complicating factor is, of course, that various theories are used as guidelines for the production of dictionaries so that the first question and answer really does not help lexicographers in assessing which competing theories are the best for supporting the production of dictionaries.

The second question assumes that FT, with its deductivist methodology, differs from other theories, and must be contrasted and evaluated in terms of lexicographic projects which all base their planning, production and publication of dictionaries on time-consuming and expensive user research. Whether or not this is so, is an empirical question that has to be verified/validated by empirical research. However, the mere fact of Function Theory's opposition to dictionaries produced under the guidance of linguistic theories (of which there are many) would suggest that not all other dictionaries are produced on the basis of time-consuming and expensive user research. In addition, nothing has stopped dictionary producers from freely utilising whatever user research has already been done by others in the planning, production and publication of their dictionaries. One could also consult the many manuals available to guide lexicographers through the theoretical, practical and methodological aspects of dictionary production, which try and consolidate the best of practices in lexicography. The question of whether or not the use of a specific method or methods can be used as a criterion for the evaluation of competing theories must

clearly be answered by a simple "no".

By conceding that one will need all kinds of quantitative methods to answer the third question, the whole exercise of Tarp's (2014) and Fuertes-Olivera and Tarp's (2014) attempts to come up with a new set of guidelines for the validation of dictionaries produced under guidance of FT, or of comparing and evaluating alternative theories, collapses.

To summarize: All of the criteria proposed in Tarp (2014) for the evaluation of FT as a functional theory based on practice collapse when scrutinized for their usability as an evaluation tool for theories.

In Section 2 it was mentioned that two other strategies to address the problems posed by claiming FT to be a lexicographic theory are (1) to define theory in such vague terms that any set of pronouncements could be classified as a theory, or (2) to define theory in such a reductionist way that only FT could be classified as a theory (thus excluding any other competing theories). Both of these strategies are used in Tarp (2009a, 2014). An example of a very vague definition of theory is a reference to a theory as "a systematically organized set of statements about the subject treated" (Tarp 2014: 72). Another, equally vague definition of the term *lexicographic theory* is the following in as much as its key concepts ("a systematically organized set of statements" and "specific types of social need") are not defined and do not exclude all linguistic theories as alternatives:

> … in the following discussion the term *lexicographic theory* is taken to mean *a systematically organized set of statements about the conception, production, usage and history of dictionaries and their relationship with specific types of social need.* (Tarp 2009b: 157)

The central tenet of FT (as provided in Section 2) is a reductionist definition of a theory, and it clearly excludes any linguistic theories as competing alternatives. Tarp (2009b: 158), for example, rejects all linguistic theories "from which linguistic concepts, theories, and methods are frequently transferred uncritically to lexicography" (Tarp 2009a: 158) unless they have gone through the mill of the tenets of FT.

Another questionable strategy used in FT is to derive its tenets from other existing theories.

Examples would be the way in which FT is derived from a theory of L2 acquisition and used in Tarp (2004a,b), its derivation from a theory of translation in Tarp (2008) and "a simple model of communication" referred to in Tarp (2014). Tarp (2014: 72) himself mentions that at the origin of FT, "Tarp … deduced relevant functions from a simple communication model using mother tongue and foreign language as variables". In all cases, however, the theory from which FT is deduced is nothing more than Tarp's own rendition of what these theories in fact are all about. Tarp's discussion of these theories, for example, has no references to the vast literature on theories of L2 acquisition and use, translation theory (cf. Piotrowski 2009), and, especially, the short-

comings of early theories of communication. It is rather disturbing that Tarp presents these theories as facts about L2 acquisition and use, translation and communication, and not as theories of these phenomena, that is, in empirical research, as a (systematic) set of *hypotheses* about these phenomena that still have to be tested in order to accumulate support for them.

The same approach is taken in Leroyer (2011), where the guidelines of FT (seen as "the triangulation of three interrelated sets of parameters: the user, the access and the data parameters" (Leroyer 2011: 128)) are embedded (although not discussed in any detail) in a "use and gratification theory in a functional framework" (Leroyer 2011: 128). This "new" combination of a (rather old) theory plus the functional approach to the design of dictionaries (cf. Leroyer 2011: 128-129) provides a new definition of FTL and, more generally, of lexicography, which combines the theoretical approach of a reinterpreted theory of use and gratification with the practical guidelines of FT for the design of dictionaries and other lexicographic tools: "Lexicography ... (is an — PHS) interdisciplinary discipline concerned with the study, design and development of functional tools aimed solely at the gratification of human information needs and problems" (Leroyer 2011: 129). The functionalist aims of this new FTL is clearly stated at the end of Leroyer (2011: 139) where, however, the development of relevant, usable lexicographic theories has completely vanished from lexicography as an interdisciplinary science: "The new science of lexicography is devoted to the development of unique, functional tools to match and satisfy a great variety of needs for information and experience in modern human societies."

The name of FT ("Function Theory") and Leroyer's (2011) link of FT to the users and gratification theory makes these two theoretical paradigms (functionalism and users and gratification theory) obvious candidates to try and analyse, describe, and explain the theoretical, methodological and practical aspects of FT, but none of the proponents of FT makes this link between FT and these two theoretical paradigms clear. Literature on these two theories does, however, suggest possible links between them and FT, but also a number of differences with regard to their theoretical, methodological and practical assumptions. (Cf. in this regard, for example, Jansen 1989 on the tenets of functionalism and Pitout 2009 on those of the users and gratification theory.)

Leroyer's (2011) suggestion of a link between FT and users and gratification theory has one important implication worth pursuing in further research. While FT (as in Functionalism) still clings to a view of a static, passive user who is influenced by outer sources (dictionaries designed according to FTL), the major change that the users and gratification theory brought was the view of users as dynamic participants who strategically select and scrutinize sources using different techniques to gratify their needs for information. As argued, nothing of this is raised in Leroyer (2011) or in FT literature. The view of the user clearly opposes the single set of criteria FT proposes for all users: optimizing access and comprehensibility and minimization of data. Although these

guidelines could benefit all types of users, they should not be elevated as the sole criteria for the presentation and evaluation of data in dictionaries.

The further point to be made with regard to the theoretical foundation of FT is that Tarp (2014) utilises deduction and axiomatization as methods for the formulation of the tenets of FT. This is in itself a widely used method for the formulation of theories, but it does not follow that the central tenet of FT states precisely of what FT is a theory.

Tarp (2014: 71-73) is very clear on the methods used in the formulation of FT and those needed for the practical execution of the tenets of FT: FT is a product of the application of three methods: the axiomatic method, the deductive method and (expert) observation. Although Tarp does not define these methods or provide any reference to sources used in their application — they are simply presented as "well-known methods developed long ago within other disciplines" (Tarp 2014: 71), he summarizes their use in the generation of FT as a general theory and its sub-theories as follows:

> The basic statement that dictionaries are utility products was formulated axiomatically based on observation of practice and the study of relevant literature including pre-theoretical reflection as other more or less developed theories … the axiomatic formulation of the complementary statements on the main functions were the result of a combination of observation and deduction.

Tarp (2014) also refers to sub-theories of FT (not yet developed as theories, e.g. the description of meaning in dictionaries) which are the result of observation, analysis of empirical data and the use of text corpora (cf. Tarp 2014: 72).

Tarp (2014) does not discuss in any detail the axiomatic method or the deductive method as methods for the formulation of lexicographic theories, but as presented in FT literature (cf. the references above), it is also clear what method proponents mostly follow in the case of FT: the lexicographer's/linguist's/researcher's subjective interpretation of theory is presented as a theoretical framework from which (s)he deduces a number of guidelines (or a general theory and/or sub-theories) to serve as bases for the deduction of lexicographically appropriate methods and practical guidelines for the design of L1/L2 dictionaries. What we are thus confronted with is a lexicography theory, method and practice driven solely by the subjective impressions, interpretations, gut feelings etc. of individuals, despite the fact that it is also claimed that these views are based on empirical perception, a study of a vast body of literature, dictionaries etc. These sources are seldom if ever mentioned, and the views are presented without acknowledgement of how fraught with problems personal interpretations of these sources could be.

From the discussion of empirical lexicographic inquiry above, however, it is evident that it really does not matter precisely how lexicographers come upon their theories (inductive or deductive, or whether a proposed theory is axiomatized or not) as long as its status as a set of to-be-validated hypotheses is acknowledged. This is made clear in Botha's (1981) discussion of the empirical

method. In empirical inquiry a theory is defined as a set of hypotheses, where a hypothesis refers to two aspects of this kind of statement:

> the functional aspect: the function of a hypothesis is to give a certain amount of information about an aspect of reality which is unknown and which cannot be known in a direct way. In this context, a direct way of getting to know something is, for instance, that of elementary sense perception … the epistemological one: the idea embodied in a hypothesis, or the information contained in a hypothesis is by nature tentative. In other words, a hypothesis represents a unit of thought of which the correctness has not been established. (Botha 1981: 108)

As Botha (1981: 108) notes, calling something a "hypothesis" gives no indication of precisely how someone arrived at the idea embodied in that hypothesis (for example, in a rational or nonrational way, by deriving axioms by way of deduction, etc.). Scientists, or for that matter lexicographers and linguists, arrive at their discoveries or knowledge of some aspect of reality (e.g. dictionary users, user needs in specific contexts of use etc.) by

(a) making assumptions about the regularity, pattern, structure, mechanism or cause underlying that which is problematic,
(b) regarding these assumptions (or *hypotheses* — PHS) as tentative, and
(c) attempting to test and justify these assumptions.

Efforts of testing and justifying the assumptions of FT are, however, excluded from FT: the methods (see below) of empirical enquiry are simply rejected and substituted by Tarp (2014) by the subjective assessments of all aspects of FT by an experienced language teacher or lexicographer. The status of these assumptions is not discussed and not presented as a systematic set of hypotheses. What Tarp (2014) in fact does is to substitute the empirical method of the status of theories with a set of assumptions of which the status is unclear and a methodology for research that substitutes the subjective interpretation of lexicographers and teachers for a very long tradition of empirical research.

If one defines a theory as a set of hypotheses about some empirically unobservable phenomenon or process, one has the suspicion that FT as currently formulated is guided by some implicit theoretical assumption (that one can indeed verify). A guess would be that the planning, production and publication of a L1/L2 (language) dictionary, which is guided by the practical guidelines of FT, will result in a better dictionary to assist/address the users' information needs in text production, reception and language learning. Formulated as such, we have a theoretical hypothesis that could provide guidelines for the practical design of a dictionary, in which case one can try and find support for this hypothesis by comparing the effects of a dictionary designed according to the guidelines of FT on text production, reception and language learning, and those of a dictionary designed according to the guidelines of LL.

Whether or not FT could be interpreted as a theory is much more complicated. Firstly, because the term can be defined in many ways; secondly, the

problem of precisely what the term *theory* means becomes even more problematic when you start using terms like *action theory* as label for Wiegand's version of function theory or refer to the FT as a theory built on practice-based evidence (cf. Tarp 2014: 72). Do these terms, for example, all refer to a theory defined as a set of hypotheses, where a hypothesis refers in empirical inquiry to two aspects of this kind of statement?

FT methodology

The term *method* is defined in Tarp (2014: 59), following Buhr and Klaus (1971), as

> a system of (methodological) rules determining classes of potential operation systems which may lead to a specific goal from a specific point of departure.

As Tarp (2014: 58-59) indicates, methods apply both to the elaboration and formulation of the general (such as FT) and specific lexicographic theories (a theory of L2 acquisition and use) and to each of the phases of practical lexicography work to accomplish the various tasks outlined in each step of planning and production of dictionary models.

The FT approach to the second kinds of methods is to reject all methods of empirical linguistic inquiry (typical of LL) to determine the variables of FT on the basis that they are too costly and time consuming, besides the fact that they, in the view of Bergenholtz and Tarp (cf. for example, Bergenholtz 2011; Tarp 2009a,b; 2014) have hardly produced anything to support practical lexicography.

Tarp (2014: 62) notes, for example, that the methods developed within linguistics "in order to analyse, describe or normalize language may be perfectly adequate for this purpose", but they may not be the most appropriate given the goals/tenets of FT, viz. when "it is a question of solving the concrete types of information need which different types of users may experience in different communicative situations", for example, when these linguistic methods are not "sufficiently focused on the specific tasks to be solved by lexicography" (Tarp 2014: 62). (Cf. Tarp 2014: 62 for a brief discussion of a number of concepts, theories and methods from Linguistics which proponents of FT have reanalysed to suit the goals of FT.)

Tarp (2009a) provides an in-depth and critical discussion of the methods commonly used in user research (in LL): questionnaires, interviews, observation, protocols, experiments and log files. The method he uses in this article is (1) to support up front the opinion of others that the use and the results of such methods of user research are in fact a waste of time and money (cf. Tarp 2009a: 276-277) — a method carried through to his final evaluation of most well-known methods of user research (cf. Tarp 2009a: 290-293); (2) by explicating each method, followed (3) by an evaluation of a method in terms of its "advantages, disadvantages and limitations" (Tarp 2009a: 277) (as mainly seen, however, from what is required for FT goals as spelled out up front in the article, cf.

Tarp 2009a: 277-283).

As questionnaires are the most common methods used in user research, and as this is the method criticized by Bergenholtz and Gouws (2013), it could serve as example of Tarp's (2009a) analysis.

Tarp (2009a: 284) classifies questionnaires as consisting typically of

— closed questions to which respondents have to give the answers to predetermined categories, and

— open questions to which respondents may add answers to other categories than the predetermined ones

Following Welker (2006), Tarp (2009a: 284) indicates that the questions asked in questionnaires can be divided into three types (with the example questions changed to reflect questions):

(a) about facts that can be easily remembered by the respondents (for instance: How many dictionaries do you have? When did you buy them? Why did you buy them?)

(b) about the usage of dictionaries, and

(c) about the user's opinion (for instance: Are you satisfied with the dictionaries? What type do you prefer? What improvements would you like?); what Tarp (2009a: 284) calls "retrospective questions".

According to Tarp (2009a: 284), Welker is of the opinion that answers to type (a) questions offer data that are reasonably reliable, while answers to the type (c) retrospective questions are all subjective, demanding a lot from respondents' memories, but that "they may guide lexicographers and publishers to an extent". Tarp, however, contrasts Welker's evaluation of type (c) questions with the opinion of Hansen and Andersen (2000: 146), i.e. that they can cause reliability problems. Welker also mentions that type (c) questions (even such as those of FT listed below) only reveal users' perception of their consultation of dictionaries, not their real usage — criticism also levelled by Hatherall (1984: 184).

According to Tarp (2009a: 284) the only advantages of questionnaires are that they can be used to involve a large number of respondents and that it is relatively easy to analyse the answers (especially closed questions that require no coding). Questionnaires, according to Tarp (2009a: 284), have a number of "serious disadvantages" of which the reliability of type (c) question is his main objection.

Tarp's (2009a: 285) final evaluation of the typical questionnaire method in dictionary user research is stated (rather bluntly) as follows:

> many lexicographers still carry out research by means of questionnaires, arriving at conclusions which even a modest sociological knowledge would show to have no scientific warranty.

Tarp (2009a: 284-285) does, however, accept the following type of research questions put by Welker, questions, it should be obvious, that probe the answers to questions which would typically be put by proponents of FT:

> What are you looking for in the dictionaries?
>
> Under what conditions do you consult it (a/the relevant dictionary/dictionaries under scrutiny — PHS) most frequently?
>
> In which percentage of the look-ups does the dictionary provide assistance?
>
> Which part of the information (provided in the dictionary — PHS) is most useful?

It also comes as no surprise that according to Tarp (2009a: 291) the best user research has been that which probes the variables of FT: the usage situation, different look-up strategies and search routes of users, where users look for information and what information they look for.

It is also important to mention that Hansen and Andersen in their discussion of the problem of the reliability of the answers to (c) type questions also mention that one need not avoid using such questions as long as the reliability problems can be anticipated and taken into consideration when interpreting the answers to these questions (cf. Tarp 2009a: 284). That is exactly the kind of warning (and presentation of solutions) one finds, for example, in Babbie and Mouton's (2001) discussion of the way researchers must build the necessary precautionary measures into the design and application of any method to avoid possible problems in the use of the many methods discussed by Tarp (2009a, 2014) (cf., for example, Babbie and Mouton 2001: 239-249 on the design/construction of questionnaire). The point has to be emphasized that the possible problems researchers may experience with the application and interpretation of the results of any of these methods is well-known and that (good) manuals on social research build the necessary precautionary measures into the design and application of various methods, and in the interpretation of the results of the research done with these methods. They do not see these possible problems merely as disadvantages that justify the abandon of the use of these methods. (For an alternative approach to the kinds of methods used in LL/lexicography research, also on the functional variables distinguished in FT, see, for example, Geyken 2014, Mann and Schierholz 2014, and Töpel 2014.)

Tarp (2009a: 291) laments the lack of research on the objective and subjective dictionary needs of users, and criticizes attempts to conduct such research by trying to deduce these needs from the linguistic information users search for. As Tarp (2009a: 292) notes, this approach is inherently circular. To do so would require that researchers move from a focus on the dictionary usage situation to the extra-lexicographic situation. For research on the latter a number of methods can be used, including: the deductive method, tests and inter-

views (How much have readers understood of a text? What reception problems did they have during reading?), text revision and marking (Which non-recognised needs occur during text reproduction and translation?) (Tarp 2009c: 293). However, according to Tarp (2009c: 293), most of these methods (other than deduction) are time consuming and too expensive to use. He also advises (cf. Tarp 2009c: 293) that (1) meticulous planning of research projects is needed as well as a critical approach to the usability of quantitative methods, i.e. whether they are "at all relevant to lexicography", and (2) research be based on FT with its distinction of scientific categories of user needs, user typology, user situation, usage situations, access routes etc.

Tarp (2014: 65) explicitly rejects (empirical) user research as a method for determining the relevant range of users, needs, functions and situations on the grounds that although it may lead to the detection of relevant needs (if based on "scientific principles": Tarp 2014: 65) they are too costly and time consuming "to be used in every dictionary project with its own characteristics" (Tarp 2014). Given Tarp's (2009a) method of the exposition and evaluation of a number of methods, his final conclusion is to be expected, viz. that quantitative user research, in particular, suffers from the following: bias in the selection of respondents (e.g. mostly or only students), lack of random selection of respondents, the size of the respondent group being too small to provide statistically significant and relevant results, problems with the formulation of questions (questions are ambiguous, informants do not understand questions, they do not have the memory to answer questions) and so on.

Instead, Tarp (2014) opts for a method where users are themselves completely omitted from the exercise, and it is up to experts on the topic of the dictionary, situation(s) and foreseen users, i.e. lexicographers, alone or together with experts (who, according to the long quotation in Tarp (2014: 66), could be experienced L1/L2 teachers or teachers of translation), to use their expertise to deduce the needs that have to be covered by a given dictionary. The main argument for this methodology is that it is "relatively easy and quick to apply" (Tarp 2014: 65), and that it has proven to be effective in the planning and production of a number of FT conceptualized dictionaries. Tarp (2014: 65) adds:

> The method may not be perfect but it is capable of determining the huge majority of relevant user needs, even some occurring only very seldom. The results can easily compete with those obtained by user research but using only a fraction of the time and resources required to base the dictionary concept on such research in each case.

There are, however, a number of pitfalls in Tarp's methodological assumptions. The first is that he expects his users to accept his pronouncements of the success of his methodology, but provides no references to relevant sources that could back up his claims. The second is that, by accepting the FT-based questions of Welker as guidelines for empirical research, Tarp exposes himself to the same critique Hatherall (1984: 184) has voiced against the questions used in

questionnaires: they could reveal users' perception of their consultation of dictionaries rather than their real usage. A third problem is the questions Tarp proposes to be answered by lexicographers or experienced teachers. Tarp (2009a) does not problematize the status of the answers to these questions. Heid (2011), however, in testing the usability of dictionaries, indicates clearly the limitations of a single expert to identify problems of usability — one needs a group of 5 to 10 experts plus 30 lay testers of a homogeneous user group to identify a significant number of problems with a given piece of software (for example, an *e*-dictionary). In addition, quantitative and qualitative research is necessary to identify such problems. Data such as that provided by Heid (2011) should warn one to be very careful in elevating the answers of lexicographers and experienced teachers as being representative of the problems and information needs of (different groups of) users of dictionaries in different situations of use.

One can only wonder what the possible source(s) for such a view of appropriate methods for lexicographic inquiry could be. Perhaps the following view of Gouws (2014: 23) on the planning, compilation and production of *e*-dictionaries sanctions this methodology in which the determination of the large category of users, situations of use and functions of monolingual language dictionaries is seen as a once-off exercise (already completed for past and future research needed for FTL):

> Lexicographers working in the field of electronic dictionaries should realise that the planning and compilation of this medium need not go through all the same phases that crossed the way of the development of printed dictionaries. We have already identified the user, the needs of the user and the functions to ensure the satisfaction of these needs.

Gouws (2014) does not reference the sources of empirical research which could/should support this FTL methodology, and nor does he indicate, for example, whether or not it implicitly refers to the use of the same methodology in Tarp (2004a,b), which provides Tarp's view of the complex variables of users, functions, information needs and users' ability to access linguistic data/MD.

To add to these pitfalls, one must also note that Tarp (2014), in defining theory, method and the interaction between them, follows the positivistic views of his sources in as much as he notes (Tarp 2014: 59) that methods are embedded in or based upon theories or constituent parts of different theories in scientific disciplines such as lexicography, which means that methods are anchored in theories by a set of "objective laws", or, if interpreted correctly, that theory-based methods (for the tasks of practical lexicography) can be deduced from lexicographic theories by means of objective laws. The existence of such "objective laws" also governing reality (and the reality of theories and dictionaries, one would assume — PHS) is suggested by Tarp (2014: 71) in the following:

> theory and method … have evolved together in a fruitful dialectic relationship, in this case mediated by practice and the observation of the objective law-governed reality.

"Practice" as the basis of determining appropriate methods for practical lexicography may in this context refer to what has been and is done in practical lexicography (irrespective of how these methods may be justified), but it is unclear how lexicographers could get from their theories to their methods based on the kind of laws proposed as governing reality.

Finally, one must also note one of the major problems of the FT's criterion that monofunctional (*e*-)dictionaries have to be produced to assist users with their information needs in different situations of use — in effect, in text reception, production and L2 language acquisition. Given that the linguistic categories that L2 dictionaries would have to contain to assist users in text production and L2 use, as outlined in Tarp (2004b: 308, 312-317), overlap to such an extent that the production of separate dictionaries for these two functions can hardly be justified. Although text reception requires, first and foremost, an explanation of the meaning of L2 words (plus a few other linguistic features), this hardly justifies the production of a separate dictionary for dynamic and experienced users for this function, given that L2 dictionaries use all kinds of strategies to make the meaning explanations in articles as accessible and comprehensible as possible (for example, signposts to differentiate polysemous distinctions, the use of dedicated information zones in articles etc.).

The plea for the design of monofunctional dictionaries in the case of electronic dictionaries is understandable, as electronic mechanisms allow for this. *E*-dictionaries can be designed in such a way that users have the opportunity of defining both themselves and relevant user situations to get access to exactly the data they need in each situation, and to change their navigation options between monolingual and bilingual access in a consultation. The design of a multifunctional lexicographic database from which to generate these monofunctional options does not necessitate that it has to be designed according to the tenets of FT. The possibilities for customization of user interfaces are also utilised in L2 dictionaries designed according to the tenets of LL.

Given these considerations, it is understandable that Tarp (2004b: 317-322) argues, and illustrates abundantly from a number of existing monolingual L2 language dictionaries, that it is possible and preferred in the case of printed dictionaries to produce only one L2 dictionary that can serve text production, reception and L2 acquisition.

4. Back to LL and FT

Hass and Schmitz (2010: 12-13) note in their overview of recent developments in electronic lexicography that the strong focus on the possibilities of the design of *e*-dictionaries has led to an abolition of the theoretical, methodological and

design issues that took centre stage in the eighties and nineties, as if the key issues raised by linguistic theories then were all resolved long ago. However, the discussion of the theoretical, methodological and practical aspects of FTL and LL in the growing FTL literature testifies to the fact that these issues are not forgotten but still take centre stage in discussions of lexicography theory, method and practice.

The argument that has been made in this article is that the theoretical, methodological and practical assumptions of FT are fraught with difficulties and that it is hardly a viable contender for LL. There is also no doubt that LL, if thoroughly scrutinized using justifiable methods, also has a number of shortcomings in its theoretical, methodological and practical assumptions. This points to an approach in which these two lexicographic approaches should not be evaluated in a dichotomous usable/unusable way, especially when the criteria for such an evaluation prove to be problematic. FT should, therefore, not be presented as an alternative to the linguistic and Wiegandian structuralism that preceded FTL (cf., for example, Gouws 2011) but as complementary to it.

As acknowledged in Tarp (2004a,b), a concern that the functional variables of FT should play a major role in the design of L2 dictionaries is not new. Historically, dictionaries have always been designed with explicit users and functions in mind, although the methods used for this purpose may be called "artisanal" methods (Tarp 2014); but since the inception of empirical research with scientific methods, research on the functional variables has certainly received a large boost. This is not to deny that the application of these methods has faced many problems, but research on these methods and the tweaking of the requirements for their application have helped to establish a large body of knowledge on the functional variables that are of importance in the design of usable L2 dictionaries.

Lexicographers have in linguistic lexicography access to a vast body of theoretical, methodological and practical research to support the design of morphological data in L1/L2 (language) dictionaries. Furthermore, Function Theory has established the importance of functional variables in the design of (language) dictionaries, but what lexicography needs now is a truly multidisciplinary approach to lexicography, and not an approach that reduces lexicography to the status of a hand-maiden of another discipline, for example, information science, or of a reductionist Function Theory of lexicography.

Bibliography

Apel, K. 2014. A Comprehensive Definition of Morphological Awareness: Implications for Assessment. *Topics in Language Disorders* 34(3): 197-209.

Apel, K., E. Diehm and L. Apel. 2013. Using Multiple Measures of Morphological Awareness to Assess its Relation to Reading. *Topics in Language Disorders* 33(1): 42-56.

Apel, K. and K. Werfel. 2014. Using Morphological Awareness Instruction to Improve Written Language Skills. *Language, Speech, and Hearing Services in Schools* 45: 251-260.

Atkins, B.T. Sue and M. Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.

Babbie, E. and J. Mouton. 2001. *The Practice of Social Research*. Oxford: Oxford University Press.

Barz, I., M. Schröder and U. Fix (Eds.). 2000. *Praxis- und Integrationsfelder der Wortbildungsforschung.* Heidelberg: Winter.

Bauer, L., R. Lieber and I. Plag. 2013. *The Oxford Reference Guide to English Morphology*. Oxford: Oxford University Press.

Bergenholtz, Henning. 2011. Access to and Presentation of Needs-adapted Data in Monofunctional Internet Dictionaries. Fuertes-Olivera, Pedro A. and Henning Bergenholtz (Eds.). 2011: 30-53.

Bergenholtz, H. and R.H. Gouws. 2013. The Presentation of Word Formation in General Monolingual Dictionaries. *Lexikos* 23: 59-76.

Botha, R.P. 1981. *The Conduct of Linguistic Inquiry. A Systematic Introduction to the Methodology of Generative Grammar.* The Hague: Mouton.

Bothma, Theo JD. 2011. Filtering and Adapting Data and Information in an Online Environment in Response to User Needs. Fuertes-Olivera, Pedro A. and Henning Bergenholtz (Eds.). 2011: 71-102.

Bowers, P.N. and J.R. Kirby. 2010. Effects of Morphological Instruction on Vocabulary Acquisition. *Reading and Writing: An Interdisciplinary Journal* 23: 515-537.

Buhr, M. and G. Klaus (Eds.). 1971. *Philosophisches Wörterbuch*. Band 1-2. Berlin: Das europäische Buch.

Carlisle, J.F. 2010. Effects of Instruction in Morphological Awareness on Literacy Achievement: An Integrative Review. *Reading Research Quarterly* 45(4): 464-487.

Carlisle, J.F. and C.A. Stone. 2005. Exploring the Role of Morphemes in Word Reading. *Reading Research Quarterly* 40(4): 428-449.

De Pauw, G. and G.-M. de Schryver. 2008. Improving the Computational Morphological Analysis of a Swahili Corpus for Lexicographic Purposes. *Lexikos* 18: 303-318.

Evert, S. and A. Lüdeling. 2001. Measuring Morphological Productivity: Is Automatic Preprocessing Sufficient? Rayson, P., A. Wilson, T. McEnery, A. Hardie and S. Khoja (Eds.). 2001: 167-175.

Felbaum, C. 1998. *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.

Fourie, Pieter J. (Ed.). 2009. *Media Studies.Volume 3. Media Content and Media Audiences.* Cape Town: Juta.

Fuertes-Olivera, Pedro, A. and Henning Bergenholtz (Eds.). 2011. *e-Lexicography. The Internet, Digital Initiatives and Lexicography*. London/New York: Continuum.

Fuertes-Olivera, Pedro A. and Sven Tarp. 2014. *Theory and Practice of Specialised Online Dictionaries: Lexicography versus Terminology.* Berlin/Boston: De Gruyter.

Geyken, A. 2014. Methoden bei der Wörterbuchplanung in Zeiten der Internetlexikographie. *Lexicographica* 30: 77-111.

Gouws, R. 2011. Learning, Unlearning and Innovation in the Planning of Electronic Dictionaries. Fuertes-Olivera, Pedro A. and H. Bergenholtz (Eds.). 2011: 17-29.

Hansen, E.J. and B.H. Andersen. 2000. *Et sociologisk værktøj. Introduktion til den kvantitative metode*. Copenhagen: Hans Reitzels Forlag.

Hartmann, R.R.K. (Ed.). 1984. *LEXeter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9–12 September 1983.* Lexicographica. Series Maior 1. Tübingen: Max Niemeyer.

Hass, U. and U. Schmitz. 2010. Lexikographie im Internet 2010 — Einleitung. *Lexicographica* 26: 1-18.

Hatherall, G. 1984. Studying Dictionary Use: Some Findings and Proposals: Hartmann, R.R.K. (Ed.). 1984: 183-189.

Heid, U. 2011. Electronic Dictionaries as Tools: Towards an Assessment of Usability. Fuertes-Olivera, Pedro A. and Henning Bergenholtz (Eds.). 2011: 287-304.

Jansen, N. 1989. Functionalism. Jansen, N. 1989. *Philosophy of Mass Communication Research*: 14-28. Cape Town: Juta.

Jiang, N. 2004. Morphological Insensitivity in Second Language Processing. *Applied Psycholinguistics* 25: 603-634.

Klosa, A. 2013a. Einleitung. Klosa, A. (Ed.). 2013: 7-11.

Klosa, A. (Ed.). 2013b. *Wortbildung im elektronischen Wörterbuch*. Studien zur Deutschen Sprache 63. Tübingen: Narr.

Leroyer, P. 2011. Change of Paradigm: From Linguistics to Information Science and from Dictionaries to Lexicographic Information Tools. Fuertes-Olivera, Pedro A. and Henning Bergenholtz (Eds.). 2011: 121-167.

Lew, R. 2008. Lexicographic Functions and Pedagogical Lexicography: Some Critical Notes on Sven Tarp's *Lexicography in the Borderland between Knowledge and Non-Knowledge. Pedagogika. Politologia. Filologia* 2008: 114-123.

Mann, Michael and Stefan J. Schierholz. 2014. Methoden in der Lexikographie und Wörterbuchforschung. Ein Überblick mit einer Auswahlbibliographie. *Lexicographica* 30: 3-57.

Nielsen, S. and R. Almind. 2011. From Data to Dictionary. Fuertes-Olivera, Pedro A. and Henning Bergenholtz (Eds.). 2011: 141-167.

Piotrowski, T. 2009. [Review of] Sven Tarp. 2008. *Lexicography in the Borderland Between Knowledge and Non-knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography. International Journal of Lexicography* 22(4): 480-486.

Pitout, M. 2009. Media Audience Theory. Fourie, Pieter J. (Ed.). 2009: 390-420.

Rayson, P., A. Wilson, T. McEnery, A. Hardie and S. Khoja (Eds.). 2001. *Proceedings of the Corpus Linguistics 2001 Conference.* Lancaster: UCREL.

Rundell, M. 2012. *'It Works in Practice but will it Work in Theory?' The Uneasy Relationship between Lexicography and Matters Theoretical.* http://www.euralex.org/elx_proceedings/Euralex2012/pp47-92%2520Rundell.pdf.

Schmid, H., A. Fitschen and U. Heid. 2004. SMOR: A German Computational Morphology Covering Derivation, Composition, and Inflection. *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004). Volume 4*: 1263-1266. Lisbon, Portugal: European Language Resources Association.

Spohr, D. 2011. A Multi-Layer Architecture for "Pluri-Monofunctional" Dictionaries. Fuertes-Olivera, Pedro A. and Henning Bergenholtz (Eds.). 2011: 103-120.

Swanepoel, P. 2001. Dictionary Quality and Dictionary Design: A Methodology for Improving the Functional Quality of Dictionaries. *Lexikos* 11: 160-190.

Tarp, Sven. 2004a. Basic Problems of Learner's Lexicography. *Lexikos* 14: 222-252.

Tarp, Sven. 2004b. Reflections on Dictionaries Designed to Assist Users with Text Production in a Foreign Language. *Lexikos* 14: 299-325.

Tarp, Sven. 2008. *Lexicography in the Borderland between Knowledge and Non-Knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography.* Lexicographica Series Maior 134. Tübingen: Max Niemeyer.

Tarp, Sven. 2009a. Reflections on Lexicographical User Research. *Lexikos* 19: 275-296.

Tarp, Sven. 2009b. The Foundations of a Theory of Learners' Dictionaries. *Lexicographica* 25: 155-168.

Tarp, Sven. 2014. Theory-based Lexicographical Methods in a Functional Perspective. An Overview. *Lexicographica* 30: 58-76.

Templeton, S. 2011/2012. Teaching and Learning Morphology: A Reflection on Generative Vocabulary Instruction. *Journal of Education* 192(2/3): 101-107.

Töpel, A. 2013. Die Wortbildungsangaben im Online-Wörterbuch und wie Nutzer sie beurteilen — eine Umfrage zu *elexiko*. Klosa, A. (Ed.). 2013b: 197-214.

Töpel, A. 2014. Methoden zur Erstellung van Bedeutungsparaphrasenangaben. . *Lexicographica* 30: 291-322.

Tono, Y. 2010. A Critical Review of the Theory of Lexicographical Functions. *Lexicon* 40: 1-26.

Welker, H.A. 2006. O uso de dicionários: Panorama geral das pesquisas empíricas. Brasilia: Thesaurus.

http://lexikos.journals.ac.za

Collocations and Grammatical Patterns in a Multilingual Online Term Bank

Elsabé Taljard, *Department of African Languages, University of Pretoria, Pretoria, South Africa (elsabe.taljard@up.ac.za)*

Abstract: This article considers the importance of including various types of collocations in a terminological database, with the aim of making this information available to the user via the user interface. We refer specifically to the inclusion of empirical and phraseological collocations, and information on grammatical patterning. Attention is also given to provision of information on semantic prosody and semantic preferences — aspects which have been rather neglected in South African terminological databanks and language for special purposes (LSP) dictionaries. Various strategies for the effective semi-automatic extraction of collocational data from specialized corpora are explored. Possibilities regarding access to and presentation of collocational information to the user are briefly considered. It is argued that users should have direct access to collocational information, and that collocations should not only be accessible via the lemmatic address of the term appearing as part of the collocation. The research is done within the context of the establishment of an Open Access Resource Term Bank, which is developed as a pedagogical tool to support students whose language of learning and teaching is not the L1.

Keywords: COLLOCATIONS, GRAMMATICAL PATTERNING, MULTILINGUAL TERMINOLOGY DATABASE, SEMANTIC PROSODY, SEMANTIC PREFERENCE, OPEN ACCESS RESOURCE TERM BANK, CORPUS-BASED TERMINOLOGY

Opsomming: Kollokasies en grammatikale patrone in 'n veeltalige aanlyntermbank. Hierdie artikel ondersoek die belangrikheid van die insluiting van verskillende tipes kollokatiewe inligting in 'n veeltalige terminologiese databank met die oog daarop om hierdie inligting aan die gebruiker via die gebruikerskoppelvlak beskikbaar te stel. Ons verwys spesifiek na die insluiting van emipriese en fraseologiese kollokasies, en ook na inligting oor grammatiese patrone. Aandag word ook gegee aan die verskaffing van inligting oor semantiese prosodie en semantiese voorkeur — aspekte wat nie veel aandag in Suid-Afrikaanse terminologiese databanke en vakwoordeboeke geniet nie. Verskeie strategieë vir die effektiewe semi-outomatiese onttrekking van kollokatiewe data uit vakgerigte korpora word ondersoek. Verskillende moontlikhede met betrekking tot die toegang tot en aanbieding van kollokatiewe inligting aan die gebruiker word oorweeg. Daar word geargumenteer dat gebruikers direkte toegang tot kollokatiewe inligting moet hê, en nie alleenlik via die lemmatiese adres van die term wat as deel van die kollokasie optree nie. Die navorsing word gedoen teen die agtergrond van die daarstel van 'n terminologiese hulpbron met vrye toegang, wat as opvoedkundige hulpmiddel ontwerp word ter ondersteuning van studente wat nie deur middel van hul L1 studeer en/of onderrig word nie.

Sleutelwoorde: KOLLOKASIES, GRAMMATIESE PATRONE, VEELTALIGE TERMINOLOGIEDATABASIS, SEMANTIESE PROSODIE, SEMANTIESE VOORKEUR, VRY-TOEGANKLIKE TERMINOLOGIESE HULPBRON, KORPUS-GEBASEERDE TERMINOLOGIE

Contextualization and introduction

The research reported on is done within the context of the establishment of a multilingual, open education resource term bank (OERTB). Establishment of such a term bank is part of a Department of Higher Education funded project, awarded jointly to the University of Pretoria and the University of Cape Town. It is envisaged as a collaborative effort between all South African universities and the aim is to create a terminological tool which can serve as pedagogical support tool to South African students. This tool will be made available to participating universities under a creative commons licence. Access to the user interface will be via the online learning systems of the various universities. The assumption is that the majority of South African students are exposed to a tertiary education system where the language of teaching and learning is not the strongest language, i.e. the L1. It is furthermore assumed that giving these students access to an internet-based term bank, which contains not only term equivalents for key concepts in the African languages, but also additional conceptual information, e. g. definitions, and guidance on usage of terms, should aid in the conceptualisation of subject content. The tool is planned as an organic one, with terminology being developed within actual pedagogical situations.

In terms of Bergenholtz and Bothma (2011: 61, 62) we envisage that our terminological tool will be used in cognitive and communicative situations. They describe cognitive situations as knowledge seeking situations which are unrelated to specific usage situations such as text reception. Within a cognitive situation the user simply wants to find knowledge, which can be stored for later use. Term banks are listed as one of the most commonly used tools in these situations. Communicative situations deal with problems or doubts that the user may have regarding the process of oral or written communication, and with issues such as text reception, text production, translation and text correction, of which the first three are possibly the most important in our specific usage situation. The practical implementation of our terminological tool deals directly with the communicative function in that it can provide a starting point for translanguaging, a practice which is described by Park (2013: 50) as assisting "multilingual speakers in making meaning, shaping experiences, and gaining deeper understandings and knowledge of the languages in use *and even of the content that is being taught*" (my emphasis). Within the context of tertiary education in South Africa forums such as tutorials would present ideal opportunities for translanguaging practices. In these pedagogical situations, students can discuss threshold concepts in their L1/strongest language, and draw on the terminological tool for African language equivalents of the English terms and explanations of key concepts in their L1.

The aim of this article is first to critically consider the importance of the inclusion of two types of information in a multilingual terminological database, with the aim of making this information available to the user via the user interface. We refer here to various kinds of collocational information, which contribute to the conceptual information provided in the term bank, and grammatical patterning, which is more usage-oriented. Pending a detailed discussion of the notion of collocation (see below), it can provisionally be described as frequently recurring word combinations. Secondly, various strategies are investigated for the effective extraction of collocational information and grammatical patterning from electronic text corpora, within the time and resource constraints of the project. Reference to time and resource constraints here may seem redundant, but within the specific context these constraints are indeed relevant. We are aware of the fact that very sophisticated procedures and tools for, inter alia, computational identification and extraction of collocations exist; however, the level of computational expertise required to utilize these resources is far beyond what is realistically available within the constraints of the project. The project is funded for three years and expenditure is limited to a fixed budget. The project team therefore has no choice but to make use of commercially available software, even though we are aware of their limitations. In the last instance, different possibilities regarding access to and presentation of collocational information to the user are briefly considered.

For the sake of clarity, a few remarks concerning the terminology used within the context of any kind of electronic terminology activity, tools or products are necessary. Perhaps somewhat ironically, the terminology used in this regard is rather confusing — the terms 'terminological/terminology data bank', 'term bank' and 'terminological database' being used rather indiscriminately and sometimes interchangeably, both generally and in scholarly work.

A trawl through the literature has brought the following to light: the terms 'term bank' and 'terminology/terminological data bank' are treated as synonyms and refer to a collection of different, but usually related databases that can be accessed by users with common software via a user interface (UI). A term bank usually belongs to or refers to an institution. Examples of well-known term banks are TERMIUM Plus® (https://www.btb.termiumplus.gc.ca) and *Grand dictionnaire terminologique* (GDT) (https://www.granddictionnaire.com) — two Canadian term banks — and InterActive Terminology for Europe (IATE). The OERTB would be an example of a term bank.

Definitions of the term 'database' emphasize the notion of a structured collection of terminological data, cf. the definition provided by the ISO Online Browsing Platform (OBF) (https://www.iso.org/obp/ui/, accessed 18-05-2015), according to which a terminology database consists of structured sets of terminological records in an information processing system. It is important to understand that users do not have access to the terminological database itself; they only have access to the information that the terminologist/databank manager chooses to make available via the user interface. A second important point is that the database may (and usually does) contain many more data categories

than the ones which are accessible to the user. The current research investigates the inclusion of collocational information and information on grammatical patterning in the OERTB database with the aim of making this information available to the user through the user interface.

On defining collocations and grammatical patterning

The concept of collocation is notoriously difficult to define, even though, as Evert (2007) points out, it is based on a widely-shared intuition that certain words have the tendency to co-occur in natural language. From the literature it is clear that collocation is a multi-dimensional phenomenon. Evert (2007) distinguishes between the Firthian and the phraseological interpretations of the notion of collocation. Within the Firthian tradition collocation is the recurrent combinations of words that co-occur more often than expected by chance and that correspond to arbitrary word usages, cf. Smadja (1993). They are observable facts about language and thus present primary data. When working with raw, i.e. un-annotated corpora, the notion of collocation as recurrent word combinations implies a kind of mechanistic lexical co-occurrence where the presence of a node primes the presence of a collocate or collocates: *salt* and *pepper*, *cow* and *milk,* and *day* and *night* are prototypical examples of what Evert calls empirical collocations; for the verb *request* object nouns that can be expected to co-occur are *information*, *permission*, *assistance* and *help*, the collocational pairs being *request* and *information*, *request* and *permission*, etc. The phraseological interpretation on the other hand, describes collocations as being semi-compositional and lexically determined word combinations, such as *make an appearance* and *give a talk*. This kind of collocation is also known as multiword expressions and includes a whole range of subcategories, from completely opaque idioms to combinations which are subject to arbitrary lexical restrictions, e.g. *take medicine* rather than *drink medicine*.

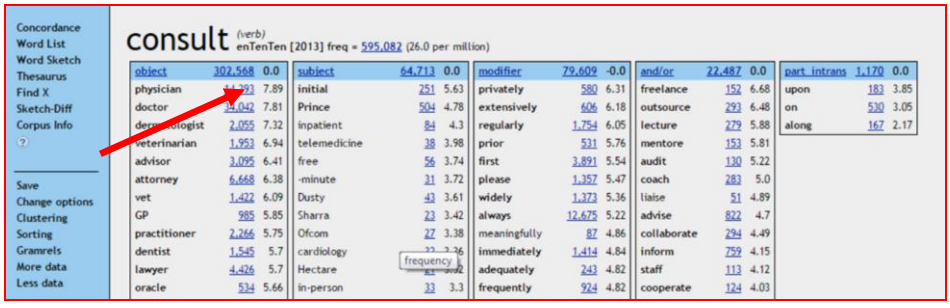
Although interesting from a theoretical point of view, the distinction between empirical collocations and multiword expressions is not of primary importance for our project, the reason being twofold. First, the available software programmes i.e. *SketchEngine* (http://www.sketchengine.co.uk) and *WordSmith Tools* (http://www.lexically.net/wordsmith) which are used for the computational processing of terminological data, do not distinguish between these two kinds of collocations. Secondly, both kinds are relevant to the current project. It would seem that empirical collocations are relevant on the conceptual level, in that members of a collocational set could be conceptually related, whereas phraseological collocations seem to be relevant more on a usage-related level. This issue will be further investigated below.

From a semantic point of view, collocation is represented by two related phenomena, i.e. semantic preference and semantic prosody, which both describe the statistically significant co-occurrence of a word with a group of other words, cf. Kübler and Pecman (2012: 188). Semantic prosody refers to the

measure with which the (affective) meaning of a word is coloured by its typical collocates, whereas semantic preference is described as the measure of co-occurrence between a word and a set of semantically related words. To briefly illustrate the notion of semantic prosody: the fact that the typical (abstract) objects that collocate with the verb *tolerate* in the *enTenTen* corpus are *dissent*, *disrespect*, *nonsense*, *intolerance, mediocrity* and *harassment*, the overall negative implication of the collocates taints the meaning of *tolerate* as being negative, resulting in a negative prosody for *tolerate*. Since it is assumed that the meaning of a term has been previously delimited by means of a definition, and that terms are supposed to be emotionally neutral, the perception that has hitherto prevailed that semantic prosody would play a lesser role within terminological work is understandable.

The following example — where the collocates of the verb *consult* are called up in a WordSketch, using the *enTenTen* corpus — is a good illustration of semantic preference. The top collocates for objects appearing with *consult* are *physician*, *doctor*, *dermatologist*, *veterinarian*, *advisor* and *attorney*, revealing the semantic preference of the verb *consult* to appear with objects sharing the semantic feature 'professional individual'. Kübler and Pecman (op. cit.) point out that semantic preference has recently aroused more interest in specialized languages, i.e. language for special purposes and has resulted in collocations being more commonly taken up in LSP dictionaries and term bases. They furthermore state that 'phenomena such as semantic prosody and preference would provide the user with complete and necessary information'.

Grammatical patterning represents another kind of collocation, i.e. the grammatical company that a word keeps. This is typically the kind of information which one would find in the WordSketches in SketchEngine. A WordSketch of the verb *consult* in the English *enTenTen* corpus, for example reveals that the grammatical pattern in which it most frequently appears is that of transitive verb, in which case it is followed by an object: the overall frequency of 'consult' as a transitive verb makes up almost 50% of its total occurrences. It also shows that the second most frequent grammatical pattern (24%) for 'consult' is to be followed by a prepositional phrase in which the preposition is most frequently 'with', followed by 'on', 'for', 'in' and 'to'. Compare the following screen shots from SketchEngine:



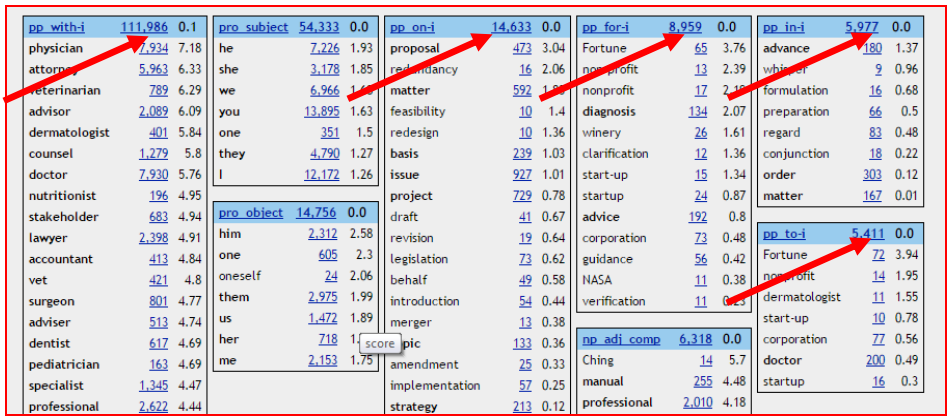


Figure 1: WordSketch of *consult*

Collocations and grammatical patterns in LSP dictionaries and terminological databases

In traditional paper LSP dictionaries collocations and information on the grammatical company that terms prefer are generally rather neglected, although as pointed out by L'Homme and Leroyer (2009), the addition of collocational information is regarded as extremely useful for specialized reference works, such as LSP dictionaries. Space constraints, coupled with the traditional view on terminology, i.e. that terms are context independent, are probably major contributing factors to the absence of these two information types in terminological products. With the advent of electronic lexicography, space is no longer of primary concern, but more importantly, the transition to corpus-based terminology not only provides access to huge amounts of data, but also opens up the possibility of semi-automatically extracting terminologically relevant data from corpora by means of corpus-query tools. The so-called modern approach to terminology furthermore places more emphasis on usage, with the use of real texts as primary sources of data, thus drawing on the importance of contextual information — which per definition includes collocational information — in satisfying the information needs of the user. The increased attention to collocational information in different kinds of terminology tools, whether it be online databases or electronic LSP dictionaries, can also possibly be ascribed to the influence of the work of Sinclair (2004) on extended units of meaning, and that of Hanks (2006), who argues that words only have meanings when they are put into context, thus establishing an association between word meaning and word use. A lexicographic application of Hanks' theory on Norms and Exploitation is the *Pattern Dictionary of English Verbs* (www.pdev.org.uk) in which users are offered prototypical syntagmatic patterns of meaning and use of each verb covered. This trend has also started to spill over into

the design of a variety of terminology tools. The *E-Advanced Learner's Dictionary of Verbs in Science* (DicSci) project, reported on by Alonso et al. (2011), the ARTES project of which one outcome is the compilation of an online bilingual LSP dictionary, see Kübler and Pecman (2012), and the *Dictionnaire fundamental de l'informatique et de l'Internet* (DiCoInfo), described in L'Homme et al. (2012), are all examples of terminological tools in which collocational information is provided. Lastly, Fuertes-Olivera also pays extensive attention to collocations in his set of online Spanish–English *Accounting Dictionaries.*

The importance of collocations in terminology in general, and for the OERTB project specifically

First, as pointed out by L'Homme (2006: 186), collocations are often unpredictable combinations, even in specialized language, and should therefore be treated in LSP dictionaries and/or term banks. This becomes especially important in a bilingual or multilingual situation where translation is one of the envisaged functions of the terminological tool. Collocations can potentially pose problems in translation, since they are often language specific and idiosyncratic. In Afrikaans for example, one 'picks up weight' (*tel gewig op*), whereas in English the verb which collocates with *weight*, is *gain*.

Secondly, collocations are domain dependent, which furthermore implies that collocations in general language with which the user may be familiar, may not apply in a specific subject field. For illustrative purposes, two small internet-based LSP corpora were compiled, one on climate change and one on film and drama studies. These were then queried by using SketchEngine (https://the.sketchengine.co.uk). In both these corpora, the term 'atmosphere' was thrown up as a keyword and is thus regarded as a key concept in both these subject fields. A list of collocate candidates was then drawn up for 'atmosphere' in a language for general purpose (LGP) corpus (*enTenTen* corpus, 40 mil sample), and in the two LSP corpora respectively. The emphasis here is on finding conceptually related collocates, therefore grammatical formatives or function words which can hardly be said to represent subject specific concepts were disregarded, even though they make out a sizable portion of the top collocate candidates. Therefore, only collocates with lexical and therefore conceptual content are listed in the table. The collocational span was set at 5 positions to left and right of the search node; the statistical measure used is T-score, one of the options offered by the SketchEngine.

LGP corpus (enTenTen)	**LSP corpus$_1$ (Climate change)**	**LSP corpus$_2$ (Drama and Film)**
create	*carbon*	*creates*
friendly	CO	*create*
relaxed	dioxide	upper
great	*Earth*	play (n)
place	greenhouse	render

Earth	gases	gothic
carbon	increase (v)	studying
warm	methane	filming
creating	oceans	setting (n)
people	released	research (n)

Table 1: Top ten raw collocates for 'atmosphere' in three different corpora

From the above, it is clear that the collocations for 'atmosphere' are indeed domain specific — there is no overlap between the collocates for 'atmosphere' in the two LSP corpora, and only a small overlap between the collocates found in the LGP corpus on the one hand, and those found in the two LSP corpora respectively. This is especially clear when looking for example at the verbal collocates of 'atmosphere' in the two LSP corpora. Verbs collocating with 'atmosphere' in the Climate change corpus are 'increase' and 'released', whereas 'create(s)', 'render', 'studying' and 'filming' are typical collocates in the Drama and Film corpus.

A third reason why collocations are important in an LSP environment is the fact that empirical collocations are assumed to be useful elements for conceptualizing a knowledge domain, as Fuertes-Olivera (2012) points out. Since the OERTB is especially aimed at assisting with conceptualization of key terms in different subject fields, this is an issue which needs special attention. Apart from providing straightforward collocational information, serious consideration should be given to provide information on the conceptual relationships existing between collocates by means of collocational networks, as described by Alonso, Millon and Williams (2011) and Williams (1998). Collocational networks are described as "statistically based chains of collocations, a web of interlocking conceptual clusters realised in the form of words linked through the process of collocation" (Alonso et al. 2011: 15). Williams (1998) argues that concepts central to a specific subject field are related, and that similar relational patterns can be identified in their surface constructs, i.e. words, or in our case, terms. Therefore, the frame of reference for any term is to be found in the lexical environment within which it appears, and which is revealed through collocation. It is assumed that concepts can be better grasped when they are presented within the environment in which they are used.

Collocational networks can be revealed by further processing of collocational information. As a starting point the strength of the association between each of the top x number (according to the keyness score) of single and multiword terms and their collocates are calculated, using e.g. Mutual Information (MI) as statistical measure. MI determines the strength of the association between two words: in a given finite corpus MI is calculated on the basis of the number of times the pair is observed together versus the number of times they appear separately. Each of the top ranking terms forms the node of a collocational network, and each collocate in turn is regarded as a node of a new collocational network. The collocational network is thus extended up to the point

where no more significant collocates are found. The end result would then be a network of related concepts, positioning each concept within the conceptual framework of at least a particular section of the special subject field.

The issue of empirical collocates is complicated by the fact that different statistical measures result in vastly different results. Compare the following table in which collocates for 'atmosphere' were drawn from the LSP corpus on climate change using MI (Mutual Information) score for the left-hand column and T-score for the right-hand column. Briefly explained, the difference between these two measures is as follows: The t-score is a measure not of the strength of association but the confidence with which we can assert that there is an association. MI is more likely to give high scores to totally fixed phrases whereas t-score will yield significant collocates that occur relatively frequently (http://wordbanks.harpercollins.co.uk/Docs/Help/statistics.html). The right-hand column is the raw collocate list, i.e. one in which the function words were retained:

	Freq	MI
travels	3	8.761
thicker	3	8.498
traps	3	8.275
amplify	4	8.176
heat-trapping	7	8.057
constantly	3	7.913
Winds	3	7.761
inert	3	7.761
coupled	4	7.591
overlying	3	7.498
composition	6	7.439

	Freq	T-score
the	428	19.512
in	214	14.02
.	191	12.447
of	168	11.855
,	164	11.108
and	106	9.178
into	76	8.648
to	90	8.48
2	66	7.917
carbon	64	7.876
is	73	7.85

Table 2: Collocates for 'atmosphere' in the Climate change LSP corpus according to MI and T-scores

According to the literature, cf. Clear (n.d.) and http://wordbanks.harpercollins.co.uk/Docs/Help/statistics.html, the choice of the measuring instrument for measuring the strength of collocational relationships depends also on the frequency of the items concerned. Clarifying the merits and the suitability of measuring instruments does not fall within the scope of this article, but is nevertheless something that needs to be clarified with experts. From the literature however, it would seem that MI is generally preferred in situations similar to the current one. Furthermore, it needs to be kept in mind that any list of collocates is only as good as the corpus it is based on, and it is possible that the two

LSP corpora compiled for the purposes of this article, namely to illustrate the value of collocational networks in LSP information sources are simply too small to render statistically significant results. (It needs to be mentioned here that within the parameters of the project we are planning on having at least 1 million word LSP corpora for each subject field.)

If, despite the concerns raised above, it is assumed that collocates which are generated by whatever statistical measure are indeed conceptually related to the search word, it is clear that sophisticated statistical processing such as reported on by Williams et al. (2012) would be necessary, and possibly also human intervention in the form of expert confirmation of conceptual relationships, to eventually present users with collocates in a format that satisfies their information needs. The typical format in which such a collocational network would be presented to the user, is a visual network, similar to what is found in Visuwords, a visual dictionary (http://visuwords.com/). The advantage of using such a visual network is that collocates can be visually represented, not in isolation, but as a complex network of semantic relationships which ultimately reveals their meaning, and thus aids with cognition — which is one of the main aims of the term bank. Currently, the possibility of adapting the software (TlTerm) which is used for the project to enable it to also generate such networks is being investigated. For the purposes of illustration, the term 'permafrost' and its collocational network as generated from the LSP corpus on climate change is used as an example:

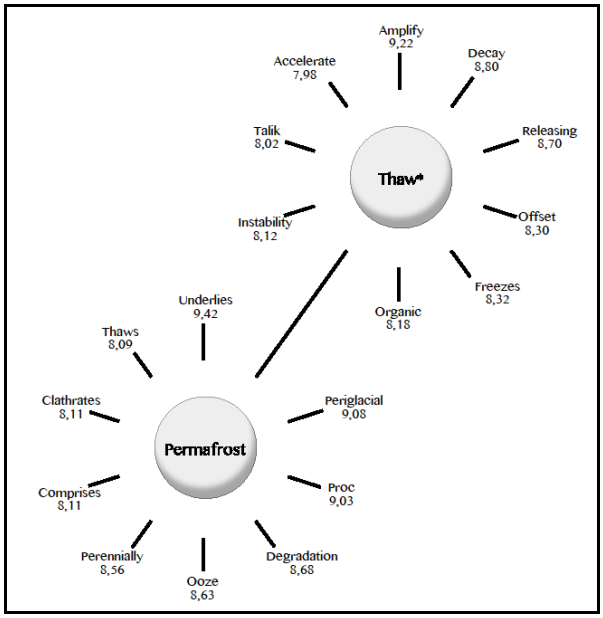


Figure 2: Collocational network for 'permafrost'

In the diagram above, the length of the lines connecting the different nodes are a reflection of the MI score, thus indicating the strength of association between the nodes. MI scores are displayed for illustrative purposes in the figure.

Further value can be added by specifying the type of relationship, e.g. 'kind of', 'part of', 'opposes' and 'is similar to' between collocates and a node and between different collocates. So, for example can the semantic relationship between 'thaw' and 'freeze' be indicated as being an antonymous one, etc. However, within the capacity, skills and time constraints of the OERTB project, this would be a rather ambitious undertaking. Even so, even if such an endeavour is not currently feasible, provision should be made for it in the conceptualization of the project, with a view to possible future implementation.

The last type of collocation to be discussed is grammatical patterning. Once again, this is important because (a) grammatical patterning in LGP is different from that of LSP, and (b) grammatical patterning also seems to be domain specific. Compare the following excerpt from a WordSketch generated for 'atmosphere' where the preference for specific grammatical patterns are revealed:

LGP corpus				**LSP_1 corpus**				**LSP_2 corpus**		
	Freq	Stat sign. score			Freq	Stat sign. score			Freq	Stat sign. score
object_of	427	2.5		object_of	52	1.3		object_of	23	4.2
subject_of	148	1.3		subject_of	85	2.5		subject_of	2	0.5
pp_obj_into	50	19.2		pp_obj_into	65	74.3		pp_obj_into	2	16.6

Table 3: WordSketch for 'atmosphere' across the three corpora

From the above it is clear that there is a bigger preference for 'atmosphere' to appear as the object of a verb in the Drama and Film corpus (LSP_2 corpus) than in any of the other two corpora. On the other hand, there does not seem to be a big tendency for it to appear as the subject of verbs in this corpus. Perhaps most significant is the preference for 'atmosphere' to appear as a prepositional object after 'into' in the Climate change corpus (LSP_1 corpus) — obviously because things are released or emitted into the atmosphere, or they escape into the atmosphere.

Apart from reflecting preference for grammatical patterns WordSketches also provide insight into which specific lexical items appear in these grammatical patterns, introducing a further level of collocational information. This excerpt reflects on a lexical level which modifiers tend to co-occur with 'atmosphere' in the various corpora:

It is clear that a relatively restricted set of modifiers co-occur with 'atmosphere' in the climate change corpus, with double the number in the drama and film corpus. Some of the modifiers appearing in the latter corpus also reveals something about the semantic prosody — the modifiers 'Gothic', 'mystical' (?), 'agoraphobic', 'chilling', 'deathly', 'cold', 'dark' reveal a clustering of modifiers

which lend a negative prosody associated with atmosphere — information that may be useful to users, not necessary on a conceptual level, but rather on the pragmatic side.

LGP corpus				LSP_1 corpus				LSP_2 corpus		
	Freq	Stat sign. score			Freq	Stat sign. score			Freq	Stat sign. score
modifier	676	1.5		modifier	47	0.3		modifier	19	1.2
---				---				---		
cozy	10	8.27		Martian	6	11.03		upper	3	11.63
relaxing	8	8.08		upper	7	10.95		Gothic	2	10.75
Martian	5	7.81		inert	2	10		genteel	1	10.68
homely	5	7.78		standard	2	9.66		mystical	1	10.68
friendly	25	7.76		low	7	8.78		agoraphobic	1	10.6
festive	6	7.76		warm	4	8.07		liberal	1	10.54
laid-back	4	7.45		global	2	6		chilling	1	10.47
controlled	5	7.39						deathly	1	10.47
casual	7	7.23						cold	1	10.14
intimate	6	7.22						intense	1	10.09
oxygen-deficient	3	7.18						northern	1	9.95
IDLH	3	7.17						British	1	9.71
inert	3	7.05						necessary	1	9.02
vibrant	6	7.05						general	1	9.02
cosy	3	6.99						dark	1	8.89
calming	3	6.9								
↓ ↓ ↓										

Table 4: Modifiers co-occurring with 'atmosphere'

In the last instance, WordSketches provide yet another relatively simple procedure to retrieve collocations. In a WordSketch, items with a high frequency of occurrence within a particular grammatical pattern are clickable, thus revealing the concordance lines in which the search word, in this case 'atmosphere' appears together with its collocate. This could potentially provide additional information with regard to usage. Compare the following example of concordance lines containing 'create', which is one of the verbs which prefers 'atmosphere' as an object in the LSP_2 corpus:

Word sketch item 9 (30.2 per million)

file2035688	the lines he reveals character, *creates* atmosphere , indicates spiritual action, and advances
file2035688	is, if it re- veals character, *creates* atmosphere , or advances the story. Obviously, Paolo
file2035688	summer ever came here," go far to *create* an atmosphere of gloom. In this first act is the creation
file2035688	can at once supply environment, *create* atmosphere , paint word-pictures. That sharp contrast
file2035688	play. Ob- serve the methods used to *create* atmosphere . An effective bit of " living stage-direction
file2035688	suggestion of witchcraft, all tend to *create* an atmosphere proper for the unfolding of the plot and
file2035682	is attempting to unravel. This *creates* an atmosphere that is more befitting Bram Stoker's work
file2035665	the "fog" and the "filthy air", *creates* an atmosphere of despair and Shakespeare is essentially
file2035643	displayed an early talent for *creating* atmosphere with imagery, including a photograph (

Figure 3: Collocates for 'create' and 'atmosphere' in the LSP_2 corpus

What is noticeable in this example, is that when 'create' is used as the verb selecting 'atmosphere' as an object, the object is never preceded by a definite article. Apart from providing collocational information, these concordance lines therefore also give guidance with regard to usage. This specific example is particularly relevant for users who have an African language as home language. The use of articles in English is often problematic for such users, since the African languages do not distinguish the grammatical category 'article', consequently the use or non-use of articles in English often causes confusion due to language interference. Having access to this kind of information would assist users specifically with text production and translation.

Presentation of and access to collocational information

With regard to the presentation of collocational information, two options are identified, i.e. implicit or explicit presentation. Implicit presentation would imply that the collocational data extracted from the LSP corpora according to the strategies described above would only be of value to the terminologist populating the database. In the case of explicit presentation, collocational information would be presented as such to the user, who will have the option of accessing this information by means of a search option. When collocational information is implicitly utilized, the terminologist would typically use these data to select usage examples which incorporate as much of the collocational data retrieved from the LSP corpus as possible. To illustrate: when selecting a usage example for the term 'atmosphere' within the drama and film domain, the terminologist will need to take the following into consideration:

— It is most often used as the object in a sentence (grammatical patterning)
— It frequently appears without an article, or with an indefinite one (grammatical patterning)
— The verb 'create' is one of the verbs which collocate with 'atmosphere'
— Many of the modifiers which co-occur with 'atmosphere' contribute to the expression of a negative semantic prosody.

These data would assist the terminologist in selecting an appropriate example from the corpus, e.g.

> The homes are cast with an unfriendly sterility that can create a chilling, agoraphobic ***atmosphere***

The disadvantage of treating collocations implicitly is that is does not provide sufficient collocational guidance — the user does not know whether *create a(n) atmosphere* is a frequent combination, or whether it is a mere coincidence. Users are oblivious to the fact that the example sentence illustrates both common usage of the term 'atmosphere' and a frequent collocation. Furthermore, no

guidance is provided as to the negative semantic prosody which is often associated with the term 'atmosphere' in a drama and film context.

When presenting collocational information in an explicit manner, due consideration should be given as to whether this information should be displayed by default, in other words, whether on carrying out an initial search collocational information will automatically be displayed to the users, or whether they will have the option of accessing the information by means of a further search. Taking the target user of the OERTB as well as the function of the term bank into consideration, the designer of the user interface would be well-advised to heed the possibility of data overload. Collocational information should therefore be made available as an optional, additional search function which can be accessed by means of clicking on a dedicated button or tab. Care should furthermore be taken that the name of the button or tab which gives access to the collocational information should be transparent, making it clear to the users what kind of information they would find by clicking on it. Labelling such a button or tab as 'Collocations' would probably have little meaning for our envisaged users. Choosing a transparent label, such as 'Term in context' or 'Frequent combinations' rather than for example the neutral 'See more' has the advantage of providing the user with additional guidance to finding the required information.

Consideration should furthermore be given to allow users to directly access collocations, i.e. they must be able to search for a collocation, without having to access the collocation via the lemmatic address of the term appearing as part of the collocation. In the data base, collocations would then in effect be treated as multi-word terms, and can therefore be given the full treatment also given to single word lemmata.

Conclusion

Provision of collocational data in any kind of terminological environment, whether it be LSP dictionaries or term banks has been sadly neglected, especially within the South African context. In this article it has been illustrated that in order to fulfil the information needs of the envisaged user, due consideration must be given to the provision of two kinds of collocational information, i.e. semantic prosody and semantic preference on the one hand, and grammatical collocation on the other. It has been illustrated that collocations are often unpredictable combinations, and since we envisage our term bank to also provide for translation needs, collocational information would be necessary. Collocations are furthermore domain specific and can assist with conceptualization within a particular subject field, thus forming a necessary component of information to be presented to the user. In order to reach this goal, full utilization must be made of data that can be extracted semi-automatically from corpora, and presented in such a way that it maximally satisfies the needs of the user. Easy access to collocational information is therefore of primary importance.

Bibliography

A Guide to Statistics. n.d. Retrieved from http://wordbanks.harpercollins.co.uk/Docs/Help/statistics.html.

Alonso, A, C. Millon and G. Williams. 2011. Collocational Networks and their Application to an E-Advanced Learner's Dictionary of Verbs in Science (DicSci). Kosem, I. and K. Kosem (Eds.). 2011. *Electronic Lexicography in the 21st Century: New Applications for New Users. Proceedings of eLEX 2011, November 2011, Bled, Slovenia*: 12-22. Ljubljana: Trojina, Institute for Applied Slovene Studies.

Bergenholtz, H. and T.J.D. Bothma. 2011. Needs-adapted Data Presentation in e-Information Tools. *Lexikos* 21: 53-77.

Clear, J. n.d. *t-score and Mutual Information Score from Birmingham Corpus Website*. http://wordbanks.harpercollins.co.uk/Docs/Help/statistics.html.

Evert, S. 2007. *Corpora and Collocations*. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.159.6220&rep=rep1&type=pdf.

Fuertes-Olivera, P. 2012. Lexicography and the Internet as a (Re-)source. *Lexicographica* 28(1): 49-70.

Hanks, P. 2006. Metaphoricity is Gradable. Stefanowitsch, A. and S.Th. Gries (Eds.). 2006. *Corpus-based Approaches to Metaphor and Metonymy*: 17-35. Berlin: Mouton de Gruyter.

Kübler, N. and M. Pecman. 2012. The ARTES Bilingual LSP Dictionary: From Collocations to Higher Order Phraseology. Granger, S. and M. Paquot (Eds.). 2012. *Electronic Lexicography*: 187-210. Oxford: Oxford University Press.

L'Homme, M.-C. 2006. The Processing of Terms in Dictionaries: New Models and Techniques. A State of the Art. *Terminology* 12(2): 181-188.

L'Homme, M.-C. and P. Leroyer. 2009. Combining the Semantics of Collocations with Situation-driven Search Paths in Specialized Dictionaries. *Terminology* 15(2): 258-283.

L'Homme, M.-C., B. Robichaud and P. Leroyer. 2012. Encoding Collocations in DiCoInfo: From Formal to User-friendly Representations. Granger, S. and M. Paquot (Eds.). 2012. *Electronic Lexicography*: 211-236. Oxford: Oxford University Press.

Park, Mi Sun. 2013. Code-switching and Translanguaging: Potential Functions in Multilingual Classrooms. *Working Papers in TESOL & Applied Linguistics* 13(2): 50-52.

Sinclair, J. 2004. *Trust the Text Language, Corpus and Discourse.* London: Routledge.

Smadja, F. 1993. Retrieving Collocations from Texts: Xtract. *Computational Linguistics* 19(1): 143-177.

Williams, G. 1998. Collocational Networks: Interlocking Patterns of Lexis in a Corpus of Plant Biology Research Articles. *International Journal of Corpus Linguistics* 3(1): 151-171.

Williams, G., C. Millon and A. Alonso. 2012. *Growing Naturally: The DicSci Organic E-Advanced Learner's Dictionary of Verbs in Science*. Fjeld, Ruth Vatvedt and Julie Matilde Torjusen (Eds.). 2012. *Proceedings of the 15th Euralex International Congress, 7–11 August 2012*: 1008-1013. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo.

Electronic term banks

InterActive Terminology for Europe. The EU's Multilingual Term Base. [Online] Available at: http://iate.europa.eu/. [Accessed 10 August 15].

Le grand dictionnaire terminologique. Office québécois de la langue française. [Online] Available at: http://www.granddictionnaire.com/. [Accessed 18 August 15].

TERMIUM Plus®. The Government of Canada's Terminology and Linguistic Data Bank. [Online] Available at: http://www.btb.termiumplus.gc.ca. [Accessed 30 August 15].

Software

SketchEngine. Kilgarriff, A., V. Baisa, J. Bušta, M. Jakubíček, V. Kovář, J. Michelfeit, P. Rychlý and V. Suchomel. 2014. The Sketch Engine: Ten Years On. *Lexicography* (2014): 1-30. https://www.sketchengine.co.uk/.

TlTerm. TshwaneDJe. Human Language Technology. http://tshwanedje.com.

WordSmith Tools. Scott, M. 2012. WordSmith Tools version 6, Stroud: Lexical Analysis Software. http://www.lexically.net/wordsmith/.

http://lexikos.journals.ac.za

Preparing an Online Dictionary of Business Communication: From Idea to Design*

Sven Tarp, *Department of Afrikaans and Dutch, University of Stellenbosch, Stellenbosch, South Africa* and *Centre for Lexicography, Aarhus University, Aarhus, Denmark (st@bcom.au.dk)*

Abstract: This article discusses the plan for an online English–Spanish dictionary of business communication based on the function theory. After a short overview of lexicographical traditions within this field, it focuses on the philosophy behind the overall dictionary concept, the methods applied to develop this concept, some of the requirements for the corresponding database, as well as the foreseen Internet user interface which may lead the user to ten completely different articles for each word entered in the search field.

Keywords: ONLINE DICTIONARY, BUSINESS COMMUNICATION, BUSINESS DICTIONARY, SPECIALIZED LEXICOGRAPHY, FUNCTION THEORY

Opsomming: Die voorbereiding van 'n aanlyn sakekommunikasiewoordeboek: Van idee tot ontwerp. In hierdie artikel word die plan bespreek vir 'n Engels–Spaans aanlyn sakekommunikasiewoordeboek gebaseer op die funksieteorie. Ná 'n kort oorsig oor leksikografiese tradisies binne hierdie veld word daar gefokus op die filosofie agter die algehele woordeboekkonsep, die metodes wat toegepas is om hierdie konsep te ontwikkel, sommige van die vereistes vir die ooreenstemmende databasis, sowel as die voorsienbare Internet-gebruikerskoppelvlak wat die gebruiker kan lei na tien verskillende artikels vir elke woord wat in die soekveld ingetik word.

Sleutelwoorde: AANLYN WOORDEBOEK, SAKEKOMMUNIKASIE, SAKEWOORDEBOEK, VAKLEKSIKOGRAFIE, FUNKSIETEORIE

1. Introduction

In this article, we will discuss the plan for an online dictionary designed to assist English and Spanish business communication, a dictionary which will be produced in collaboration with experts and scholars from various countries. We will focus on the philosophy behind the overall dictionary concept, the methods applied to develop this concept, some of the requirements for the corresponding database, as well as the envisaged user interface which may lead the user to 10 completely different articles for each word (string) entered in the search field.

* This article was presented as a paper at the Twentieth Annual International Conference of the African Association for Lexicography (AFRILEX), which was hosted by the University of KwaZulu-Natal, Durban, South Africa, 6–8 July 2015.

The dictionary project is based on the functional principle combined with the option to individualize, and is inspired by the experience obtained from a similar project including four printed bilingual dictionaries between Danish, on the one hand, and English and Spanish. Before explaining the project in further details, we will give a brief overview of the history of business dictionaries involving English and Spanish.

2. Lexicographical overview

In the European context, there is a several hundred years' old tradition of compiling dictionaries related to the broad field of business and economics, cf. Besomi (2011). Among the early dictionaries are those produced by Savary des Bruslons (1723), Postlethwayt (1751–55), Rolt (1756) and Mortimer (1766) in the 18th century. These dictionaries were generally designed to transmit knowledge about trade and commerce and explain the many terms related to these activities, cf. Tarp and Bothma (2013: 224). This purpose was also the main idea behind Virio's *Collección alfabética* of British duties, tariffs, law extracts, mandates, regulations and rulings, published in Madrid in 1792, cf. Astigarraga and Zabalza (2007: 36-38).

Although trade and commerce between the European countries — and overseas — developed relatively fast in the 18th century (a fact reflected in the above mentioned dictionaries), it was not until the 19th century that the first dictionaries designed to assist interlingual communication saw the light. Strangely enough, from their very appearance these communicative dictionaries seem to have developed along a separate line "with no or only little interconnection" to the increasing number of cognitive dictionaries designed to transmit knowledge about business and economics, cf. Tarp (2015a: 188).

Among the first known dictionaries designed to assist business communication between the two languages in question is Veitelle's trilingual English–Spanish–French *Mercantile Dictionary* from 1864, which, according to the subtitle, includes "a Complete Vocabulary of the Technicalities of Commercial Correspondence, Names of Articles of Trade, and Marine Terms". An extract from this dictionary where the three languages are treated in separate columns can be seen in Figure 1.

Abandon.	abandonner; délaisser.	abandonar; dejar.
Abandonment.	délaissement, m.	abandono, m.; dejacion, f.
Abatement.	réduction; remise, f.	rebaja; reduccion, f.
Abide the consequences.	subir les conséquences.	sufrir las consecuencias.
Aboard.	à bord.	á bordo.
About.	environ.	cerca de.
to go about (*naut.*).	virer de bord.	virar de bordo.
Abstract.	exrait; relevé, m.	extracto; resúmen, m.
— of account sales.	relevé de compte de vente.	extracto de cuenta de venta.
Absynth.	absinthe, f.	agenjo, m.
Abundant (rice is).	les riz sont abondants.	abunda el arroz.

Accept.	accepter.	aceptar.
Acceptance.	acceptation, f.	aceptacion, f.
general —.	acceptation totale.	aceptacion total.
partial —.	acceptation partielle.	aceptacion parcial.
absolute —.	acceptation pure et simple	aceptacion pura y simple.
— in blank.	acceptation à découvert; en blanc.	aceptacion á descubierto; en blanco.

Figure 1: Extract from the trilingual Mercantile Dictionary (1864)

Another early plurilingual business dictionary to be mentioned here is Graham and Oliver's *The Foreign Traders' Dictionary of Terms and Phrases in English, German, French, and Spanish*, published by MacMillan in 1906. As can be seen in Figure 2, the four languages treated in this dictionary are also presented in separate columns which cover two opposite pages here.

Trade, commerce, business	Der Handel, das Geschäft	Le commerce, les affaires (f.pl.)	El comercio, los negocios
barter	der Tauschhandel	le commerce d'échange, le troc	la permuta, el trueque
exchange, interchange	der Austausch, Tauschhandel	l'échange (m.), le troc	el trueque, el cambio
traffic	der Handelsverkehr, der Handel	le trafic	el tráfico
monopoly	das Monopol	le monopole	el monopolio
Home trade	Der inländische Handel, Binnen-	Le commerce (à l') intérieur	El comercio interior [del país]
foreign trade	der ausländische Handel [handel]	le commerce (à l') étranger	el comercio exterior
import trade	der Einfuhrhandel	le commerce d'importation	el comercio de importación
export trade	der Ausfuhrhandel	le commerce d'exportation	el comercio de exportación
carrying [shipping] trade	das Rhederei-Geschäft	les affaires (f.) maritimes	los negocios marítimos
transit [transhipment] trade	der Transitverkehr	le commerce de transit	el comercio de tránsito
coasting trade	der Küstenhandel	le cabotage	el cabotaje
Wholesale trade	Der Grosshandel, das Engrosgeschäft	Le commerce de [en] gros	El comercio al por mayor
retail trade	der Kleinhandel, das Detailgeschäft	le commerce de [en] détail	el comercio al por menor

Figure 2: Extract from the quadrilingual *The Foreign Traders' Dictionary* (1906)

Figures 1 and 2 clearly indicate that the lexicographical content and architecture of the first interlingual business dictionaries were rather simple compared to modern standards. Since then various business dictionaries have been published covering both English and Spanish, most of them now bilingual. Noteworthy in this respect is López and Watt's bilingual *Oxford Business Spanish Dictionary*, published by Oxford University Press in 2002 and now only available online. Figure 3 shows an article from the Spanish–English part of the printed version of this dictionary:

cerrar 1 *vt* (tienda, negocio) close;
(definitivamente) close down; (salida) seal off;
~ un acuerdo close a deal; **~ una posición** close a position; **~ un trato** make a deal; **~ una venta** complete *o* close a sale
2 *vi* close down

Figure 3: Article from the *Oxford Business Spanish Dictionary*

This article reflects the lexicographical development that has taken place since the dictionaries compiled by Veitelle (1864) and Graham and Oliver (1904). Especially useful is the provision of meaning discrimination allowing the users to select the needed equivalent among the ones proposed. However, based on the experience from students' use of a similar Spanish–Danish business dictionary, the syntactic data (*vt* and *vi*), even if written in full form, is probably too abstract to understand for many users without background in linguistics. A more user-friendly solution would therefore have been that these data were exemplified and explained in greater details as the words' syntactic properties constitute one of the main challenges in the redaction of foreign-language texts.

Today the *Oxford Business Spanish Dictionary* is only available for subscribers on the Internet. Here it competes with a growing number of free-access business dictionaries, many of which reflect the same poor article architecture employed in the first printed dictionaries more than hundred years ago. One such example is Andy D. Miles' *Business English to Spanish Glossary* which, according to the webpage, contains 10,000 lemmata. An extract from this online dictionary can be seen in Figure 4.

bill	(1) s. billete (2) s. factura, cuenta (3) s. pagaré, letra, libranza (4) s. anteproyecto de ley (5) v. facturar, cobrar.
bill book	libreta de facturas, talonario de facturas, facturero (Arg.).
bill broker	corredor de letras, casa de descuento de letras.
bill diary	registro de letras, libro diario de letras.
bill for collection	(1) factura a cobrar (2) letra a cobrar, pagaré a cobrar.
bill market	mercado de letras.
bill of debit	pagaré.

Figure 4: Extract from *A-D. Miles' Business English to Spanish Glossary*

What strikes the eye here is that the dictionary does not offer meaning discrimination and thus creates a new and completely unnecessary problem for the user. Neither does it provide data on gender, inflection, syntactic properties or collocations, all of them data that are absolutely necessary to assist an English first-language speaker translating or producing a correct business text in Spanish. This dictionary is emblematic and typical for a large number of specialized dictionaries that are now freely available on the Internet and compiled by people who may or may not be specialists within their respective fields, but who don't seem to have the lexicographical background needed to produce dictionaries of a quality required by the users in the 21st century, cf. Caruso (2011).

To finish the picture, it should be mentioned that there also are a number of monolingual dictionaries designed to assist reception and production of business texts. One example of this is the *Cambridge Business English Dictionary* which is available both in a printed version and online. According to the advertising, this dictionary "has over 35,000 words, phrases, and meanings, and includes business-specific vocabulary" as well as "help with how to use English naturally in business situations like meetings, conference calls, and emails". However, this and other monolingual business dictionaries will not be discussed here.

3. Motivation

Various facts indicate the need for a new online dictionary for business and other types of professional communication between English and Spanish:

— In a still more internationalized world there is a growing demand for a high-quality dictionary within this field.

— English and Spanish are among the three most spoken languages in the world and used as official languages in more countries than any other language.

— Only part of the required lexicographical information can be retrieved from general dictionaries in the two languages.

— The many specialized dictionaries designed to provide knowledge about disciplines relevant to this field do, as a rule, not cater for their users' communicative needs.

— Most of the existing online dictionaries designed to assist English and Spanish business communication are of very low lexicographical quality.

— None of the existing online dictionaries have completely adapted to the Internet and made full use of the new technologies and techniques available to deliver a more personalized lexicographical product.

With this in mind, the author of this article was contacted by a publishing house with a view to preparing a project for an online English–Spanish business dictionary. The project is inspired by the experience obtained from a similar project including four printed bilingual dictionaries between Danish, on the one hand, and English and Spanish, on the other, as well as the experience from more than 25 years of teaching foreign-language business communication. However, although inspired by past experience, the new dictionary concept has to be adapted to the Internet and make use of the new technologies and techniques, and therefore needs to be designed from scratch. The main features of the new concept will be presented in the following sections.

4. Subject field and foreseen user group

Business communication is an amorphous field and, as such, very difficult to delimit lexicographically. This is reflected in the fact that existing business dictionaries contain lemma stocks that vary greatly, much more than can be explained by "normal" competition. It seems that each author has his or her own idea of how the field should be embraced. In this respect, the project will be based on the experience accumulated at the Aarhus School of Business where interlingual business communication has been taught as an academic discipline for more than half a century. Here, the field or "discipline" of business communication is traditionally understood as divided into several genres like mailing, sales letters, enquiries, offers, orders, deliveries, payments, complaints, etc. To this can also be added more recent web-based "about us" texts like company history, visions, mission and corporate social responsibility.

All these genres have their own special vocabulary, characteristics and features which may vary from "daily" language in terms of technicality, formality and requirements to politeness. The corresponding texts may include terms related to any product or business matter, e.g. those Veitelle (1864) called "Names of Articles of Trade" (see above). Such terms can obviously not be incorporated into a modern dictionary focussing on business communication as each of them may appear only in very few texts, whereas they, in their totality, can be counted in tens of thousands, if not more. The tradition from past centuries of lemmatizing the names of the products that are traded cannot be maintained in today's industrial consumer society. However, the above-mentioned texts also contain a large number of frequently used terms related to technicalities such as packaging, transport, payment, banking, insurance, sponsoring, contracts, legal matters, advertising, etc. These terms have to be treated in a dictionary of business communication.

As such, the planned dictionary will cover a specific vocabulary somewhere in the grey zone between general and specialized language as it has been defined by Fuertes-Olivera and Tarp (2014: 7). On the one hand, it will contain frequently used specialized terms belonging to various relevant disciplines, and on the other hand, it will include general-language words and expressions frequently employed in business communication, sometimes with a specific usage.

The main purpose of the dictionary is to assist business people, secretaries, civil servants, translators, students and other possible users who engage in business and other types of professional foreign-language communication in Spanish and English. This group of envisaged users, which is basically composed of first-language speakers of English and Spanish, is very heterogeneous and may have different lexicographical needs that should be catered for in the dictionary design.

5. Basic philosophy and main functions

Taking into account the heterogeneous user group described in the previous

section, the dictionary project is based on the functional principle combined with the option to individualize, i.e. a lexicographical Model T Ford taking the first modest steps towards a future Rolls Royce according to the classification proposed by Fuertes-Olivera and Tarp (2014: 16).

Today, the main relevant user situation is the production of business texts directly in the foreign language or based upon an outline in the source language (mostly done by secretaries). Direct translation of this type of text is mainly restricted to the education system, but as students are also part of the foreseen user group, the support to both foreign-language (L2) text production and L1–L2 translation make up the two main functions of the planned dictionary. Lexicographically, these two functions require very similar treatment. However, as the corresponding data, with only a few additions, may also serve L2 text-reception, L2–L1 translation and L1 text-production, these functions will also be covered by the project, although considered of secondary importance. Taking into account that the user group is composed of first-language speakers of both English and Spanish, it adds up to a totality of 10 lexicographical functions expressed in a combination of monolingual and bilingual lexicographical solutions.

The basic philosophy underpinning the project is that the traditional bilingual dictionary which contrasts two different languages is inconvenient. This kind of dictionary usually provides too little data to fully assist the users in foreign-language text production. And if it does furnish the needed data, the inevitable result would be that many articles would be filled with too much data, thus creating a new problem for its users, namely information stress due to data overload.

With this in mind, the planned project will basically be composed of two monolingual dictionaries with the possibility of bilingual access from L1 to L2 and explication of L2 (by means of equivalents and notes in L1). It can therefore be used not only to assist L2 production and L1–L2 translation, but also L2–L1 translation, L2 reception and even L1 production, with the two former functions being the prioritized ones.

It is foreseen that the dictionary will contain explications of "difficult" technical terms like *letter of credit*, *incoterms*, and *bill of lading* as well as short definitions in case of polysemy but generally it will not be designed to serve L1 text reception. The reason for this decision is that most first-language users will generally know the meaning of the majority of words belonging to the "grey zone" between general and specialized language. However, if this proves to be false or only partly correct, additional definitions can always be included in a later phase of the compilation process.

6. How to satisfy foreign-language text production

As already indicated, the needs related to L1–L2 translation and L2-text production require almost the same lexicographical response, although there may

be some minor differences as the latter, in contrast to the former, is not source-language dependent and, hence, is subjected to more linguistic variation, cf. Tarp (2004). In the following, the two user situations will be treated as one, exemplified with the problems and needs related to L2 production. In this respect, a person writing a business text in a foreign language may experience three types of problems eventually giving rise to a dictionary consultation:

1. The person does not know the word to be used in L2.
2. The person knows a L2 word but is not sure whether it can be used with the specific meaning and in the specific context.
3. The person knows the word but has doubts about orthography, gender, inflection, syntactic properties, collocations, synonyms, antonyms, fixed expressions, etc.

Any first-language speaker of L1 may experience problems related to one of the three scenarios described, but the person with the highest L2 proficiency level is most likely to have problems of type 3 whereas learners at beginners' level most often will face problems of type 1, etc. The three types of problem require three different lexicographical solutions:

> Problem 1: An **L1-L2 solution** offering L2 equivalents and meaning discrimination.
>
> Problem 2: An **L2 solution** with meaning explications which could be even more advantageous to the user if they were given by means of L1 equivalents, i.e. a **L2-L1 solution.**
>
> Problem 3: An **L2 solution** including data on orthography, gender, inflection, syntactic properties, collocations, synonyms, antonyms, fixed expressions, etc.

The user with problems of type 3 may be completely satisfied with the proposed solution. The users with problems of type 1 and 2, however, may experience additional needs once they have chosen or confirmed the word to be used, as they may have doubts about gender, inflection, syntactic properties, collocations, synonyms, antonyms, fixed expressions, etc. Such needs can best be solved by:

— A **L2–L1 solution** including data on gender, inflection, syntactic properties, collocations, synonyms, antonyms, fixed expressions, etc.

The reason why an L2–L1 solution with this content would be the best additional solution here is that L1 data, apart from meaning explication, will make it easier for users with low and intermediate L2-proficiency level to navigate in the displayed article and choose the relevant collocations or opt for the most convenient syntactic combinations, among others. In addition, although a user with high proficiency level may be perfectly satisfied with the L2 solution pre-

sented above, an alternative L2-L1 solution as the one proposed as additional to the other user categories would not be bad either (except in case of data overload). For the convenience of this user type, a special button to hide (and unfold) the L1 data could easily be included in the user interface. The article appearing when hiding the L1 data would then be the same as the one a first-language speaker of L2 would get when looking for assistance in text production in this language.

Consequently, the overall concept will, in accordance with the underlying philosophy, consist of two monolingual dictionaries in English and Spanish, each of them with a bilingual dimension expressed in the possibility of access from L1 to L2 and explication of some L2 items in L1. In this way, the project will with elegance and flexibility serve both Spanish and English first-language users of various proficiency levels, not only when they need help in connection with L1–L2 translation and L2 production, but also in relation to L2 reception, L2–L1 translation and L1 production.

7. Lexicographical database

Once the situations to be covered by the planned business dictionary had been established, the next step was to determine the possible types of need which the foreseen user group may experience in these situations. There are various methods to perform this task, cf. Fuertes-Olivera and Tarp (2014: 53-57). In this project, the method chosen was the functional one, i.e. deduction based on the lexicographer's own experience and knowledge of the subject field and the envisaged users' problems when working with foreign-language business texts, cf. Tarp (2013: 148-149). Deduction should not be confused with introspection as it requires a combination of lexicographical expertise and real subject-field knowledge based upon academic studies and research. In this particular case it includes knowledge obtained from teaching business communication and marking thousands of tasks and exercises in searching information as well as subsequent discussions with the students.

The result of the deduction process was a detailed list of user needs related to the different situations covered by the dictionary. Taking this list as the point of departure, it was relatively easy to establish which data categories and which relations between them would be required to cover the detected needs, a precondition for preparing the lexicographical database to be employed in the project.

Each dictionary project has its own personality and it is therefore recommendable to design a lexicographical database from scratch for each new project. The pre-fabricated dictionary writing systems available on the market are, as a rule, too inflexible to be used in projects like the one described here. It is especially important that the corresponding user interface employed by the lexicographers to compile the dictionary is correctly understood, i.e. as a means of production. This requires, among other things, 1) that it guarantees high produc-

tivity and high quality (allowing the preparation of as many correct data as possible in the shortest possible time) and 2) that it, in order to achieve these goals, is as comfortable and easy as possible for the lexicographer to work with, cf. Tarp (2015b).

Apart from this, the overall requirement from the database is, on the one hand, that it (and the dictionary as such, understood as the totality of data that may be displayed in the totality of hypothetic consultations) should include *as much data as possible* relevant to the topic as well as to the detected user needs; and on the other hand, that the individual articles to be displayed on the screen should contain *as little data as possible* in order to avoid information overload, i.e. just as much data as required to satisfy the user's needs in each case, no more no less.

In order to achieve this, the classical method of analysis and synthesis was used. By means of this method, the data categories to be covered were first separated into their smallest relevant parts (i.e. those data items which eventually may be displayed separately on the screen) and then combined in order to establish their relevant mutual relations. All this had then to be communicated to the programmer of the lexicographical database (and of the lexicographers' interface). As this communication takes places between specialists from two different disciplines with different terminologies, it entails a lot of possible misunderstandings and may take unnecessarily long if a common language is not found. In this respect, the main terminology used was *data fields* and their mutual *relations*, i.e. the structures to be represented in the lexicographical database, cf. Bergenholtz and Nielsen (2013) and Tarp (2015b). No specific terms for the corresponding structures were ever used.

In this particular case, the communication was both oral and written using a diagram as the one reproduced below:

1. English lemma
    - 1.1. Homonymy number
    - 1.2. Polysemy number
    - 1.3. UK/US
    - 1.4. Full form
        - 1.4.1. Full form note in English
        - 1.4.2. Full form note in Spanish
    - 1.5. Abbreviation
        - 1.5.1. Abbreviation note in English
        - 1.5.2. Abbreviation note in Spanish
    - 1.6. English definition
        - 1.6.1. Link to English Internet page
        - 1.6.2. English text to link

1.7. Spanish definition
    1.7.1. Link to Spanish Internet page
    1.7.2. Spanish text to link
1.8. Part of speech in English
1.9. Part of speech in Spanish
1.10. Inflection
1.11. Syntactic mini-rule(s)
    1.11.1. Example sentence in English
    1.11.2. Example sentence translated into Spanish
    1.11.3. Mini-rule note in English
    1.11.4. Mini-rule note in Spanish
1.12. Synonym(s)
    1.12.1. UK/US
    1.12.2. Synonym note in English
    1.12.3. Synonym note in Spanish
1.13. Antonym(s)
    1.13.1. UK/US
    1.13.2. Antonym note in English
    1.13.3. Antonym note in Spanish
1.14. Spanish equivalent(s)
    1.14.1. Equivalent note in English
    1.14.2. Equivalent note in Spanish
1.15. Collocation(s)
    1.15.1. UK/US
    1.15.2. Spanish translation
    1.15.3. Collocation note in English
    1.15.4. Collocation note in Spanish
1.16. Fixed expression(s)
    1.16.1. UK/US
    1.16.2. Spanish translation
    1.16.3. Fixed expression note in English
    1.16.4. Fixed expression note in Spanish
1.17. Memo field
1.18. Reserve field

1.19. Reserve field

1.20. Reserve field

1.21. Reserve field

Diagram 1: Data fields in the English part of Business Dictionary

The diagram presented above was initially given to the programmer in charge of designing the lexicographical database. As can be seen, to support the English part of this dictionary the database requires a total of 50 fields of which 45 are defined data fields, one is a memo field (to be used for notes and internal communication between the lexicographers), and four are maintained as reserve fields (if something new shows up during the compilation process). The Spanish part of the dictionary contains a similar number of fields.

It should, however, be noted that the relations reflected in this diagram represent only a fraction of the many relations relevant to the project. In order to design the database needed to have the final product exactly as required, all relevant relations between the data types had to be described and communicated to the programmer. A diagram reflecting all the relevant data and their mutual relations would be extremely complex and the risk of misunderstanding or missing some of these relations would be considerable. The written diagram was therefore combined with oral explications in a complex interdisciplinary communication.

8. Dictionary portal and search options

The dictionary portal to be accessed via the Internet is, at this point, foreseen to include a traditional search field and eleven search options which can be seen in Figure 5.

A user of the Business Dictionary who has accessed the web-portal shown in Figure 5 will first have to enter a word (or string) in the search field and then chose one of the eleven search options (buttons). A click on the first of these buttons (the one next to the search field) initiates a general search in the database, whereas a click on any of the other ten buttons starts specific searches in a pre-selected number of database fields, thereby filtering the data to be displayed. Although these ten specialized buttons, by first glance, may seem to represent each of the ten functions covered by the dictionary, this is actually not the case as it will be argued in the following.

The first thing to be noted in the dictionary portal shown in Figure 5 is that users can carry out specific searches based upon their first language. This possibility provides a first important adaptation to each of the different characteristics of the heterogeneous user group described above. After deciding the preferred language (and without performing any action), the users can choose

between five search options in accordance with the activity giving rise to the consultation.

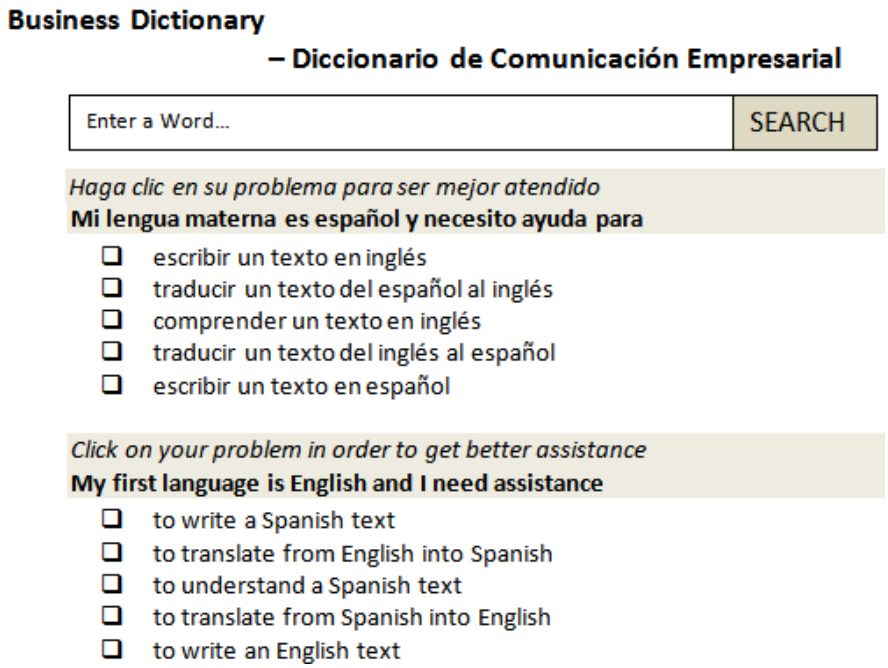


Figure 5: Planned dictionary web-portal

If English first-language speakers, for instance, enter a Spanish word in the search field and click on the first option, i.e. *to write a Spanish text,* then a Spanish article with English equivalents and notes will appear on the screen (with the aforementioned possibility of hiding the English items changing it into purely monolingual). By contrast, if the same users write an English word in the search field and activate the second option, i.e. *to translate from English into Spanish*, they will be directed to an "in-between" article with no other data than Spanish equivalents, meaning discrimination, and part of speech. Here, they can decide to stop the consultation if an equivalent is all they need or, if they also require additional information, they can click on the preferred equivalent and continue to a Spanish article identical to the aforementioned one in order to get data on gender, inflection, syntactic properties, collocations, synonyms, antonyms and fixed expressions (see the corresponding argumentation in Section 6).

However, a little "trick" will be incorporated in the search buttons. If the users, instead of a Spanish word, enter an English word and activate the first option — or a Spanish word (instead of an English one) and click on the second

option — then the immediate search results would be reversed. The idea behind this little "trick" is that the dictionary should be as easy as possible to operate without reading any lexicographical instructions. Hence, if English first-language speakers were experiencing a problem when writing a Spanish text, and if this problem consisted in that they did not know the Spanish word to use in a specific context, then they may — just may — enter a known English word in the search field and click on "to write a Spanish text" because this activity is actually the one they are engaged in. A similar thing could happen if they were translating an English text into Spanish. In order to reduce the number of frustrated or superfluous consultations, the system (dictionary) simply takes such situations into consideration and performs "intelligent searches" based on the language to which the search word belongs. The guiding principle is that the users, as a rule, should not perform more than two actions (entering a search string and clicking on a specific option) before getting quick and easy access to the data required in each case — and one more action (a click) if they need additional data.

The "duality" hidden in the two first search options clearly indicates that these options — although the accompanying text may insinuate it — do not correspond directly to specific functions, but to possible steps required to fully serve the two main and prioritized functions covered by the dictionary, i.e. L2 text production and L1–L2 translation. As to the three other search options in the portal presented in Figure 5, they do correspond directly to the functions with the same name. Clicking on them will lead the users to three different articles. Finally, if they, instead of the specialized search options, choose the general search button placed on the right of the search field, they will get a "traditional" article with unfiltered data, i.e. all the data attached to the word in question, including various homonyms. However, whether or not this general option will be maintained will depend on the users' reaction once the dictionary is available for access via the Internet.

9. Two examples of master articles

Apart from the web-portal, it is foreseen that the completed dictionary project will include eleven other user interfaces, ten for each of the search options discussed in the previous section and one to be used in the case of homonymy. Each of these eleven interfaces will represent a master article containing fields for all the data categories to be presented in each case. In this section we will briefly discuss the two master articles directly related to the prioritized functions of the dictionary.

The following master article is the one planned to serve an English user consulting the dictionary in order to get assistance when having problems in the production of Spanish business texts:

Spanish lemma homonymy number *part of speech in English*
<Inflection>
polysemy number

DEFINITION
- English definition
  hyperlink to external source
  English text to hyperlink

ABBREVIATED FORM
- abbreviation
  (abbreviation note in English)

FULL FORM
- full form
  (full form note in English)

EQUIVALENTS
- (equivalent note in English) **English equivalent(s)** (UK/US)

SYNTACTIC PROPERTIES
- **mini-rule(s)**
  example sentence in Spanish
  example sentence translated into English
  (mini-rule note in English)

COLLOCATIONS
- **Spanish collocation(s)**
  English translation
  (collocation note in English)

FIXED EXPRESSIONS
- **Spanish fixed expression(s)**
  English translation
  (fixed expression note in English)

SYNONYMS
- Spanish synonym(s)

ANTONYMS
- Spanish antonym(s)

Example 1: Master article for English user writing a Spanish text

This master article includes all the data categories that an English first-language speaker may need when having problems related to text production directly in Spanish. The items underlined represent hyperlinks 1) referring the users to an external source, 2) directing them to another article in the dictionary, or 3) activating a pop-up window (which is a more convenient way of showing, for instance, the large number of inflected forms of Spanish verbs). The plural forms of some items in Example 1 indicate that there could be more than one item of this data category included in the displayed article. Apart

from this, one will note that there is an extensive use of metatexts written with small caps (DEFINITION, ABBREVIATED FORM, etc.) which serve as article-internal indicators or headlines separating sections with data of the same category.

The data to be lemmatized will include words, word combinations, terms, abbreviations, acronyms, names of relevant institutions, fixed expressions, and routine formulae. As indicated above, it is envisaged that a number of "difficult terms" will be defined. In some of these cases the definitions will be accompanied by links to relevant external sources. For instance, the users consulting one of the *incoterms* will, apart from a definition, be offered a link to the corresponding webpage of the *International Chamber of Commerce* which is the organization in charge of regulating these terms, and where they can get updated information. Users consulting the term *letter of credit* will be referred to a website of a financial institution where they, apart from additional information, can download a letter-of-credit form to be filled in, and so on.

As it is presented in example 1, the master article prepared for an English first-language speaker writing Spanish business texts may appear relatively overcrowded with data categories. In this respect, it is important to remember that it is a master article, and that the concrete articles displayed on the screen in all cases will include less data categories. As already explained above, definitions are only foreseen in relatively few articles (apart from cases of polysemy). In addition, external links, abbreviations, full forms and fixed expressions will also be irrelevant in most cases. The data on abbreviation and full form will be mutually exclusive, etc. Hence, if the English users referred to in Example 1 choose to hide the Spanish items, the most typical data categories to appear on the screen in this case would be the ones shown in Example 2:

Spanish lemma *part of speech in English*
<Inflection>
polysemy number definition

SYNTACTIC PROPERTIES
- **mini-rule(s)**
  example sentence in Spanish
  (mini-rule note in Spanish)

COLLOCATIONS
- **Spanish collocation(s)**
  (collocation note in Spanish)

SYNONYMS
- Spanish synonym(s)

ANTONYMS
- Spanish antonym(s)

Example 2: Typical article for English user writing a Spanish text

If the English users instead write an English word in the search field and click

on the button *to translate from English into Spanish*, an article based upon the following master article will be displayed:

English search word(s)
(meaning discrimination) **Spanish equivalent** PART OF SPEECH
(meaning discrimination) **Spanish equivalent** PART OF SPEECH
(meaning discrimination) **Spanish equivalent** PART OF SPEECH
(meaning discrimination) **Spanish equivalent** PART OF SPEECH
etc. ...

Example 3: Master article for an English user translating an English text into Spanish

As discussed in Section 6, an article of this type may solve the problem for the English first-language users who only need a Spanish equivalent and nothing more. For other users who will need additional data, the article will serve as an "in-between" page, from where a simple click on the preferred equivalent will lead them to an article based on the master article shown in Example 1.

Finally, it should be noted that none of the master articles shown in the above three examples reflects the graphical design of the articles eventually to be presented to the users. No decision has yet been taken in this respect. With a well-prepared database, the use of the available technologies allows almost any design — in this case the one considered most appropriate, taking into account the characteristics of the target users — and it also permits its modification or complete replacement with another design if the first one does not live up to the expectations. This is one of the many advantages brought to lexicography by the introduction of new technologies and media.

10. Some techniques to be employed

It is impossible to treat all features and technicalities related to the project in the framework of this contribution. Here we will limit ourselves to a brief discussion of a few techniques that may be used either to make the displayed articles more readable or to further reduce the amount of data displayed in the first instance in order to avoid information overload and adapt the dictionary to smaller devices like tablets and smartphones. Most of the techniques will be known from already published online dictionaries though they may still not be sufficiently generalized.

One of the many challenges in modern online lexicography is to handle the contradiction between the almost unlimited freedom from space restraints in the database and the undeniable space restraints at the level of the screenshots, i.e. a contradiction very similar to the one between storage space and presentation space in the terminology proposed by Lew 2012). This contradic-

tion is also present in the planned dictionary project which applies techniques that, one the one hand, are more space-consuming, and on the other, reduce the total amount of data on the screen, i.e. only the data displayed on the screen at a given moment but of course not all the data needed by the users in each consultation.

Among the space-consuming techniques are 1) less compact article formats with different data placed in separate lines, 2) abolition of metalexicographical abbreviations and their replacement with full forms, and 3) use of metatexts to introduce sections with specific data categories. These techniques are employed in order to improve readability and are therefore justified in spite of being more space-consuming.

Among the data-reducing techniques are 1) data filtering based on identification of the user's first language and the type of activity where the need occurs, 2) use of hidden data to be unfolded when required, and 3) use of pop-up windows to present additional data. These techniques are applied in order not to burden the users with too many data which, if presented simultaneously on the screen, may increase the consultation time as well as the risk that the users retrieve incorrect information or eventually abort the consultations.

The challenge is to balance the two types of technique for the benefit of the users. This implies that all lexicographical data required by the latter in the first instance should be included in the screenshot without the need to scroll down or sweep the screen to the right or left in order to find out whether or not the required data can be found there. Of course, as data underload can never the alternative, this may frequently imply that the users have to perform an extra action in order to access the data needed to get the right information, but it should be absolutely clear from the data presented in the first screenshot that the needed data are included in the dictionary and can be accessed by clicking or moving the mouse over a specific item, or even by scrolling down or sweeping the screen (in this case with the knowledge that the data are there). This principle is especially relevant in dictionaries for communicative purposes which are usually consulted by people experiencing problems while doing something else and therefore demand a quick answer in order to continue with their main activity, e.g. foreign-language text production.

Data-reducing techniques are important in all types of online dictionary used on all types of screen but they are especially relevant when the dictionaries are accessed from devices with small screens. In such devices there are two ways to avoid scrolling down or sweeping the screen, namely to make the letter types smaller or to reduce the amount of displayed data. The first of these solutions cannot be recommended as "it is quite convenient to see a lengthy entry on a computer screen, but it is not convenient to see it in a smartphone screen" (Kwary 2015: 203-204). The Business Dictionary therefore opts for the second solution which is achieved by a flexible combination of data-reducing and space-consuming techniques, i.e. hidden data combined with metatexts:

Spanish lemma

INFLECTION

DEFINITION

ABBREVIATED FORM

FULL FORM

EQUIVALENTS

SYNTACTIC PROPERTIES

COLLOCATIONS

FIXED EXPRESSIONS

SYNONYMS

ANTONYMS

Example 4: Possible first-instance article to appear on smartphone screen

This master article foreseen to be used in smartphones is based on the one presented in Example 1. The metatexts will only be shown when they actually indicate the presence of data that are hidden; in case of zero data they will disappear. The idea is to inform the users where there may be relevant additional data which can be unfolded "on demand" by simply touching the screen over the respective metatexts. This master article still has to be refined and tested, as its convenience or inconvenience in the last instance depend on the users' reaction. But as indicated in the previous section, it can easily be modified or even replaced by another if this appears to be relevant. A special smartphone option for people experiencing problems in oral communication and only offering equivalents is also a possibility that is considered.

Apart from the mentioned space-consuming and data-reducing techniques, a few other techniques with other purposes also deserve to be highlighted here. These include improved search algorithms which, among other things, will allow the users to access words and word combinations that are not lemmatized. It is not envisaged, for instance, that all equivalents (used to explain L2), synonyms and antonyms are treated as lemmata, neither are the collocations, but the users will nevertheless be able to search for them and, thus, get access to useful data. Finally, the use of log files and interactive options is also foreseen in order to get feedback from the users with a view to continuously improving the dictionary both in terms of available lexicographical data and functionality.

11. Conclusions

This paper aims at sharing some reflections and discussing some steps in the process from idea to design of an online dictionary of Spanish and English

business communication. A meticulous, theory-based approach in this preparatory phase is crucial for guaranteeing a successful lexicographical product when it eventually is placed at the disposal of users on the Internet. Major mistakes and shortcomings will necessarily affect the final product as any major reprogramming of the underlying database will delay the project and challenge the project budget once the compilation process has started, and therefore only can be recommended in exceptional cases.

However, a well-considered dictionary design, or concept, is not in itself a guarantee of success. Just as important are meticulous efforts in the remaining phases of the overall dictionary project, cf. Tarp (2014). These phases include choosing the empirical basis and determining the different methods which the lexicographers should employ to select and prepare the data to be incorporated in the database as well as the compilation process itself. The selection of lemmata, equivalents, collocations, etc., will be based on a combination of existing business dictionaries, glossaries published by various relevant institutions, indexes in textbooks, and corpora — an empirical basis recommended by Bergenholtz and Tarp (1995) as the most appropriate to be used in specialized lexicography. By contrast, the definitions will be written based on expert knowledge and sometimes also on the material published by normative bodies such as the International Chamber of Commerce. All this will hopefully be treated in a later contribution if everything goes as planned.

12. References

12.1 Dictionaries

Combley, Roz. 2011. *Cambridge Business English Dictionary*. Cambridge: Cambridge University Press.

Graham, James and George A.S. Oliver. 1906. *The Foreign Traders' Dictionary of Terms and Phrases in English, German, French, and Spanish*. London/New York: Macmillan.

López, Sinda and Donald Watt. 2002. *Oxford Business Spanish Dictionary. Spanish–English, English–Spanish*. Oxford: Oxford University Press.

Miles, Andy D. 2012. *Business English to Spanish Glossary*. [http://www.andymiles.com]

Mortimer, Thomas. 1766. *A New and Complete Dictionary of Trade and Commerce*. London: Printed for the author, and sold by S. Crowder, and J. Coote, and J. Fletcher.

Postlethwayt, Malachy. 1751–1755. *The Universal Dictionary of Trade and Commerce.* London: John and Paul Knapton.

Rolt, Richard. 1756. *A New Dictionary of Trade and Commerce*. London: T. Osborne and J. Shipton.

Savary des Bruslons, Jacques. 1723. *Dictionnaire universel de commerce*. Paris: Chez J. Etienne.

Tarp, Sven in cooperation with Birger Andersen, Karin Balsgart and Morten Pilegaard. 1999. *Dansk–engelsk erhvervsordbog*. Aarhus: Systime.

Tarp, Sven in cooperation with Birger Andersen and Karin Balsgart. 2006. *Engelsk–dansk. Erhvervsordbog*. Copenhagen: Gyldendal.

Tarp, Sven in cooperation with Virginia Hvid and Anne Lise Laursen. 1999. *Dansk–spansk erhvervsordbog*. Aarhus: Systime.

Tarp, Sven in cooperation with Virginia Hvid and Anne Lise Laursen. 2003. *Spansk–dansk erhvervsordbog*. Copenhagen: Gads Forlag.

Veitelle, I. de. 1864. *Mercantile Dictionary: A Complete Vocabulary of the Technicalities of Commercial Correspondence, Names of Articles of Trade, and Marine Terms, in English, Spanish, and French, with Geographical Names, Business Letters and Tables of the Abbreviations in Common Use in the Three Languages.* New York: D. Appleton & Co.

Virio, Juan Bautista de. 1792. *Colección alfabética de los aranceles de la Gran Bretaña, y extractos de las Leyes, Reglamentos, Órdenes y Providencias expedidas en aquel Reyno para el régimen de sus Aduanas, y fomento de su comercio.* Vol. 1-4. Madrid: J. Ibarra.

12.2 Other literature

Astigarraga, Jesús and Juan Zabalza. 2007. Los diccionarios de Comercio y Economía en el siglo XVIII español. *Revista de Historia Industrial* 35: 13-46.

Bergenholtz, Henning and Jesper Skovgård Nielsen. 2013. What is a Lexicographical Database? *Lexikos* 23: 77–87.

Bergenholtz, Henning and Sven Tarp (Eds.). 1995. *Manual of Specialised Lexicography: The Preparation of Specialised Dictionaries.* Amsterdam/Philadelphia: John Benjamins.

Besomi, Daniele (Ed.). 2011. Crises and Cycles in Economic Dictionaries and Encyclopaedias. London/New York: Routledge.

Caruso, Valeria. 2011. Online Specialized Dictionaries: A Critical Survey. Kosem, Iztok and Karmen Kosem (Eds.). 2011. *Electronic Lexicography in the 21st Century: New Applications for New Users*: 66-75. Ljubljana: Institute for Applied Slovene Studies.

Fuertes-Olivera, Pedro A. and Sven Tarp. 2014. *Theory and Practice of Specialised Online Dictionaries: Lexicography versus Terminography.* Berlin/Boston: De Gruyter.

Kwary, Deny A. 2015. Challenges of Online Learner's Dictionaries to be Used in Smartphones. *Estudios de Lexicografía* 4: 199-206.

Lew, Robert. 2012. How Can We Make Electronic Dictionaries More Effective? Granger, Silvia and Magali Paquot (Eds.). 2012. *Electronic Lexicography*: 343-361. Oxford: Oxford University Press.

Tarp, Sven. 2004. Reflections on Dictionaries Designed to Assist Users with Text Production in a Foreign Language. *Lexikos* 14: 299-325.

Tarp, Sven. 2013. What Should We Demand from an Online Dictionary for Specialized Translation? *Lexicographica* 29: 146-162.

Tarp, Sven. 2014. Theory-based Lexicographical Methods in a Functional Perspective: An Overview. *Lexicographica* 30: 58-76.

Tarp, Sven. 2015a. On the Disciplinary and Functional Status of Economic Lexicography. *Ibérica* 29: 179-200.

Tarp, Sven. 2015b. Structures in the Communication between Lexicographer and Programmer: Database and Interface. *Lexicographica* 31. Forthcoming.

Tarp, Sven and Theo J.D. Bothma. 2013. An Alternative Approach to Enlightenment Age Lexicography: The Universal Dictionary of Trade and Commerce. *Lexicographica* 29: 222-284.

Diasystematic Information in the "Big Five": A Comparison of Print Dictionaries, CD-ROMS/ DVD-ROMS and Online Dictionaries

Marjeta Vrbinc *(marjeta.vrbinc@ff.uni-lj.si), Faculty of Arts, University of Ljubljana, Ljubljana, Slovenia*
and
Alenka Vrbinc *(alenka.vrbinc@ef.uni-lj.si), Faculty of Economics, University of Ljubljana, Ljubljana, Slovenia*

Abstract: The information provided by labels is called diasystematic information, which gives restrictions and limitations concerning the use of a lexical item. The focus of the study, the findings of which are presented in this contribution, is five British monolingual learner's dictionaries (OALD9, LDOCE6, COBUILD7, CALD4, MED2), which are often referred to as the "Big Five". The aims of the study are to compare the print edition and the electronic versions (CD-ROM/DVD-ROM and online dictionaries) of the same dictionary to see whether the lists of labels used in one particular dictionary coincide across versions of one and the same dictionary. Parallels are then drawn between dictionaries to determine similarities and differences in the use of labels providing different types of diasystematic information. Some of the most important findings of the study are that lists of labels differ in all three versions of one and the same dictionary and that some labels enumerated in the lists either are not used in the A–Z section at all or are used in a different form. Apart from that, some labels used in the dictionaries are so close in interpretation that the intended user will probably experience difficulty in distinguishing between them.

Keywords: DIASYSTEMATIC INFORMATION, TAXONOMIES OF DIASYSTEMATIC INFORMATION, LABELS, MONOLINGUAL LEARNER'S DICTIONARIES, PRINT DICTIONARIES, DICTIONARIES ON CD-ROMS/DVD-ROMS, ONLINE DICTIONARIES, RESTRICTIONS AND LIMITATIONS CONCERNING USE, LISTS OF LABELS IN FRONT MATTER, ACTUAL USE OF LABELS

Opsomming: Diasistematiese inligting in die "Groot Vyf": 'n Vergelyking van gedrukte woordeboeke, CD-ROMS/DVD-ROMS en aanlyn woordeboeke. Die inligting wat deur etikette verskaf word, word diasistematiese inligting genoem. Dit dui die beperkings en begrensings rakende die gebruik van 'n leksikale item aan. In die studie, waarvan die bevindings in hierdie bydrae aangebied word, word gefokus op vyf Britse eentalige aanleerderswoordeboeke (OALD9, LDOCE6, COBUILD7, CALD4, MED2) wat dikwels die "Groot Vyf" genoem word. Die doel van die studie is om die gedrukte uitgawe en die elektroniese weergawes (CD-ROM/DVD-ROM en aanlyn woordeboeke) daarvan te vergelyk om vas te stel of die lyste etikette wat gebruik word in een spesifieke woordeboek ooreenstem met ander weergawes van die-

selfde woordeboek. Parallelle word dan tussen woordeboeke getrek om ooreenkomste en verskille te bepaal in die gebruik van etikette wat verskillende soorte diasistematiese inligting verskaf. Van die belangrikste bevindings van die studie is dat lyste etikette in al drie weergawes van dieselfde woordeboek verskil en dat sommige etikette wat in die lyste genoem word, glad nie in die A–Z-afdeling gebruik word nie of in 'n ander vorm gebruik word. Afgesien hiervan, kan sommige etikette wat in die woordeboeke gebruik word, so eenders geïnterpreteer word dat die gebruiker vir wie die woordeboek bedoel is, waarskynlik probleme sal ervaar om tussen die etikette te onderskei.

Sleutelwoorde: DIASISTEMATIESE INLIGTING, TAKSONOMIEË VAN DIASISTEMATIESE INLIGTING, ETIKETTE, EENTALIGE AANLEERDERSWOORDEBOEKE, GEDRUKTE WOORDEBOEKE, WOORDEBOEKE OP CD-ROM/DVD-ROM, AANLYN WOORDEBOEKE, BEPERKINGS EN BEGRENSINGS RAKENDE GEBRUIK, LYSTE ETIKETTE IN DIE VOORWERK, WERKLIKE GEBRUIK VAN ETIKETTE

1. Introduction

As one of the information categories in the dictionary entry, labels have been present in dictionaries for a very long time. The information provided by labels is called diasystematic information, which gives restrictions and limitations concerning the use of a lexical item. A comparison of different dictionaries reveals certain similarities and differences in the inclusion and treatment of diasystematic information that will be addressed in more detail in this contribution. Since labels give information on the connotative value of lexical items, they are of particular importance to non-native speakers of a language; thus, the focus of the study is five British monolingual learner's dictionaries (MLDs), i.e., *Oxford Advanced Learner's Dictionary of Current English* (OALD9), *Longman Dictionary of Contemporary English* (LDOCE6), *Collins COBUILD Advanced Dictionary of English* (COBUILD7), *Cambridge Advanced Learner's Dictionary* (CALD4) and *Macmillan English Dictionary for Advanced Learners* (MED2), which are often referred to as the "Big Five". All these dictionaries are either accompanied by a CD-ROM or DVD-ROM or they have an online version (free access and/or access by a unique PIN code). This study examines whether a user of different versions of one and the same dictionary gets the same information as regards the meaning and the use of labels, or whether one version provides information, whereas the other one lacks it. Before taking a closer look at the labels used in the "Big Five", it is necessary to outline the theoretical background to diasystematic information provided in the form of labels.

1.1 Taxonomies of diasystematic information

The existing literature offers various taxonomies of diasystematic information proposed by different researchers. Mostly, they vary in degree of complexity, but there is also a great deal of overlapping. Jackson (2002: 109-115), for instance, enumerates seven types of usage labels, i.e., dialect; formality; status;

effect; history; topic or field and disputed usage. Landau (2001: 217-272), however, classifies eight common kinds of usage information, i.e., currency or temporality; regional or geographic variation; technical or specialized terminology; restricted or taboo sexual and scatological usage; insult; slang; style, functional variety, or register and status or cultural level. Atkins and Rundell (2008: 182-186) distinguish nine types of linguistic labels: domains; region; dialect; register; style; time; slang and jargon; attitude and offensive terms. The most detailed classification can be found in Hausmann (1989: 651), who identifies as many as eleven types of labels. His classification was also adopted by various other scholars such as Bergenholtz and Tarp (1995: 131-134) and Svensén (2009: 326-332) and is also taken as a theoretical background in our research. Below, the classification proposed by Hausmann (1989: 651) is presented, and parallels with other classifications enumerated above are drawn:

1. diachronic information: this associates a word or one of its senses with a particular period in the history of language. This dimension includes a range of labels that can be arranged chronologically from archaic, via obsolete to contemporary words or senses and neologisms. The most common temporal labels found in contemporary dictionaries are *old-fashioned, obsolete, archaic, old use* or *dated*. This group is referred to as history by Jackson (2002), currency or temporality by Landau (2001) and as time by Atkins and Rundell (2008).

2. diatopic information: this associates a word or one of its senses with a particular regional dialect or national variety. Most British dictionaries nowadays include words or senses typically used in different varieties of English. Sometimes regional areas within a country are specified; thus, *regional* or *dialect* is used as a label. Jackson (2002) refers to this group as dialect, but says that "dialect labels refer to geographical restrictions, and we can take this to include both national varieties and regional dialects within a national variety" (ibid: 110). Landau (2001) calls diatopic information regional or geographic variation, while Atkins and Rundell (2008) divide it into two separate groups: region and dialect.

3. diaintegrative information: this associates a word or one of its senses with the dimension of integration into the native stock of words of a language. Monolingual dictionaries usually provide information on the language of origin, mostly for words that have retained their original form (e.g., from *Latin,* from *French,* from *Italian*). Interestingly, diaintegrative information is not included in any of the other classifications and can also be regarded as belonging to etymology as a different information category.

4. diamedial information: this associates a word or one of its senses with a particular medium of communication. The most common labels are *written* and *spoken*. Apart from Hausmann, Landau (2001) is the only scholar who includes diamedial information in his classification of diasystematic

information and refers to it as style, functional variety, or register.

5. diastratic information: this associates a word or one of its senses with a particular social group, consequently referring to sociolects, such as slang and different kinds of jargon. The most common labels are *slang, vulgar* and *taboo*. Diastratic information is included in all the taxonomies: Jackson (2002) calls it status, Landau (2001) divides it into two groups, i.e., restricted or taboo and scatological usage and slang, and Atkins and Rundell (2008) also have two groups for diastratic information, i.e., slang and jargon and offensive terms.

6. diaphasic information: this associates a word or one of its senses with a particular register of a language, the most common labels being *formal* and *informal*. Diaphasic information is also included in all the taxonomies: Jackson (2002) calls it formality, Landau (2001) refers to it as style, functional variety, or register, and Atkins and Rundell (2008) call it register.

7. diatextual information: this associates a word or one of its senses with a particular discourse type or genre. The most common labels are *poetic* and *literary*. Diatextual information is disregarded by Jackson (2002) but is referred to as style, functional level, or register by Landau (2001) and as style by Atkins and Rundell (2008).

8. diatechnical information: this associates a word or one of its senses with a particular subject field. In monolingual dictionaries, subject-field labels, field labels or domain labels indicate that a certain word or one of its senses belongs to technical or scientific vocabulary. The large number of sublanguages typical of different subject fields present a problem even for educated native speakers, since each subject field has its own vocabulary. Consequently, some dictionaries, especially learner's dictionaries, often use the general label *technical/specialized/specialist* instead of giving detailed information on specific subject fields. Diatechnical information is included in all classifications: Jackson (2002) calls it topic or field, Landau (2001) technical or specialized terminology and Atkins and Rundell (2008) refer to it as domains.

9. diafrequent information: this associates a word or one of its senses with a particular frequency of occurrence. Labels used to indicate frequency are *less frequent* and *rare*. It should be stressed that dictionaries label only less frequent items, which means that unmarked items are more common. Hausmann's classification is the only classification that includes diafrequent information — all other scholars disregard it altogether.

10. diaevaluative information: this associates a word or one of its senses with a particular attitude or evaluation or the speaker's mood. Labels used to denote diaevaluative information are *appreciative, derogatory, offensive, humorous, ironic* and *euphemistic*. Diaevaluative information is one of the categories included in all classifications studied: Jackson (2002) calls it

effect, Atkins and Rundell (2008) attitude, while Landau (2001) divides it into two groups depending on the value a particular label expresses. Labels denoting positive connotation (*approving, humorous*) belong to style, functional variety, or register, while those expressing negative connotation (*offensive, derogatory, disapproving*) belong to the group referred to as insult.

11. dianormative information: this associates a word or one of its senses with a certain degree of deviation from a cultural standard. Labels expressing dianormative information are *non-standard, substandard* and *disputed.* In other words, the acceptability of items marked with one of these labels is questionable as regards linguistic correctness. Disputed usage is the term used by Jackson (2002) to describe dianormative information, while Landau (2001) refers to it as status or cultural level. Atkins and Rundell do not include this group in the classification of diasystematic information.

Table 1 below summarizes these four classifications.

Hausmann (1989)	**Jackson (2002)**	**Landau (2001)**	**Atkins & Rundell (2008)**
diachronic	history	currency or temporality	time
diatopic	dialect	regional or geographic variation	— region; — dialect
diaintegrative	/	/	/
diamedial	/	style, functional variety, or register	/
diastratic	status	— restricted or taboo and scatological usage; — slang	— slang and jargon; — offensive terms
diaphasic	formality	style, functional variety, or register	register
diatextual	/	style, functional variety, or register	style
diatechnical	topic or field	technical or specialized terminology	domains
diafrequent	/	/	/
diaevaluative	effect	— insult;	attitude

		— style, functional variety, or register	
dianormative	disputed usage	status or cultural level	/

Table 1: Summary of classifications.

The label *figurative* is used as a label in many dictionaries, but it does not fit any of the above-mentioned categories of labels, since it refers to the meaning extension of a lexical item rather than expressing any restriction on usage. In the process of figurative extension, it is possible for a word to acquire semantic features that might not have been present in the original meaning (Atkins and Rundell 2008: 289). This label typically marks examples illustrating the use of the lemma or one of its senses and can thus be said to denote secondary senses that have the status of conventional metaphors (Hanks 2006: 28). That is why it must be treated separately as a label expressing certain shades of semantic meaning as well as a certain degree of stylistic level.

The label *trademark* is also used quite frequently. According to Landau (2001: 218), this label provides diatechnical information, which he calls technical or specialized terminology, but such a classification of this label can be disputed, since it expresses the origin of the lexical item without any connection whatsoever with technical or specialized terminology, for example, Kleenex, Levi's, Lycra, Polaroid, Rolex, Sellotape, Skype, Tupperware, Teflon, Thermos, etc. On the other hand, lexical items marked as trademarks can also belong to terminology typical of a specific subject field, for example, AZT or Prozac (medical or pharmaceutical terms). In such cases, the user would profit from getting the information concerning the subject field rather than getting the information on the origin of the lexical item.

Against this theoretical background, the aims of this study are: (1) to present a more detailed view of current practices employed in the British monolingual learner's dictionaries under investigation, in which labels of various kinds are abundantly provided; (2) to compare the print edition and the electronic versions (CD-ROM/DVD-ROM and online dictionaries) of the same dictionary to see whether the lists of labels used in one particular dictionary coincide across versions of one and the same dictionary; (3) to draw parallels between dictionaries to determine similarities and differences in the use of labels providing different types of diasystematic information.

2. Labels in print, CD-ROM/DVD-ROM and online versions of the "Big Five"

2.1 Labels in OALD9

In the print edition of OALD9, the labels are listed on the inside front cover

under the title "Labels used in the dictionary". They can be found in two columns: the first one contains twelve labels "used with words that express a particular attitude or are appropriate in a particular situation" (OALD9: inside front cover), i.e., *approving, disapproving, figurative, formal, humorous, informal, ironic, literary, offensive, slang, specialist* and *taboo*, whereas the second one includes five labels that "show other restrictions on the use of words" (ibid), i.e., *dialect, old-fashioned, old use, saying* and ™. A closer look at the first twelve labels reveals that they belong to different classes of diasystematic information. The labels *approving, disapproving*, *humorous, ironic* and *offensive* express diaevaluative information; *formal* and *informal* provide diaphasic information; *literary*, diatextual information; *slang* and *taboo*, diastratic information and *specialist* expresses diatechnical information, whereas *figurative* belongs to none of the categories mentioned in Section 1.1. The labels enumerated in the OALD9's second column again provide different types of diasystematic information: the labels *old-fashioned* and *old use* express diachronic information and *dialect*, diatopic information, whereas *saying* expresses the type of lexical item rather than its connotative value; it is thus not taken into account in any of the taxonomies discussed in Section 1.1. The last label listed in the second column is ™, which gives information on the origin of the lexical item (cf. also Section 1.1). Apart from that, fifteen labels expressing diatopic information are listed under "Abbreviations used in the dictionary": *AustralE, BrE, CanE, EAfrE, IndE, IrishE, NAmE, NEngE, NZE, SAfrE, ScotE, SEAsianE, US, WAfrE, WelshE*. As has been mentioned, the label *dialect*, which also belongs to the group of labels providing diatopic information, is, however, listed under "Labels used in the dictionary" and not under "Abbreviations used in the dictionary".

Contrary to the CD-ROM of the previous, i.e., the 8th edition, the DVD of the 9th edition does not contain the Advanced Search option, which makes it impossible to compare the list of labels on the DVD with those given in the print and online editions. Among the labels expressing diatechnical information, the print dictionary lists only one very general label, i.e., *specialist*. The label *specialist* has been introduced into the 9th edition, while the label *technical* was used in previous editions. The list of labels in the online edition, however, still includes *technical* instead of *specialist*, although the label *specialist* is used in the A–Z section, an error that should be corrected in the online edition. A closer observation of the A–Z section, however, shows that besides the very general label *specialist*, numerous other subject-field labels are abundantly provided, though not listed in the front matter. For instance:

antilogarithm — *mathematics*
polygon — *geometry*
antimatter — *physics*
melanoma — *medical*
anus — *anatomy*
antivirus — *computing*
meltdown, sense 2 — *economics*

niche, sense 2 — *business*
polymer — *chemistry*
polyphony — *music*
polysemous — *linguistics*
magnitude, sense 2 — *astronomy*
magnitude, sense 3 — *geology*
chromosome — *biology*
fiduciary — *law*

It has to be stressed that the online version "Oxford Learner's Dictionaries" is the only online dictionary among the dictionaries studied that includes information on the labels. This can be accessed via a tab entitled "About", where the "Guide to Symbols and Labels" section can be selected and the section "Labels used in Oxford Learner's Dictionaries" provides information on the labels. Interestingly, the lists of labels as well as explanations of each individual label are identical and the examples illustrating each label are almost identical in the print dictionary and in the online dictionary. The online dictionary, however, does not include the section Abbreviations, which can be found on the inside front cover of the print dictionary, which means that this information is completely lacking in this version of OALD. Although these labels are not provided in the "Guide to Symbols and Labels", they are used in the online dictionary in exactly the same way as in the print dictionary.

2.2 Labels in LDOCE6

Similar to OALD9, LDOCE6 also lists labels on the very first page of the dictionary. Here, the labels are subdivided into three categories: Words which are used only or mainly in one region or country (*BrE, AmE, AusE*), Words which are used in a particular situation, or show a particular attitude (*approving, disapproving, formal, informal, humorous*) and Words which are used in a particular context or type of language (*biblical, law, literary, medical, not polite, old-fashioned, old use, spoken, taboo, technical, trademark, written*). The first category of labels is homogeneous, since all the labels express diatopic information; this cannot be claimed for the second and third categories. The second category includes labels providing diaphasic (*formal, informal*) and diaevaluative information (*approving, disapproving, humorous*). The third category is the most heterogeneous one, since the labels *law, medical* and *technical* provide diatechnical information; *literary* and *biblical*, diatextual information; *not polite*, diaevaluative information; *old-fashioned* and *old use*, diachronic information; *spoken* and *written,* diamedial information; and *taboo*, diastratic information. For the classification of *trademark*, see Section 1.1.

The total number of labels included in LDOCE6 is twenty, which makes it the dictionary with the fewest labels among all the five dictionaries under consideration. Among the labels expressing diatechnical information, the generic

label *technical* and two more specific labels, i.e., *law* and *medical*, are used. In the online version offered by the access code, on the other hand, all labels are listed in the Advanced Search option under Register. The labels that are listed in the print dictionary under "Words which are used in a particular situation, or show a particular attitude" and "Words which are used in a particular context or type of language" coincide with the labels under Register in the online version except for the labels *approving* and *disapproving*, the labels which were added in the print edition of LDOCE6 but were not included in the previous (i.e., the 5th) edition. The labels expressing diatopic information provided in the print edition under "Words which are used only or mainly in one region or country" cannot be found among the labels in the online version, which does not, however, imply that they are not used in this version. On the contrary, they are used in the same way as in the print dictionary. The free online dictionary seems promising at first sight, since it offers the tab "How to use the LONGMAN DICTIONARY OF CONTEMPORARY ENGLISH ONLINE". Here, the user would most certainly expect to find instructions as to the use of all features found in the dictionary, including labels. Disappointingly, the labels cannot be found, nor are they explained, but they are used in the dictionary in the same way as in the print dictionary and in the online version offered by the access code.

2.3 Labels in COBUILD7

The first thing one notices when searching for metalinguistic information in COBUILD7 is that the online dictionary accessed by a unique code provides no information on diasystematic information or any other kind of metalinguistic information. The consequence is that no comparison can be made, but at the same time, it can be claimed with a high degree of certainty that labelling is done in much the same way in both versions of the COBUILD dictionary. The labelling described in this contribution is therefore based on the print edition.

If compared to the other four MLDs, COBUILD7 is clearly the only dictionary with quite extensive front matter. In the other four dictionaries, front matter is reduced to a list of labels and/or other abbreviations used in the dictionary and a graphic presentation of the dictionary entry taken from each individual dictionary. COBUILD7, however, differs greatly in this respect. On pages xiii-xv of the front matter, the user gets precise information on Style and Usage (p. xiii-xiv) and Pragmatics (p. xiv-xv). The Style and Usage section has two subsections: Geographical labels and Style labels. The Geographical labels subsection lists two labels, i.e., *Brit* and *Am*, and provides brief explanations of them. Then follows a list of other geographical labels used to mark a lexical item typically used in other varieties of English, such as *Australian, Irish, Northern English* and *Scottish*. Interestingly, these labels are introduced by 'e.g.', implying that not all the labels used in the A–Z section are listed here. Nevertheless, all these labels provide diatopic information. The Style labels section gives twenty labels in total (*business, computing, dialect, formal, humorous, infor-*

mal, journalism, legal, literary, medical, military, offensive, old-fashioned, rude, spoken, technical, trademark, very offensive, very rude, written), but a close inspection of these labels shows that not all the labels listed belong to "style labels" as they are referred to in the title of this section. Among these labels, we can find *business, computing, legal, medical, military* and *technical*, which are obviously subject-field labels, i.e., labels expressing diatechnical information. The label *dialect* belongs to the same group as *Brit* and *Am* mentioned above; the labels *formal* and *informal* express diaphasic information; *humorous, offensive* and *very offensive*, diaevaluative information; *journalism* and *literary*, diatextual information; *rude* and *very rude*, diastratic information; *spoken* and *written*, diamedial information and *old-fashioned*, diachronic information. For the classification of *trademark*, see Section 1.1.

The Pragmatics section starts by explaining what pragmatics is and expands on how pragmatic information is included in the dictionary. The "pragmatics labels" include *approval, disapproval, emphasis, feelings, formulae, politeness* and *vagueness*, and it is obvious that some of these coincide with "style labels" as used in other dictionaries, i.e., *approval* with *approving*, *disapproval* with *disapproving*, *politeness* with *polite*.

2.4 Labels in CALD4

In the print edition of CALD4, the labels are found under the title "Style and usage labels used in the dictionary". The list includes 31 labels (*abbreviation, approving, Australian English, child's word/expression, disapproving, female, figurative, formal, humorous, informal, Indian English, Irish English, legal, literary, male, Northern English, not standard, offensive, old-fashioned, old use, polite word/expression, saying, Scottish English, slang, South African English, specialized, trademark, UK, US, written abbreviation, A1, A2, B1, B2, C1, C2*). The labels express different types of diasystematic information: diatopic information (*Australian English, Indian English, Irish English, Northern English, Scottish English, South African English, UK, US*); diaevaluative information (*approving, disapproving, humorous, offensive, polite word/expression*); diachronic information (*old-fashioned, old use*); diaphasic information (*formal, informal, child's word/expression*); diastratic information (*slang*); diatextual information (*literary*); diatechnical information (*legal, specialized*); dianormative information (*not standard*), and diafrequent information (*A1, A2, B1, B2, C1, C2*). For the classification of *trademark* and *figurative*, see Section 1.1, and for the classification of *saying*, see Section 2.1.

Several labels stand out as they are not labels proper, meaning that they do not fall into any of the categories of labels expressing diasystematic information. Two of the labels peculiar to CALD4 are *female* and *male*, which are listed but not explained in the front matter. This means that the user cannot find any information about what they mark. It should be stressed that *female* and *male* are actually not labels proper, since they introduce the feminine or the masculine form of the lemma: for example, *female* lioness in sense 1 of the lemma lion. On the CD-ROM, we are faced with the opposite situation. The

two labels are not enumerated in the list of labels, but if the user comes across them in the A–Z section, s/he gets a brief explanation about their meaning by simply placing the cursor over the label: for example, *female* is explained as 'only applies to women'. The labels *abbreviation* and *written abbreviation* should also not go unmentioned, since they indicate the type of lemma rather than functioning as labels proper and are usually treated as part-of-speech indicators.

As in OALD9, there are discrepancies between the print edition and the CD-ROM version of CALD4 in that the print edition gives two labels not included in the list of labels on the CD-ROM, i.e., *female* and *male*. An obvious discrepancy concerns the labels expressing diatopic information. In the print dictionary, this category is represented by the eight labels mentioned above, while the CD-ROM version offers two labels (*British English only* and *American English only*) and a third choice 'other regions'. The search results for *British English only* show that the lemmata or their senses marked with the label *UK* are obtained, and similarly the search for *American English only* yields those lexical items labelled *US*. As is to be expected, the possibility 'other regions' yields all other regional varieties, which are more precisely listed in the print dictionary. With respect to labels expressing diatopic information, it can be established that the A–Z section shows certain inconsistencies when compared with the lists of labels: labels not found in any of the lists are used in the A–Z section. If we compare the following two lemmata

anyroad NORTHERN FOR **anyway**
bairn SCOTTISH ENGLISH OR NORTHERN ENGLISH a child

we can see that "anyroad" is marked *Northern,* which is a label not found in any of the lists, while "bairn" is marked *Northern English*, which is a label included in the list in the print edition; besides that, it is self-explanatory; thus any explication seems to be redundant.

A "label" found on the CD-ROM but not in the print edition is *short forms* used to indicate the contracted forms (e.g., could've SHORT FORM OF could have). Needless to say, this is not a label, neither is it used in the plural form as it appears in the list of labels. Typographically, however, it belongs to labels, as it appears in block capitals. In the print edition, *short forms* is not to be found among the labels and their typography also differs from that used for labels proper.

As far as the online version of CALD is concerned, there is no list of labels with their explications, but the user learns what a label denotes by positioning the cursor over the label. In this way, the online version resembles the CD-ROM of this dictionary, and the explications in both electronic versions are the same as in the front matter of the print dictionary.

2.5 Labels in MED2

The print edition of MED2 lists the labels under three titles:

— Style and attitude labels (twelve labels: *formal, humorous, impolite, informal, literary, offensive, old-fashioned, showing approval, showing disapproval, spoken, very formal, very informal*);

— Subject labels (eighteen labels: *art, astronomy, biology, business, chemistry, cinema, computing, economics, legal, linguistics, literature, maths, medical, music, physics, science, theatre, tourism*);

— Regional labels (three labels, i.e., *British, American, mainly American,* are enumerated and explained; in contrast, eleven labels, i.e., *Australian, Irish, Scottish, Canadian, Caribbean, East African, Indian, New Zealand, South African, Welsh, West African,* are listed but no explanation is provided).

These titles are to be found on the inside front cover, as is the case in most of its competitor dictionaries. The labels *formal, informal, very formal* and *very informal* express diaphasic information; *showing approval, showing disapproval, offensive, humorous* and *impolite*, diaevaluative information; *old-fashioned*, diachronic information; *spoken*, diamedial information; *literary*, diatextual information, whereas diatechnical information is provided by all the labels listed under Subject labels and diatopic information by all the labels enumerated under Regional labels. In the print edition, the labels under Style and attitude labels are explained, whereas all others are considered to be obvious and self-explanatory. The same holds true of the treatment of labels on the CD-ROM accompanying the print edition, where the labels can be found under Study pages, About the dictionary, Style and attitude labels. The online version, on the other hand, lacks lists of labels as well as explanations of labels.

It is important to point out that this is the only dictionary which gives a list of subject-field labels and does not use the generic label *technical/specialized/specialist*. A comparison of the lists in the print dictionary and the lists in the Super Search under the Advanced Search option on the CD-ROM shows only minor divergences. Apart from the labels listed in the print edition under Style and attitude labels, the electronic version offers the following additional labels: *often humorous, journalism, mainly journalism* and *mainly literary*. As regards labels expressing diatopic information, two labels enumerated in the print dictionary are absent from the list on the CD-ROM, i.e., *East African* and *Welsh*. Apart from that, one and the same label is given as *Indian English* on the CD-ROM and as *Indian* in the print dictionary. The variant given on the CD-ROM, i.e., *Indian English*, is also used in the A–Z section of the print dictionary. The subject-field labels listed in the print dictionary and on the CD-ROM mostly coincide, the only difference being that the label *trademark* is listed only on the CD-ROM but is lacking in the print edition.

The online dictionary does not provide any guidance whatsoever as to the labels used in the dictionary.

3. Labels across the dictionaries studied

In this section, we would like to make a survey of labels included in our research in relation to whether they can be found in all five dictionaries under investigation or whether they appear in one or two dictionaries, but are not used by the lexicographers of other dictionaries. It seems logical to start with labels that can be found in all five dictionaries. There are seven such labels: *approving, disapproving, formal, humorous, informal, literary* and *old-fashioned*. In MED2, *approving* and *disapproving* are expressed as *showing approval* and *showing disapproval* and in COBUILD7 as *approval* and *disapproval*. One label appears in all dictionaries except LDOCE6: *offensive*. Five labels can be found in three dictionaries: *law/legal* (*legal* in CALD4 and COBUILD7, *law* in LCODE6), *old use* (OALD9, LDOCE6, CALD4), *spoken* (LDOCE6, COBUILD7, MED2), *technical/specialized/specialist* (*technical* in LDOCE6, *specialized* in CALD4, *specialist* in OALD9) and *trademark* (OALD9, LDOCE6, CALD4). As many as ten labels are used in two dictionaries: *figurative* (OALD9, CALD4), *not polite/impolite* (*not polite* in LDOCE6, *impolite* in MED2), *polite/politeness* (*polite* in CALD4, *politeness* in COBUILD7) *medical* (LDOCE6, COBUILD7), *saying* and *slang* (OALD9, CALD4), *taboo* (OALD9, LDOCE6), *written* (LDOCE6, COBUILD7), *dialect* (OALD9, COBUILD7) and *journalism* (CALD4, MED2). If we disregard labels typical of MED2 and COBUILD7 which only express intensification or frequency of one and the same label (MED2: *very formal, very informal, often humorous, mainly journalism, mainly literary, mainly spoken*; COBUILD7: *very offensive, very rude*), all other labels listed in the front matter of the dictionaries under consideration are used in one dictionary only:

OALD9: *ironic;*
LDOCE6: *biblical;*
COBUILD7: *business, computing, military, rude;*
CALD4: *not standard, abbreviation, written abbreviation, child's word/expression, female, male, A1, A2, B1, B2, C1, C2*.

As far as the subject-field labels are concerned, a more extensive list can be found only in MED2 (see Section 2.5), whereas other dictionaries give just a few subject-field labels, a situation which does not facilitate an apt comparison:

OALD9: *specialist* (the A–Z section, however, includes several subject-field labels, see Section 2.1);
LDOCE6: *law, medical, technical*;
COBUILD7: *business, computing, legal, medical, military, technical*;
CALD4: *legal, specialized*.

If we compare the labels expressing diatopic information, we can see that there is a great deal of overlapping, since the labels *British English, American English* and *Australian English* are used in all five dictionaries; *Irish English* and *Scottish English* in four (OALD9, COBUILD7, CALD4, MED2); *South African English*

(OALD9, MED2, CALD4) as well as *Northern English* (OALD9, COBUILD7, CALD4) appear in three dictionaries; whereas *New Zealand English, West African English* and *Welsh English* are used in two dictionaries (OALD9, MED2) and the same holds true for *Canadian English* and *East African English* (OALD9, MED2); *South-East Asian English* and *English from the United States* (OALD9), *Indian English* (CALD4) as well as *Caribbean English* (MED2) are used in only one dictionary under consideration.

4. Discussion

4.1 Interpretation and intelligibility of labels easily confused

If we refer back to Section 3, we can see that five labels (*formal, informal, literary, humorous* and *old-fashioned*) are used to mark lexical items of various kinds in all dictionaries under consideration. Among these labels, the labels *formal* and *literary* may pose some problems for the intended dictionary users, i.e., advanced learners of English, especially if they are used together, which is often the case. This only adds to the complexity of the problem, since the connotation they mark may not be obvious to every user. The reason can also be sought in the users' mother tongue or more precisely, in the method of labelling that is familiar to the user from monolingual dictionaries written in his/her native language. In Slovene lexicography, for example, the label *knjižno* at least roughly corresponds to the English label *formal*. The front matter of the *Slovar slovenskega knjižnega jezika*, 2nd edition (Dictionary of Standard Slovene 2014: §133) explains that the label *knjižno* is assigned to 'words, senses or phraseological units used particularly in fiction or in scientific texts'. As a matter of fact, *knjižno* can also be used to mark the same connotation as the English label *literary*, which means that dictionary users familiar with a system of labelling similar to that for the native speakers of Slovene may be faced with the problem of the correct interpretation of the difference between *formal* and *literary*. A comparison of the explications of these two labels in the front matter of MLDs reveals that these are very simple and sometimes even overlap. In CALD4, for instance, the distinction between these two labels is explained as follows:

formal — "used in serious or official language or when trying to impress other people"
literary — "formal and descriptive language used in literature"

It is questionable whether these two explications are sufficiently clear for a foreign learner to distinguish accurately between these two labels, because *literary* is explained as labelling formal language. The addition of 'in literature' to the explication of *literary* may also puzzle the user who may not know the connotation of this restriction. Does this label mark expressions that the user is not supposed to use without sounding poetic? Do such expressions express meta-

phorical and metonymic transfer? Another problem that should be mentioned in connection with *literary* is that in some dictionaries (though not in the MLDs under investigation), there is the seemingly similar label *literature*, which is a domain rather than a style label. Users familiar with the label *literature* are likely to confuse these two labels, thus interpreting the label *literary* as being a domain label indicating terms from the field of literature instead of words found in literary and poetic texts. This confusion is also touched upon by Atkins and Rundell (2008: 229).

Another label used only in LDOCE6 and COBUILD7 is *written*. The explanation provided in the print edition of LDOCE6 says that it refers to "a word or phrase that is used only, or nearly always, in written English". If we compare this with the explanation given for the label *formal*, which states that it refers to "a word that is suitable for formal speech or writing, but would not normally be used in ordinary conversation", we can see that they are both explained in a similar way. Therefore, the question can be posed whether a learner of English would spot the difference between these two labels. Apart from the difference between *written* and *formal*, another pair is used in LDOCE6 and COBUILD7 as well as in MED2, which presents the same difficulty in interpretation as *written* and *formal*, namely, *informal* and *spoken*. A comparison of the description of these two labels offers no solution to this problem, since in MED2, the explanations provided are extremely close:

informal — "more common in speech than in writing and not used on a formal occasion"
spoken — "used in speech rather than writing"

Slightly more precise, but still not sufficiently clear are the explanations in LDOCE6:

informal — "a word or phrase that is used in normal conversation, but may not be suitable for use in more formal contexts, for example in writing essays or business letters"
spoken — "a word or phrase used only, or nearly always, in conversation"

The same holds true for the explanations found in COBUILD7:

informal — "used mainly in informal situations, conversations, and personal letters, e.g. **pep talk**"
spoken — "used mainly in speech rather than in writing, e.g. **pardon**"

Another problem concerns the interpretation of the labels *old-fashioned* (used in all five dictionaries) and *old use* (used in OALD9, LDOCE6 and CALD4). The difference is that *old-fashioned* marks lexical items that are not often used nowadays but are used by older people or have been used by people in the recent past, while *old use* implies that the lexical item marked with this label is no longer in current use. This distinction may be recognized by a native speaker of

English, but it may not be so obvious to a foreign learner, who is the target user of MLDs. COBUILD7 and MED2 use only the label *old-fashioned*, and considering the fact that the intended user may not be able to understand the difference between the two labels, this seems a sensible decision. The labels *old-fashioned* and *old use* are definitely labels that should be used with care in dictionaries if both labels are used in one and the same dictionary.

4.2 Labels expressing opposing connotative values

Sets of labels expressing opposing connotative values (*formal* and *informal*, *polite* and *impolite*, *spoken* and *written*) are also worth considering. The labels *formal* and *informal* are used in all dictionaries, whereas *polite* and *impolite*, *spoken* and *written* are not. *Impolite* (or *not polite* as used in MED2) is a label found in LDOCE6 and MED2, but interestingly, neither of these dictionaries has the opposing label *polite*. *Polite*, on the other hand, is used in CALD4 as well as in COBUILD7 (here, the label *politeness* is used), but neither of them has the label *impolite*. This means that the entire pair *polite/impolite* is not used in any of the dictionaries included in our study. From the point of view of dictionary users, this may indicate inconsistency or a failure to mark the opposing value of the lexical item in question. It seems sensible to label lexical items that imply politeness as well as those that imply impoliteness. This is especially true of dictionaries for foreign learners, who need precise guidance on lexical items expressing a certain degree of politeness/impoliteness, since labelling such words is aimed at warning those unfamiliar with them that they may be either polite or impolite. Doubtlessly, pragmatic information supplied by labels is as vital as the semantic information supplied by the definition (cf. also Norri 2000: 93). Similarly, the label *spoken* can be found in MED2, and one would rightly expect the label *written* to be used in the same dictionary, but this is not the case. LDOCE6 and COBUILD7, on the other hand, also use the label *spoken*, but the label *written* can also be found. It seems just as illogical as in the case of the pair *polite/impolite* to find *spoken* without its opposing label *written*, which is in line with Fedorova (2004: 269). What is more, MED2, which lacks the label *written*, has a very precise labelling of spoken connotation, since it uses two labels to mark that a lexical item or one of its senses belongs to spoken language: *spoken* and *mainly spoken*. Obviously, the question can be posed why it is necessary to label one concept so precisely while omitting the other concept completely.

4.3 Labels expressing diatechnical information

Another interesting feature is that MED2 is the only dictionary that gives a list of subject-field labels (cf. Section 2.5). In OALD9, for example, the only subject-field label enumerated in the list of labels is the generic label *specialist*, but as mentioned in Section 2.1, other subject-field labels are also used in the diction-

ary. Here, we are faced with a problem of higher-level domain markers (*technical/specialized/specialist*), on the one hand, and lower-level domain markers (*mathematics, physics, biology, business, phonetics*, etc.), on the other. This means that some lexical items are marked with a generic label without giving further details as to the specific subject-field of a particular lexical item, while others are labelled with very specific subject-field labels giving the user precise information about the subject-field where the lexical item functions as a term (cf. Vrbinc and Vrbinc 2013: 449, 454). The question is firstly, what is the criterion for deciding which label to use (a generic or a specific one) and secondly, what level of precision would benefit the end user. Is it enough to simply use the label *technical/specialized/specialist* just to indicate that the lemma or one of its senses thus labelled belongs to terminology? Or would the user expect to find the information about the precise subject-field or sub-field? These are the questions that cannot be answered without carrying out a user survey, which is beyond the scope of this study.

The opposite extreme is the use of labels denoting sub-fields (e.g., *business* vs. *economy, medical* vs. *anatomy, mathematics* vs. *geometry*). A general dictionary user cannot be expected to recognize the subtle differences between such closely related labels. If such labels are used, one would expect an explanation of the distinction between them, but taking account of the type of dictionary and the target audience, it can be claimed with a high degree of certainty that this is an unnecessary complication. In MLDs, one would expect that the subject-field labels would refer to fields of science only and would disregard the sub-fields. That means it is recommendable to use generic subject-field labels rather than more specific subject-field labels.

4.4 Labels expressing diatopic information

The labels indicating diatopic information are numerous, which is to be expected, given that the dictionaries claim to cover the vocabulary of the entire English-speaking world. The print edition of OALD9 gives three labels that need to be discussed in more detail: *NAmE* (North American English), *US* (English from the United States) and *Canadian English*. Although the difference between them might be obvious to a linguist or a native speaker of English, it is questionable whether a learner makes any real difference between these labels. The labels *US* and *Canadian English* may be understandable, but the label *NAmE* is most certainly not because it can be interpreted as a label encompassing both *US* and *Canadian English*. We dare to assume that an average user of a learner's dictionary would equate the label *NAmE* with the label *US* or *AmE* (as used in other dictionaries); thus, the label *NAmE* would not do the job it intends to do. Similarly, the CD-ROM version of CALD4 offers two labels (*British English only* and *American English only*) and a third choice 'other regions'. The search results for *British English only* show that the lemmata or their senses marked with the label *UK* are obtained; similarly, the search for *American English only* gives

those lexical items labelled *US*. The use of 'other regions' instead of giving a full range of labels found in the dictionary is also a feature typical of the CD-ROM accompanying the print edition of CALD4. A similar situation can be observed in MED2, which lists two variants of the same label, one in the print dictionary, i.e., *Indian*, and one on the CD-ROM, i.e., *Indian English*. In the A–Z section, however, the label *Indian English* is used in both versions of the dictionary. Such a discrepancy between the labels used in any of the lists (either in the front matter of the print dictionary or in the list provided within the advanced search options on the CD-ROM) and those actually used in the A–Z section should by all means be avoided. Interestingly, COBUILD7 does not list all the labels indicating the varieties of English but uses 'eg' to indicate that the list is incomplete. This policy is far from ideal, since a user cannot be supposed to know which other labels apart from those listed s/he can encounter in the A–Z section of the dictionary.

4.5 Register or style labels?

A general observation is that some MLDs list very different labels among what they call register or style labels, although they are far from being register or style labels, e.g., *biblical, law, medical, technical, trademark* (LDOCE6); *abbreviation, female, male, saying, specialized, trademark* (CALD4); *business, computing, dialect, legal, medical, military* (COBUILD7). This practice should be avoided and two possible solutions can be put forward. The first one is that labels should be grouped according to the connotative value they express, where the theoretical considerations proposed by metalexicographers within the framework of the classification of diasystematic labelling could be made use of. It is, however, true that considering the very specific target audience of MLDs, lexicographers try to simplify all dictionary components in order to make them more approachable and user-friendly. Taking this into consideration, all the labels used in a particular dictionary could be listed together in alphabetical order without any further division, and the title could simply be "Labels used in the dictionary".

4.6 Labels or not?

In COBUILD7, some "pragmatics labels" coincide with labels proper as used in other dictionaries, while some of them show no parallel with labels proper. Among the latter, three labels should be highlighted: *emphasis, feelings* and *vagueness*. The question is whether these convey sufficient information for the dictionary user to make full use of them. The print edition offers explanations for them, explanations which are lacking in the online dictionary. The following are the explications found in the front matter of COBUILD7 (xv):

emphasis — "The label EMPHASIS indicates that you use the word or expres-

sion to show that you think something is particularly important or true, or to draw attention to it. An example of a word with this label is *absolutely*."

feelings — "The label FEELINGS indicates that you use the word or expression to show how you feel about a situation, a person, or a thing. An example of a word with this label is *unfortunately*."

vagueness — "The label VAGUENESS indicates that you use the word or expression to show lack of certainty. People often use vague language to make statements 'softer', so that what they say does not appear too direct or too strongly stated. Examples of vague language are *presumably …*, *Do you know what I mean?*, *kind of …*, and *sort of …*"

The labels *saying, TM/trademark, abbreviation* and *written abbreviation* also deserve attention, since they cannot be considered labels proper; consequently, they do not fall into any of the categories for the classification of diasystematic information. These labels give information on the type of lexical item (*saying, abbreviation, written abbreviation*) or the origin of the lexical item (*TM/trademark*), but they themselves do not give any information on the connotative aspect of the lexical item they are used to mark. The question can therefore be posed why some dictionaries (print dictionaries or their CD-ROMs) list these labels under style or register labels. It is misleading for the user to find labels that cannot be regarded as style or register labels among other labels that address the connotative aspect of lexical items. Apart from that, phraseological units classified as sayings are included in the special idioms sections in CALD4. The usability of the label *saying*, however, remains questionable, since it is doubtful whether the user needs the information about the type of phraseological units. Besides that, it seems somehow illogical to label sayings and omit the labelling of formulae, catchphrases, collocations and other types of phraseological units. The precise classification of phraseological units affects neither the decoding nor the encoding process; it can thus be regarded as superfluous for the learner who is the intended user of a learner's dictionary.

4.7 Too precise labelling of the same connotation

If we address the issue concerning the labels used in one dictionary only, we can see that MED2 and COBUILD7 stand out in that they both include as many as six labels that occur in no other dictionary (MED2: *often humorous, very informal, journalism, mainly journalism, mainly literary* and *mainly spoken*; COBUILD7: *very offensive, very rude, emphasis, feelings, vagueness* and *formulae*). It should, however, be emphasized that the majority of these labels appear in pairs: *humorous* vs. *often humorous, informal* vs. *very informal, literary* vs. *mainly literary* and *spoken* vs. *mainly spoken, offensive* vs. *very offensive, rude* vs. *very rude*. *Mainly*, which precedes *journalism, literary* and *spoken*, implies that the lexical item thus marked is chiefly used in journalistic language, in literary works or in spoken language. *Very*, modifying *informal*, expresses intensification and is close to the

label *slang*, which is not used in this dictionary. Intensification is also expressed by *very rude* and *very offensive* in COBUILD7. *Often*, however, restricts the meaning of *humorous*, implying that the lexical item with this label is not necessarily used with humorous connotation in all contexts. Such labels, therefore, provide more detailed information on the connotative value of the lexical item in question, but do not in any way contribute to clearer labelling and easier interpretation by the intended dictionary user.

4.8 Labels included in the dictionary front matter but absent from the A–Z section

In LDOCE6, the labels *approving* and *disapproving* are listed in the print version but are lacking in the online version. This means that the Advanced search option yields no results as regards these two labels. A manual search in the print edition was rather unsuccessful, which means that we were unable to locate a lemma labelled *approving* or *disapproving*. Several lemmata, however, were found where the definition of a lemma or one of its senses said: "use this to show approval" (e.g., the lemma *succinct* is defined in the following way: clearly expressed in a few words — use this to show approval) or "used to show disapproval" (e.g., the lemma *agitator* is defined in the following way: someone who encourages people to work towards changing something in society — used to show disapproval). This means that the labels proper are not used, since labelling is integrated into the definition part. The question can therefore be raised why the dictionary front matter includes the labels even though they are most probably not used to label the (dis)approving connotative values of the lemmata or their senses.

5. Conclusion

As opposed to print dictionaries, online dictionaries (the only exceptions being OALD and CALD) are the most deficient as regards information on diasystematic labelling as well as other metalexicographic information, which is a finding that cannot be viewed with much optimism. Nowadays, in the age of modern technology, one can logically expect an increasing number of users for online dictionaries. It is true that currently, not everyone in the world enjoys good internet connectivity, but it is also true that access to the internet is becoming increasingly widespread, and this is expected to improve in the future. Also, the future of print dictionaries seems to be questionable (Macmillan, for example, announced in November 2012 that, in future, only online dictionaries will be available; http://www.macmillandictionaryblog.com/bye-print-dictionary), which means that the future of lexicography probably lies in online dictionaries. Consequently, users of online dictionaries need guidance as to certain aspects of dictionary use, and diasystematic information is doubt-

lessly an information category of the dictionary entry that needs to be explained in a comprehensive, clearly formulated, understandable and consistent way. Not all labels are self-explanatory; thus, it is essential for every type of dictionary, regardless of the medium, to provide its users with the necessary information. This is also in line with Norri (2000: 93), who says that "[i]n many cases greater care could be taken when explaining the method of labelling to the reader. All too often, the preface provides scant guidance in this important matter".

Another very important aspect is that all labels listed in the front matter of print dictionaries or in the CD-ROM or online versions should be used in the same form in the A–Z section. To have one form of a label in a list of labels and another form in the A–Z section is undesirable and most user-unfriendly and should certainly be avoided by compilers of all dictionaries, regardless of the target audience. Apart from that, all labels should be followed by a detailed explanation — one that the intended user will understand. In cases, where two different labels express similar connotative values, special care should be taken to explain the subtle differences as precisely as possible. The explanations should be short, concise and to the point, since a user should understand them immediately. Lexicographers should also avoid listing labels that are not used in the A–Z section (e.g., the labels *approving* and *disapproving* in LDOCE6), or omit labels that are used only sparsely in the entire A–Z section (e.g., the label *biblical* is used only seven times in LDOCE6).

Information about the restrictions and limitations provided by diasystematic information is vital for decoding and even more so for encoding. It should therefore be tailored to the needs and skills of the intended dictionary users, so that they can use this type of information correctly and efficiently. We should bear in mind that the intended user is a foreign learner who should receive clear guidance as to the use of lexical items s/he looks up in a dictionary. Consequently, dictionaries would benefit from theoretical classification of labels for their presentation in the front matter or in the electronic versions. To conclude, the possible problems that deficient (treatment of) diasystematic information could have for users should be regarded as a set of hypotheses that still have to be tested empirically, which should be a logical follow-up to this study.

Bibliography

Dictionaries

Carroll, Katherine (Ed.). 2012[7]. *Collins COBUILD Advanced Dictionary of English*. Glasgow: HarperCollins. (COBUILD7)

Delacroix, Laurence (Ed.). 2014[6]. *Longman Dictionary of Contemporary English*. Harlow, Essex: Pearson Education. (LDOCE6)

Deuter, Margaret, Jennifer Bradbery and Joanna Turnbull (Eds.). 2015[9]. *Oxford Advanced Learner's Dictionary of Current English*. Oxford: Oxford University Press. (OALD9)

Mayor, Michael (Ed.). 2009[6]. *Longman Dictionary of Contemporary English*. Harlow, Essex: Pearson Education. (LDOCE6)

McIntosh, Colin (Ed.). 2013[4]. *Cambridge Advanced Learner's Dictionary*. Cambridge: Cambridge University Press. (CALD4)

Rundell, Michael (Ed.). 2007[2]. *Macmillan English Dictionary for Advanced Learners*. Oxford: Macmillan Education. (MED2)

2014[2]. *Slovar slovenskega knjižnega jezika*. Ljubljana: Cankarjeva založba.

Other literature

Atkins, B.T. Sue and Michael Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.

Bergenholtz, Henning and Sven Tarp (Eds.). 1995. *Manual of Specialised Lexicography. The Preparation of Specialised Dictionaries*. Amsterdam/Philadelphia: John Benjamins.

Fedorova, Irina V. 2004. Style and Usage Labels in Learner's Dictionaries: Ways of Optimization. Williams, Geoffrey and Sandra Vessier (Eds.). 2004. *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004, Lorient, France, July 6–10, 2004*: 265-272. Loriet: Faculté des Lettres et des Sciences Humaines, Université de Bretagne Sud.

Hanks, Patrick. 2006. Metaphoricity is Gradable. Stefanowitsch, Anatol and Stefan Thomas Gries (Eds.). 2006. *Corpus-based Approaches to Metaphor and Metonymy*: 17-36. The Hague: Mouton de Gruyter.

Hausmann, Franz Josef. 1989. Die Markierung in einem allgemeinen einsprachigen Wörterbuch: eine Übersicht. Hausmann, Franz Josef, Oskar Reichmann, Herbert Ernst Wiegand and Ladislav Zgusta (Eds.). 1989. *Wörterbücher. Ein internationales Handbuch zur Lexikographie. Erster Teilband/Dictionaries. An International Encyclopedia of Lexicography. Volume 1/Dictionnaires. Encyclopédie internationale de lexicographie. Tome 1*: 649-657. Berlin/New York: Walter de Gruyter.

Jackson, Howard. 2002. *Lexicography: An Introduction.* London/New York: Routledge.

Landau, Sidney I. 2001[2]. *Dictionaries: The Art and Craft of Lexicography*. Cambridge: Cambridge University Press.

Norri, Juhani. 2000. Labelling of Derogatory Words in Some British and American Dictionaries. *International Journal of Lexicography* 13(2): 71-106.

Svensén, Bo. 2009. A Handbook of Lexicography: The Theory and Practice of Dictionary-making. Cambridge: Cambridge University Press.

Vrbinc, Marjeta and Alenka Vrbinc. 2013. Differences in the Inclusion and Treatment of Terminology in OALD3, OALD4 and OALD8. *Lexikos* 23: 440-455.

Internet Reference

http://www.macmillandictionaryblog.com/bye-print-dictionary (Accessed on 25 November 2015.)

http://lexikos.journals.ac.za

Eine neue Adressierungsart: Positionsadressierung bei indexikalischen Angaben und funktionalen Angabezusätzen

Herbert Ernst Wiegand, *Germanistisches Seminar, Universität Heidelberg, Heidelberg, Bundesrepublik Deutschland (wiegand.h.e.oberurff@t-online.de)*

Zusammenfassung: Nachdem die wichtigsten Begrifflichkeiten aus dem metalexikographischen Diskurs über Adressierung in den letzten 25 Jahren kurz in Erinnerung gerufen wurden, wird als neue Adressierungsart die Positionsadressierung von indexikalischen Angaben, wie z.B. Silbentrennungsangaben und Angaben der Abtrennbarkeit bei Partikelverben, sowie die von binnenerweiternden funktionalen Angabezusätzen, wie z.B. Auslassungs- und Versendekennzeichnungen, eingeführt. Weiterhin wird die Hypothese vertreten, dass auch nichttypographische Strukturanzeiger positionsadressiert sind. Die Annahme der Positionsadressierung trägt zum genaueren Verständnis der Informationsgewinnung des Benutzers-in-actu bei. Abschließend wird darauf hingewiesen, dass die Berücksichtigung der Positionsadressierung zur Differenzierung von Artikeltextstrukturen beiträgt, auf deren Trägermenge eine Adressierungsrelation definiert ist.

Stichwörter: INDEXIKALISCHE ANGABE, ANGABEADRESSIERUNG, ANGABEZUSATZ-ADRESSIERUNG, POSITIONSADRESSIERUNG

Abstract: A New Type of Addressing: Positional Addressing at Indexical Items and Functional Item Additions. Following a brief reminder of the most important concepts from the metalexicographic discourse of the past twenty five years, a new type of addressing is introduced. This is the positional addressing of indexical items, e.g. items indicating syllable division and items indicating the detachability of particle verbs as well as that of functional item additions, which expand items internally, e.g. omission and end of verse indicators. The hypothesis is further presented that non-typographical structural indicators are also positionally addressed. The supposition of positional addressing contributes to a more precise understanding of the information retrieval of the active user. Finally it is shown that the consideration of positional addressing contributes to the differentiation of article structures where an addressing relation is defined with regard to their structure-carrying set.

Keywords: INDEXICAL ITEM, ITEM ADDITION ADDRESSING, ITEM ADDRESSING, POSITIONAL ADDRESSING

1. Vorbemerkung zum Adressierungs-Diskurs von 1989–2013

Der lexikographietheoretische Diskurs zur Adressierung in ein- und zwei-

sprachigen Printwörterbüchern beginnt mit Hausmann/Wiegand (1989: 349ff.) und reicht derzeit bis Wiegand/Gouws (2013: 273-314). In dem Zeitraum, der zwischen dem Erscheinen dieser beiden Handbuchartikel liegt, wurden zahlreiche Arbeiten publiziert, die die Kenntnisse zur Adressierung erweitert und vertieft haben. Zu nennen sind hier — neben kleineren Darstellungen (vgl. dazu Wiegand 2011: 109-110) — vor allem die folgenden Beiträge, die in der Reihenfolge ihres Erscheinungstermins aufgeführt werden:

Hausmann/Werner 1991; Gouws 1994; Louw/Gouws 1996; Wiegand 2000; 2002; 2007; 2007a, 2011; Wiegand/Gouws 2013; Wiegand 2014: 421-442.

In rund drei Jahrzehnten wurden Forschungsergebnisse erzielt, die dazu geführt haben, dass man in der Metalexikographie das für große Teile der Lexikographie charakteristische Phänomen der Adressierung inzwischen wohl in den allermeisten Aspekten recht gut verstanden hat. Damit der Leser die Positionsadressierung als eine neue Adressierungsart leicht und angemessen in das gegebene metalexikographische Wissenskorpus einordnen kann, werden zunächst zur Erinnerung im folgenden Abschnitt einige gezielt ausgewählte Wissensausschnitte stark gerafft präsentiert.

2. Ausgewählte Essentials aus dem Wissenskorpus zur Adressierung

Nicht alle Texte, die zu einem Wörterbuch gehören, weisen Adressierung auf, sondern nur vollständig oder partiell kondensierte. Das sind vor allem die akzessiven Einträge und bei diesen insbesondere die Wörterbuchartikel. Ein vollständig kondensierter Artikeltext weist auf der Ebene der Wörterbuchform keine natürlichsprachliche Syntax auf, weil er nicht mittels der Grammatik irgendeiner natürlichen Sprache, sondern nach den Instruktionen eines lexikographischen Instruktionsbuches erzeugt wurde. Während bei natürlichsprachlichen Texten besonders die Grammatik der Sprache, in der ein Text geschrieben ist und einige Textbildungsregeln sowie gegebenenfalls einige extratextuelle Sachverhalte dem kompetenten Leser, der sie kennt, ermöglichen, anhand der textuellen Daten sprachliche Informationen als kognitive Gegebenheiten zu gewinnen, indem er — um nur einen einfachen Sachverhalt zu erwähnen — das Prädikat eines Satzes auf dessen Subjekt bezieht, ist die Sachlage für einen Benutzer-in-actu selbst dann, wenn er zu den kundigen Benutzern zählt und somit die Benutzungshinweise gut kennt, völlig anders. Denn ein kondensierter Wörterbuchartikel eines Sprachwörterbuches ist ein semiotisches Artefakt, in dem Teile und Eigenschaften einer oder mehrerer natürlicher Sprachen und/oder gegebenenfalls natürlicher Variäten mit Teilen und Eigenschaften anderer semiotischer Systeme zu einem komplexen Textgebilde mit einem genuin lexikographischen Textformat instruktionenbasiert zusammengefügt sind. Besonders die artikelinternen und gegebenenfalls die artikelexkurrenten Adressierungsbeziehungen muss ein Benutzer-in-actu kennen, wenn er nicht

nur formbezogene Informationen sucht, die die Rechtschreibung betreffen und die dadurch erhältlich sind, dass er bestimmte Angaben einfach aufmerksam wahrnimmt. Eine Adressierungsbeziehung führt immer von einem adressierten Textsegment zu einer anderen Textgegebenheit, die die Bezugsadresse trägt. Warum hier von *Textgegebenheit* gesprochen wird, wird später genauer erklärt. Jeder Benutzer-in-actu, der anhand eines Wörterbuchartikels Informationen erhalten möchte, die sich nicht nur auf die Form geschriebener Sprache beziehen, muss artikelinterne Daten korrekt auf andere Daten beziehen gemäß den Adressierungsbeziehungen, die im jeweiligen Wörterbuch gelten. Da es relativ viele Datentypen und Datendistributionstypen gibt, müssen auch zahlreiche Typen von Adressierung unterschieden werden. Dies sei im Folgenden lediglich angedeutet (ausführlich dazu u.a. Wiegand 2011).

Adressierte funktionale Textsegmente sind Angaben und funktionale Angabezusätze. Ob Angabetexte als adressiert zu gelten haben oder nicht, ist derzeit kontrovers. Man unterscheidet demgemäß zwischen Angabeadressierung und Angabezusatzadressierung.

(a) *Angabeadressierung*

Die Angabeadressierung kann links- oder rechtsgerichtet sein, je nachdem in welcher Richtung die Bezugsadresse vom Benutzer-in-actu zu suchen ist. Besonders die Angabelinksadressierung kann adjazent oder nichtadjazent sein. Sie ist dann adjazent linksgerichtet, wenn der Bezugsadressenträger unmittelbar links der adressierten Angabe situiert ist, wobei bei der Festlegung der unmittelbaren Nachbarschaft nichttypographische Mikrostrukturanzeiger nicht berücksichtigt werden. Sie ist einfach gedehnt linksgerichtet, wenn zwischen adressierter Angabe und Adressenträger gerade eine elementare Angabe situiert ist, und sie ist mehrfach gedehnt linksgerichtet, wenn die adressierte Angabe vom Adressenträger durch mindestens zwei elementare Angaben getrennt ist. Die Anzahl der elementaren Angaben zwischen einer adressierten Angabe und ihrem Adressenträger liefert das Maß für die Adressenentfernung. Sind es z.B. sieben Angaben, hat die Adressenentfernung den Wert 7. Findet sich eine adressierte Angabe auf einer rechten Wörterbuchseite und ihre Adresse auf einer der nächsten Wörterbuchseiten, so dass die adressierte Angabe und ihre Adresse nicht in ein und demselben Wahrnehmungsraum situiert sind, liegt gebrochene Adressierung vor. Bei der Angabeadressierung wird weiterhin zwischen unbedingter, bedingender und bedingter Adressierung unterschieden. Die beiden letztgenannten Adressierungsarten treten auf, wenn im Wörterbuchgegenstandsbereich, *wenn-dann*-Beziehungen gegeben sind, die lexikographisch vertextet werden. Eine weitere Unterscheidung ist die zwischen Angabenah- und Angabefernadressierung. Bei der Angabenahadressierung ist die Adressierungsbeziehung innerhalb akzessiver Einträge gegeben, so dass zwischen artikelinterner, außentextinterner, binnentextinterner, umtextinterner,

einschubinterner und registerinterner Angabenahadressierung unterschieden wird. Bei letzterer sind die adressierte Angabe und der Bezugsadressenträger in zwei verschiedenen akzessiven Einträgen situiert, die zugriffsverschieden sind, weil sie im Zugriffsbereich zweier verschiedener äußeren Zugriffsstrukturen liegen. Die Angabefernadressierung kann artikelexkurrent, binnentext-, umtext- und einschubexkurrent sein. Je nachdem, ob der Adressenträger eine lemmatische oder eine nichtlemmatische Angabe ist, wird zwischen lemmatischer und nichtlemmatischer Adressierung unterschieden. Die artikelinterne lemmatische (adjazente oder nichtadjazente) Linksadressierung ist wohl die am häufigsten auftretende Adressierungsart und gilt daher als der klassische Fall. In kondensierten akzessiven Einträgen, die keine Wörterbuchartikel sind und somit keine Lemmazeichengestaltangabe, sondern einen Eintragseingang, wie z.B. einen Umtexteintragseingang (kurz: Umtexteingang), aufweisen, entspricht der lemmatischen die Eingangsadressierung und der nichtlemmatischen die eingangsexterne Adressierung.

Angabenahadressierung tritt nicht nur innerhalb von akzessiven Einträgen auf, sondern auch innerhalb von funktionalen Artikelteilstrecken, und zwar in Artikelnestern und in Artikelnischen. In diesen sind die Nestsublemmata an das Nesteingangslemma und die Nischensublemmata an das Nischeneingangslemma adressiert; da diese Adressierungsbeziehungen nestartikel- bzw. nischenartikelexkurrent sind, ist artikelübergreifende Nest- bzw. Nischeneingangsadressierung gegeben, die mehrere Untertypen aufweist. Es sei darauf aufmerksam gemacht, dass eine Reihe von seltener auftretenden Besonderheiten bei der Angabeadressierung nicht erwähnt wurden.

(b) *Angabezusatzadressierung*

Funktionale Angabezusätze sind Textsegmente mit Angabefunktion aber ohne Textkonstituentenstatus. Man unterscheidet die oben erweiternden funktionalen Angabezusätze, die auch *lexikographische Superscripte* heißen, von den unten erweiternden funktionalen Angabezusätzen, die auch *lexikographische Subscripte* heißen. Im Unterschied zu den phonetischen Super- und Subscripten, die Teil des Wörterbuchgegenstands sind, gehören die lexikographischen zur Wörterbuchform. Weiterhin werden die binnenerweiternden funktionalen Angabezusätze unterschieden, die entweder glossierend oder nichtglossierend sind. Die glossierenden sind entweder Glossate oder Glossatfolgen, die nichtglossierenden sind u.a. grammtikbezogen, wie z.B. Genus-, Numerus- oder Kasuskennzeichnungen, oder phonetikbezogen, wie z.B. Aussprachekennzeichnungen. Während oben und unten erweiternde funktionale Angabezusätze sowohl intern- als auch externlemmatisch auftreten, finden sich binnenerweiternde funktionale Angabezusätze nur externlemmatisch. Lexikographische Superscripte, wie z.B. der Akut über die Lemmazeichengestaltangabe „b ó h r e n“ aus Weigand (1873) sind hinabadressiert; der unten situierte Adressenträger

(hier „o“) ist ein Angabeformsegment. Lexikographische Subscripte, wie z.B. der Unterstrich in „**liber<u>a</u>l**“ (HWDG 1984) sind hinaufadressiert; der oben situierte Adressenträger (hier „**a**“) ist ebenfalls ein Angabeformsegment. Binnenglossate sind grundsätzlich adjazent linksadressiert, es sei denn, sie sind nicht das vordere Glossat in Glossatfolgen, wie z.B. in „schmutzige (*unanständige, ordinäre, gemeine*) Witze, Bemerkungen machen“ (HWDG 1984). Nur das vordere Glossat ist adjazent links, das mittlere und das hintere Glossat sind nichtadjazent linksadressiert. In nichtelementaren Glossaten können die Glossatssegmente glossatintern entweder adjazent links adressiert sein, wie z.B. die Bedeutungsidentifizierungskennzeichnung „1“ in „**Schub** … K e g e l n alle neune auf ein S. (*mit einem Wurf* 1) treffen (HWDG 1984), oder sie sind adjazent rechtsadressiert, wie die verdichtete Synonymidentifizierungskennzeichnung „SYN“ in „**Ereignis** … *den ~en* (SYN ’Dingen 2.2`) *gefasst entgegensehen*“ (DGWDaF 2000). Glossierende Glossatsegmente, wie z.B. „Dingen“ sind immer glossatexkurrent linksadressiert.

Während alle kondensierten Wörterbuchartikel einem der Typen von Adressierungskonstellationen zugeordnet werden können, z.B. dem Typ der vollständig lemmatischen mit internlemmatischer Adressierung (vgl. z.B. Wiegand 2011: 194ff), gehören alle erweiterten Glossate zu einem der Typen von glossatzugehöriger Adressierungskonstellationen, z.B. zum Typ der vollständig dreifach gleichgerichteten Glossat- und Glossatsegmentadressierung (vgl. Wiegand (2014: 432ff.).

Gibt es in einem Wörterbuchartikel einen funktionalen Angabezusatz oder mehrere, weist der Artikel hybride Artikeltextstrukturen auf. Lexikographische Super- und Subscripte bewirken, dass die Artikeltextstrukturen architektonisch angereicherte Teilstrukturen aufweisen. Binnenerweiternde funktionale Angabezusätze bewirken, dass die Artikeltextstrukturen binnenerweiterungsbedingte Teilstrukturen aufweisen. Eine Übersicht über hybride Artikeltextstrukturen findet man in Wiegand (2015).

Damit sind wichtige Wissensvoraussetzungen genannt, die gegeben sein müssen, um vollständig zu verstehen, warum die Positionsadressierung eine neue Adressierungsart ist und was unter Positionsadressierung verstanden werden soll.

3. Positionsadressierung von Angaben und funktionalen Angabezusätzen

Zu den allermeisten Angabetypen gehören Angaben, die adressiert sind, wenn sie als Textkonstituenten Teile von Artikeltexten sind. Es gibt sehr wenige Angabetypen, deren zugehörige Angaben niemals adressiert sind. Die beiden prominentesten Angabetypen dieser Art sind der Typ der Lemmazeichengestaltangabe sowie der Typ des Formkommentars. Im Folgenden seien hierzu einige Beispiele betrachtet.

(1) **Scherz**, der; -es, -e *auf Belustigung berechnete Äußerung od. Handlung, Spaß*: ein gelungener, harmloser, netter, geistvoller, anzüglicher, grober, übler S.; zu Scherzen aufgelegt sein; sich (mit jmdm.) einen schlechten S. erlauben; etw. im S. (*nicht im Ernst, nicht in böser Absicht*) sagen + S. beiseite (*im Ernst*)
(= wa_1 aus HWDG 1984)

In wa_1 ist **Scherz** eine als lemmatische Substantivangabe ausgeprägte elementare Lemmazeichengestaltangabe; sie ist elementar, da sie nicht durch eine Anwendung der Methode der funktional-positionalen Segmentation in mindestens zwei Angaben segmentierbar ist. Sie ist durch keinen funktionalen Angabezusatz erweitert und ist daher auch mit dem Lemma als Element der Trägermenge der alphabetischen Hauptzugriffsstruktur des HWDG formgleich. Ihre Angabefunktion erfüllt sie dadurch, dass mit ihr das Lemmazeichen *Scherz* erwähnt (oder: genannt) wird; damit ist zugleich die Form für den Nominativ Singular des Substantivs *Scherz* in orthographisch korrekter Schreibung gegeben, die gemäß der deutschen lexikographischen Tradition die Nennform für ein Substantiv ist, zu dem, wenn es morphologisch vollständig entwickelt ist, insgesamt acht Flexionsformen gehören. Mit der Nennung von *Scherz* ist zugleich das Thema für wa_1 gegeben, das damit gemäß dem Mikrostrukturprogramm des HWDG (1984) lexikographisch bearbeitet werden kann. Während alle anderen Angaben in wa_1 ihre genuine Angabefunktion nur dadurch erfüllen können, dass sie artikelintern entweder lemmatisch oder nichtlemmatisch adressiert sind, so dass in wa_1 die Adressierungskonstellation der partiell lemmatischen Adressierung gegeben ist, sind der eingeschränkt adressenhomogene nichtelementare Formkommentar sowie die Lemmazeichengestaltangabe nichtadressiert. Die Lemmazeichengestaltangabe ist der Bezugsadressenträger für alle lemmatisch adressierten Angaben in wa_1. Die Bezugsadresse |Scherz| ist daher in dem Sinne polyfunktional, dass bei ihr mehrere lemmatische Adressierungsbeziehungen enden. In wa_1 gibt es keine Positionsadressierung, denn kein adressierungsfähiges Textsegment ist an eine textuelle Artikelposition adressiert.

(2) **Fi|lou** < [fil<u>u</u>>] m. 6> *Spitzbube, Gauner, Schlaukopf, gerissener od. leichtsinniger Mensch* [frz., „Spitzbube“ < engl. *fellow* „ Bursche“]
(= wa_2 aus Wahrig-^{8}DW)

In wa_2 ist die Lemmazeichengestaltangabe „**Fi|lou**“ durch eine Silbentrennungsangabe binnenerweitert. Silbentrennungsangaben heißen auch *Worttrennungsangaben* (vgl. Bergenholtz [et al. 1997]. Sie werden auch durch Punkte realisiert, wie in ˊ**Ab•gleich•chung** (aus BW 1980) oder durch einen unterbrochenen senkrechten Strich wie in Fe̩l¦**sen** (aus Hollós 2001). Bei Partikelver-

ben kann ein senkrechter Strich sowohl als Silbentrennungsangabe als auch als Angabe der Abtrennbarkeit fungieren, wie z.B. in ′**ab|strah•len** (aus BW 1980).

Im Folgenden wird zunächst die binnenerweiterte Lemmazeichengestaltangabe „**Fi|lou**“ aus wa_2 betrachtet. Der senkrechte Strich in „**Fi|lou**“ ist einwandfrei als Angabe zu klassifizieren, deren Angabeform durch ein Angabesymbol realisiert ist, da „**Fi|lou**“ durch eine Anwendung der funktional-positionalen Segmentation so segmentierbar ist, dass die beiden Silbenangaben „**Fi**“ und „**lou** sowie die Silbentrennungsangabe „|“ gegeben sind. Bisher galt die Hypothese für die Angabeadressierung, dass jede Angabe an eine Angabeadresse adressiert ist, deren Bezugsadressenträger ein Textsegment, und zwar entweder eine Angabe, ein Angabesegment oder ein Angabeformsegment ist. Daher tritt nun die Frage auf: An welche Angabeadresse ist die Silbentrennungsangabe „|“ in wa_1 adressiert? An die linkssituierte Adresse |Fi|, deren Adressenträger das Angabeformsegment „**Fi**“ ist, oder an die rechtssituierte Adresse **|lou|**, deren Adressenträger das Angabeformsegment „**lou**“ ist, oder an beide adjazent situierte Adressen? Es ist hier noch einmal *expressis verbis* daran erinnert, dass ein Benutzer-in-actu um eine lexikographische Information als kognitive Entität zu erhalten, adressierte Angaben auf ihre Angabeadressen beziehen muss. Das ein Benutzer-in-actu die Silbentrennungsangabe „|“ aber auf eine der beiden Silbenangaben oder gar auf beide bezieht, um die Information zu erhalten, dass eine Silbentrennung gerade an der Stelle möglich ist, an der die Silbentrennungsangabe steht, ist mehr als kontraintuitiv. Vielmehr ist es wohl doch eher so, dass eine Silbentrennungsangabe, gleichgültig, ob sie durch einen senkrechten Strich, einen unterbrochenen senkrechten Strich oder einen Mittenpunkt (oder: Mittelpunkt) realisiert ist, dem Benutzer die Stelle zeigt, an dem eine Silbentrennung erfolgen darf. Deswegen nenne ich ab jetzt Silbentrennungsangaben und alle Angaben, die auf die gleiche Weise funktionieren, im Anschluss an den Typ des indexikalischen Zeichens *indexikalische Angaben*, womit zugleich ein bisher nicht bekannter Angabetyp gefunden ist. Für indexikalische Angaben gilt: Sie sind positionsadressiert: Der Benutzer-in-actu muss mithin z.B. die Silbentrennungsangabe „|“ in „**Fi|lou**“ auf die textuelle Position beziehen, in der die Silbentrennungsangabe steht, damit er die lexikographische Information erhält, die zu vermitteln der genuine Zweck einer Silbentrennungsangabe ist. Das Gleiche gilt für die Angabe der Abtrennbarkeit bei Partikelverben und für weitere indexikalische Angaben wie beispielsweise die Angabe zur Identifizierung der Wortbildungskonstituenten in den Lemma **aber|kauf**, **aber|könig** und **ab|ermorden** (aus FWB 1989, vgl. dazu Reichmann 1998: 138). Angaben zur Identifizierung der Wortbildungskonstituenten finden sich auch in Lernerwörterbüchern, wie z.B. in **Fahrer|er•laub•nis** und **f<u>a</u>hr•plan|mäßig** (aus ^{3}LGWDaF 1993). Bei indexikalischen Angaben führt die Adressierungsbeziehung also nicht zu einem anderen Textsegment, sondern zu einer anderen Textgegebenheit (vgl. oben), nämlich demjenigen textuellen Ort, an dem die indexikalische Angabe steht.

Auch binnenerweiternde funktionale Angabezusätze können positionsadressiert sein, wie z.B. die in eckigen Klammern stehende Auslassungskenn-

zeichnung *[…]* in wa_3. Die eckigen Klammern gelten als Angabestrukturanzeigerpaar.

(3) **abbau**, **der** (daneben noch undurchsichtig gewordenes **abau** <*â – bû*; (e) s/-e. …
AUBIN, Weist. Köln/Brühl 41, 40 (rib., 15. Jh): *sal men [...] ouch wroigen alle abouwe des boefs*
(= wa_3 aus FWB 1989)

Die Auslassungskennzeichen "..." in wa_3 ist ein indexikalischer funktionaler Angabezusatz, der den Ort des inneren Belegtextschnitts kennzeichnet.

(4) **äffer**, *der*; **– /– Ø**
1. > Gaukler, Scharlatan, Possenreißer <; …
2. > Betrüger < ; zu *affen* 2.
FASTNACHTSP. 606,9 (nobd.,15. Jh.): *Du pist ain rechter trieger | Und ain effer und ain lieger*
(wa_4 aus FWB 1989)

In wa_4 ist der senkrechte Strich „|" in der Belegtextangabe ein binnenerweiternder indexikalischer funktionaler Angabezusatz vom Typ der Versendekennzeichnung. Zwei Versendekennzeichnungen mit paarigem Angabestrukturanzeiger können zu einer Versauslassungskennzeichnung „| *[…]*|" kombiniert werden, wie in:

(5) **affenort,** *das*
phraseologisch *jn. auf das affenort setzen*
> jn. zum Narren halten < mit Öffnung zu
> jn. betrügen <
FASTNACHTSP. 855, 9 (nobd., n. 1450); *Hort, junger geselle, sie ist mein swester | Und hat mir sere geclagt gestern, | […] | Du habst ir dick und vil gesworen | […]| Und hab deinen willen getan an gelt, | Des werde sie nimmermer von dir ersetzt, | Und habst sie auf das affenort gesetzt,* SCHLOSSER, H. v. Sachsenh. 1618 (schwäb., 1453): *Sie saczt mich uff das offen ort | Und hielt mich als ein jungen goch.*
(= wa_5 aus FWB 1989)

4. Sind artikelinterne nichttypographische Strukturanzeiger positionsadressiert?

Die letzte zusammenfassende Darstellung zu Strukturanzeigern ist meines Wissens ein Abschnitt in Wiegand 2005: 333-337. Alle dort formulierten Ausführungen sind auch derzeit noch gültig. Neu hinzugekommen sind die paarigen nichttypographischen Angabestrukturanzeiger (vgl. Wiegand 2010: 514), die nicht wie die nichttypographischen Mikrostrukturanzeiger den Benutzer dabei unterstützen, die artikelinterne Angabeverteilung leichter zu erkennen, son-

dern anzeigen, dass eine elementare Angabe eine für den Benutzer-in-actu relevante interne Gliederung aufweist. Solange die Positionsadressierung nicht bekannt war, war nicht zu erwarten, dass eine Frage danach, ob nichttypographische Strukturanzeiger als adressiert zu gelten haben oder nicht, als sinnvoll gelten könnte. Denn an welche Textsegmente sollten z.B. die drei satzzeichenartigen Trennzeichen (*sensu* Wiegand 2005: 335) im Formkommentar von wa_1, nämlich das Komma nach der Lemmazeichengestaltangabe „**Scherz**“, das Semikolon nach der Artikelangabe „der“ und das Komma nach der verdichteten Singularbildungsangabe „-es“ adressiert sein? Die Annahme, dass nichttypographische Strukturanzeiger vom Typ des Trennzeichens entweder adjazent links- oder adjazent rechtsadressiert sind, kann sicherlich nichts dazu beitragen, die Informationsgewinnung eines Benutzer-in-actu genauer zu erklären. Nimmt man aber an, das nichttypographische Strukturanzeiger, deren genuine Funktion ja wörterbuchformbezogen ist, gerade auf den textuellen Ort zu beziehen sind, an dem sie stehen, weil der Benutzer-in-actu durch sie erfährt, dass links von ihnen ein anderes Formelement steht als rechts von ihnen, dann ist die Annahme, nichttypographische Strukturanzeiger seien als indexikalische Zeichen zu verstehen und damit positionsadressiert eine einleuchtende Hypothese, die bei der Erklärung hilft, wie der Benutzer verschiedene Textsegmente entweder unterscheidet oder zusammenordnet.

5. Schlussbemerkung: Hinweise zur Differenzierung von Artikelangaben- und exhaustiven Artikelangabenstrukturen

Es ist klar, dass die Berücksichtigung der Positionsadressierung bei der Strukturbildung und Strukturdarstellung von Artikelangabenstrukturen und exhaustiven Artikelangabenstrukturen geringfügige Differenzierungen bewirken. Dies sei an der Darstellung der abstrakten hierarchischen Kernstruktur eines einfachen Formkommentars (fk_1) als Substruktur einer Artikelangabenstruktur und einer exhaustiven Artikelangabenstruktur erläutert, ohne dass der Strukturbildungsprozess eingehend erläutert wird.

fk_1: **Hoo|li|gan** [' hu:lign], der; -s, -s
(aus: Duden-^{7}DUW 2011)

Die exhaustive funktional-positionale Segmentation von fk_1 erbringt die folgenden Textsegmente, die bei der nachfolgenden Aufzählung mittels des Symbols „$\in$“ für die Element-Klassenrelation ihren Textsegmentklassen zugeordnet werden.

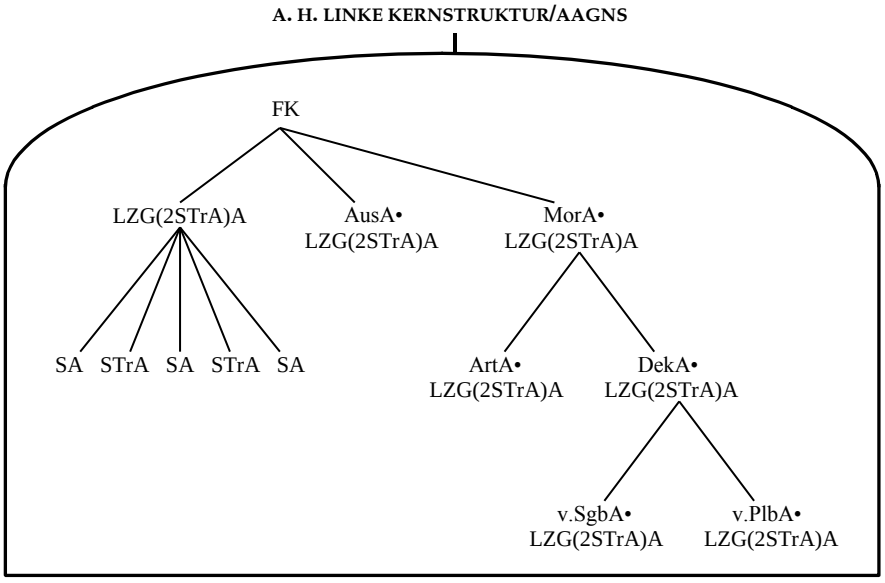
Hoo $\in$ SA (= Silbenangabe)
| $\in$ StrA (= Silbentrennungsangabe)
li $\in$ SA (= Silbenangabe)
gan $\in$ SA (= Silbenangabe)

[∈ *vZoZ* (= vorderes Zusammenordnungszeichen)
„hu:lign“∈ AusA (= Ausspracheangabe)
] ∈ *hZoZ* (= hinteres Zusammenordnungszeichen)
, [= Komma] ∈ *TZ* (= Trennzeichen)
der; -s, -s ∈ MorA.S (= Morphologieangabe bei Substantiven)
der ∈ ArtA (= Artikelangabe)
-s ∈ v.SgbA (= verdichtete Singularbildungsangabe)
; [= Semikolon] ∈ *TZ* (= Trennzeichen)
-s ∈ v.PlbA (= verdichtete Pluralbildungsangabe)

Im Folgenden werden zuerst zwei abstrakte hierarchische linke Kernstrukturen als Substrukturen der zugehörigen Artikelangabenstrukturen gebildet; Es sei daran erinnert, dass Artikelangabenstrukturen um Adressierungsstrukturen erweiterte Artikelmikrostrukturen sind. Bei der Struktur (1) wird die Positionsadressierung nicht, bei der Struktur (2) wird sie dagegen berücksichtigt; beide Strukturen finden sich in Abb. 1. Zunächst wird eine Trägermenge gebildet; sie heiße $\mathbb{M}^{a}_{liKS}(fk_1)$ und kann wie folgt mit einer Mächtigkeit von |10| notiert werden:

$\mathbb{M}^{a}_{liKS}(fk_1)$ = {FK, LZG(2STrA)A, SA, ST_1A, AusA, MorA.S, ArtA, DekKA, v.SgbA, v.PlbA}

Auf $\mathbb{M}^{a}_{liKS}$ (fk_1) werden dann drei strukturprägende Relationen definiert: eine Präzedenzrelation, eine partitive Relation und eine Angabeadressierungsrelation. Das Ergebnis dieser Definitionsoperationen ist die Struktur (1) in Abb. 1.



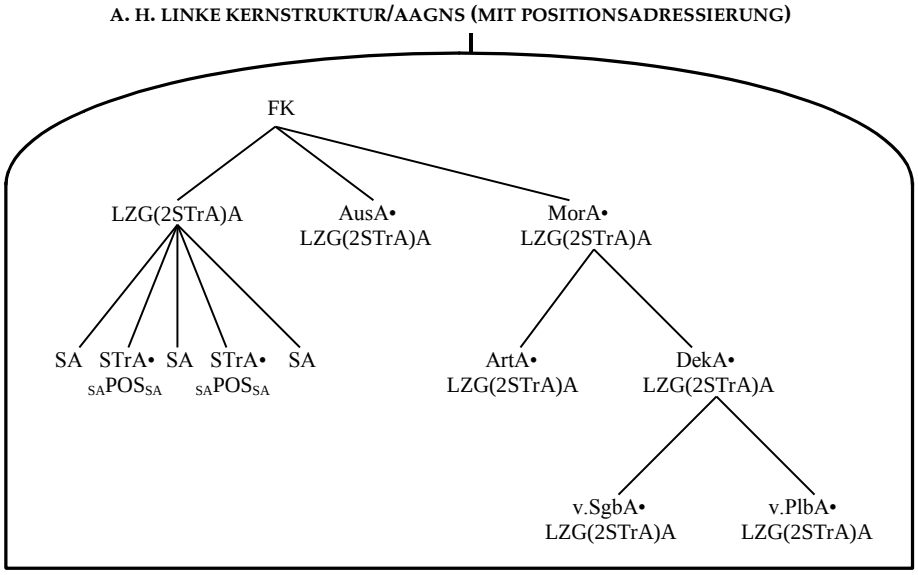
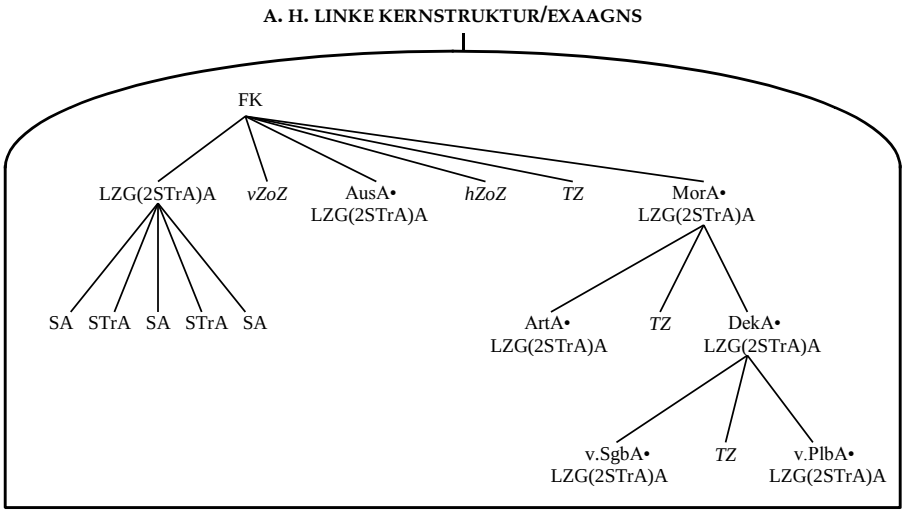


Abb. 1: Strukturgraphe für die abstrakte (a) hierarchische linke Kernstruktur (liKS) als Substruktur der abstrakten hierarchischen Artikelangabenstruktur, die zu fk_1 gehört; /AAGNS ist zu lesen wie *als Substruktur der Artikelangabenstruktur*; X•Y ist zu lesen wie *X ist adressiert an Y*

Im Folgenden wird die abstrakte hierarchische linke Kernstruktur als Substruktur der abstrakten hierarchischen Artikelangabenstruktur so gebildet, dass die Positionsadressierung der Silbentrennungsangaben Berücksichtigung findet. Dazu muss zunächst die Trägermenge $M^a_{liKS}(fk_1)$ um die Klasse der Positionen zur Trägermenge $erw^{M^a_{liKS}}(fk_1)$ erweitert werden. Da die Positionsadressierungsrelation und die Angabeadressierungsrelation zu demselben Relationstyp gehören, nämlich zum Typ der zweistelligen asymmetrischen und irreflexiven Adressierungsrelation, muss lediglich festgelegt werden, dass die Variable „*y*" im Relationsterm *x ist andressiert an y* auch mit Positionen (bzw. Klassen von Positionen) belegbar ist. Eine andere Möglichkeit besteht darin, dass man zwei Adressierungsrelationen auf $erw^{M^a_{liKS}}(fk_1)$ definiert, eine Angabe- und eine Positionsadressierungsrelation und die zugehörigen Belegungsvorschriften demgemäß spezifiziert. Als Klassensymbol für adressentragende Positionen wird „POS" gewählt. Das Symbol kann unten indiziert werden (z.B. $_xPOS_y$), um eine Unterklasse von Positionen zu spezifizieren, z.B. $_{SA}POS_{SA}$ als Symbol für Positionen zwischen zwei Silbenangaben. Definiert man auf $erw^{M^a_{liKS}}(fk_1)$ die drei strukturprägenden Relationen, ist das Ergebnis die Struktur (2) in Abb. 1.

Analog zum bisherigen Vorgehen werden nun zwei abstrakte hierarchische linke Kernstrukturen als Substrukturen der zugehörigen exhaustiven Artikelangabenstrukturen gebildet. Es sei daran erinnert, dass exhaustive Artikelangabenstrukturen erweiterte Artikelkonstituentenstrukturen sind (vgl. z.B. Wiegand 2015), so dass alle nichttypographischen Mikrostrukturanzeiger als Textkonstituenten berücksichtigt werden. Bei der Struktur (1) wird die Positionsadressierung nicht, bei der Struktur (2) wird sie dagegen berücksichtigt. Damit die Struktur (1) erhältlich ist, muss zunächst $\mathrm{M}^{\mathrm{a}}_{\mathrm{liKS}}$ (fk_1) um die Klassen der nichttypographischen Mikrostrukturanzeiger zu $\mathrm{M}^{\mathrm{a}}_{\mathrm{liKS}}$ (fk_1) erweitert werden, so dass sie wie folgt mit einer Mächtigkeit von |13| notiert werden kann: $\mathrm{erw}_1\mathrm{M}^{\mathrm{a}}_{\mathrm{liKS}}$ (fk_1) = {FK, LZG(2STrA)A, SA, STrA, AusA, MorA.S, ArtA, DekKA, v.SgbA, v.PlbA, *TZ*, *vZoZ*, *hZoZ*}. Definiert man auf dieser Trägermenge eine Relation vom Typ der Präzedenzrelation, eine vom Typ der partitiven Relation und eine vom Typ der Angabeadressierungsrelation, ist die Struktur (1) in Abb. 2 erhältlich. Die Struktur 2 in Abb. 2 ist erhältlich, wenn man die Trägermenge $\mathrm{erw}_1\mathrm{M}^{\mathrm{a}}_{\mathrm{liKS}}$ (fk_1) um das Symbol POS erweitert und die Symbole für die Klassen von nichttypographischen Strukturanzeigern *vZoZ*, *hZoZ* und *TZ* um einen unteren Index P erweitert, der anzeigt, das alle Elemente der Klasse positionsadressiert sind. TZ_p beispielsweise ist dann zu lesen wie *Klasse der Trennzeichen, die positionsadressiert sind*.



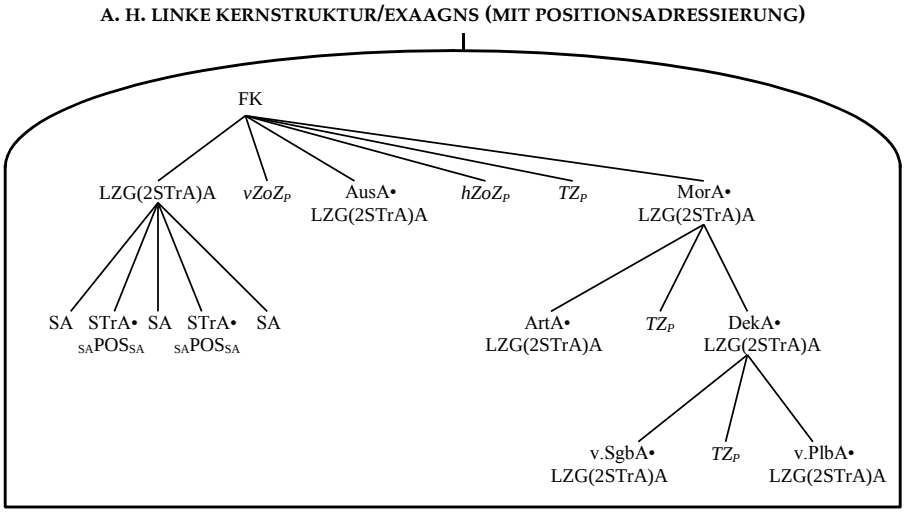


Abb. 2: Strukturgraph für die abstrakte hierarchische linke Kernstruktur als Substruktur der abstrakten hierarchischen exhaustiven Artikelangabenstruktur; /EXAAGNS ist zu lesen wie *als Substruktur der exhaustiven Artikelangabenstruktur*.

Literatur

Bergenholtz, Henning, Ilse Cantell, Ruth Vatvedt Fjeld, Dag Gundersen, Jón Hilmar Jónsson und Bo Svensén. 1997. *Nordisk Leksikografisk Ordbok ...* Oslo: Universitetsforlaget.

BW 1980 = Wahrig, Gerhard, Hildegard Krämer und Harald Zimmermann (Hrsg.). 1980. *Brockhaus-Wahrig. Deutsches Wörterbuch in sechs Bänden*. 1. Bd. *A–BT*. Wiesbaden/Stuttgart: F.A. Brockhaus/Dt. Verlagsanstalt.

DGWDaF 2000 = Kempcke, G. (Hrsg.). 2000. *De Gruyter Wörterbuch Deutsch als Fremdsprache.* Berlin/New York: Walter de Gruyter.

Duden-[7]DUW 2011 = Dudenredaktion (Hrsg.). 2011. *Duden. Deutsches Universalwörterbuch.* 7., überarb. u. erw. Aufl. Mannheim/Zürich: Dudenverlag.

FWB 1989 = Anderson, R.R., U. Goebel und O. Reichmann (Hrsg.). 1989. *Frühneuhochdeutsches Wörterbuch*. Bd. 1. Einführung *a–äpfelkern*. Reichmann, O. (Bearb.). Berlin/New York: Walter de Gruyter.

Gouws, Rufus H. 1994. Ostensiewe adressering in vertalende woordeboeke. *Lexikos* 4: 61-85.

Gouws, Rufus H., Ulrich Heid, Wolfgang Schweickard und Herbert Ernst Wiegand (Hrsg.). 2013. *Dictionaries. An International Encyclopedia of Lexicography*. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography. Berlin/Boston: Walter de Gruyter.

Hausmann, Franz Josef, Oskar Reichmann, Herbert Ernst Wiegand und Ladislav Zgusta (Hrsg.). 1989–1991. *Wörterbücher. Ein internationales Handbuch zur Lexikographie/Dictionaries. An International Encyclopedia of Lexicography/Dictionnaires. Encyclopédie internationale de lexicographie.* 3 Teilbände. Berlin/New York: Walter de Gruyter.

Hausmann, Franz Josef und Reinhard Otto Werner. 1991. Spezifische Bauteile und Strukturen zweisprachiger Wörterbücher: eine Übersicht. Hausmann, Franz Josef, Oskar Reichmann, Herbert Ernst Wiegand und Ladislav Zgusta (Hrsg.). 1989–1991: 2729-2769.

Hausmann, Franz Josef und Herbert Ernst Wiegand. 1989. Component Parts and Structures of General Monolingual Dictionaries: A Survey. Hausmann, F.J., O. Reichmann, H.E. Wiegand und L. Zgusta (Hrsg.). 1989–1991: 328-360. [Wieder abgedruckt in: Hartmann, R.R.K. (Hrsg.). 2003. *Lexicography: Critical Concepts. Vol. III: Lexicography, Metalexicography and Reference Science*: 207-254. London/New York: Routledge].

Hollós, Zita. 2001. *Német-magyar Suliszótar*. Szeged: Grimm Könyv kiadó.

HWDG 1984 = *Handwörterbuch der deutschen Gegenwartssprache*. 1984. In zwei Bänden. Von einem Autorenkollektiv unter der Leitung von Kempcke, G. […] Bd. 1: *A-K*; Bd. 2: *L-Z*. Berlin: Akademieverlag.

[3]LGWDaF 1993 = Götz, D., G. Haensch und H. Wellmann (Hrsg.). 1993. *Langenscheidts Großwörterbuch Deutsch als Fremdsprache. Das neue einsprachige Wörterbuch für Deutschlernende.* Berlin [u.a.]: Langenscheidt.

Louw, Phillip Adriaan und Rufus H. Gouws. 1996. Lemmatiese en nielemmatiese adressering in Afrikaanse vertalende Woordeboeke. *Suid-Afrikaanse Tydskrif vir Taalkunde* 14(3): 92-100.

Reichmann, Oskar. 1989. Lexikographische Einleitung. Anderson, R.R., U. Goebel und O. Reichmann (Hrsg.). 1989. *Frühneuhochdeutsches Wörterbuch*. Bd. 1. Einführung *a–äpfelkern*. Reichmann, O. (Bearb.). Berlin/New York: Walter de Gruyter.

Wahrig-[8]DW 2006 = Wahrig-Burfeind, Renate (Hrsg.). 2006. *Deutsches Wörterbuch. Mit einem Lexikon der Sprachlehre.* 8. vollständig neu bearb. Aufl. Gütersloh/München: Wissen Media Verlag GmbH.

Weigand, Friedrich Ludwig Karl. 1873. *Deutsches Wörterbuch*. 2. verb. u. verm. Aufl. 1. Bd. *A–L*. Gießen: Ricker.

Wiegand, Herbert Ernst. 2000. Adressierung in der ein- und zweisprachigen Lexikographie. Eine einführende Übersicht über die Forschungs- und Problemlage. *Lexikos* 10: 32-74.

Wiegand, Herbert Ernst. 2000. Wissen, Wissensrepräsentation und Printwörterbücher. Heid, U., St. Evert, E. Lehmann und Ch. Rohrer (Hrsg.). 2000. *Proceedings of the Ninth EURALEX International Congress. EURALEX 2000. Stuttgart, Germany, august 8–12.2000*: 15-38. Stuttgart: Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.

Wiegand, Herbert Ernst. 2002. Adressierung in zweisprachigen Printwörterbüchern. Wiegand, H.E. (Hrsg.). 2002. *Studien zur zweisprachigen Lexikographie mit Deutsch VIII*: 111-175. Germanistische Linguistik 166. Hildesheim/Zürich/New York: Georg Olms Verlag.

Wiegand, Herbert Ernst. 2005. Angaben, funktionale Angabezusätze, Angabetexte, Angabestrukturen, Strukturanzeiger, Kommentare und mehr. Ein Beitrag zur Theorie der Wörterbuchform. *Lexicographica* 21: 202-379.

Wiegand, Herbert Ernst. 2007. Il riferimento indirizzato nei dizionari bilingui a stampa. *Studi Italiani di Linguistica Teorica e Applicata* 36: 7-87.

Wiegand, Herbert Ernst. 2007a. Adressierung in Printwörterbüchern. Präzisierungen und weiterführende Überlegungen. *Lexicographica* 22: 187-261.

Wiegand, Herbert Ernst. 2010. Hybride textuelle Strukturen und hybride textuelle Einheiten. Ein Beitrag zur Theorie der Wörterbuchform. *Lexikos* 20: 495-539.

Wiegand, Herbert Ernst. 2011. Adressierung in der ein- und zweisprachigen Lexikographie. Eine zusammenfassende Darstellung. Kürschner, W. (Hrsg.). 2011. *Miscellanea Linguistica. Arbeiten*

zur Sprachwissenschaft: 109-234. Frankfurt a.M. [u.a.]: Peter Lang.

Wiegand, Herbert Ernst. 2014. Kotextspezifische Semantik, Pragmatik und Wörterbuchform. *Lexicographica* 30: 379-487.

Wiegand, Herbert Ernst. 2015. Funktionale Angabezusätze und hybride textuelle Strukturen in Wörterbuchartikeln. Korrigierende Ergänzungen, weiterführende Vertiefungen, zusammenfassende Übersichten und mehr. *Lexicographica* 31: 249-329.

Wiegand, Herbert Ernst und Rufus H. Gouws. 2013. Addressing and Addressing Structures in Printed Dictionaries. Gouws, R.H., U. Heid, W. Schweickard, H.E. Wiegand (Hrsg.). 2013: 273-314.

http://lexikos.journals.ac.za

Centre for Political and Related Terminology in Southern African Languages (CEPTSA) — Translating and Explanatory Dictionaries*

Mariëtta Alberts, *Research Unit for Languages and Literature in SA Context, North-West University, Potchefstroom Campus, Potchefstroom, South Africa (albertsmarietta@gmail.com)*

Abstract: The Centre for Political and Related Terminology in Southern African Languages (CEPTSA) published the revised and amended bilingual translating version of the Modern Political Dictionary (MPD) in 2011, and immediately started to define core political terms. The phases of the project, consisting of different translating and explanatory versions, are discussed.

The aim of CEPTSA is to promote the usage of political and related terminology in Southern Africa. Research is being done on these subject fields, and relevant concepts and terms are harvested, defined and translated. The source language is English and Afrikaans was initially the target language. The Centre, however, already compiled a list of 1000 core terms of the subject area in Tswana, Northern Sotho, Zulu and Xhosa. The Centre is currently busy with defining a further 1500 core terms and as soon as this process is finalised, the English/Afrikaans core terminology list and definitions will be published and the data translated into the official African languages.

The Centre provides a terminological and subject-related service to lecturers and under- and postgraduate students in international politics, political studies and governance, public administration, municipal government and administration, development studies and strategic studies. A service is also rendered to members of parliament, provincial legislature and local authorities, language practitioners and the media.

The Centre received several awards for the bilingual dictionaries already published, namely the Dictionary category of the South African Translators' Institute (SATI) (2003), the 'Woordfees' award of the *Afrikaanse Taal- en Kultuurvereniging* (ATKV) (Afrikaans Language and Cultural Association) (2006), and the Stals award of the *Suid-Afikaanse Akademie vir Wetenskap en Kuns* (South African Academy for Science and the Arts) (2010).

Keywords: BILINGUAL DICTIONARY, BI-DIRECTIONAL DICTIONARY, CORE TERMS, DICTIONARY, EXPLANATORY DICTIONARY, POLITICS, SOURCE LANGUAGE, SUBJECT FIELD, TARGET GROUP, TARGET LANGUAGE, TERMINOLOGY, TRANSLATING DICTIONARY

* This article was presented as a paper at the Twentieth Annual International Conference of the African Association for Lexicography (AFRILEX), which was hosted by the University of KwaZulu-Natal, Durban, South Africa, 6–8 July 2015.

Opsomming: Die Sentrum Vir Politieke en Verwante Terminologie in Suider-Afrikaanse Tale (CEPTSA) — Vertalende en verklarende woordeboeke. Die Sentrum Vir Politieke en Verwante Terminologie in Suider-Afrikaanse Tale (CEPTSA) het die hersiene en bygewerkte tweetalige vertalende weergawe van die *Nuwerwetse Politieke Woordeboek* (NWP) in 2011 gepubliseer en dadelik begin om kernterme te definieer. Die fases van die projek wat verskillende vertalende en verklarende weergawes behels, word bespreek.

Die doel van CEPTSA is om die gebruik van politieke en verwante terminologie in Suider-Afrika te bevorder. Navorsing word dus op hierdie vakterreine gedoen en relevante konsepte en terme word versamel, gedefinieer en vertaal. Die brontaal is Engels en aanvanklik was die doeltaal Afrikaans. Die Sentrum het egter reeds 'n lys met 1000 kernterme van die vakgebied in Tswana, Noord-Sotho, Zulu en Xhosa saamgestel. Die Sentrum is tans besig om 'n verdere 1500 kernterme te verklaar en sodra die proses afgehandel is, sal die Engels/Afrikaanse kernterme en definisies gepubliseer word en die data in die amptelike Afrikatale vertaal word.

Die Sentrum lewer 'n terminologies-vakkundige diens aan dosente en voor- en nagraadse studente in die internasionale politiek, staatkunde, publieke administrasie, plaaslike regering en administrasie, ontwikkeling- en strategiese studies. Die diens aan lede van die parlement, provinsiale wetgewers en plaaslike bestuur, taalpraktisyns en die media kan nie onderskat word nie.

Die Sentrum het verskillende toekennings gekry vir die tweetalige publikasies wat reeds die lig gesien het, naamlik die Woordeboekkategorie van die Suid-Afrikaanse Vertalersinstituut (SAVI) (2003), die Woordfeestoekenning van die Afrikaanse Taal- en Kultuurvereniging (ATKV) (2006) en die Stalprys van die Suid-Afikaanse Akademie vir Wetenskap en Kuns (SAAWK) (2010).

Sleutelwoorde: BRONTAAL, DOELTAAL, KERNTERME, POLITIEK, TEIKENGROEP, TERMINOLOGIE, TWEERIGTING WOORDEBOEK, TWEETALIGE WOORDEBOEK, VAKGEBIED, VERKLARENDE WOORDEBOEK, VERTALENDE WOORDEBOEK, WOORDEBOEK

1. Introduction

One of the main problems facing terminology development in a multilingual dispensation such as that of South Africa is the lack of subject specialists with an interest in terminology development. A further problem is that not all subject specialists have sufficient terminographical or linguistic background to be able to conceptualise in both source and target language(s). Nor do they have functional linguistic expertise to demarcate the concept and denote the concept in source and target languages. Functional linguistic expertise also entails knowledge in word-forming principles and spelling and orthography conventions. The various editions of the political dictionaries are excellent examples of subject specialists with expert knowledge in all the relevant political domains, a desire to develop the terminology of the subject field and with sufficient linguistic expertise to conceptualise in source and target languages. These subject specialists were also ably assisted throughout the compilation process by trained and experienced terminologists. The various editions of the dictionaries are excellent examples of interdisciplinary collaboration.

The Centre for Political and Related Terminology in Southern African lan-

guages (CEPTSA) has as main goal to facilitate a proper understanding of the political realm by compiling dictionaries in the various languages used in Southern Africa and particularly the official languages of South Africa. It is argued that political dictionaries assist with conceptualisation within the subject area and enhance the process of developing the official South African languages into functional languages (e.g. for higher learning and in the working environment). The Centre therefore provides assistance to subject specialists and language practitioners. The disciplines involved and which are broadly referred to as the political sciences, include politics, international politics, international relations, African politics, strategic studies, as well as aspects of political development, administration and political economy. The political sciences as discipline is a dynamic field with subject specialists, language practitioners and the ordinary public often being confronted with new terms almost on a daily basis. These terms come from all over the world (cf. Botha 2011).

The purpose with the current bilingual translating publications of the Centre is to facilitate the comprehension of the subject-related terminology of the political sciences, to assist with text reception and text production as well as to standardise translation of texts from English to Afrikaans and vice versa. Many of the dictionary entries are, however, from foreign languages and it also includes terms from the other official languages of South Africa. The database is continuously being updated and several updated reprints have seen the light between official editions.

An additional project of CEPTSA is the compilation of explanatory dictionaries. These dictionaries will contain definitions in all the official languages of South Africa. One thousand core terms have already been translated into an additional four of the official languages of South Africa, namely Northern Sotho (Sesotho sa Leboa/Sepedi), Tswana (Setswana), Zulu (isiZulu) and Xhosa (isiXhosa) and a draft list was disseminated to target users for their feedback.

The members of the Centre are currently finalising a project of 2500 core definitions in English. The members have already translated the terms and definitions into Afrikaans and are busy editing the source and target languages. The English–Afrikaans/Afrikaans–English translating and explanatory dictionary will be published while the committee compiles the multilingual dictionary containing all official African languages.

2. Background

The original bilingual bidirectional English–Afrikaans/Afrikaans–English translating political dictionary, namely *Political and Related Terminology/Staatkundige en Verwante Terminologie* (Government Printer, Pretoria) (cf. DAC 1989), was the result of a needs assessment study that was done in 1983–84 by the National Terminology Services (NTS) of the Department of National Education (later the Department of Arts, Culture, Science and Technology; now: Terminology Coordination Section, National Language Service, Department of Arts

and Culture). The needs assessment study indicated a need for a dictionary consisting of terminology of the political and related sciences. A Technical Committee for Political Sciences was established consisting of experts in the relevant fields. This committee functioned until 1997 under the auspices of the National Terminology Services.

In 1991 the Technical Committee for Political Sciences conducted another needs assessment study to determine the need for a revised and updated version of the 1989 publication. This committee started the revision process by excerpting terminology related to the international and national political scene in order to capture the enormous growth of new concepts in this and related fields. There was indeed a dire need for a revised and updated version.

The history of the *Centre for Political and Related Terminology in Southern African Languages* (CEPTSA), however, goes back to 1998 when the Department of Arts, Culture, Science and Technology (DACST) ceased all support for the project. CEPTSA was established in the Department of Politics and Governance, Rand Afrikaans University (now: University of Johannesburg), and continued with the work of the Technical Committee for Political Sciences under the auspices of this Department.

The Centre originally consisted of subject specialists from various institutions who were not only interested in the study field but also keen to compile a comprehensive political dictionary. The members were assisted in compiling the dictionary by skilled terminologists.

The Centre has a non-profit policy with the emphasis on addressing the needs of the various disciplines involved, as well as of society. The university does not fund the Centre, which is required to generate its own income. Since the proceeds from the sales of the dictionary/dictionaries and glossaries are totally insufficient to cover the costs of their production, the Centre is dependent on grants and donations for its work. While fulfilling a real need the products of the Centre are subject-specialised with a limited circulation and thus do not provide high visibility for sponsors, which makes it extremely difficult to obtain financial contributions (cf. Le Clus 2011).

The Centre is currently managed by a part time Director and Governing Committee to control the activities of the Centre under the supervision of the University Council. The work of CEPTSA is done by research fellows who are recognised experts in the various disciplines and languages represented. They work on a voluntary and part-time basis. The research team is supported by an experienced terminologist who handles the database. A locally developed database system, TshwaneTerm developed by TshwaneDJE, is used for the compilation of the various dictionaries and technical support is provided by TshwaneDJE throughout all the projects.

The overall purpose of the Centre is to promote political and related terminology in Southern African languages. The project is done within the South African and world context. The Centre fulfils the purpose by doing research with regard to political and related terms and how they may best be named

and systematised in the relevant languages. Moreover it compiles, makes available and promotes the use of bilingual translating dictionaries of political and related terms. The Centre is also in the process of compiling multilingual explanatory dictionaries and related products in Southern African languages and assists with the creation and use of political and related terms on request. The Centre aims to obtain the co-operation of experts and interested parties and to promote common interests with regard to political and related terms. Finally, the Centre actively endeavours to contribute to the linguistic and terminological empowerment of the community (cf. Le Clus 2011).

The centre renders a service to:

— Under- and postgraduate students in International Politics, Political Studies and Governance, Public Administration, Municipal Government and Administration, Development Studies, and Strategic Studies.
— Academic teachers of these subjects.
— Members of Parliament, Provincial Legislatures and Local Authorities.
— Language practitioners in the public sector.
— Language practitioners in the private sector.
— The media.
— Related institutions in other states.
— Institutions involved in the empowerment of the community (cf. Le Clus 2011).

The Centre has as its ultimate goal the compilation and publishing of political dictionaries in all the official languages of South Africa. As a starting point the Centre excerpted some 16 000 subject specialist terms in English and translated these terms into Afrikaans. This project was completed and the first edition of the full bilingual bi-directional Modern Political Dictionary was published in the course of 2002. An updated reprint appeared in 2006 and the third revised edition was published in 2011.

The next phase was to start with the African languages on the basis of the need from the various linguistic communities and the disciplinary expertise available. A start was made by defining 250 core terms in English and Afrikaans and translating them into Zulu and Northern Sotho as a trial run (cf. CEPTSA n.d.); this was followed by a further 750 terms with definitions that were translated into the same languages. These terms and their definitions were subsequently translated into Xhosa and Tswana also and published as a draft glossary of 1000 core political terms (cf. Le Clus 2011).

The glossary of 1000 core political terms is being expanded to 2500 core political terms and definitions in the same six languages. It was decided to complete and publish the English–Afrikaans version of the 2500 core political terms as soon as possible to allow subject specialists and language practitioners to utilise the data and to give feedback to the compilers of the dictionary before finalising the multilingual version.

The bi- and multilingual political dictionaries compiled by CEPTSA are good examples of a perfect terminographical process. The various stages and the execution of the project are discussed in the following sections. The current database contains many more terms than the previous publications and various additional annexures are also planned. The bilingual translating dictionaries and the envisaged multilingual explanatory political sciences dictionaries within South African and world context receive attention since they are valuable assests to the subject field and language communities.

3. The terminographical process

Political and societal changes in the new democratic South Africa have largely transformed the functional role of the indigenous languages. Section 6 of the Constitution provides for multilingualism and the development of the country's linguistic heritage (Department of Justice 1996). This system of official multilingualism has produced a substantial demand for terminology creation as various business matters (civil service departments, local administrative bodies, courts of law, amongst others) need to be conducted in the different official languages (cf. Government Gazette 2012, 2013). Multilingualism in South Africa is a sociolinguistic fact to be taken seriously (Alberts 1998: 230). A large proportion of South Africa's inhabitants can only be reached by means of indigenous languages. In a subject-related situation (i.e. politics), it is furthermore accepted that specialised information is conveyed and assimilated best through the mother-tongue or first language (cf. Alberts and Mollema 2013: 8). However, information flow is prevented from being established by factors such as low literacy rates, low levels of proficiency in English and terminologically poorly developed African languages (cf. Fourie 1994: 11-15).

A problem facing multilingual terminology development in South Africa is a shortage of subject specialists with an interest in terminology development. Furthermore very few of the subject specialists with such interest have a functional linguistic knowledge of the respective official languages. Terminology development is an interdisciplinary activity where the point of departure is subject related, but without linguistic input it is impossible to compile dictionaries for special purposes. The various bilingual and multilingual translating and explanatory political dictionaries compiled by CEPTSA will serve as examples of interdisciplinary collaboration where subject matter and language(s) successfully meet.

Terminology is a tool for communication in languages for special purposes, e.g. politics. It addresses user needs and language functions such as text reception (decoding) and text production (encoding). Terminology allows for subject-related communication between subject specialists and between subject specialists and laypeople. The documentation and standardisation of terms will ensure exact communication. Standardisation, however, is a process and terms

can only be standardised when they are systematised and documented, and through frequent usage are able to penetrate the subject field and languages.

The need for unambiguous communication in the theoretical and applied fields of human activity is constantly increasing. Subject specialists and terminologists are focused on the provision of unambiguous source- and target-language terms for well-defined concepts. In order to achieve this aim, it is necessary to determine the precise meanings of terms which enable users to comprehend and use them in a universally accepted manner (cf. Alberts and Mollema 2013: 1). Exact communication can only be achieved if the sender of the message and its receiver attach the same meaning to the given message. In the normal communication process, **specialised information** is **encoded** to be conveyed/transmitted from a **sender** (S1) as communication source to a **receiver** (R1) as recipient of information who will **decode** the **message** and then react on the stimulus received, i.e.:

S1 → specialised information → **R1**

Ambiguity in the specialised information may give rise to confusion and distortion of the communication process. One way to ascertain the exact meaning of a message conveyed through the medium of language is to document and standardise the terminology of languages for special purposes (LSP). When everyone in a specific language group working in a similar working environment understands the same message conveyed by a specific concept denoted by a specific linguistic label (the term), one can consider the term to be standardised. This is precisely what terminology is all about. It is the task of the terminologist to make sure that basic terminological principles, user needs, cultural differences and language attitudes of the professional group are taken into account when denoting concepts and coining terms (cf. Alberts and Mollema 2013: 8-9)

As background to vocabulary development and the acquisition of related terminology, a brief explanation of terminological theory is required. According to Sager (1990: 4), terminology is concerned with "the study and use of the systems of symbols and linguistic signs employed for human communication in specialised areas of knowledge and activities". Terminology is "a representation of an equally coherent, but possibly differently structured system of concepts" (Sager 1990: 114). A term is created when various linguistic labels are used to describe or name a specific object or concept. It therefore refers to a definite concept which is clearly defined within specific parameters. A term is as such the linguistic representation of a mental construct. There is a special interrelationship between the symbol, the concept (that is its mental representation in one's brain) and the various linguistic labels used in different languages to describe the object and concept. If this does not exist, a misunderstanding or miscommunication will result (Sager 1990: 57; Alberts and Mollema 2013: 9):

Concept: 

Definition: That branch of government which is charged with the making of authoritative and enforceable rules (laws) for a society

SL term: **eng:** legislative authority

TL term equivalents:

afr: *wetgewende gesag*
zul: *umkhandlu wesishayamthetho*
nso: *lekgotlatheramelao*
xho: igunya lomthetho
tsw: pusotheramelao (cf. CEPTSA database 2015)

Diagram 1: Conceptualisation and term creation (cf. Alberts and Mollema 2013: 9)

Bilingual and multilingual dictionaries act as facilitators in scientific or technical communication processes. The process of compiling terminology lists/glossaries/subject related dictionaries, however, is a long and tedious endeavour that requires input from subject specialists, language practitioners and linguists. Subject specialists often are not expertly versed in the rules of a particular language, but likewise language specialists such as terminologists, translators and linguists are not experts on the terminology of the discipline. Thus, the compilation of subject-related dictionaries should not involve the mere translation of terms and definitions by language practitioners and linguists. Such a project requires the input from experts in the discipline that are also mother-tongue speakers of the target language, as well as being quite knowledgeable of the source language (cf. Botha 2011).

The political sciences dictionary project was originally planned in three phases. The first phase was supposed to be the publication of the bilingual bidirectional translating dictionary (i.e. the 1989 edition). The second phase consisted of the defining of core terms and the third phase entailed the inclusion of terminology equivalents in the African languages (cf. Alberts 2003: 272). In practice, the first phase, however, consisted of the the revision and update of the 1989 bilingual dictionary and the phases were later replanned as:

Phase 1:	Bilingual Translating	**Phase 2**:	Bilingual Explanatory	**Phase 3**:	Multilingual Explanatory

Due to the new demanding political situation in South Africa in the 1990's seri-

ous planning was done to decide on new policies and strategies and these initiatives created numerous political and related terms. Although the members of CEPTSA started defining terms and were fully aware of the need for an explanatory dictionary, the process was slow. It was decided to rather revise the 1989 edition of the dictionary.

CEPTSA almost trebled the number of entries of the original dictionary and finalised and published its first bilingual bidirectional translating dictionary in 2002 (CEPTSA 2002). A third, revised edition of the full bilingual Modern Political dictionary, containing 16 000 political and related terms, was published in May 2011 (CEPTSA 2011).

The CEPTSA dictionary project with all its phases is an excellent example of how a terminographical project should be conducted. The concepts are excerpted and researched within their conceptual and contextual relationships. The terms are excerpted in the different subject fields in various languages. The members of the Centre consulted subject-specific sources such as textbooks, periodicals, technical and general dictionaries, and also popular sources such as magazines and papers. Personal contributions by subject specialists and queries by language practitioners also played a major role in the harvesting of terms.

The terms are then depicted in English, which was determined to be the source language (SL). Only after ascertaining the correctness of the SL terms in relation to their different concepts, the SL terms are supplied with target language (TL) equivalents. In the case of the bilingual translating dictionaries Afrikaans is the TL. This is a result of the bilingual policy of the previous dispensation.

In the case of explanatory dictionaries the excerpted terms are representative of core concepts in the political and related fields of study. The committee members, in collaboration with other experts, define these terms in the SL and only when there are absolute consensus on the definitions, these definitions are translated into the TL. This is also the process being followed for the multilingual version. CEPTSA supply the SL and TL terms and definitions to language practitioners to translate them into the various official African languages making use of the available disciplinary and linguistic expertise.

A start for a multilingual product was made by utilizing 250 core terms and definitions in English and Afrikaans and translating them into Zulu (isiZulu) and Northern Sotho (Sesotho sa Leboa/Sepedi) as a trial run (cf. CEPTSA n.d.). The members of CEPTSA selected a further 750 core terms to be defined and translated into the same four languages. These terms and definitions were subsequently translated into Xhosa (isiXhosa) and Tswana (Setswana) and published as a draft multilingual explanatory term list of 1000 core political concepts (cf. Le Clus 2011). The draft lists with six of the official South African languages were submitted to the relevant National Language Bodies of the Pan South African Language Board for verification and authentication by language experts. This term list was not officially published.

CEPTSA decided to expand the explanatory version of the dictionary by adding another 1500 concepts. The terms and definitions relating to these concepts are already defined in English and Afrikaans and the process to include the African language equivalents has started. It was decided to first publish the English–Afrikaans/Afrikaans–English version of the explanatory dictionary consisting of 2500 core concepts while the African language versions are being compiled.

The ultimate project will be a multilingual explanatory Modern Political Dictionary consisting of 16 000 translated terms and definitions in all official languages and in all relevant subject fields (cf. Le Clus 2011). The end-result will contain relevant annexures and will benefit the various target groups.

CEPTSA's computerised database gives total flexibility in printing the dictionary in totality or desired selections from it according to demand, i.e. a glossary in specific languages only or an *ad hoc* glossary for a particular purpose (cf. Le Clus 2011). The TshwaneTerm database is managed by an experienced terminologist. The members of the committee revise the content regularly and the database is then updated by the terminologist.

4. CEPTSA's bilingual translating dictionary projects

The *Political and Related Terminology/Staatkundige en Verwante Terminologie* published in 1989 was revised and updated considerably by the Technical Committee Political Sciences and later by CEPTSA, before publishing the 2002 and 2011 editions of the *Modern Political Dictionary/Nuwerwetse Politieke Woordeboek*. The Centre has already received three awards for the editions published, namely the Dictionary category of the South African Translators' Institute (SATI) (2003), the 'Woordfees' award of the *Afrikaanse Taal- en Kultuurvereniging* (ATKV) (Afrikaans Language and Cultural Association) (2006), and the Stals award of the *Suid-Afikaanse Akademie vir Wetenskap en Kuns* (South African Academy for Science and the Arts) (2010).

CEPTSA decided to divide the project into various phases. The publication of the 2002 and 2011 editions *Modern Political Dictionary* ends the first phase of the project. The latest (third) revision of the bilingual English–Afrikaans/ Afrikaans–English *Modern Political Dictionary* published in 2011 contains considerably more terms than the previous publications of 1989 and 2002. The central list contains the bilingual (English–Afrikaans) terms in alphabetical order. The content is reversed (Afrikaans–English) and various annexures were added as back matter. Some terms are also placed in their various conceptual relationships in annexures. The terms contained in the annexures are also listed in their alphabetical places in the central list, but the conceptual ordering in the annexures adds valuable assistance to the target user. User guidance is provided in an extensive front matter of the dictionary.

The 2011 edition contains subject-related material from international politics, political studies and governance, public administration, municipal gov-

ernment and administration, development studies and strategic studies. In excerpting the concepts the members harvested terms from textbooks, subject-related and popular magazines and newspapers and consulted general and technical dictionaries, considering the political sense of each term, e.g.:

administration: **administrasie**
administration {*of an act*}: **uitvoering** {*van 'n wet*}
administration {*customary in the USA, eg Clinton Administration*}: **administrasie** (**bewind**) {*gebruiklik in die VSA, bv Clinton-administrasie*}

The latest edition of the dictionary also contains several neologisms while problematic terms received attention and were in some cases amended. Gains from the indigenous languages as well as from foreign languages were added. The committee dealt with orthography and transliteration problems that developed from the Afrikaansification of borrowed terms. Another valuable asset is the variety of functional suffixes that were used productively in the *Modern Political Dictionary*. The dictionary also contains quite a number of abbreviations for full forms.

4.1 Neologisms

Various concepts were sourced as English terms and these terms needed Afrikaans equivalents:

brain gain: kundigheidswins
fat cat: roomvraat (geiljan, vetkat <af te keur>)
gutter education: flenteronderwys (afskeeponderwys)
impeachment {*charge against incumbent of a specific office*}: ampsaanklag {*aanklag teen bekleër van 'n bepaalde amp*}
impeachment: staat van aanklag
kickback: gunsloon

4.2 Problematic terms

It sometimes poses a problem when a SL term needs to be translated. In some cases these terms cannot easily be translated since specific connotations (even sometimes emotional connotations) are either attached to the term which cannot be translated in the TL without the meaning being lost or term equivalents already exist in the TL which do not convey the same meaning as that of the SL terms:

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 movements 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}*

4.3 Gains from indigenous languages

Several terms originating from the indigenous languages are already part of the English and Afrikaans vocabulary, e.g. **shebeen**: sjebien; **spaza shop**: spaza-winkel; **stokvel**: stokvel.

Other terms gained from the official African languages are:

4.3.1 Afrikaans

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}

4.3.2 Sotho (Northern Sotho/Sesotho sa Leboa/Sepedi and South Sotho/Sesotho

difaqane <Sotho> (**mfecane** <Nguni> {displacement of population caused by wars — e.g. those by Shaka}: **difaqane** <Sotho> (**mfecane** <Nguni> {bevolkingsverplasing agv oorloë — bv dié deur Shaka}

4.3.3 Tswana (Setswana)

Kgosi <pl diKgosi; Setswana> {leader of a group in Tswana communities}: **Kgosi** <mv diKgsoi; Setswana> {leier van 'n groep in Tswana-gemeenskappe}

4.3.4 Zulu (isiZulu)

amakhosi <Zulu — pl of inkosi> {traditional leaders by birth}: **amakhosi** <Zulu — mv van inkosi> (tradisionele leiers deur geboorte)

imbongi <Zulu> (praise singer): **imbongi** <Zulu> (lofsanger)

inkatha <Zulu> (literally — a coil that Zulu women use to carry loads on their

heads}: **inkatha** <Zulu> (letterlik — 'n klos waarmee Zuluvrouens vragte op hul koppe dra)

inkatha <Zulu> {figuratively — unity is strength}: **inkatha** <Zulu> (figuurlik — eendrag maak mag)

mukhuku <Zulu> (**squatter shack, shack**): **mukhuku** <Zulu> (plakkershut)

umKhonto weSizwe <Zulu> *{spear of the nation}* **(MK)**: **umKhonto weSizwe** <Zoeloe> *{spies van die nasie}* **(MK)**

4.4 Gains from foreign languages

— **emir** <Arabic>: emir <Arabies>; **kalam** <Arabic>: kalam <Arabies>
— **émigré** <French>: émigré <Frans> **force majeur** <French>: force majeur <Frans>; **gendarme** <French>: gendarme <Frans>; **grande bourgeoisie** <French>: hoër middelklas
— **Abwehr** <German>: Abwehr <Duits>; **Anschluss** <German>: Anschluss <Duits>; **Bundeswehr** <German>: Bundeswehr <Duits>
— **anabasis** <Greek: anabasis <Grieks>; **boule** <Greek>: boule <Grieks>; **katabasis** <Greek>: katabasis <Grieks>; **enosis** <Greek>; enosis <Grieks>
— **Irgun Zva'i Leumi** <Hebrew>: Irgun Zva'i Leumi <Hebreeus>
— **condottieri** <Italian>: condottieri <Italiaans>
— **Kokkai** <Japanese>: Kokkai <Japannees>
— **aldeamentos** <Portugese>: aldeamentos <Portugees>; **favela** <Portugese>: favela <Portugees>; **indigena** <Portugese>: indigena <Portugees>; **infanta** <Portugese>: infanta <Portugees>; **infante** <Portugese>: infante <Portugees>
— **instituchiki** <Russian>: instituchiki <Russies>
— **barrio** <Spanish>: barrio <Spaans>; **conquistador** <Spanish>: conquistador <Spaans>; **desaparecidos** <Spanish>: desaparecidos <Spaans>; **infanta** <Spanish>: infanta <Spaans>; **infante** <Spanish>: infante <Spaans>
— **Laban ng Demokratikong Pilipino** <Tagalog/Filipino>: Laban ng Demokratikong Pilipino <Tagalog/Filippyns>
— **khan**<Turkomenic>: khan <Turkmeens>

4.5 Orthography and transliteration

The members of CEPTSA consulted with the Dictionary of South African English, i.e. the National Lexicography Unit for English, regarding the English orthography. They also checked the spelling of source language terms, especially the terms originating from foreign languages. Various embassies were consulted regarding the spelling of foreign terms. In the case of Afrikaans orthography the *Taalkommissie* (Language Commission for Afrikaans) and the *Afrikaanse Woordelys en Spelreëls* (2002 and 2009 editions) were consulted. The Bureau of the Woordeboek van die Afrikaanse Taal, i.e. the National Lexicog-

raphy Unit for Afrikaans, was also consulted. Examples of spelling or orthographic conventions that were problematic during the transliteration process are:

— **faqih** <Arabic>: faki; **fedayeen** <Arabic>: fedajien
— **Chung Kuo** <Chinese>: Tjong Gwo; **ganbu** <Chinese>:ghanboe; **Kuo-min ta-hui** <Chinese>: Gwo-min ta-hê
— **Ashkenazim** <Hebrew>: Asjkenasim; **kibbutz** <Hebrew>: kibboets; **Hasidim** <Hebrew>: Gasidim
— **Kshatriya** <Indian languages>: Tsjatrija
— **jasus-khane** <Iranian>: jasoes-gane
— **keterbukaan** <Indonesian>: kieterboekaan
— **bushido** <Japanese>: boesjidô; **keiretsu** <Japanese>: keiretsoe; **kinken seiji** <Japanese>: kinken sêji; ko <Japanese>: kô; **koenkai** <Japanese>: kôênkai
— **chaebol**<Korean>: **tjêbil**; juche <Korean>: joetje
— **kulak** <Russian>: koelak

In the case of borrowings from the official African languages the National Language Bodies and the National Lexicography Units for these languages were consulted. Care was taken to ensure correct translations and/or pronunciation in cases of transliteration, cf.:

— **kaya** <Zulu>: kaia; **induna** <Zulu>: indoena
— **difaqane** <Sotho>: difaqane <Sotho>
— **Kgosi** <Tswana>: Kgosi <Tswana>

4.6 Functional suffixes

Certain suffixes were very functional in the 2011 version of the dictionary, i.e. -acy; -archy; -cracy; -ism:

-acy {"state of"/"plan"/"expertise"/"argument"/"effect"/"system"} cf. conspir**acy**, diplom**acy**, effic**acy**, fall**acy**, liter**acy**, pap**acy**, suprem**acy**

-archy {"rule by"} cf. an**archy**, aut**archy**, call**archy**, di**archy**, mon**archy**, olig**archy**, patri**archy**

-cracy {"exercise of power by"} cf. aristo**cracy**, auto**cracy**, bureau**cracy**, demo**cracy**, ergoto**cracy**, geronto**cracy,** hagio**cracy,** hiero**cracy**, iso**cracy**, kakisto**cracy**, nomo**cracy**, ochlo**cracy**, pluto**cracy**, sham**cracy**, techno**cracy**, theo**cracy**

-ism {"action or convention"/"nature or condition"/"principle, doctrine or supporting organization"} cf. commun**ism**, dual**ism**, elit**ism**, fanatic**ism**, global**ism**, hol**ism**, Chart**ism**, Confusian**ism**, Fasc**ism**, Gaull**ism**, Hertzog**ism**

4.7 Acronyms and Abbreviations

Acronyms and abbreviations commonly used in this field are incorporated in

the dictionary. The full forms as well as the abbreviated versions thereof are entered in the main part of the dictionary. Complete information such as contextual references and elucidation of terms appear at the full form entry, and the acronym or abbreviation refers to the full form. e.g.:

IDSEO ⇨Investigating Directorate for Serious Economic Offences
Investigating Directorate for Serious Economic Offences <formerly Office for Serious Economic Offences> **(IDSEO)**: **Ondersoekdirektoraat vir Ernstige Ekonomiese Misdrywe** <voorheen Kantoor vir Ernstige Ekonomiese Misdrywe> **(ODEEM)**

targeted procurement (TP) *{preferential public purchasing from designated groups in order to achieve policy objectives}***: beteikende verkryging (BV)** *{voorkeur openbare aankope by aangewese groepe ter wille van beleidsdoelwitte}*
TP ⇨targeted procurement

This list is not meant to be exhaustive and the main focus of the acronyms and abbreviations selected for inclusion is directed at South Africa and Africa. The list also contains the most important ones occurring internationally.

In addition to the abbreviations and acronyms appearing at their alphabetical place in the main part of the dictionary, they are also listed in easily visible form in a separate annexure to facilitate quick reference.

Examples of acronyms in the English annexure:

CEDAW: Convention on the Elimination of all forms of Discrimination Against Women
CODESA: Convention for a Democratic South Africa
CONTRALESA: Congress of Traditional Leaders of South Africa
FAPLA: Forças Armadas Populares para a Libertaçao de Angola
NEHAWU: National Education, Health and Allied Workers' Union
NUMSA: National Union of Metalworkers of South Africa
OAPEC: Organisation of Arab Petroleum Exporting Countries

Examples of abbreviations in the English annexure:

CCMA: Commission for Conciliation, Mediation and Arbitration
EAC: East African Community
EC: European Community
KGB: Komitet Gosudarstvennoi Bezopasnosti
OBE: outcomes based education
OBE: Order of the British Empire
SPQR: Senatus Populusque Romanus

Examples of acronyms in the Afrikaans annexure:

FAPLA: Forças Armadas Populares para a Libertaçao de Angola

IDASA: Instituut vir 'n Demokratiese Alternatief vir Suid-Afrika
KONTRALESA: Kongres van Tradisionele Leiers van Suid-Afrika
kwango: kwasi-outonome owerheidsinstelling
NAVO: Noord-Atlantiese Verdragsorganisasie
NEHAWU: National Education, Health and Allied Workers' Union
OPUL: Organisasie van Petroleumuitvoerlande

Examples of abbreviations in the Afrikaans annexure:

IOA: Internasionale Ontwikkelingsagentskap
KVSOSA: Konferensie oor Veiligheid, Stabiliteit, Ontwikkeling en Samewerking in Afrika
OBE: Orde van die Britse Ryk
PFP: Progressiewe Federale Party
SAOB: Suider-Afrikaanse Ontwikkelingsbank
SPQR: Senatus Populusque Romanus
WVK: Waarheid-en-Versoeningskommissie

4.8 Annexures

Various annexures are added as back matter to enhance the value of the dictionary.

4.8.1 Parliaments of the world

The table has been arranged alphabetically according to the popular names of the states in English. The popular name is followed by the formal name of the state, which in turn is followed by the name of the state in its original language when that language is not English. Where parliaments are bicameral the two chambers are given, but the sequence does not indicate which one has decisive legislative power. Examples of the parliaments of the world are:

STATE: STAAT	**PARLIAMENT: PARLEMENT**	**CHAMBER: KAMER**	**CHAMBER: KAMER**
Australia: Australië Commonwealth of Australia: Gemenebes Australië	Federal Parliament: Federale Parlement	House of Representatives: Huis van Verteenwoordigers	Senate: Senaat
Austria: Oostenryk **Republic of Austria: Republiek Oostenryk** ***Republik Österreich***	Federal Assembly: Bondsvergadering ***Bundesversammlung***	National Council: Nasionale Raad **Nationalrat**	Federal Council: Bondsraad *Bundesrat*

Canada: Kanada	Parliament: Parlement Parliament/Parlement	House of Commons: Laerhuis House of Commons/Chambre des Communes	Senate: Senaat Senate/Sénat
Czech Republic: Tsjeggiese Republiek Ćeská Republika	Parliament of the Czech Republic: Parlement van die Tjeggiese Republiek Parlament České Republiky	Chamber of Deputies: Kamer van Afgevaardigdes Sněmovna Poslancu	Senate: Senaat Sénate
Ethiopia: Etiopië Federal Democratic Republic of Ethiopia: Federale Demokratiese Republiek Etiopië Ye Ethiopia Faderalawe Democracyawe Republic	Parliament: Parlement	House of People's Representatives: Huis van Volksverteenwoordigers Yehizib Twekayoch Mekir Ena	House of Federation: Huis van Federasie Yefedereshn Mekir Bet
Malaysia: Maleisië	Parliament: Parlement	House of Representatives: Huis van Verteenwoordigers Dewan Rakyat	Senate: Senaat Dewan Negara
Netherland: Nederland Kingdom of the Netherlands: Koninkryk Nederland Koninkrijk der Nederlanden	States General: State-generaal Staten-Generaal	First Chamber: Eerste Kamer Eerste Kamer	Second Chamber: Tweede Kamer Tweede Kamer
United Kingdom: Verenigde Koninkryk United Kingdom of Great Britain and Northern Ireland: Verenigde Koninkryk Groot Brittanje en Ierland	Parliament: Parlement	House of Commons: Laerhuis	House of Lords: Hoërhuis
United States of America: Verenigde State van Amerika	Congress: Kongres	House of Representatives: Huis van Verteenwoordigers	Senate: Senaat

4.8.2 Titles and offices of the nobility, e.g.

Titles and offices of the nobility are extraneous to South Africa but the list was included since they play a role in past and present day politics, e.g.

Beefeater ⇨Yeoman of the Guard
count n. *{European nobleman whose rank is equivalent to a British earl}*: **graaf** *{Europese edelman wie se rang gelykstaande aan 'n Britse graaf is}*
dame *{British; the title of a woman awarded the Order of the British Empire or several other orders, equivalent to a knight}*: **dame** <uitspraak daam> *{Brits; titel van 'n vrou aan wie die Orde van die Britse Ryk of etlike ander ordes toegeken is, gelykwaardig aan 'n ridder}*
dauphin <French> *{eldest son of French monarch}*: **dauphin** <Frans> *{oudste seun van Franse monarg}*
dauphine <French> *{eldest daughter of French monarch}*: **dauphine** <Frans> *{oudste dogter van Franse monarg}*
Ladyship <preceded by Your or Her> *{title used to address or refer to any peeress, except a duchess}*: **Ladyskap** <voorafgegaan deur U of Haar> *{titel wat gebruik word om enige edelvrou, behalwe 'n hertogin, aan te spreek of na te verwys}*
Master of the Horse: **Opperstalmeester**
prince royal *{eldest son of eg British monarch}*: **vorsteprins** *{oudste seun van bv Britse monarg}*
Yeoman of the Guard *{ceremonial; Tower of London}* **(Beefeater)**: **Lid van die Koninklike Lyfwag** *{seremonieel; Tower of London}*

4.8.3 Classical Latin and Greek political terms, e.g.

Terms from classical Rome and Greece are still relevant to the study of politics and were included in the dictionary and, again for ease of reference, grouped together in an annexure, e.g.

aedilis <Latin> *{magistrate in classical Rome entrusted with public works}*: **aedilis** <Latyn> *{magistraat in klassieke Rome belas met openbare werke}*
archon *{title of a high political leader in ancient Greece}*: **argon** <mv argonte> *{titel van 'n hoë politieke leier in antieke Griekeland}*
centuria <Latin> *{Roman military unit}*: **centuria** <Latyn> *{Romeinse militêre eenheid}*
cohortes praetoriae <Latin> *{praetorian guard}*: **cohortes praetoriae** <Latyn> *{pretoriaanse wag}*
concilium plebis <Latin> *{assembly from the people of classical Rome empowered to approve decrees affecting plebeians}*: **concilium plebis** <Latyn> *{vergadering uit die volk van klassieke Rome gemagtig om dekrete wat die plebejers raak, goed te keur}*
ecclesia <Greek> *{Athens, meeting of the demos}*: **volksvergadering** *{Athene, vergadering van die demos}*

gerusia <Greek> *{classical Greece}*: **stamraad** *{klassieke Griekeland}*
nomos <Greek> *{law; classic Greece}*: **nomos** <Grieks> *{wet; klassieke Griekeland}*
Senatus Populusque Romanus <Latin> *{senate and population of Rome}* **(SPQR)**: **Senatus Populusque Romanus** <Latyn> *{senaat en bevolking in Rome}* **(SPQR)**
SPQR ⇨Senatus Populusque Romanus

4.8.4 British, American and South African ministerial hierarchy and administrative offices

The terminology applicable to ministerial and senior administrative offices differs considerably between South Africa, the United Kingdom and the United States of America which makes it difficult for the layperson to relate to these terms. This causes translation and interpretation problems. CEPTSA included these terms in the main part of the dictionary as well as in a separate annexure to assist with these problems. Examples are:

BRITISH OFFICES: BRITSE AMPTE

Ministerial offices: Ministeriële ampte
secretary of state {also member of cabinet}: **minister** {ook lid van kabinet}
minister of state {usually not member of cabinet}: **staatsminister** {gewoonlik nie lid van kabinet nie}
parliamentary secretary: **parlementêre minister**
parliamentary under-secretary {junior minister}: **parlementêre onderminister** {junior minister}
parliamentary private secretary {backbencher in parliament who assists a minister}: **parlementêre privaatsekretaris** {agterbanker in parlement wat 'n minister bystaan}
law officer {eg Solicitor General}: **regsbeampte** {bv Sollisiteur-generaal}

The Cabinet: Die Kabinet
Prime Minister, First Lord of the Treasury and Minister for the Civil Service: Eerste Minister, Eerste Lord van die Tesourie en Minister van die Staatsdiens
Deputy Prime Minister and First Secretary of State: Adjunk Eerste Minister en Eerste Kabinetsminister
Chancellor of the Exchequer: Minister van Finansies
President of the Council and Leader of the House of Commons: President van die Raad en Leier van die Laerhuis
Lord Chancellor: Lord Kanselier
Secretary of State for Foreign and Commonwealth Affairs: Minister van Buitelandse en Statebondsake
Etc.

AMERICAN OFFICES: AMERIKAANSE AMPTE

Office of the President: Kantoor van die President
President: President
Vice-President: Visepresident
Executive Office of the President: Uitvoerende Kantoor van die President
Chief of Staff to the President: Stafhoof vir die President
Assistant to the President and Cabinet Secretary: President se Assistent en Sekretaris van die Kabinet
Assistant to the President and Director of Legislative Affairs: President se Assistent en Direkteur: Wetgewende Aangeleenthede
Note. There are some twelve such assistants.
Nota. Daar is sowat twaalf sulke assistente.
National Security Adviser: Nasionale Veiligheidsadviseur

Members of the Presidential Cabinet and cabinet-rank officers: Lede van die President se Kabinet en beamptes met kabinetsrang
Secretary of Agriculture: Minister van Landbou
Secretary of Commerce: Minister van Handel
Secretary of Defence: Minister van Verdediging
Etc.

SOUTH AFRICAN OFFICES: SUID-AFRIKAANSE AMPTE

Office of the President: Kantoor van die President
President: President
Deputy President: Adjunkpresident

The Cabinet: Die Kabinet
Minister of Agriculture and Land Affairs: Minister van Landbou en Grondsake
Minister of Arts, Culture, Science and Technology: Minister van Kuns, Kultuur, Wetenskap en Tegnologie
Minister of Communications: Minister van Kommunikasie
Minister of Correctional Services: Minister van Korrektiewe Dienste
Minister of Defence: Minister van Verdediging
Minister of Education: Minister van Onderwys
Minister of Environmental Affairs and Tourism: Minister van Omgewingsake en Toerisme
Minister of Finance: Minister van Finansies
Minister of Foreign Affairs: Minister van Buitelandse Sake
Etc.

4.8.5 Typical military ranks

The rank structure added to the annexure is intended to acquaint users with

the way in which military forces are typically structured, but the rank nomenclature does vary among the armed forces of the world. The list features military ranks in the Army, Airforce and the Navy and contains information on officers such as **Field Marshall/General of the Army/General of the Airforce/Admiral of the Fleet** to non-commissioned officers such as **ordinary soldiers/airmen/seamen**.

4.8.6 Intelligence and security services

The information supplied in the table on intelligence and security services has been compiled from published sources and the intention with this annexure is to facilitate the study of politics, e.g.

(Former) Yugoslavia: (Voormalige) Joegoslawië	OZNA	Otsek Zaseita Naroda	Bureau of People's Protection	Buro vir Volksbeskerming	Intelligence and security service: Inligtings- en veiligheidsdiens
	SDB	Sluzba Drazavne Bezbednosti	State Security Service	Staatsveiligheidsdiens	Intelligence and security, successor to UDBA: Inligting en veiligheid, opvolger van UDBA
	UDBA	Uprava Drazavne Bedbeznosti	Administration of State Security	Administrasie van Staatsveiligheid	Intelligence and security, successor to OZNA and predecessor to SDB: Inligting en veiligheid, opvolger van OZNA en voorganger van SDB

4.8.7 Nicknames of political personalities, e.g.

In politics personalities who become well known often acquire nicknames, some of an endearing nature and some less flattering. Some of the better known nicknames are listed, e.g.

Clever Trevor *{satiric name for Trevor Manuel, minister of finance in Mandela and Mbeki cabinets, 1997–}*: **Clever Trevor** *{satiriese naam vir Trevor Manuel, minister van finansies in Mandela- en Mbeki-kabinette, 1997–}*

Groot Krokodil *{abusive name for former South African state president PW Botha}*: **Groot Krokodil** *{skeldnaam vir voormalige Suid-Afrikaanse staatspresident PW Botha}*

Henry the Kiss *{nickname for Henry Kissinger, former security adviser to President Nixon 1969–1973 and secretary of state of the USA 1973–1977}*: **Henry the Kiss** *{bynaam vir Henry Kissinger, voormalige veiligheidsraadgewer vir president Nixon 1969–1973 en minister van buitelandse sake van die VSA 1973–1977}*

Iron Lady *{name used for former British prime minister Margaret Thatcher}*: **Ystervrou** *{naam gebruik vir voormalige Britse eerste minister Margaret Thatcher}*

Kortbroek *{name used for Marthinus van Schalkwyk, leader of the New National Party; 1996–}*: **Kortbroek** *{naam gebruik vir Marthinus van Schalkwyk, leier van die Nuwe Nasionale Party; 1996–}*

Lang Hendrik *{name used for General Hendrik Johannes van den Bergh, a confidant of former South African prime minister John Vorster and head of BOSS}*: **Lang Hendrik** *{naam gebruik vir generaal Hendrik Johannes van den Bergh, 'n vertroueling van voormalige Suid-Afrikaanse eerste minister John Vorster en hoof van die BSV}*

Madiba *{traditional name of former South African president Nelson Mandela}*: **Madiba** *{tradisionele naam van voormalige Suid-Afrikaanse president Nelson Mandela}*

Monty *{name used for Bernard Law Montgomery, British field marshal}*: **Monty** *{naam gebruik vir Bernard Law Montgomery, Britse veldmaarskalk}*

Oubaas ⇨Slim Jannie

People's Princess *{pet name for Diana, Princess of Wales}*: **People's Princess** *{troetelnaam vir Diana, Prinses van Wallis}* **(Prinses van die Mense)**

Slim Jannie *{names used for Jan Christiaan Smuts, a prime minister of the former Union of South Africa}* **(Oubaas)**: **Slim Jannie** *{name gebruik vir Jan Christiaan Smuts, 'n eerste minister van die voormalige Unie van Suid-Afrika}* **(Oubaas)**

4.8.8 Rank structure of the SA Police Service

The rank structure of the South African Police (SAP) prior to 1994 was changed when the SAP was transformed into the South African Police Service (SAPS), and a new rank structure came into effect on 1 December 1995. On 1 April 2010 the SAPS reverted to the original rank structure. As a matter of interest the table annexed compares the original and 1995 SAPS rank structures with the SA Army rank structure.

4.8.9 Parliamentary Committees of South Africa

Both South African Houses of Parliament are divided into committees that play a vital role in the process of building democracy and public involvement. This annexure deals with the work of various committees.

4.9 References

A complete bibliography of literature consulted was kept but not incorporated

in the publications.

4.10 Target group

The published bilingual translating dictionaries are valuable assets to the bilingual terminology practice and subject specialists, language practitioners, academics and students are able to find solutions for terminology related problems in the fields covered by the contents of the dictionaries.

The target groups for the various dictionaries are *inter alia*:

— Members of Parliament.
— Translators in government service as well as the private sector.
— Teachers at academic institutions.
— Students whose mother tongue is not English.
— Journalists, researchers, speech writers, etc.
— Lexicographical and terminographical institutions.
— Language practitioners in Dutch and Flemish (cf. Le Clus 2011).

5. CEPTSA's bilingual and multilingual explanatory dictionaries

The first project in the execution of the second phase of this project is a bilingual bi-directional explanatory dictionary. The value of the bilingual translating dictionary would be enhanced by the bilingual explanatory dictionary and by the envisaged multilingual explanatory political sciences dictionary.

The CEPTSA committee is already in the process of compiling a complete multilingual explanatory version of the *Modern Political Dictionary*. Subject specialists, linguists and language practitioners are consulted when concepts are researched and defined.

CEPTSA is currently finalising the first 2500 core terms and definitions in English and Afrikaans and this explanatory dictionary will be published soon. In the mean time the committee will, with the collaboration of African language translators translate the terms and definitions.

It is clear that not all of the terms in the dictionary will have to be defined — adjectives derived from nouns and many combinations of terms already defined would for example not require further definition and in some cases the illustrations provided with the terms or the nature of the term would suffice. It was, as basis for planning, concluded that the Centre should initially look at defining some 5 000 terms. The final dictionary, however, would contain about 16 000 SL terms (cf. Le Clus 2011).

The initial definitions require considerable effort, since they have to be sufficiently clear for the layman to grasp the concept, comprehensive enough to cover the scholarly essentials and as concise as possible without prejudicing clarity and comprehension. Once a researcher's definition has been approved at a working session, the translation of the definition requires less time than its

initial crafting.

The bilingual explanatory English–Afrikaans political dictionary
— uses short scientific definitions to explain basic political and related facts, events and trends
— uses low register and simplified language in definitions
— enhances insight into complex South African and international developments

and moreover, wishes to promote the academic development of all South African languages (cf. Le Clus 2011).

Examples of entries in the bilingual explanatory dictionary:

baby boom. *A dramatic increase in the fertility rate, and thus the total number of births, of a country; usually with reference to the period 1946–1964 after World War II.*
geboortegolf. *'n Dramatiese toename in die fertiliteitskoers, en dus in die totale aantal geboortes, van 'n land; gewoonlik met verwysing na die tydperk 1946–1964 na die Tweede Wêreldoorlog.*

non-associational interest group. *An interest group distinguished by its intermittent articulation of interests and absence of a continuous organisational structure, eg kinship, class or economic factions, in a society that lobby their interests on an ad hoc basis.*
nie-assosiatiewe belangegroep. *'n Belangegroep gekenmerk deur die onderbroke artikulering van belange en die afwesigheid van 'n aaneenlopende organisatoriese struktuur, bv verwantskap-, klasse- of ekonomiese faksies in 'n samelewing wat hulle belange op 'n ad hoc-grondslag bevorder.*

CEPTSA treated homonyms as separate entries seeing that these are terms with the same spelling form and with the same pronunciation, but with different meanings. Homonyms do not get homonym numbers, they are merely entered in succession and dealt with as separate articles, e.g.:

judiciary n. *The branch of government that deals with the administration of justice, courts of law and the decisions of judges.*
regsprekende gesag. *Die owerheidsvertakking wat die administrasie van regspleging behartig, geregshowe en die uitsprake van regters hanteer.*

judiciary n. <the judiciary>. *The judges and magistrates of a state as a collective institution.*
regbank <die regbank>**.** *Die regters en landdroste van 'n staat as 'n kollektiewe instelling.*

king maker. *Someone who uses his influence and power to put a king or queen of his own choice on the throne.*
koningmaker. *Iemand wat sy invloed en mag gebruik om 'n koning of koningin*

van sy eie keuse op die troon te plaas.

king maker. *Someone who uses his influence and power to have his own choice of leader elected in a position of authority.*
koningsalwer. *Iemand wat sy invloed en mag benut om sy eie keuse van leier in 'n posisie van gesag verkies te kry.*

Polysemes on the other hand have two or more related meanings. These meanings apply in all the languages and are captured in the same article but numbered separately, e.g.

bywoner <from Afrikaans>**.** *1. Historically in South Africa a non-landowning white man on a farm, who in exchange for his services, was accorded privileges such as free lodging, use of land, or even a share in the crop. The man and his dependants were collectively referred to as bywoners. 2. When figuratively used a person who, because of limited resources, is dependent on the support of a patron to whom allegiance is owed.*
bywoner. *1. Histories in Suid-Afrika 'n wit man op 'n plaas, wat nie grond besit het nie en in ruil vir sy dienste sekere voordele soos gratis inwoning, grondgebruik en selfs deelsaai geniet het. Die man en sy afhanklikes is gesamentlik bywoners genoem. 2. Wanneer dit figuurlik gebruik word, 'n persoon wat vanweë beperkte hulpbronne afhanklik is van die steun van 'n begunstiger aan wie hy dan trou verskuldig is.*

faction fighting. *1. Rivalries and conflict among sections and communities within a larger formation; also fighting among various clans and families in a tribe. 2. In-fighting among divisions within a particular political party.*
faksiegevegte. *1. Wedywering en konflik tussen seksies en gemeenskappe binne 'n groter samestelling; ook bakleiery tussen verskeie sibbes en families van 'n stam. 2. Binnegevegte tussen afdelings binne 'n bepaalde politieke party.*

lekgotla <Sotho>**.** *1. Court where men sit; troop of warriors. 2. Originally a court or council where men meet; now often used for high-level consultative meetings in general.*
lekgotla <Sotho>**.** *1. Hof waar mans sit; troep krygers. 2. Oorspronklik 'n hof of raad waar mans ontmoet het; nou dikwels gebruik vir hoëvlak raadplegende vergaderings in die algemeen.*

Where parts of speech are indicated, the nouns are followed by verbs:

bug n. *A concealed electronic eavesdropping device commonly used by intelligence and security services to acquire information.*
oor. *'n Versteekte elektroniese afluisterapparaat wat algemeen deur inligtings- en veiligheidsdienste gebruik word om informasie te bekom.*

bug v. *To install an electronic eavesdropping device.*
ore installeer. *Om 'n elektroniese afluisterapparaat te installeer.*

Cross-referencing:

a) Abbreviations

The abbreviated form is supplied at its alphabetical place and it is cross-referenced to the full form where the relevant information is supplied, e.g.:

newly industrialised country (NIC, newly industrialised state). *A state that has experienced rapid economic growth through outward looking macro-economic policies, eg Brazil, Singapore, South Korea and Taiwan.*
nuut geïndustrialiseerde staat (NGS). *'n Staat wat vinnige ekonomiese groei ervaar deur uitwaartse makro-ekonomiese beleid te volg, bv Brasilië, Singapoer, Suid-Korea en Taiwan.*
NGO => nongovernmental organisation
NIC => newly industrialised country
nongovernmental organisation (NGO). *An institution that delivers a public service using private funding.*
nie-regeringsorganisasie (NRO). *'n Instelling wat 'n openbare diens lewer deur die gebruik van privaat befondsing.*

b) Synonyms

Synonyms appear in brackets next to the preferred term and are also entered at their alphabetical places. The dictionary article is not repeated at the synonym and the synonym refers to the preferred term, e.g.:

constituent state (federal territorial unit). *A constituent political entity of a federal state, variously known as state (USA), province (Canada), land (Germany) and canton (Switzerland).*
deelstaat (federale territoriale eenheid, konstituerende staat). *'n Samestellende politieke entiteit van 'n federale staat, met verskillende benamings soos staat (VSA), provinsie (Kanada), land (Duitsland) en kanton (Switserland).*
federal territorial unit => constituent state

c) Relevant information

The user is cross-referenced to related information, e.g.

frigate. <*see also* ***battleship, corvette, destroyer***> *A warship displacing between 1 000 and some 3 600 tonnes and mainly designed for antisubmarine or anti-air warfare.*
fregat. <*kyk ook* **slagskip, korvet, torpedodraer>** *'n Oorlogskip wat tussen 1000 en sowat 3 600 ton verplaas en hoofsaaklik vir duikbootjag- of lugafweeroorlogvoering ontwerp is.*

fundamentalism <*see also* **religious fundamentalism**>**.** *A deeply held conviction in the primary and most basic doctrine of a creed, often linked to an intense commitment and obsessive passion.*

fundamentalisme. *'n Diepgesetelde opvatting in die primêre en mees basiese leerstellings van 'n oortuigingsbelydenis, dikwels gekoppel aan 'n intense verwante toewyding en obsessiewe passie.*

CEPTSA published a draft list of 250 core English terms that were defined with the translation equivalents and definitions supplied in Afrikaans, Zulu and Northern Sotho, (cf. CEPTSA n.d.)

Examples of entries in the draft list containing four languages:

ballot n. (poll). *Any voting, whether secret or not, for a candidate or a motion.*
stemming. *Enige stemming vir 'n kandidaat of mosie, of dit geheim is al dan nie.*
balote (kgetho). *Kgetho efe goba efe; e ka ba ya sephiri goba ye e sego ya sephiri; ya nkgetheng goba ya tšhišinyo.*
ukuvota. *Ukuvota kwanoma yiluphi uhlobo, kungaba okuyimfihlo noma okungeyona okubhekiswe kulowo ongenele ukhetho noma kumbono othile.*

cabinet. *An executive committee tasked with the governmental functions of a state.*
kabinet. *'n Uitvoerende komitee wat met die regeringsfunksies van 'n staat belas is.*
kabinete. *Komitikhuduthamaga yeo e filwego mediro ya mmušo ya naga.*
ikhabinethi. *Isigungu esiphakeme esinikezwe igunya lokuphatha imisebenzi kahulumeni ezweni.*

majority rule. *The exercise of power according to the will of the majority.*
meerderheidsheerskappy. *Magsuitoefening volgens die wil van die meerderheid.*
ukubusa ngezwi leningi. *Ukusetshenziswa kwamandla ngokuya ngentando yeningi.*
pušo ya bontši. *Tiragatšo ya maatla go ya ka thato ya bontši.*

squatter. *A person that occupies land illegally and erects a shack or some other temporary shelter on it.*
plakker n. *'n Persoon wat grond onwettig beset en 'n hut of ander tydelike skuiling daarop oprig.*
moipei. *Motho yoo a dulago lefaseng le a dulago go lona e se ka molao gomme a aga mokhukhu goba mokutwana ofe goba ofe wa nakwana go lona.*
ohlala ngokungemthetho. *Umuntu ohlala endaweni ngokungemthetho abuye akhe umkhukhu noma indawo yokukhosela yesikhashana.*

CEPTSA managed to define and translate 1000 core SL terms and definitions into Afrikaans, Zulu, Xhosa, Northern Sotho and Tswana. Subject specialists and translators working in the field of political sciences will therefore soon be able to utilize subject related matter in six of the eleven official languages. This is CEPTSA's endeavour towards multilingualism (cf. paragraph 3).

Examples of entries in the glossary containing six languages:

abdication. *The renunciation of an office or responsibility, eg a throne.*
abdikasie. *Die afstanddoening van 'n amp of verantwoordelikheid, bv 'n troon.*

tholo ya modiro. *Tlogelo ya ofisi goba mošomo, goba ona maikarabelo, mohlala, sedulo sa bogoši goba yona terone.*
ukudela. *Ukushiya phansi isikhundla esithile noma umsebenzi, isib. ubukhosi.*
ukurhoxa. *Ukurhoxa e-ofisini okanye esikhundleni, umz. ukushiya itrone.*
tholo-tiro. *Go itatola tiro kgotsa go tlogela tiro ya ofisi le maikarabelo a yona, jaaka setilo sa bogosi/terone.*

abolition. *The act of doing away with a practice such as capital punishment or slavery.*
afskaffing. *Die handeling om 'n praktyk soos die doodstraf of slawerny tot niet te maak.*
phedišo. *Modiro wa go fediša mokgwa wa go swana le kotlo ya lehu goba bokgoba.*
ukuchithwa. *Isenzo sokuchitha isejwayezi esithile njengokujeziswa ngakwezo-mnotho noma ubugqila.*
ukubhangiswa. *Isenzo sokuphelisa inkqubo enjengeyesigwebo sentambo okanye ubukhoboka.*
phediso. *Kgato ya go fedisa/khutlisa tsamaiso/tiragalo jaaka katlholelo leso kgotsa bokgoba.*

brinkmanship. *In politics, the art of taking big risks, even to the verge of war, to test the resolve of the adversary in the hope that the adversary will back down.*
waagpolitiek. *In die politiek is dit die kuns om groot risiko's te waag, selfs tot die randjie van oorlog, ten einde die vasbeslotenheid van 'n teenstander te toets met die hoop dat die teenstander sal toegee.*
sepolotiki sa bophonkgedi. *Mo dipolotiking, bokgabo goba mokgwa wa go lekelela kudu, le ge e le mo lebakeng la go ka tsoga ga ntwa neng goba neng, go lekelela morero wa lenaba goba legogelathoko ka kholofelo ya gore legogelathoko le tlo boela morago.*
idelakufa. *Kwezepolitiki, lokhu kusho ubuciko bokuzifaka ezimweni ezinobungozi, okungaba ngisho nayimpi ukuze kuxazululeke umbango ngethemba lokuthi isitha kumbe imbangi izogoba uphondo.*
idelakufa. *Kwipolitiki, ubugcisa bokuthatha umngcipheko omkhulu, nokude kuye elwangcibeni lomlo, bokuvavanya isisombululo sobutshaba ngethemba lokuphelisa ubutshaba.*
bokwakwabidi ba polotiki. *Mo dipolotiking, go tewa botswerere ba bogatlhamela masisi, le fa eka nna ba go fetlha ntwa, ba go kgopakgopetsa mmaba ka tsholofelo ya gore o tla katakatela morago.*

log rolling <USA>**.** *The practice of collusion between politicians involving favours for mutual gain.*
wedersydse politieke bevoordeling <VSA>**.** *Die praktyk van samespanning tussen politici wat gunste tot onderlinge voordeel behels.*
kamogetšano ya thekgano <USA>**.** *Tiragalo ya bogwaragwara goba bomenemene gare ga boradipolotiki yeo e akaretšago thušo yeo go yona bohle ba tlo boelwago ka go lekana.*

ukusebenzisana kwezombusazwe <USA>**.** *Yisejwayezi sokusebenzisana phakathi kwabezombusazwe okufaka phakathi ukusizana ngenhloso yokuzuza okuthile nhlangothi zombili.*
ukusebenzisana kwezopolitiko <USA>**.** *Ukusebenzisana kwabezopolitiko bencedana injongo iyinzuzo macala omabini.*
go dikisana <USA>**.** *Mokgwa wa tirisanommogo gareng ga boradipolotiki e akaretsang go direlelana molemo gore botlhe ba tle ba ungwe molemo.*

majoritarian democracy. *Parliamentary practice in which the majority party takes full control of all executive powers and does not share power with coalition partners.*
meerderheidsdemokrasie. *Parlementêre praktyk waarin die meerderheidsparty volle beheer van alle uitvoerende bevoegdheid oorneem en nie mag met koalisievennote deel nie.*
temokrasi ya bontši. *Tshepedišo ya palamente ye ka yona lekoko la bontši le tšeago taolo ye e feletšego ya maatla-taolo, gomme le sa abelane maatla le badirišani ka lona ka tshwaraganelo le lona.*
intando yeningi yabaningi. *Inqubo yasephalamende lapho iningi linamandla okulawula ngokugcwele futhi lingabelani amandla nawozakwalo ababumbene nalo.*
idemokhrasi yesininzi. *Inkqubo yepalamente ekuthi ngayo iqela elisisininzi lilawule yonke into kwaye lingabelani nawo namaqela elisebenzisana nawo.*
demokerasi ya pusontsi. *Tsamaiso ya sepalamente e mo go yone lekoko la bontsi le tsayang taolo e e tletseng ya dithata tsotlhe tsa tsamaiso, mme le sa arolelane maatla le maphata a semphato.*

New Partnership for Africa's Development (NEPAD). *Vision and programme of action for the development of the African continent; initiated by SA President Mbeki.*
Nuwe Vennootskap vir Afrika se Ontwikkeling (NUVAO). *Visie en aksieprogram vir die ontwikkeling van die Afrikavasteland; deur SA president Mbeki van stapel gestuur.*
Tirišanommogo ye Mpsha ya Tšwelopele ya Afrika (NEPAD). *Ponelopelo le lenaneo la modiro la kgolo goba tšwelopele ya kontinente ya Afrika, ye e tlilego ka mopresidente wa Afrika-Borwa Mopresidente Thabo Mbeki.*
Inhlangano yobumbano ethuthukisa i-Afrika (NEPAD). *Umbono kanye nezinhlelo okuhloswe ngazo ukuthuthukisa izwekazi lase-Afrika; lo mbono waqalwa ngumongameli Mbeki waseNingizimu Afrika.*
Intsebenziswano eNtsha yoPhuhliso lweAfrika (NEPAD). *Umbono neenkqubo zamanyathelo ophuhliso lwelizwekazi iAfrika owasungulwa ngu-Mongameli Mbeki woMzantsi Afrika.*
Kgolagano e Ntšhwa ya Ntshetsopele ya Aferika (NEPAD). *Ponelopele le lenaane la dikgato tsa togamaano go kgontsha ntshetsopele ya kontinente ya Aferika; kgato e e akantswe ke Moporesidente Mbeki wa Aferika Borwa.*

zoning. *Directions setting out the purpose for which land may be used.*
sonering. *Voorskrifte wat die doel waarvoor grond gebruik mag word, uiteensit.*

kago ya magora. *Dikaelo tšeo di hlathollago maikemišetšo ao naga goba lefase le ka šomišwago ka gona.*
ukwehlukanisa izindawo. *Izinkomba ezibekwayo ngenhloso ekhomba ukuthi umhlaba ungasetshenziselwani.*
ukucandwa komhlaba. *Izikhokelo ezibonisa ukuba umhlaba mawusetyenziselwe ntoni na.*
kageletso. *Ditaelo tse di tlhalosang maikaelelo a gore lefatshe le ka dirisiwa.*

The committee is already working on a consolidated list containing 2500 core terms and definitions. Examples of terms and definitions in this list are:

accredited diplomatic representative. *A person whom the receiving state recognises as a representative of the sending state and is accorded diplomatic privileges and immunities but not necessarily full diplomatic status.*
geakkrediteerde diplomatieke verteenwoordiger. *'n Persoon wat deur die ontvangerstaat as 'n verteenwoordiger van die sendstaat erken word en aan wie diplomatieke voorregte en immuniteit, maar nie noodwendig volle diplomatieke status nie, toegeken word.*
moemedi wa merero ya ka ntle yo a dumeletšwego. *Motho yo e lego gore mmušo wo o amogelago o mo lemoga bjalo ka moemedi wa mmušo wo o mo romelago e bile o fiwa ditshwanelo tša boemedi le tšhireletšo eupša e sego maemo ao a tletšego a boemedi.*
umkhulumeli wezangaphandle osemthethweni. *Umuntu ongomkhulumeli wezwe lakhe abe ekwelinye izwe, anikezwe amagunya obunxusa nokuvikeleka kodwa abe engenasikhundla esiphelele kwezombusazwe.*

breach of the peace. *1. Belligerent action by one or more powers in contravention of behaviour required to maintain the nonviolent conduct of international relations. 2. The criminal offence of disrupting the public order through rowdy behaviour.*
vredesbreuk. *1. Strydlustige optrede deur een of meer moondhede in stryd met die gedrag wat vereis word om die niegewelddadige voer van internasionale betrekkinge te handhaaf. 2. Die kriminele oortreding om die openbare orde deur rumoerige gedrag te skend.*
tlolo ya khutšo. *1. Kgato ya ntwa ke mmušo goba e mentši ka go tlola maitshwaro ao a nyakegago go boloka boitshwaro bjo e sego bja ntwa bja dikamano tša boditšhabatšhaba. 2. Molato wa bosenyi wa go šitiša taolo ya setšhaba ka maitshwaro a go se laolege.*
ukwephula uxolo. *1. Isenzo sokuthanda ukulwa esenziwa umbuso owodwa noma ngaphezulu ukuphikisana nokuziphatha okumisiwe ukunqanda izimo zodlame zobudlelwane bomhlaba jikelele. 2. Ukwephulwa komthetho ngokuletha inxushunxushu emphakathini ngokuziphatha ngobudlova.*

no-go area. *An area that is controlled by criminals or insurgents and bars normal movement to the extent that the police or armed forces can only enter it by force.*
wegblygebied (moeilikheidsoekgebied). *'n Gebied wat deur misdadigers of insurgente beheer word en normale beweging in so 'n mate versper dat die polisie*

of gewapende magte dit slegs met geweld kan binnegaan.
lefelo leo le sa tsenwego. *Lefelo leo le laolwago ke disenyi goba marabele, gomme ba thibela mesepelo ya batho ya tlwaelo go fihla moo e lego gore maphodisa goba mašole ba ka e tsena fela ka go šomiša maatla.*
indawo ekungabhadwa kuyo. *Indawo elawulwa yizigelekeqe noma abashokobezi bavimbele iminyakazo ejwayelekile ngendlela yokuthi amaphoyisa noma umbutho ohlomile ungene kuyo ngokusebenzisa amandla.*
indawo ekungahanjwayo kuyo. *Indawo elawulwa zizaphula-mthetho okanye ngabarhubuluzi nto eyenza ukuba kungahanjwa kuyo ngendlela eqhelekileyo kangangokuba amapolisa okanye amajoni asuka angene ngetshova.*
lefelo ga-go-tsenwe. *Lefelo le le laolwang ke disenyi kgotsa batsoolodi mme le ganetsa tsamao e e tlwaelegileng mo e leng gore le ka tsenwa fela ke sepodisi le sesole ka kgapeletso.*

judicial review: *A re-examination by judges, eg of the proceedings of a lower court.*
regterlike hersiening: *'n Herondersoeking deur regters, bv van die verrigtinge van 'n laer hof.*
tshekatsheko-leswa ya tshepetšo ya molao: *Tebelelo-leswa ke baahlodi, mohlala; ga ditshepetšo tša kgorotsheko ya fase.*
ukubu tlhomo ya ditiro <go tlhola ditiro> **(yekezwa kokuphathelene nomthetho):** *Ukuhlolwa kwesibili okwenziwa ngamajaji, isib. Kwezinqubo zezinkantolo ezisezingeni eliphansi.*
ukuhlolwa ngokwasemthethweni: *Ukuhlolwa kwetyala lihlolwa ziijaji, umz.ityala ebeliviwe kwinkundla esezantsi.*
tshekatshekoseswa ya molao: *Tshekatsheko seswa ke baatlhodi, jaaka ya ditsamaiso tsa kgotla-tshekelo e kwa tlase.*
ukuhlolisisa ezobulungiswa: *Amandla wekhotho ukuhlolisisa imithetho, ihlelo elinikelwe mbusi nezenzo zempande ezinye zombuso, nokutjho bona azikalungi.*
tekolo ya melao: *Matla a boahlodi a ho sheba hape melao le diketso tsa makala a mang a mmuso, e be bo di phatlalatsa jwalo ka tse sa amohelehang.*
kubukisisiswa kwekuhlulelwa: *Emandla ekwahlulela ekubukisisweni kwemtsetfo, kunciphisa netento kwalamanye emagatja ahulumende, nekukhipha sincumo sekutsi kutsi akukaciniseki.*
nxiyanxiyo wa vuavanyisi: *Matimba ya huvo ya milawu ku xiyaxiya milawu, swileriso na swiendlo swa marhavi man'wana ya mfumo no hlambanya leswaku a ya ha ri na mpfumelelo wo tirha.*

The subject and language community are, however, in need of political and related terminology in the additional official languages and future publications by CEPTSA will address this need. An attempt towards supplying the terms and definitions in all the official languages was already made. Seeing that CEPTSA is not funded, and has to generate its own funding, the current members of CEPTSA are not remunerated for any work done. CEPTSA was thus far successful in obtaining translation assistance from the Language Unit, University of Johannesburg (Zulu, Northern Sotho and Tswana) and from the Lan-

guage Centre, University of Stellenbosch (Xhosa). The Centre will soon have to find amicable financial solutions in order to translate the 2500 core terms and definitions into the other official languages.

The main concern is to find subject specialists who are recognised experts in the fields of politics and with a sound linguistic background in the related target languages and terminologists and translators with a keen interest in the subject field and with a superb knowledge of the intricacies of the relevant target languages to assist with the compilation of the multilingual explanatory dictionaries.

Translating a *term* into ones own language by a mother tongue speaker when both the term and its elucidation are available generally requires less time than constructing a definition, therefore a slightly lower rate of pay would apply to the translation of terms. Translating a *definition* is somewhat more complicated, however, since it must not only be technically correct but must be rendered in the correct idiom of the target language, therefore a higher rate of pay would be applicable to the translation of definitions. In selecting a method of remuneration for the writing of definitions and translations into African languages the Centre had to choose between a system of payment for time worked at an approved hourly rate on the one hand or payment per product delivered on the other hand. After considering the methodology of the programme it was decided to remunerate researchers for product delivered and accepted, at a fixed amount per term and definition, taking into consideration the average time that could reasonably be expected to be required for creating the definition/translation in relation to a reasonable rate of pay for work of this nature.

The factors affecting the cost of research and editing are thus the number of terms to be defined and translated, the rate of remuneration for definitions and translations, and the number of languages involved plus the number of working sessions for evaluation and approval, the number of researchers involved in these sessions and the rate of remuneration for such working sessions. It should be noted that such a budget does not include salaries; the researchers are part time collaborators who are remunerated per approved product and whose reasonable subsistence and travel expenses are reimbursed.

The working sessions will evaluate and approve the initial definitions as well as the translations of terms and definitions received from the individual researchers. The cost of this is determined by the number of members of the Centre as well as additional researchers and translators involved plus the time required for each working session, the frequency of the sessions, the duration of the programme and the reasonable rate of remuneration for the participating members. Juta and Company (Pty) Ltd indicated its interest in publishing the various bi- and multilingual explanatory dictionaries and in assisting with the process and financing of translating terms and definitions into the official African languages.

The ultimate product will be a Modern Political Dictionary in English, Afrikaans and the official African languages, featuring:

— An explanatory, multilingual dictionary covering the fields of International Politics, Political Science, Public Administration (including municipal government and administration), Development Studies and Strategic Studies.
— 16 000 terms with source language terms and definitions (English) and translation equivalents and definitions in Afrikaans, Zulu, Xhosa, Ndebele, Swati, South Sotho, Tswana, Northern Sotho, Venda and Tsonga.

Although target users might still use English they will at least have the definitions and terms available in their first languages if the concepts are defined in SL and TLs and they will then be able to conceptualise, deduct and internalise the relevant information. There is no reason why the official languages could not develop into functional languages for higher learning or the working environment if LSP dictionaries are (made) available.

6. Conclusion

The terminographical results of the Centre for Political and Related Terminology in Southern African Languages (CEPTSA) receive attention. The bilingual publications are discussed and special emphasis is placed on aspects such as neologisms, problematic terms, orthography and transliterations. The borrowed terms and gains from the indigenous languages as well as from foreign languages are scrutinised. An overview is given of a variety of functional suffixes that were used productively in the *Political and Related Terms.* The latest dictionary contains quite a number of abbreviations for full forms. An indication is given of the type of abbreviations entered in the technical dictionary and examples are given of the kind of annexures added to the dictionary. The article discusses the envisaged explanatory dictionaries, i.e. a bilingual bi-directional English–Afrikaans explanatory dictionary and examples are given of term equivalents and definitions already available in some of the official African languages. CEPTSA's plan regarding work on a multilingual explanatory dictionary in all official South African languages is discussed. Translating definitions and coining terms in the various official languages requires subject knowledge and linguistic expertise. A collaborative working relationship between subject specialists, linguists and translators is needed to compile the multilingual explanatory political dictionary in order to provide standardised terminology in the political and related fields. The article ends with an overview on future plans.

7. Bibliography

Alberts, M. 1998. The Harmonisation of Nguni and Sotho: Terminological Perspectives. Prah, K.K. 1998. *Between Distinction and Extinction. The Harmonisation and Standardisation of African Languages*: 229-242. Johannesburg: Witwatersrand University Press.

Alberts, M. 2003. 'n Vertalende vakwoordeboek van politieke terme. *Lexikos* 13: 271-286.

Alberts, M. and N. Mollema. 2013. Developing Legal Terminology in African Languages as Aid to the Court Interpreter: A South African Perspective. *Lexikos* 23: 29-58.

Botha, S. 2011. *Report on the work by CEPTSA (Centre for Political and Related Terminology in Southern African Languages).* Unpublished document.

Centre for Political and Related Terminology in Southern African Languages (CEPTSA). n.d. *Multilingual Modern Political Dictionary: 250 Trial Definitions. Meertalige Nuwerwetse Politieke Woordeboek: 250 Proefdefinisies. Sesilimi ezahlukene samagama anamuhla apathelene nezepolitiki: Incasiselo yamagama angu-250 angumzamo. Pukuntšu ya dipolelo tše dintši ya sebjalebjale ye e amanago le mantšu a politiki: Ditlhalošo tše 250 tša mantšu tša boitekelo.* Johannesburg: Centre for Political and Related Terminology in Southern African Languages (CEPTSA), Auckland Park: Rand Afrikaans University.

Centre for Political and Related Terminology in Southern African Languages (CEPTSA). 2002. *Nuwerwetse Politieke Woordeboek/Modern Political Dictionary*. (Compiled by the Committee for Political and Related Terminology, CEPTSA.) Auckland Park: Rand Afrikaans University (RAUpers).

Centre for Political and Related Terminology in Southern African Languages (CEPTSA). Botha, S., P. le Clus and A. Venter (Eds.). 2011. *Modern Political Dictionary/Nuwerwetse Politieke Woordeboek*. CEPTSA. Auckland Park: University of Johannesburg.

Centre for Political and Related Terminology in Southern African Languages (CEPTSA). 2015. *Database of the Centre for Political and Related Terminology in Southern African Languages*. CEPTSA.

Department of Justice. 1996. *Constitution of the Republic of South Africa, 1996. (Act No. 108 of 1996).*

Department of National Education. 1989. *Staatkundige en verwante terminologie/Political and Related terminology*. Compiled by the National Terminology Services, Department of National Education in collaboration with the Committee for Political and Related Terminology. Pretoria: Government Printer.

Fourie, D.J. 1994. The Rejection of Terminology: Observations from the African Languages. *South African Journal for African Languages* 12(1): 11-15.

Government Gazette. 2012. Use of Official Languages Act 12 of 2012. *Government Gazette*, Vol. 568, No. 35742, 2 October 2012. Cape Town: The Presidency.

Government Gazette. 2013. SA Use of Official Languages Act — Proposed Regulations. *Government Gazette* 26, Notice 411 of 2013, No. 36392, April 2013: 56-67. Cape Town: National Parliament.

Le Clus, P. 2011. *Information Brochure. Centre for Political and Related Terminology in Southern African Languages (CEPTSA).* Auckland Park: Department of Politics, University of Johannesburg.

Sager, J.C. 1990. *A Practical Course in Terminology Processing.* Amsterdam/Philadelphia: John Benjamins.

http://lexikos.journals.ac.za

Semi-automatic Term Extraction for an isiZulu Linguistic Terms Dictionary*

Langa Khumalo, *Linguistics Program, School of Arts, University of KwaZulu-Natal, South Africa (khumalol@ukzn.ac.za)*

Abstract: The University of KwaZulu-Natal (UKZN) is compiling a series of Language for Special Purposes (LSP) dictionaries for various specialized subject domains in line with its language policy and plan. The focus in this paper is the term extraction for words in the linguistics subject domain. This paper advances the use of frequency analysis and the keyword analysis as strategies to extract terms for the compilation of the dictionary of isiZulu linguistic terms. The study uses the isiZulu National Corpus (INC) of about 1,2 million tokens as a reference corpus as well as an LSP corpus of about 100,000 tokens as a study corpus. The study is analyzed through the use of a software tool called WordSmith Tools (version 6). WordSmith Tools (hence forth WS Tools) is an integrated suite of three main programs, which include the WordList, Concord and Keywords, used in analysing words and word patterns in any given text. Using the WS Tools software a lot of qualitative and quantitative research can be done in the language. Central to this study is a computational determination of which words are typical of the linguistic domain in isiZulu and therefore stand out as preferred candidates for headword selection. Thus the study uses the corpus linguistics method as a basis for theoretical analysis. The advantage of such a theoretical approach is that a corpus is stored and queried by means of computer and computer software, which makes it easy to find, sort and count items, either as a basis for linguistic description or for addressing language-related issues and problems. Using the WS Tools software, the study shows that term extraction for the isiZulu dictionary of linguistic terms is done following reliable computational techniques in corpus lexicography.

Keywords: TERM EXTRACTION, LGP CORPUS, LSP CORPUS, WORDSMITH TOOLS, FREQUENCY, WORDLIST, CONCORD, KEYNESS, LEXICOGRAPHY, CORPUS LEXICOGRAPHY, HEADWORD SELECTION, LSP DICTIONARY

Opsomming: Semi-outomatiese term-onttrekking vir 'n isiZulu taalkundige termwoordeboek. Die Universiteit van KwaZulu-Natal (UKZN) is besig met die samestelling van 'n reeks Taal vir Spesiale Doeleindes (TSD)-woordeboeke vir verskeie gespesialiseerde vakgebiede wat strook met hul taalbeleid en -plan. Die fokus van hierdie artikel is die termontrekking vir woorde in die vakgebied taalkunde. Die gebruik van frekwensieanalise en sleutelwoordanalise as strategieë in die samestelling van die isiZulu taalkundige termwoordeboek word bevorder. Die studie gebruik die isiZulu National Corpus (INC) van ongeveer 1,2 miljoen items as 'n verwysingskorpus asook 'n TSD-korpus van ongeveer 100,000 items as 'n studiekorpus. Die studie is ontleed

* This article was presented as a paper at the Twentieth Annual International Conference of the African Association for Lexicography (AFRILEX), which was hosted by the University of KwaZulu-Natal, Durban, South Africa, 6–8 July 2015.

met behulp van 'n sagteware nutsprogam, WordSmith Tools (weergawe 6). WordSmith Tools (voortaan WS Tools) is 'n geïntegreerde programsuite bestaande uit drie hoofprogramme, wat WordList, Concord en Keywords insluit, en wat gebruik word in die analise van woorde en woordpatrone in enige gegewe teks. Met behulp van die WS Tools-sagteware kan baie kwalitatiewe en kwantitatiewe navorsing in die taal gedoen word. Sentraal in hierdie studie is 'n rekenaarmatige bepaling van watter woorde verteenwoordigend is van die isiZulu-taalkundige domein en daarom voorkeur geniet by trefwoordseleksie. Sodoende word die korpuslinguistiekmetode as basis vir teoretiese analise gebruik. Die voordeel verbonde aan so 'n teoretiese benadering is dat 'n korpus gestoor en geraadpleeg word deur middel van 'n rekenaar en rekenaarsagteware, wat dit maklik maak om items te vind, te sorteer en te tel, óf as basis vir taalkundige beskrywing óf om taalkundig verwante kwessies en probleme aan te spreek. Deur gebruik te maak van WS Tools-sagteware, toon die studie dat term-ontrekking vir die isiZulu taalkundige termwoordeboek gedoen word deur betroubare rekenaarmatige tegnieke in korpusleksikografie te volg.

Sleutelwoorde: TERM-ONTTREKKING, TAD-KORPUS, TSD-KORPUS, WORDSMITH TOOLS, FREKWENSIE, WOORDELYS, KONGRUENSIE, SLEUTELSTATUS, LEKSIKOGRAFIE, KORPUSLEKSIKOGRAFIE, TREFWOORDSELEKSIE, TSD-WOORDEBOEK

1. Introduction

The University of KwaZulu-Natal (UKZN) is compiling a series of Language for Special Purposes dictionaries for various specialized subject domains in line with its language policy and plan (Khumalo 2014: 1). The Language Policy and Plan of the University of KwaZulu-Natal (UKZN) is wholly informed by the country's widely acclaimed constitution, which enshrines multilingualism and provides that every official language must enjoy parity of esteem and must be treated equitably. In line with the provisions enshrined in the South African constitution section 6 (subsection 2 and 4), the Language in Education Policy of 1997, and consistent with the framework as set out in the Language Policy for Higher Education of 2002, and congruent with the Use of Official Languages Act of 2012, UKZN identifies with the goals of South Africa's multilingual language policy and seeks to be a key player in the successful implementation of this policy. Consequent to these statutory provisions UKZN has articulated this commitment through its Language Policy and Plan, which was first approved by Senate on the 2nd of August 2006. The Language Policy and Plan was recently revised and approved by Senate in November 2014.

UKZN has further taken a conscious and practical decision to develop isiZulu through its framework of functional bilingualism. Through this framework it recognizes English as the primary language of its academic program, and commits itself to the development and intellectualization of isiZulu to be a language of administration, teaching and learning, innovation and science. To this end, a detailed Language Plan monitored and evaluated by the University Language Board (ULB) is in place, and a practical Language Program has been set in motion by the University Language Planning and Development Office (ULPDO) in order to fully operationalize the University's Language Policy.

One of the major aims of the UKZN language policy is to achieve for isiZulu the institutional and academic status of English through providing facilities to enable the use of isiZulu as a language of learning, instruction, research and administration in the long term. As a result of these and other language policy objectives there has been a massive language development program, which is isiZulu corpus building and isiZulu terminology development, which are germane in the intellectualization of isiZulu. Work on the building of the isiZulu National Corpus (INC) started in the last quarter of 2014. The INC was piloted in November 2014 at 1, 1 million tokens and now stands at just under 2 million. Terminology development has taken place through arduous resource intensive statutory processes of consultation, verification, authentication and standardization. The terminology that has been standardized and approved by the isiZulu National Language Body include terminology for architecture, anatomy, computer science, corporate relations, environmental science, law, and nursing. A total of 1863 terms are now in the isiZulu Term Bank. The imperative to provide teaching and learning tool in the form of discipline specific dictionaries has thus been voiced. These will enhance cognitive capacity of both the staff and students in accessing otherwise complex scientific phenomenon, which hitherto have been contributing to the negative student performance. Specialized dictionaries are the ones that cover a relatively restricted set of phenomena. This type of dictionary covers the terminology of a particular subject field or discipline. It is also known as an LSP dictionary, which is short for Language for Special Purposes. In this paper we discuss term extraction for an isiZulu linguistic terms dictionary using a corpus linguistics method.

2. Corpus linguistics method

The study uses the corpus linguistics method as a basis for theoretical analysis. According to Sinclair (2005) a corpus is "a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research." The advantage of such a theoretical approach is that "[…] a corpus [is] stored in a computer, it is easy to find, sort and count items, either as a basis for linguistic description or for addressing language-related issues and problems" (Kennedy 1998: 11). A corpus is thus a collection of naturally occurring texts derived from real life language use in either written or spoken form, which is then processed, stored and accessed by means of computers. Such a corpus is then useful as a basis for investigating language use and for developing dictionaries, spell checkers and other human language technologies (HLTs).

The approach we espouse in this study is a corpus linguistic one. We use a language for general purposes corpus (aka LGP) as a *reference corpus* (RC) and a language for special purposes (aka LSP) as an *analysis corpus* (AC). The RC is a non-technical corpus while the AC is a domain-specific, technical corpus. The

LSP corpus used in this study comprises of the two main isiZulu grammar textbooks *Uhlelo lwesiZulu*, and *Izikhali zabaqeqeshi nabafundi*, a collection of isiZulu grammar lecture notes from academics in the School of Arts and the School of Education at UKZN, and online linguistic documents in isiZulu. Using these two corpora that are quite different in terms of content, we compare the behavior of lexical units and identify lexical units that are specific to the AC.

In order to explicate the LSP corpus further, Lynne Bowker (2002: 45) states that the LSP corpus is one that "focuses on a particular aspect of a language. It could be restricted to the LSP of a particular subject field, to a specific text type, to a particular language variety or to the language used by members of a certain demographic group (e.g. teenagers). Because of its specialized nature, such a corpus cannot be used to make observations about language in general. However, general reference corpora and special purpose corpora can be used in a comparative fashion to identify those features of a specialized language that differ from general language …"

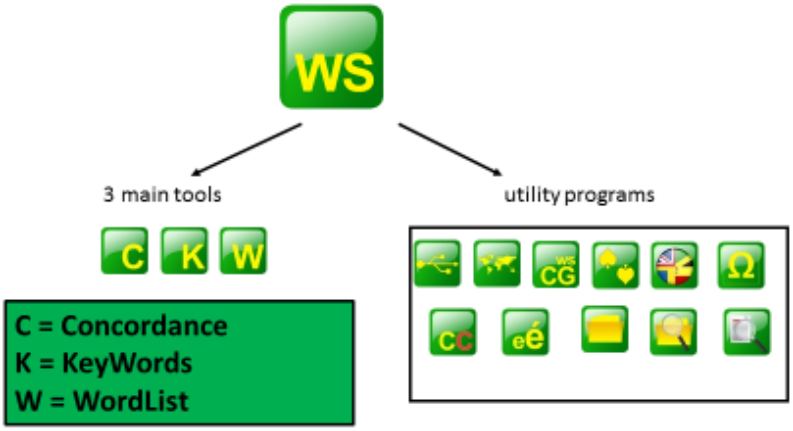
The advantages of LSP corpora are that they contain a wealth of authentic usage information. Since LSP corpora comprise of texts that have been written by subject field experts, the researchers have before them a body of evidence pertaining to the function and usage of words and expressions in the LSP of the field. With the help of corpus analysis tools, it becomes possible to sort these contexts so that meaningful patterns are revealed. An LSP corpus basically contains thousands of words that have been written by subject field experts and, as such, it can be seen to represent distilled expert knowledge.

The RC used in this study is an LGP corpus with 1 186 675 running words. The size of the RC, although still modest, can guarantee that the articles cover a wide range of subjects and that their content is heterogeneous. In contrast the AC is discipline-specific with an exclusive focus on linguistics. The AC is an LSP corpus with 111 922 running words, which comprises two isiZulu basic grammar textbooks and a collection of notes on the teaching of isiZulu grammatical structure.

Our study is analyzed through the use of a software tool called WordSmith Tools (version 6). WordSmith Tools (hence forth WS Tools) is an integrated suite of three main programs, which include the WordList, Concord and Keywords, used in analysing words and word patterns in any given text. WS Tools was developed by Mike Scott, who had earlier worked with Tim Jones to develop MicroConcord. WS Tools was first released in 1996 and the current version (version 6.0) was released in 2011. The Wordlist tool can be used to produce wordlists or word-cluster lists from a text and render the results alphabetically or by frequency order. It can also calculate word spread across a variety of texts. The Concord tool can give any word or phrase in context – so that one can study its co-text, i.e. to see what other words occur in its vicinity. The Keywords tool calculates words which are key in a text, i.e., used much more frequently or much less frequently in a given corpus (e.g. the LSP corpus) than expected in terms of a general corpus of the language (e.g. the INC). Using

the WS Tools software a lot of qualitative and quantitative research can be done in the language. Table 1 below shows the main features of the WS Tools as described above.

Table 1: Wordsmith Tools (version 6)



Central to this study is thus a computational determination of which words are typical of the linguistic domain in isiZulu and therefore stand out as preferred candidates for headword selection. Using the WS Tools software, the study will perform the following. The author will run a frequency list to determine the most frequent words in the LSP corpus. A frequency list provides an array of different types of words, tokens, or forms which make up a corpus. These can be listed from the most frequent token to *hapax legomena* (i.e. those forms that occur only once in a given corpus) or vice versa. Frequency lists are a powerful tool in corpus lexicography. They guide lexicographers on which words to include in a dictionary. Frequency lists also provide developers of second language teaching material with the most relevant words, phrases, and expressions to teach. In this study a frequency list sheds more light on the most common words in isiZulu linguistic domain. These words may be the ones which characteristically typify the domain. According to Kilgarriff (1997: 135) "The more common it is, the more important it is to know it."

3. Term extraction

The focus in this study is the term extraction for words in the linguistics subject domain. Term extraction means the automatic mining or retrieval of relevant terms from a given corpus. Term extraction remains a challenge to anyone interested in domain-specific information retrieval (Jacquemin 2001); (Bouri-

gault et al. 2001); (Drouin n.d.). The goal in this study is to extract words that are typical for the isiZulu linguistic domain. We use the keyword tool in WS version 6 to extract linguistic terms. The main goal is to reduce (not eliminate) the amount of noise in the list of candidate terms.

4. Frequency analysis

It is crucial to affirm the observation by Summers (1996: 261) that "all aspects of lexicography are influenced by frequency." This is particularly crucial in selecting word candidates for inclusion in a dictionary. Headword selection becomes informed by the frequency through a statistical analysis. We premise our analysis on the most frequent 100 words on the assumption that this would be the most typically used words. The word list flows from the most frequent word to the least frequent in a descending order. The most frequent words in the AC are given in Table 2. ***N*** stands for the number the word occupies in the list of words in the word list, and ***Freq.*** is the number of times a word occurs in the corpus.

Table 2: Most frequent 100 tokens

N	Word	Freq.	N	Word	Freq.
1	ukuthi	861	51	bona	67
2	noma	812	52	emva	67
3	bese	512	53	mina	66
4	kodwa	481	54	kubo	64
5	lapho	421	55	ziye	63
6	futhi	419	56	indawo	62
7	ngoba	409	57	kule	62
8	nje	353	58	kwezinye	62
9	ke	342	59	nayo	62
10	ukuba	296	60	kusho	59
11	lokhu	279	61	ngenhla	59
12	khona	262	62	nokuthi	59
13	phela	255	63	yini	59
14	naye	236	64	ala	58
15	ngo	236	65	izakhi	58
16	kanti	231	66	nazo	58
17	kanye	213	67	wena	57
18	ngaye	190	68	leli	56
19	lapha	189	69	isimo	55
20	kahle	187	70	lesi	54
21	no	178	71	laba	53
22	zonke	157	72	zona	53
23	njengoba	152	73	ngazo	52

24	ake	148
25	sithi	148
26	kuye	147
27	na	138
28	ukuze	137
29	lezi	132
30	kanje	131
31	ngokuthi	130
32	lusizo	121
33	usuke	117
34	ngayo	116
35	kube	115
36	kuthi	110
37	ngabe	89
38	lo	87
39	ngu	87
40	manje	85
41	uye	82
42	ba	80
43	kanjani	80
44	lokho	76
45	yakhe	75
46	yonke	73
47	njalo	72
48	lowo	71
49	bonke	70
50	baye	67

74	uhlelo	52
75	wonke	52
76	enye	51
77	lezo	51
78	zakhe	51
79	lolu	50
80	nga	50
81	thina	50
82	yona	49
83	nazi	48
84	ngaso	48
85	ngakho	46
86	yena	45
87	kuze	44
88	kude	43
89	kulo	43
90	kuwo	43
91	nabo	43
92	aba	42
93	kepha	41
94	uzobe	41
95	konke	40
96	siye	40
97	kuzo	38
98	labo	38
99	sakhe	38
100	sika	38

Table 2 shows that the ten most frequent words in the AC are *ukuthi*, *noma*, *bese*, *kodwa*, *lapho*, *futhi*, *ngoba*, *nje*, *ke*, and *ukuba*. All these words are function or grammatical words, which belong to a closed word class. The closed word classes include concords, pronouns, numerals, connectives etc. This top ten word list is not unique as function words commonly dominate all frequency lists. It is therefore the case that functional words are normally removed from the word list in order to retain content words. Table 3 below shows the list of the most frequent 100 tokens after excluding the function words.

Table 3: Most frequent 100 tokens excluding function words

N	Word	Freq.
1	u	829
2	e	550
3	lapho	421

N	Word	Freq.
51	lusizo	121
52	usuke	117
53	ngayo	116

4	ngoba	409
5	isibonelo	387
6	nje	353
7	ke	342
8	ukuba	296
9	ulimi	290
10	lokhu	279
11	khona	262
12	amagama	260
13	o	257
14	phela	255
15	naye	236
16	kanye	213
17	indlela	204
18	umuntu	201
19	kukhona	196
20	ubunye	191
21	ngaye	190
22	njll	190
23	isigaba	189
24	lapha	189
25	kahle	187
26	unkamisa	180
27	kakhulu	173
28	abantu	163
29	zonke	157
30	ubuningi	154
31	njengoba	152
32	ake	148
33	sithi	148
34	kuye	147
35	isenzo	143
36	amabizo	142
37	kusuke	142
38	phakathi	139
39	na	138
40	ibhola	137
41	igama	137
42	ukuze	137
43	lezi	132
44	kanje	131
45	ibizo	130
46	ngokuthi	130

54	kube	115
55	la	114
56	le	111
57	onkamisa	111
58	kuthi	110
59	isakhi	104
60	ndlela	101
61	umntwana	101
62	izibonelo	100
63	kolimi	100
64	leyo	100
65	abanye	99
66	isuke	99
67	kuphela	99
68	yolimi	98
69	izenzo	96
70	izib	96
71	ezinye	95
72	isabizwana	95
73	ngaphandle	95
74	into	94
75	iziqu	94
76	umakoti	94
77	zisuke	90
78	ngabe	89
79	abe	88
80	umusho	88
81	lo	87
82	ngu	87
83	imisindo	86
84	izintombi	86
85	ana	85
86	manje	85
87	ongwaqa	85
88	ubaba	84
89	umoya	84
90	kuba	83
91	kufanele	83
92	uye	82
93	ekhaya	81
94	eqondisayo	81
95	ongenazwi	81
96	ba	80

47	umfana	129	97	kanjani	80
48	ingane	127	98	ukusetshenziswa	80
49	emshweni	126	99	izivumelwano	79
50	inkathi	122	100	isib	77

Table 3 shows the same data as Table 2 with the exclusion of function words. The removal of function words reveals content words that could define the genre. The list of content words reveals clearly the genre of linguistics. For example *u, e, o;* (vowels); *isibonelo* (example); *ulimi* (language), *amabizo* (nouns); *indlela* (mood), *ubunye* (singular) etc. are typical linguistic words. The frequency list has somewhat helped to isolate words that are typical. Other words on the top 100 wordlist are not particular to the discipline. Such words include *ngoba*, *umuntu*, *ngaye* and others. This is not unusual since the top 100 words are not isolated on any measure that isolates words that are typical to a text. In order to achieve this we use the keyword analysis.

5. Keyword analysis

We use the keyword analysis in order to identify words particular to the isi-Zulu linguistics domain. This is done through the calculation of keyness, which isolates words which are key to the AC. According to Mike Scott (2006: 92) keyness is "calculated by comparing the frequency of each word in the word list of the text under investigation with the frequency of the same word in the reference word list." Calculations are done using the Keyword tool of WS Tools. The output is a list of keywords, or words whose frequencies are higher in the AC than in the RC. Table 4 below shows the top 100 words most typical in the linguistic domain extracted through the Keyness tool.

Table 4: Top 100 linguistic tokens

N	Keyword	English gloss	Freq.	Keyness
1	isibonelo	example	387	1515,82
2	i	vowel *i*	1002	1424,26
3	a	vowel *a*	1005	1172,94
4	bese	and	512	875,18
5	ulimi	language	290	773,57
6	uma	if	1179	659,00
7	—	—	—	—
8	—	—	—	—
9	unkamisa	vowel	180	557,61
10	phela	finish	255	510,56
11	e	vowel *e*	550	488,01
12	njll	etc.	190	485,03
13	u	vowel *u*	829	473,92
14	ubunye	singular	191	465,09

15	emshweni	in sentence	126	423,19
16	isigaba	noun class	189	413,95
17	kusuke	from	142	400,56
18	ongenazwi	voiceless	81	392,36
19	ibizo	noun	130	374,68
20	amabizo	nouns	142	368,78
21	amagama	words	260	365,93
22	yolimi	linguistic	98	364,73
23	ubuningi	plural	154	361,18
24	onkamisa	vowels	111	357,17
25	izibonelo	examples	100	356,86
26	kolimi	linguistic	100	356,86
27	isakhi	morpheme	104	351,84
28	zisuke	from	90	350,20
29	isuke	from	99	349,82
30	umusho	sentence	88	341,03
31	usuke	from	117	329,89
32	inkathi	tense	122	324,55
33	isenzo	verb	143	322,38
34	noma	or	812	313,96
35	umakoti	bride	94	309,01
36	onezwi	voiced	63	303,29
37	zenkulumo	of speech	73	299,87
38	o	vowel *o*	257	295,88
39	ongwaqa	consonants	85	293,59
40	iziqu	stem	94	290,38
41	usizo	help	121	281,57
42	konkamisa	on vowels	74	280,32
43	isabizwana	substantive	95	279,64
44	imisindo	sounds	86	273,14
45	umkhongi	negotiator	54	268,64
46	intombi	girl	52	258,69
47	isib	e.g.	77	256,67
48	umfana	boy	129	246,03
49	ngaye	through him	190	239,20
50	abantu	people	48	238,79
51	iqhikiza	full-grown girl	53	235,40
52	izib.	e.gs	96	230,55
53	eqondisayo	inductive mood	81	225,76
54	ukusetshenziswa	used	80	223,45
55	izakhi	morphemes	58	223,20
56	basuke	left	76	222,65
57	izib	e.gs	93	220,98
58	inkomo	cows	70	220,66
59	izivumelwano	agreements	79	219,54
60	unsinini	alveolar	46	219,34
61	sokukhomba	demonstrative	69	218,52
62	yenkulumo	of speech	68	218,48
63	isibanjalo	copulative	68	212,35
64	ana	reciprocal suffix	85	212,06
65	izintombi	girls	86	211,83

66	ziye	gone	63	201,67
67	ingane	child	127	201,46
68	ungwaqabathwa	click sounds	42	199,62
69	zamabizo	nominal	57	196,35
70	isandiso	locative	64	195,85
71	imisho	sentences	63	195,60
72	sithi	we say	148	190,00
73	qaphela	note	65	189,02
74	isiqalo	prefix	63	188,05
75	zesenzo	of verbs	48	187,20
76	isiqu	stem	66	184,66
77	indlela	mood	204	179,69
78	onguputshu	plosive	36	179,09
79	ngonkamisa	are vowels	62	178,77
80	umgudu	cavity	54	176,52
81	ukwakhiwa	morphology	61	171,50
82	ukulandula	negation	58	171,44
83	izenzo	verbs	96	170,55
84	izilimi	languages	71	165,12
85	umkhwenyana	bridegroom	42	162,97
86	udwendwe	que	34	154,04
87	iphimbo	tone	56	153,57
88	sesenzo	verbal	48	153,35
89	izibanjalo	copulatives	47	151,25
90	zabomdabu	of tradition	33	144,03
91	baye	gone	67	142,51
92	ibhola	ball	137	141,52
93	emabizweni	in nouns	44	140,74
94	izingcezu	morphemes	44	140,74
95	sebizo	nominal	45	138,87
96	senhloko	subjectival	49	135,74
97	zezenzo	verbal	48	135,02
98	ndlela	mood	101	134,62
99	intombazane	girl	27	134,32
100	esuke	from	39	132,81

6. Discussion

The 100 keywords in Table 4 are a more typical reflection of the linguistics discipline when juxtaposed with those in Table 3. The keyness tool has successfully extracted terms which are key to the domain of linguistics from the corpus. The list includes the vowels *a, e, i, o, u,* **(3, 11, 2, 38, 13)**; language *ulimi* **(5)**; vowel *unkamisa* **(9)**; singular *ubunye* **(14)**, in a sentence *emshweni* **(15)**; noun class *isigaba* **(16)**, voiceless *ongenazwi* **(18)**; noun *ibizo* **(19)** nouns *amabizo* **(20)**; consonants *ongwaqa* **(39)**; indicative mood *eqondisayo* **(53)**; agreements *izivumelwano* **(59)**; copulative *isibanjalo* **(63)** click sound *ungwaqabathwa* **(68)**; cavity *umgudu* **(80)**; tone *iphimbo* **(87)**; subjectival *senhloko* **(96)**; **etc.**

The top 100 wordlist suggests that the keyness analysis is crucial in iso-

lating data that is domain specific. The results of these experiments are useful as potential candidates for headword selection are highlighted. The study has shown that term extraction for the isiZulu dictionary of linguistic terms is done following reliable computational techniques in corpus lexicography.

7. Conclusion

We explored frequency and keyword analysis in generating domain specific candidates for headword selection. Using such statistical approach is faster, reliable and free from human error or bias. It is clear from the study that corpora are useful in enhancing the dictionary microstructure and the keyness list will form the basis for headword selection for the isiZulu linguistics terms dictionary. Term extraction thus reduces the amount of noise in the list of candidate terms. Native speaker intuition is used to compliment this vital computational resource.

References

Bourigault, D. et al. 2001. *Recent Advances in Computational Terminology*. Amsterdam/Philadelphia: John Benjamins.

Jacquemin, C. 2001. Spotting and Discovering Terms through Natural Language Processing. Cambridge, MA: MIT Press.

Kennedy, G.D. 1998. *An Introduction to Corpus Linguistics*. London/New York: Longman.

Khumalo, L. 2014. *Developing an isiZulu Dictionary of Linguistics Terms: Challenges and Prospects.* Unpublished paper presented at the Nineteenth Annual International Conference of the African Association for Lexicography (AFRILEX), which was hosted by the Research Unit for Language and Literature in the SA Context, North-West University, Potchefstroom Campus, Potchefstroom, South Africa, 1–3 July 2014.

Kilgarriff, A. 1997. Putting Frequencies in the Dictionary. *International Journal of Lexicography* 10(2): 135-155.

Scott, M. 2004–2006. *Oxford WordSmith Tools Version 4.* Oxford: Oxford University Press.

Sinclair, J. 2005. Corpus and Text: Basic Principles. Wynne, M. (Ed.). *Developing Linguistic Corpora: A Guide to Good Practice*: 1-16. Oxford: Oxbow Books. Available online from http://ahds.ac.uk/linguistic-corpora/ [Accessed 20 October 2005].

Summers, D. 1996. Computer Lexicography: The Importance of Representativeness in Relation to Frequency. Thomas, J. and M. Short (Eds.). 1996. *Using Corpora for Language Research: Studies in Honour of Geoffrey Leech*: 260-266. London/New York: Longman.

http://lexikos.journals.ac.za

Planning and Macrostructural Elements for a Multilingual Culinary Dictionary of Gabonese Languages*

Virginie Ompoussa, *Département des Sciences du Langage, Université Omar Bongo, Libreville, Gabon (vompoussa@yahoo.com)*
and
Hugues Steve Ndinga-Koumba-Binza, *Centre for Text Technology, North-West University, Potchefstroom, Republic of South Africa and Département des Sciences du Langage, Université Omar Bongo, Libreville, Gabon (22602569@nwu.ac.za)*

Abstract: This paper is an account of an ongoing dictionary project. It aims to outline the main features of the planned dictionary. The project is a first of its kind in Gabonese lexicography. The current paper gives a description of the data collection and sources as well as the macrostructural issues of the forthcoming dictionary. The paper starts with a brief outline of the project background, needs and interests.

Keywords: CULINARY DICTIONARY, MULTILINGUAL DICTIONARY, DATA COLLECTION, MACROSTRUCTURAL ELEMENTS

Résumé: Eléments de planification et macrostructuraux pour un dictionnaire multilingue de cuisine des langues gabonaises. La présente communication est un rapport sur un projet de dictionnaire en cours. Elle vise à décrire les principales caractéristiques du dictionnaire projeté. Ce projet de dictionnaire est une première dans la lexicographie gabonaise. Dans cet article une description est faite de la collecte des données et des sources du dictionnaire ainsi que de ses caractéristiques macrostructurales relatives du dictionnaire en projet. L'article commence par un bref aperçu du background, des besoins et de l'intérêt du futur dictionnaire.

Mots-clés: DICTIONNAIRE DE CUISINE, DICTIONNAIRE MULTILINGUE, COLLECTE DES DONNEES, ELEMENTS MACROSTRUCTURAUX

* An earlier version of this article was presented as a paper at the Nineteenth Annual International Conference of the African Association for Lexicography (AFRILEX), which was hosted by the Research Unit for Language and Literature in the SA Context, North-West University, Potchefstroom Campus, Potchefstroom, South Africa, 1–3 July 2014.

1. Introduction

This article is an account of an on-going research project. The project is concerned with the compilation of a multilingual culinary dictionary for Gabonese languages. This paper primarily seeks to present the conceptualization of the intended dictionary. This should consequently attract attention and allow for the gathering of input from experts.

African cuisine is known as one of the exotic culinary pieces in the world. This explains why African restaurants are becoming more and more common in most of the world's biggest cities. In some of these restaurants, menus, dishes and recipes are often named in their original African languages. Nevertheless, a culinary dictionary in any African language hardly exists. The current project is intended to fill the gap with specific attention on Gabonese cuisine and languages.

In addition, one of the current trends in Gabonese lexicography for the past few years has been the compilation and publication of thematic dictionaries. This is an outcome of a fast-growing move to onomasiological studies in Gabonese linguistics (Raponda-Walker 2002, Malekou 2007, Kwenzi Mikala 2008, Mickala Manfoumbi 2011 and 2012). Although the forthcoming dictionary can be labeled as a thematic dictionary, it will however differ from most thematic dictionaries on various metalexicographical aspects such as a clearly-indicated typology, an identification of target users and a formulation of a genuine purpose.

The present paper contains three main sections, i.e. the data collection and sources, and the macrostructural features of the planned dictionary. These are preceded by a brief outline of the dictionary project background, needs and interests.

2. Project Background, Needs and Interests

The current planned dictionary project stems from an ethnolinguistic study that focused on the inventory, the denomination and the typology of culinary items in Gabonese languages. The ethnolinguistic project had made a distinction between names for comestibles, names for special dishes and names for kitchen tools for making those dishes. Cookware and kitchen tools can look alike in different cultures, but cooking methods and food preservation procedures vary from one region to another or from one culture to another. In fact, the main aim was to identify, inventory and classify the various traditional cookware and kitchen tools, cooking methods and food preservation techniques that exist within multicultural Gabon.

The ethnolinguistic project revealed the needs for such a dictionary in the context of multilingual Gabon. In fact, Gabon is known to be a language mosaic with French as the predominant and sole official language (cf. Ndinga-Koumba-Binza 2007). However, the population of Gabon is firmly of the Bantu

culture (cf. Kwenzi Mikala 1988 and 1989) as it is shown through their culinary tradition, habitual dressing codes, music, costumes and beliefs, etc. A multilingual dictionary will firstly reveal not only the development of the natural common Bantu lexicon (which in fact plays a role in the mutual intelligibility between a number of these languages), but also the development of African language vocabulary through linguistic adoptives, i.e. loanwords (cf. Raponda-Walker 1955). This dictionary project will contribute to proving or invalidating the hypothesis that argues that the Gabonese ethnolinguistic groups largely share a strong cultural heritage. Equally, the project will inevitably contribute to the preservation of the multi-ethnic Gabonese culture.

It should be noted that the target users of the planned dictionary are the public at large, starting at the population of Gabon, which is known to have a literacy rate of near 95% (cf. Nzang-Bie 2001).

3. Data Collection and Sources

This section is concerned with the data gathered for the planned dictionary. The first three sub-sections are concerned with the sources where data were gathered from. Data sources for a dictionary are known as the dictionary basis, which comprises the primary sources, the secondary sources and the tertiary sources (cf. Svensén 2009: 39, Ndinga-Koumba-Binza and Saphou-Bivigat 2012: 237). The different sources and their specificities are duly presented in this section. The last sub-section presents the type of data collected, which includes linguistic data and cultural data. The difference between the two types of data is also explained.

3.1 Fieldwork and primary sources

As mentioned in the previous section, this dictionary project stems from an ethnolinguistic research project. Thus, the primary data source of the planned dictionary is indeed the database resulting from research fieldwork for the ethnolinguistic project. Fieldwork and open-ended oral interviews were conducted in various villages for speech data and tools identification. There were no specific questionnaires for the interviews.

The main reason for choosing to conduct fieldwork in villages, and not in urban areas, is the fact that villages are still the custodians of culture as pointed out in most studies on Gabonese villages (cf. Lisimba 1997, Mayer 2002, Raponda-Walker 1963, Balandier and Pauvert 1952). Villagers still speak their respective native languages commonly on a daily basis (cf. Mabika Mbokou 2012, Ndinga-Koumba-Binza 2011 and 2005b). Indigenous knowledge is also still very well preserved in villages (cf. Mearns 2006 and 2007, Mearns, Du Toit and Mukuka 2006, Mwaura 2008, Lwoga, Ngulube and Stilwell 2010). Meanwhile, culinary culture is much influenced in cities where the populations are

experiencing many imported eating and food habits (e.g. fast food shops, Asian restaurants, European cuisine, etc.).

The methodology used for data collection was adopted and adapted from Medjo Mvé (2011) and from Kwenzi Mikala and Bigoundou (1999), which consists of collecting only lexical items from native speakers in the village by fieldworkers who are native speakers themselves of the language in use in a specific area. The specificity of the methodology in the current project is twofold. First, the main informants were all female. This is due to the fact that cooking is notably a women's activity in Gabonese culture. Thus, the hypothesis on the fieldwork was that women would know much more about cookware and cuisine items than men. However, a few men were also consulted as occasional informants, especially when it came to translating the lexical items into French.

The second aspect of the specific methodology is the fact that fieldwork was not conducted on the basis of a questionnaire, but solely on discussions in kitchens just before and/or after lunch and/or supper time. The discussions were mainly focused on getting names of tools used, of meals made and of recipes. The main reason for this choice was the difficulty to make a questionnaire *ex nihilo*. In fact, the existing linguistic and anthropological questionnaires appear too impressionistic for domain-specific research such as cuisine. This reason adds to another difficulty, i.e. the quasi impossibility of finding a study of the same kind in African languages.

Subsequently, various intensive fieldwork sessions were conducted by different small groups (of two to three researchers) in a number of villages of the nine provinces of Gabon. The purpose of the fieldwork was to collect both visual and spoken data. Data obtained from the fieldwork and comparative classification appear adequate to gather the necessary material for a multilingual dictionary.

The general corpus of the lexical items constitutes the dictionary basis as the primary source of the planned multilingual dictionary. This corpus was built from recorded speech data. It is in fact known that the nature of primary sources can depend on whether the language a dictionary is being made for has enough written resources. Otherwise, the primary sources will be oral sources (cf. Granberry 1993, Nzang-Bie 2002, Soami 2002 and 2010a, Ndinga-Koumba-Binza and Saphou-Bivigat 2012). This is the case of the current dictionary project whose primary source is the corpus built from speech data collected during fieldwork in Gabon where most native languages are still being reduced to writing (cf. Ndinga-Koumba-Binza and Roux 2009, Hubert and Mavoungou 2010, Ndinga-Koumba-Binza 2010a).

3.2 Secondary sources

According to Wiegand and Kučera (1981: 100), secondary sources can often be written materials which are in principle all the dictionaries consulted during

the compilation phase of the dictionary. It is known that the emerging Gabonese lexicography is fast-growing, experiencing the production of many dictionaries in the last fifteen years (cf. Ndinga-Koumba-Binza 2005a and 2010b, Ekwa Ebanega and Tomba Moussavou 2006 and 2008).

The current existing Gabonese dictionaries consist of two types. The first is made of bilingual and multilingual dictionaries which comprise dictionaries and references works made by religious missionaries and colonial administrators of the early era of dictionary production in Gabon (cf. Ndinga-Koumba-Binza 2005a: 138; Ekwa Ebanega and Tomba Moussavou 2008: 350). Among other these dictionaries: Biton (1907), Galley (1964), ILALOK (2008), Lemoine (1994), Marichelle (1902), Mickala Manfoumbi (2004), Raponda-Walker (1995), Van der Veen and Bodinga-bwa-Bodinga (2002), Mavoungou and Plumel (2010), Mintsa (2013a) and Idiata and de Nadillac (2012) to name a few.

The second type is made of monolingual dictionaries of Gabonese French (cf. Mavoungou 2013). These are Boucher and Lafage (2000), Ditougou (2009), Dodo-Bounguendza (2008, 2010 and 2013), Mavoungou *et al.* (2014), and Moussounda Ibouanga (2011).

All dictionaries consulted were the sources for not only written data, but also and mainly for a number of equivalences for lexical items in French. Other sources such monographs and studies in the fields of linguistics and anthropology were also from times to times consulted whenever necessary. A few of these studies include Mintsa (2013b), Mba Abessole (2006), Idiata (2006) Akomo-Zoghe (2010), Loembe (2005), Mavoungou and Ndinga-Koumba-Binza (2010), Mabik-ma-Kombil (2001), Etsio (1999) and Rekanga (2014). The entire series of *Rapidolangue*, a method for learning Gabonese native languages developed Raponda-Walker Foundation (Hubert 2010, 1997 and 1995) was also consulted whenever necessary.

3.3 Linguistic data and cultural data

The current study notes a difference between linguistic data and cultural data. Linguistic data refers to speech data recorded from specific language speakers during the fieldwork. This is comprehended in the traditionally limited conception of linguistic data in the field of linguistics, i.e. "the observable patterns of speech or writing, especially when recorded and gathered together in a corpus" (Crystal 2008: 128). Linguistic data recorded in Gabonese native languages for the projected dictionary were processed in lemmatization procedures to form the lemmata of the dictionary.

Table 1 below displays a sample of linguistic data recorded. The dictionary project being a multilingual project, only five Gabonese native languages are considered in the table sample below. Thus, column 1 presents linguistic terms in Fang, column 2 contains words in Lumbu (also known as Yilumbu), and column 3 displays lexical items in Nzebi (also called Yinzebi). The two last columns contain equivalences of the linguistic data in Obamba (also called

Lembaama) and Vili (also known as Civili) respectively.

Fang	Yilumbu	Yinzebi	Obamba	Civili
Akok	Ibenga	Ishighu	Kri	Cibenga
Sas	Tsafu	Tsegha	Ntshighi	Nsafu
Ngon	Teeri	Ndzaka	Ndjeghe	Ntêta
Edong	Yufu	Mbungu	Odu	Cufu

Table 1: Sample of linguistic data in different languages (see Addendum for images of these items)

A cuisine is of the most intrinsic features of a culture (cf. Albala 2011). This explains why a cuisine is always associated to a culture or a region. It is in fact often ethnically, nationally or continentally labeled. At ethnic level, we have examples such as Sotho cuisine, Nguni cuisine and Afrikaans cuisine. At national level, examples can be Gabonese cuisine, Mozambican cuisine and South African cuisine. At continental level, we have African cuisine, Asian cuisine and European cuisine as examples. At national and at continental levels, the label mainly refers to respectively the similarities between various ethnic groups within the same nation and between various nations within the same continent. For instance, although it originated from Japan, sushi has become known as an Asian cultural meal for being most often served in any Asian restaurant in the western world.

The main reason of the association to a culture or a region is the fact that a cuisine is in a pure sense a cooking style involving distinctive ingredients, techniques and dishes. And it is generally influenced by both the availability of ingredients locally and the religious or traditional laws. The fact that a cuisine is an inherent item of a culture (implying identity the same way a language does) denotes that collecting linguistic data related to it is a form of cultural data collection.

The cultural data collection consisted of four procedures. The first procedure was to see the physical item. The second procedure was the acquisition of the linguistic data of the item. The third procedure was to take a digital photograph of the item. The fourth procedure was the acquisition of knowledgeable information about each item. In some circumstances, the informants also willingly gave the items to the fieldworkers as tokens of appreciation for the visit and interest on their culture.

The speech data were also digitalized in the process of building the corpo-

ra in Microsoft Excel sheets. The speech data are included on the corpus in the form of orthographic and phonetic transcriptions.

4. Macrostructural Issues and Decisions

Like most dictionaries of Gabonese native languages, the planned dictionary has to encounter a number of macrostructural challenges. The current project is concerned with the following issues, i.e. the orthography of Gabonese languages, the lemmatization, and ordering as well as terminology equivalences in French.

4.1 Alphabet and orthography

As it has been mentioned earlier, native Gabonese languages are unwritten languages to date. Nevertheless, various proposals have been made to put Gabonese languages in writing (cf. Raponda-Walker 1932, Hombert 1990, Mayer 1990, Idiata 2002, Maloughou Mangama 2009, Ndinga-Koumba-Binza and Roux 2009 and Ndinga-Koumba-Binza 2010a). It is unfortunate that "none of the proposals have yet received government accreditation to be taught and used in schools" (Ndinga-Koumba-Binza and Saphou-Bivigat 2012: 240). In addition, there also exist proposals specific to a particular language such Civili (Ndinga-Koumba-Binza 2010c), Fang (Afane Otsaga 2010), Yilumbu (Mavoungou 2010a) and Yipunu (Soami 2010b). It even appears awkward that related languages such as Civili, Yilumbu and Yipunu have three different orthography proposals.

Following Ndinga-Koumba-Binza and Roux (2009), contributors in Hubert and Mavoungou (2010) reflected on a unified, harmonized and standardized orthographic system for all Gabonese languages. Their proposal is a revised version of the proposal adopted at the national experts' workshop in April 1999 (cf. Idiata 2002). The proposal in Hubert and Mavoungou (2010) takes into account a number of issues raised in previous proposals (especially those contained in Raponda-Walker 1932, Mavoungou 2010a, Afane Otsaga 2010, Soami 2010b and Ndinga-Koumba-Binza 2010c) and by Ndinga-Koumba-Binza and Roux (2009).

The current dictionary project makes use of the latter proposal, which is unifying and harmonizing as suggested in Hubert and Mavoungou (2010). It is a proposal that bears a maximum of community-acceptable criteria mentioned by Ndinga-Koumba-Binza and Roux (2009) for a user-friendly orthographic system.

Ndinga-Koumba-Binza and Roux (2009: 101) suggest that a writing system for the minor languages such as those of Gabon should be based on:

(i) phonological characteristics,

(ii) an alphabet for orthographic purposes, including writing, reading, punctuation and capital letters rules,

(iii) user-friendliness in learning and writing,

(iv) minimum of problems in readability,

(v) machine printability, and

(vi) uniformity for all Gabonese languages

4.2 Word tradition lemmatization

As far as the structure of the lemma is concerned, the projected dictionary follows the tradition adopted in most dictionary projects in Gabon, i.e. word tradition lemmatization (cf. Ndinga-Koumba-Binza and Saphou-Bivigat 2012: 242). This means that lexical items of the lemmata enter in their complete structural form according the word division and are lemmatized with their noun prefixes. It is seen in Mavoungou (2010b) that the custom in Gabonese lexicography with regard to word division is to write conjunctively in all lexical instances (the word is prefix + stem) and disjunctively in morphosyntactic and verbal declinations (predicates, copulas, verbs, etc. are separated from each other in a phrase or sentence).

It is our view that both word tradition and the word division adopted by this dictionary contribute more to the dictionary user-friendliness than the stem tradition and any other word division. In fact, the general public who the dictionary is targeted to is not only mostly acquainted with the French word division, but also may not understand the segmentation of the prefix from the stem.

4.3 Terminology equivalence issue

Terminology equivalence has been an issue in both bilingual lexicography and translation studies (cf. Adamska-Sałaciak 2010, Wiegand 2002 and Koller 1995). In fact, very often lexicographers are faced with the "*absence of lexical equivalence*" (Hohulin 1986: 43) or with no "*perfect interlingual correspondence*" (Adamska-Sałaciak 2011: 2) from one language to another. Adamska-Sałaciak (2011: 2) indicates that the "*reasons for this can be located on three interrelated planes: language structure itself, the extralinguistic world, and the human mind* ".

The situation appears to be the same in the context of Gabonese native languages to French and vice versa when it comes to cooking and gastronomic items. In this case, the extralinguistic world, i.e. differences on culinary habits which are founded on cultural differences, is the main reason for terminology equivalence issues of the planned dictionary.

Nevertheless, as the terminology equivalence issue has been variously dealt

with in the lexicographic literature (cf. Zgusta 1987, Gouws 1996 and 2002, Wiegand 2002, and Adamska-Sałaciak 2010 and 2011), a number of strategies exist for dealing with lexicographic equivalence cases. For the present study, all linguistic data were recorded in the mother-tongue of the informants with translation in French.

The interviewees provided the translation in French whenever possible, and the fieldworkers had to require the help of other informants when an interviewee could not provide a French equivalence of any lexical item. A third alternative for the determination of lexical equivalence was to check in existing dictionaries of Gabonese languages (including Gabonese French dictionaries such as Dodo-Bounguendza 2008, 2010 and 2013; Ditougou 2009 and Mavoungou *et al.* 2014). This alternative was also for the verification of translations obtained from interviewees.

Loanword adoption is found among other strategies for translation as part of solving the terminology equivalence issue. In the direction from French to Gabonese native languages, the equivalence issue is often solved with the fact the latter record a great number of loanwords adopted from the former. The examples below in **(1)** show a few of adoptives from French (Fr.) into Civili (Civ.), Yilumbu (Yil.), Yinzebi (Yin.) and Pove (Pov.).

(1) a. zoni (Civ.) <Fr. oignon "*onion*"
b. sukila (Yil.) <Fr. sucre "*sugar*"
c. ingata (Yin.) <Fr. gateau "*cake*"
d. nyonde (Pov.) <Fr. oignon "*onion*"

There also exist a great deal of loanwords adopted from other European languages such as English (Eng.) and Portuguese (Port.). A few examples for the word "*rice*" in Civili (Civ.), Pove (Pov.), Yipunu (Yip.) and Myene (Mye.) are contained in **(2)** below.

(2) a. loosu (Civ.) <Port. aroz
b. oresi (Mye.) <Eng. rice
c. uresi (Yip.) <Eng. rice
d. mulesu (Pov.) <Port. aroz

On the other hand, in the direction from Gabonese native languages to French, the issue is more complicated with the total absence of lexical equivalence of Gabonese culinary and food realities in the French language. In fact, when the encounter between Western civilization and African cultures took place, the latter were mainly in a receiving position and thus had less influence on the former. This explains why there are very few African loanwords in European languages compared to European loanwords in African languages. Medjo Mvé (2007) believes that the French language is in the process of evolving as it is more and more including in its lexicon African loanwords such as Gabonese native language words. This is particularly true of Gabonese French. The

examples below in **(3)** show some of the rare African loanwords in Gabonese French related to culinary lexicon. The source languages of these loanwords are Myene (Mye.), Fang, Yipunu (Yip.) and Teke.

(3) a. odika <Mye. "fruit of *Irivingia Gabonensis*"
b. folong <Fang. "a type of vegetable"
c. musungu <Yip. sugarcane wine[1]
d. safou <Civ. "fruit of *Burseraceae Dacryodes Edulis*"[2]
d. nkumu <Teke. "a type of vegetable" (*Gnetum Africanum*)
e. nyembwe <Mye. "palm nut sauce"

The second strategy for translation of items from Gabonese native languages into French was the use of local designations. This is mainly based on Gabonisms, i.e. particular Gabonese lexical items and expressions found in French the way it is spoken in Gabon (cf. Mavoungou 2013, Dodo-Bounguendza 2008, Ompoussa 2008 and 2011, Bagouendi-Bagère Bonnot 2007). Most of these lexical items and collocations are formed through calque and semantic shifts when equivalences do not exist in French. The following examples in **(4)** are noted in French from languages such as Myene (Mye.), Pove (Pov.), Civili (Civ.) and Yinzebi (Yin.).

(4) a. odika (Mye.): chocolat indigène (fruit of *Irivingia gabonensis*)[3]
b. ngoya (Pov.): sanglier (*potamochoerus suinae*)[4]
c. cilemba (Civ.): feuilles de manioc (leaves of *manihot esculenta*)[5]
d. bukulu (Yin.): oseille à petites feuilles (*Hibiscus Sabdariffa*)[6]

4.4 Thematic and alphabetic ordering

One of the key components in the nomenclature of the planned dictionary is its thematic and alphabetic ordering. This determines the type of the dictionary in planning. In fact, although it is planned to be a multilingual translation dictionary, the planned dictionary cannot avoid containing a number of encyclopedic features such as the ordering, the full definition of extralinguistic and cultural items as well as the pictorial illustrations.

Thematic and alphabetic ordering is common practice in encyclopedic dictionary compilation. According to Stark (2011: 101), the "advantages of alphabetical order are its familiarity, fixed nature, and objectivity... The benefits of alphabetical order are so enduring that some dictionaries with thematic content have opted to arrange their themes alphabetically". The planned dictionary is indeed a thematic dictionary since it is intended that the macrostructural nomenclature is to be organized in themes and the themes will be arranged alphabetically. Entries will also be presented alphabetically within each theme.

Thus, the ordering organization will presumably look as follows below in **(5)** where uppercase letters A, B and C denote the themes, the lowercase letters

a, b, c, etc. indicate the sub-theme and the Arabic numbers characterize the entries in a sub-theme.

(5)

A. Cookware
B. Dishes
C. Food and ingredients
a. Dairy
b. Drinks
c. Fish
d. Fruits
    1. apricot
    2. banana
    3. cherry
    4. lemon
    5. mango
    6. orange
e. Meat
f. Poultry
g. Sea food
h. Species
i. Vegetables
    1. broccoli
    2. cabbage
    3. carrot
    4. lettuce
    5. spinach
j. Kitchen
k. Recipes

5. Concluding Remarks

Studies on Gabonese cultures and languages have paid little attention to culinary habits and gastronomic potentialities of Gabon. This article has reported on an ongoing dictionary compilation project, which is the lexicographic component of a general ethnolinguistic project on the culinary traditions of multilingual and multicultural Gabon. This paper has focused mainly on the dictionary basis (data and sources) and on various macrostructural issues of the planned dictionary.

This project contributes to the strategic planning of Gabonese lexicography as it opens with a first of its kind perspective within the field in Gabon. A further contribution of the project is that it should reveal the development of the natural Bantu common lexicon as it tests the hypothesis on the shared cultural heritage of Gabonese ethnolinguistic groups.

Finally, the compilation and production of such a dictionary will also give sense to the perception by Emejulu (2000: 53) to "identifying possible problem areas and providing a working platform for rational and profitable/profit-making lexicographic development in a multi-cultural and multilingual set up". It is herein believed that Gabonese lexicography would succeed in such a perspective in producing dictionaries for cultural domains like the one planned in the present article.

Notes

1. **Musungu** (originally from Yipunu) and **malamba** (originally from Fang) are the lexical items for sugarcane wine in Gabonese French.
2. This fruit of *Burseraceae dacryodes edulis* is also known as **atanga** in Gabonese French.
3. In Gabonese French **odika** (which is originally a Myene word) is an alternative for the Gabonism ***chocolat indigène*** (which literally means "indigenous chocolate", but has nothing to do with normal chocolate except for the similar color).
4. In Standard French, the common name for *potamochoerus suinae* (which only exists in Africa and Madagascar) is **potamochère**. It is however called **sanglier** in Gabonese French due to its resemblance with *sus scrofa*, commonly known as **sanglier** in European French and **boar** in English. *Sus scrofa* does not live in Africa.
5. *Manihot esculenta* is known as **cassava** or **Brazilian arrowroot** in English, **manioc** in French and **tapioca** in other languages such as Spanish and Portuguese. In Gabonese cuisine, the root (known as **tubercule** in Gabonese French) is eaten as potato or as flour (a traditional bread commonly known as **baton de manioc** in Gabonese French is also made from it) and the leaves as vegetables.
6. *Hibiscus Sabdariffa* is commonly known as **roselle** in English and is eaten in Gabonese cuisine as a vegetable often mixed with sardine or any other fish. A beverage commonly known in Gabonese French as **bissap** is also made from it. The beverage is sweet when mixed with sugar or honey.

Acknowledgements

The authors which to thank the groups of the students in the Language Sciences Department at Omar Bongo University who labored as fieldworkers of the research contained in this article. The second author, holder of a Research Career Advancement (RCA) Fellowship, is also grateful to the National Research Foundation (NRF) in South Africa for this fellowship. The authors alone are responsible for any particular view or shortcoming this article may contain.

References

Dictionaries and Encyclopedias

Albala, K. (Ed.). 2011. *Food Cultures of the World Encyclopedia*. Santa Barbara, CA: Greenwood.

Biton, A. 1907. *Dictionnaire français–ndumu et ndumu–français, précédé d'éléments de grammaire.* Nantes: Imprimerie Bourgeois.

Boucher, K. and S. Lafage. 2000. *Le lexique français du Gabon (entre tradition et modernité). Le Français en Afrique* 14. Special Issue. Nice: Institut de Linguistique Française.

Ditougou, L. 2009. *On est ensemble: 852 mots pour comprendre le français du Gabon.* Libreville: Editions Raponda-Walker.

Dodo-Bounguendza, E. 2008. *Dictionnaire des gabonismes*. Paris: L'Harmattan.

Dodo-Bounguendza, E. 2010. *Diagnostic du français du Gabon. Guide pratique destiné aux journalistes, politiques, administratifs et universitaires.* Libreville: Les Editions Ntsame.

Dodo-Bounguendza, E. 2013. *Dictionnaire du parler toli–bangando. Argot des jeunes gabonais*. Libreville: Les Editions Ntsame.

Crystal, D. 2008. *A Dictionary of Linguistics and Phonetics*. Sixth edition. Malden, MA: Blackwell Publishing.

Galley, S. 1964. *Dictionnaire fang–français/français–fang*. Neuchatel: Henri Messeiller.

Idiata, D.F. and L. de Nadillac. 2012. *Dictionnaire isangu–français*. Libreville: Editions du CENAREST.

ILALOK (Institut des Langues Locales au Kouilou). 2008. *Dictionnaire vili–français/Mpisukulu bi kum' bi tshi vili ku tshi mputu*. Paris: L'Harmattan.

Lemoine, G.M. 1994. *Petit dictionnaire français–gisira–ngosi (Mandji–Ndolu) et éléments des verbes.* Libreville: Multipress.

Marichelle, C. 1902. *Dictionnaire français–vili*. Loango: Imprimerie de la Mission.

Mavoungou, P.A. and B. Plumel. 2010. *Dictionnaire yilumbu–français*. Libreville: Editions Raponda-Walker.

Mavoungou, P.A., F. Moussounda Ibouanga and J.A. Pambou. 2014. *Le dico des makaya et des mamadou. Contribution à l'étude du français du Gabon*. Libreville: Editions Odette Maganga.

Mickala Manfoumbi, R. 2004. *Lexique pove–français/français–pove*. Libreville: Editions Raponda-Walker.

Mickala Manfoumbi, R. 2011. *Dictionnaire des noms personnels punu*. Paris: Editions Alpha-Omega.

Mickala Manfoumbi, R. 2012. *Dictionnaire des noms personnels pove*. Paris: Editions Alpha-Omega.

Mintsa, J. 2013a. *Dictionnaire français–fang/fang–français*. Morrisville, NC: Lulu Press.

Moussounda Ibouanga, F. 2011. *Français du Gabon: approches sociolinguistiques et lexicographiques (le toli bangando).* Paris: Editions Universitaires Européennes.

Raponda-Walker, A. 1995. *Dictionnaire français–mpongwé*. Libreville: Editions Raponda-Walker.

Van der Veen, L. and S. Bodinga-bwa-Bodinga. 2002. *Gedandedi sa geviya/Dictionnaire geviya français*. Louvain/Paris/Sterling: Peeters Publishers.

Other Literature

Adamska-Sałaciak, A. 2010. Examining Equivalence. *International Journal of Lexicography* 23(4): 387-409.

Adamska-Sałaciak, A. 2011. Between Designer Drugs and Afterburners: A Lexicographic-semantic Study of Equivalence. *Lexikos* 21: 1-22.

Afane Otsaga, T. 2010. Propositions pour l'orthographe du fang. Hubert, J. and P.A. Mavoungou (Eds.). 2011. *Ecriture et standardisation des langues gabonaises*: 187-190. Stellenbosch: SUN Press.

Akomo-Zoghe, C.S.P. 2010. *Parlons fang. Culture et langue des fang du Gabon et d'ailleurs.* Paris: L'Harmattan.

Bagouendi-Bagère Bonnot, D. 2007. *Le français au Gabon: représentations et usages*. Unpublished Ph.D. thesis. Aix-en-Provence: Université de Provence.

Balandier, G. and J.C. Pauvert. 1952. *Les villages gabonais: aspects démographiques, économiques, sociologiques, projets de modernisation*. Mémoires de l'Institut d'Etudes Centrafricaines N^{0}5. Montpellier: Imprimerie Charité.

Ekwa Ebanega, G.M. and F. Tomba Moussavou. 2006. Des considérations sur la lexicographie et la linguistique au Gabon. *Lexikos* 16: 239-250.

Ekwa Ebanega, G.M. and F. Tomba Moussavou. 2008. A Survey of the Dictionary Use of Gabonese Students at Two South African Universities. *Lexikos* 18: 349-365.

Emejulu, J.D. 2000. Lexicography, an Economic Asset in Multilingual Gabon. *Revue Gabonaise des Sciences du Langage* 1: 51-69.

Etsio, E. 1999. *Parlons téké. Langue et culture du Congo et Gabon.* Paris: L'Harmattan.

Gouws, R.H. 1996. Bilingual Dictionaries and Communicative Equivalence for a Multilingual Society. *Lexikos* 6: 14-31.

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

Granberry, J. 1993. *A Grammar and Dictionary of the Timucua Language*. Tuscaloosa: The University of Alabama Press.

Hohulin, E.L. 1986. The Absence of Lexical Equivalence and Cases of its Assymmetry. *Lexicographica* 2: 43-52.

Hombert, J.M. 1990. Présentation de l'alphabet des langues du Gabon. *Revue Gabonaise des Sciences de l'Homme* 2: 97-103.

Hubert, J. 1995. *Rapidolangue. Méthode d'apprentissage des langues nationales. Volume 1 (fang, inzebi, lembaama, omyene, yipunu).* Libreville: Editions Raponda-Walker.

Hubert, J. 1997. *Rapidolangue. Méthode d'apprentissage des langues nationales. Volume 2 (ikota, ghetsogho).* Libreville: Editions Raponda-Walker.

Hubert, J. 2010. Etat des lieux sur l'enseignement des langues gabonaises. Hubert, J. and P.A. Mavoungou (Eds.). 2010. *Ecriture et standardisation des langues gabonaises*: 5-28. Stellenbosch: SUN Press.

Hubert, J. and P.A. Mavoungou (Eds.). 2010. *Ecriture et standardisation des langues gabonaises*. Stellenbosch: SUN Press.

Idiata, D.F. 2002. *Il était une fois les langues gabonaises*. Libreville: Editions Raponda-Walker.

Idiata, D.F. 2006. *Parlons isangu. Culture et langue des bantu-masangu du Gabon.* Paris: L'Harmattan.

Koller, W. 1995. The Concept of Equivalence and the Object of Translation Studies. *Target* 2: 191-222.

Kwenzi Mikala, J.T. 1988. L'identification des unités-langues gabonaises et leur classification interne. *Muntu* 8: 54-64.

Kwenzi Mikala, J.T. 1989. Contribution à l'inventaire des parlers bantu du Gabon. *Pholia. Revue du Laboratoire de Phonétique et Linguistique Africaine* 2: 103-110.

Kwenzi Mikala, J.T. 2008. *Les noms des personnes chez les Bantu du Gabon*. Paris: L'Harmattan.

Kwenzi Mikala, J.T. and J. Bigoundou. 1999. La pêche: une technique d'acquisition des Bapunu du Sud-Gabon. *iBoogha* 3: 133-148.

Lisimba, M. 1997. *Les noms de villages dans la tradition gabonaise.* Saint-Maur: Editions Sépia.

Loembe, G. 2005. *Parlons vili. Langue et culture de Loango.* Paris: L'Harmattan.

Lwoga, E.T., P. Ngulube and C. Stilwell. 2010. Understanding Indigenous Knowledge: Bridging the Knowledge Gap through a Knowledge Creation Model for Agricultural Development.

South African Journal of Information Management 12(1), Art. #436, 8 pages. DOI: 10.4102/sajim.v12i1.436.

Mabik-ma-Kombil. 2001. *Parlons yipunu. Langue et culture des punu du Gabon-Congo.* Paris: L'Harmattan.

Malekou, P.E. 2007. *Les anthroponymes et les toponymes gisir: proposition d'un modèle de dictionnaire.* Unpublished M.A. thesis. Libreville: Université Omar Bongo.

Maloughou Mangama, M.E. 2009. Un système alphabétique pour les langues gabonaises? Ondo Mebiame, P. (Ed.). 2007. *Quel avenir pour les langues et cultures du Gabon?*: 163-176. Libreville: Editions CUI-Gabon.

Mavoungou, P.A. 2010a. Propositions pour l'orthographe du yilumbu. Hubert, J. and P.A. Mavoungou (Eds.). 2010. *Ecriture et standardisation des langues gabonaises*: 181-186. Stellenbosch: SUN Press.

Mavoungou, P.A. 2010b. *Lexicographie et confection des dictionnaires au Gabon.* Stellenbosch: SUN Press.

Mavoungou, P.A. 2013. Gabonese French Dictionaries: Survey and perspectives. *Lexikos* 23: 255-272.

Mavoungou, P.A. and H.S. Ndinga-Koumba-Binza. 2010. *Civili, langue des baloango: esquisse historique et linguistique*. Munich: Lincom Europa.

Mayer, R. 1990. Histoire de l'écriture des langues du Gabon. *Revue Gabonaise des Sciences de l'Homme* 2: 65-91.

Mayer, R. 2002. *Histoire de la famille gabonaise*. Libreville: Editions du LUTO.

Mba Abessole, P. 2006. *Aux sources de la culture fang*. Paris: L'Harmattan.

Mabika Mbokou, L. 2012. A Survey Bilingualism in Multilingual Gabon. Ndinga-Koumba-Binza, H.S. and S.E. Bosch (Eds.). 2012. *Language Science and Language Technology in Africa: Festschrift for Justus C. Roux*: 163-175. Stellenbosch: SUN Press.

Mearns, M.A. 2006. *Conservation of Indigenous Knowledge*. Ph.D. thesis. Johannesburg: University of Johannesburg.

Mearns, M.A. 2007. The Basotho Cultural Village: Cultural Tourism Enterprise or Custodians of Indigenous Systems? *Indilinga: African Journal of Indigenous Systems* 6(1): 27-50.

Mearns, M.A., A.S.A. du Toit and G. Mukuka. 2006. Conservation of Indigenous Knowledge at Cultural Villages. *Mousaion* 24(1): 23-50.

Medjo Mvé, P. 2007. Langues gabonaises et évolution du français: aspects sociolinguistiques. *Annales de l'Université Omar Bongo* 13: 285-298.

Medjo Mvé, P. 2011. *Introduction à la langue et la culture des chasseurs-cueilleurs Bakoya (Région de Mékambo, Gabon): avec un petit dictionnaire.* Cologne: Rüdiger Köppe Verlag.

Mintsa, J. 2013b. *Le complet de la langue fang*. Morrisville, NC: Lulu Press.

Mwaura, P. (Ed.). 2008. *Indigenous Knowledge in Disaster Management*. Nairobi: United Nations Environment Programme.

Ndinga-Koumba-Binza, H.S. 2005a. Considering a Lexicographic Plan for Gabon within the Gabonese Language Landscape. *Lexikos* 15: 132-150.

Ndinga-Koumba-Binza, H.S. 2005b. Politique linguistique et éducation au Gabon: un état de lieux. *Journal of Education* 4(1): 97-78. Réduit: Mauritius Institute of Education.

Ndinga-Koumba-Binza, H.S. 2007. Gabonese Language Landscape: Survey and Perspectives. *South African Journal of African Languages* 27(3): 97-116.

Ndinga-Koumba-Binza, H.S. 2010a. Alphabet et écriture: approche historique et cas des langues gabonaises. Hubert, J. and P.A. Mavoungou (Eds.). 2010. *Ecriture et standardisation des langues gabonaises*: 29-67. Stellenbosch: SUN Press.

Ndinga-Koumba-Binza, H.S. 2010b. Trends in Gabonese Modern Lexicography. Paper presented at the 15th International Conference of the African Association for Lexicography (AFRILEX) held at the University of Botswana, Gaborone, Botswana, 19–21 July 2010.

Ndinga-Koumba-Binza, H.S. 2010c. Propositions pour l'orthographe du civili. Hubert, J. and P.A. Mavoungou (Eds.). 2010. *Ecriture et standardisation des langues gabonaises*: 191-199. Stellenbosch: SUN Press.

Ndinga-Koumba-Binza, H.S. 2011. From Foreign to National: A Review of the Status of the French Language in Gabon. *Literator* 32(2): 135-150.

Ndinga-Koumba-Binza, H.S. and J.C. Roux. 2009. On writing Gabonese Languages. Batibo, H., R. Dikole, S. Lukusa and R.O.B. Nhlekisana (Eds.). 2009. *Language, Literature and Society. Proceedings of the First International Conference of the Department of African Languages and Literature, University of Botswana, 26–28 June 2008*: 83-106. Gaborone: Associated Printers.

Ndinga-Koumba-Binza, H.S. and G. Saphou-Bivigat. 2012. Dictionary Basis and Lemmatisation for an Encyclopedic Dictionary of Yilumbu. Ndinga-Koumba-Binza, H.S. and S.E. Bosch (Eds.). 2012. *Language Science and Language Technology in Africa: Festschrift for Justus C. Roux*: 237-250. Stellenbosch: SUN Press.

Nzang-Bie, Y. 2001.Vers une éducation multilingue au Gabon: première approche. *Revue Gabonaise des Sciences du Langage* 2: 17-29.

Nzang-Bie, Y. 2002. Le corpus lexicographique dans les langues à tradition orale. *Lexikos* 12: 211-226.

Ompoussa, V. 2008. *Les particularités lexicales dans le français scolaire au Gabon: le cas de la ville de Port-Gentil.* Ph.D. thesis. Grenoble: Université Stendhal de Grenoble.

Ompoussa, V. 2011. Les fondements psychologiques des mots des jeunes à Libreville. *Mbaandza. Revue d'Etude et d'Analyse Francophones* 1: 75-86.

Raponda-Walker, A. 1932. L'alphabet des idiomes gabonais. *Journal de la Société des Africanistes* 2(2): 139-146. Reprinted in Raponda-Walker, A. (Ed.). 1998. *Les langues du Gabon*: 7-15. Libreville: Editions Raponda-Walker.

Raponda-Walker, A. 1955. Idiomes gabonais: similitudes et divergences. *Bulletin l'Institut d'Etudes Centrafricaines* 10: 211-236. Reprinted in Raponda Walker, A. (Ed.). 1998. *Les langues du Gabon*: 90-136. Libreville: Editions Raponda Walker.

Raponda-Walker, A. 1963. Toponymie de l'Estuaire du Gabon et de ses environs. *Bulletin de l'Institut de Recherches Scientifiques au Congo* 2(2): 87-121.

Raponda-Walker, A. 2002. *Notes d'histoire du Gabon. Suivi de toponymie de l'Estuaire de Libreville et toponymie du Fernan-Vaz et Port-Gentil.* Libreville: Editions Raponda-Walker.

Rekanga, J.P. 2014. *Grammaire du myènè-nkomi*. Libreville: Editions Odette Maganga.

Soami, L.S. 2002. Principes de base de collecte et traitement des données lexicographiques en fonction des types de dictionnaires. Emejulu, J.D. (Ed.). 2002. *Eléments de lexicographie gabonaise*: 85-103. New York: Jimacs-Hillman Publishers.

Soami, L.S. 2010a. *Towards the Development and Application of Representative Lexicographic Corpora for the Gabonese Languages.* Ph.D. thesis. Stellenbosch: University of Stellenbosch.

Soami, L.S. 2010b. Propositions pour l'orthographe du yipunu. Hubert, J. and P.A. Mavoungou (Eds.). *Ecriture et standardisation des langues gabonaises*: 200-204. Stellenbosch: SUN Press.

Stark, M. 2011. *Bilingual Thematic Dictionaries*. Berlin/Boston: Walter de Gruyter GmbH.

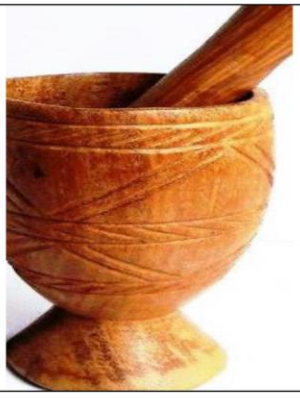
Svensén, B. 2009. The Handbook of Lexicography: The Theory and Practice of Dictionary-making. Cambridge, UK: Cambridge University Press.

Wiegand, H.E. 2002. Equivalence in Bilingual Lexicography: Criticism and Suggestions. *Lexikos* 12: 239-255.

Wiegand, H.E. and A. Kučera. 1981. Brockhaus-Wahrig: Deutsches Wörterbuch auf dem Prüfstand der praktischen Lexikologie. 1. Teil: 1. Band (A-BT); 2. Band (BU-FZ). *Kopenhagener Beiträge zur Germanistischen Linguistik* 18: 94-206.

Zgusta, L. 1987. Translational Equivalence in a Bilingual Dictionary: *Bahukosyam*. *Dictionaries* 9: 1-47.

Addendum: Images of the items presented in the sample of linguistic data (Table 1)

Fang	*Yilumbu*	*Yinzebi*	*Obamba*	*Civili*	
Akok	Ibenga	Ishighu	Kri	Cibenga	
Sas	Tsafu	Tsegha	Ntshighi	Nsafu	
Ngon	Teeri	Ndzaka	Ndjeghe	Ntêta	
Edong	Yufu	Mbungu	Odu	Cufu	

Jana Luther, Fred Pheiffer en Rufus H. Gouws (Redakteurs). *HAT. Handwoordeboek van die Afrikaanse Taal.* Sesde uitgawe. 2015, xviii + 1618 pp. ISBN 978-1-77025-700-9. Kaapstad: Pearson. Prys: R490.

Indien jy bevrediging geput het uit die gebruik van die vorige uitgawes van die *HAT*, sal jy nog meer bevrediging put uit die gebruik van die sesde uitgawe van die *HAT* wat in Julie 2015 verskyn het.

Die sesde uitgawe is die eerste werklik omvattend herbewerkte en uitgebreide uitgawe van die *HAT* en dit is gepas dat dit 50 jaar na die verskyning van die eerste uitgawe (1965) gepubliseer is. Resensies van *HAT 6* deur Daniel Hugo (2015) en Joan Hambidge (2015) het reeds in onderskeidelik *Rapport* en *Die Burger* verskyn.

Die sentrale teks van die nuwe *HAT* beslaan 1 614 bladsye, 204 bladsye meer as die vyfde uitgawe wat tien jaar gelede in 2005 verskyn het. Elke bladsy bevat gemiddeld 50 lemmas volgens die *Gids vir die Woordeboekgebruiker*, wat beteken dat die *HAT* 'n totaal van 80 700 lemmas bevat. Baie het in die afgelope tien jaar in Afrikaans gebeur en *HAT 6* slaag goed daarin om die ontwikkeling van die afgelope tien jaar te weerspieël. Meer as 3 200 nuwe lemmas is bygevoeg en meer as 5 300 nuwe betekenisonderskeidings. Ongeveer 3 000 ouer woorde en uitdrukkings, waarvan baie glad nie meer gebruik word nie, is uit die gedrukte woordeboek gelaat. Hierdie woorde en uitdrukkings word egter behou in 'n nuwe aanlyn moederwoordeboek. Hulle is dus nie verlore vir die nageslag nie.

Die bondige en leesbare *Gids vir die woordeboekgebruiker* is 'n nuttige hulpteks vir die leksikografiese leek. Terme soos *lemma (trefwoord), wisselvorme, homonieme, poëme, sitate* en *kollokasies* word in verstaanbare Afrikaans verhelder. 'n Interessante vernuwing is dat kollokasies afsonderlik verklaar word, bv.:

> **sa̲l·do** *s.nw.* [**~'s; ~'tjie**] *(<It., rekeningkunde)* verskil tussen debiet en krediet; oorskot: *die saldo in jou bankrekening nagaan* | **'n batige saldo** 'n winsbalans.

Die oordeelkundige aangawe van etimologiese inligting, dikwels slegs die herkomstaal van 'n lemma, verhoog die waarde van die artikels. Etimologiese inligting word aan die einde van 'n artikel verskaf (vet gedruk tussen blokhakies met die herkomswoord gekursiveer) en indien slegs die herkomstaal verstrek word, geskied dit gekursiveer aan die begin van die artikel:

> **a·o̲r·ta** s.nw. [**~s**] (Gr., anat.) … ; **sa̲l·do** s.nw. [**~'s; ~'tjie**] (<It., rekeningkunde)
> **an·ti·pa·tie** s.nw. [**~ë**] afkeer; teësin: 'n sterk antipatie teen(oor) leegleêrs voel. [**anti + Grieks pathos gevoel**]

Meer morfologiese en grammatikale inligting word in *HAT 6* verstrek as in vorige uitgawes, bv.:

nul *s.nw., telw.* [**~le; ~letjie**] teenoor **nul** s.nw. en telw. (-le)
on·be·skr<u>e</u>·we *b.nw.* [~] [voltooide deelwoord van **onbeskryf**] teenoor **on·be·skre'·we** b.nw.

'n Wins by die aanduiding van die aksent in die nuwe *HAT* is die onderstreping van die beklemtoonde lettergreep, bv. **b<u>o</u>t·tel** en selfs by eenlettergrepige lemmas soos **b<u>ou</u>d** en **b<u>ou</u>l** wat 'n diftong bevat (alhoewel *y*-gevalle oorgeslaan word). Dit laat geen twyfel oor watter lettergreep die hoofklem ontvang nie. Uitspraakleiding met behulp van die gewone alfabet (tussen ronde hakies, voorafgegaan deur die merker *uitspr.*) word slegs by vreemde woorde gegee: **b<u>ou</u>les** (*uitspr.* boel).

Afkortings word volgens die moderne konvensie alfabeties as lemmas opgeneem en nie in 'n aparte lys van afkortings nie. 'n Knap vernuwing is dat afkortings van trefwoorde ook (tussen blokhakies) by die voluit geskrewe lemma opgeneem word:

TB *afk.* tuberkulose.
tu·ber·ku·l<u>o</u>·se *s.nw.* [afk. **TB**] [geen mv.] *(patol.)* besmetlike bakteriële siekte gekenmerk deur die vorming van tuberkels/knoppies/knobbeltjies in …

Die taak van 'n standaard- sinchroniese woordeboek soos die *HAT* is om die standaardvariëteit van 'n taal te beskrywe soos dit ten tyde van die samestelling van die woordeboek gebruik word. In die voorwoord tot *HAT 6* word die volgende stelling gemaak: "As Standaardwoordeboek beskryf die *HAT* veral die verskuiwende standaardvariëteit van Afrikaans. Dit sluit 'n beperkte dog groeiende versameling in van leksikale items uit van die ander variëteite (wat volledige in die *WAT* bewerk word)." Pheiffer (2015) maak die stelling dat Kaaps die een variëteit van Afrikaans is wat in die nuwe uitgawe van die *HAT* veel groter aandag as voorheen gekry het en dat *HAT 6* waarskynlik een van die eerste publikasies is wat Kaaps op so 'n skaal opneem. Dit is prysenswaardig dat op hierdie wyse erkenning gegee word aan Kaaps, maar die vraag ontstaan of die Kaapse lemmas as deel van die "verskuiwende standaardvariëteit van Afrikaans" beskou kan word. Word *mang* (tronk) en *sharp* (reg/goed so) derhalwe as Standaardafrikaans beskou of is dit nog deel van die variëteit Kaaps? Woorde uit Kaaps soos *mang, skanghagha* en *sharp* word as (*sl.)* geëtiketteer. Ander Kaapse woorde soos *poenankies* word net soos woorde soos *skarminkel* (wat ongeëtiketteer in die *AWS* opgeneem is) as (*infml.*) geëtiketteer. Die opname van woorde uit Kaaps in *HAT 6* is 'n wins, maar die opname van die variëteite van Afrikaans in die *HAT* as 'n standaardwoordeboek sal voortaan duidelik verantwoord moet word.

Engelse woorde soos *stupid* en *cool* word as (*Eng., sl.*) gemerk, maar die volgende lemmas in *HAT 5* is weggelaat in *HAT 6*: *stunning* (*E., geselst.*) *suffer* (*E., geselst.*), *clue* (*E.*) en *casual* (*E.*). Dit beteken dat daar besluit is die woorde is

nie deel van die "verskuiwende standaardvariëteit van Afrikaans" nie. Dit is nie duidelik wat die maatstaf vir hierdie kategorie is nie.

Die verbreking van 'n leksikografiese konvensie is waarskynlik vir almal aanvaarbaar indien dit tot voordeel van die gebruiker is. Dit is egter moeilik om enige voordeel vir die woordeboekgebruiker daarin te sien dat definisies nie volgens konvensie met 'n hoofletter begin nie. Die definisie staan nie uit as die belangrikste inligtingstipe van 'n verklarende woordeboek nie en is veral moeilik uitkenbaar waar die definisie ook besonder bondig is. Ook poëme begin nie altyd met 'n hoofletter nie, met dieselfde gevolg:

> **I·be·ries** (*geogr.*) b.nw. **[~e]** van/uit Iberië: *die Iberiese ryk/skiereiland*.

Die hoë frekwensie van hakies en die solidus belemmer die leesbaarheid van definisies. Dit lyk soms asof daar nie gekies kon word tussen verskillende sinonieme of wisselvorme as definiërende woorde nie:

> **om·to·wer** *ww.* [**tower om, het omgetower**] Ook **om·to·wer** [**het ~**] (soos) deur/met 'n towerslag iets anders (laat) word:
> **kwyl** *s.nw.* [geen mv.] speeksel/spoeg wat uit die mond loop: *'n baba se kwyl afvee.*
> **ka·fe·ïen** *s.nw.* [geen mv.] Ook *(verouderd)* **ka·fe·ï·ne, kaf·fe·ïen, kaf·fe·ï·ne** wesen(t)like stimulerende bestanddeel van koffie, tee …
> **ras·se·kaart** *s.nw.* [gew. ekv.] *Ook* **ras·kaart** (SA) verwysing na velkleur. | **die ras(se)kaart uithaal/speel/gebruik** in die beoordeling/verduideliking/verdediging van 'n saak (onnodig/moedswillig) na iemand se velkleur/ras verwys.

Soos in vorige uitgawes (Botha 2001 en 2005) word die bewerking van lemmas nie altyd deur konsekwentheid gekenmerk nie. Dit blyk onder andere uit die hantering van sinonieme: by *leksikograaf* word *woordeboekmaker* as sinoniem vermeld, maar *woordeboekmaker* word slegs as 'n onverklaarde samestelling by *woordeboek* opgeneem. Die gebruiker kan dus nie van *woordeboekmaker* na *leksikograaf* gelei word nie.

By die nuut toegevoegde lemmas en by die beledigende leksikale items kom soms 'n ongelykheid van bewerking voor. Die definisies van *leeuloop* en *papsak* hel oor na die ensiklopediese en val in hierdie kategorie. Ongelyke behandeling kom ook by rassismes voor. Die *k*-woord word uitvoerig behandel met 'n definisie en 'n waarskuwing dat iemand gedagvaar kan word vir die gebruik van die woord. By die *h*-woord word egter geen definisie verstrek nie en die behandeling is baie bondiger met geen vermaning dat iemand gedagvaar kan word vir die gebruik van die woord nie. *Outa* en *meid* word ook ongelyk bewerk.

Ten spyte van tekortkominge staan dit tans, nog meer as in die verlede,

buite kyf dat die *HAT* die leier is in die geledere van Afrikaanse handwoordeboeke.

Verwysings

Botha, W. 2001**.** Moderner 'HAT' steeds gewortel in sterk tradisie. *Die Burger,* 24 Januarie 2001: 14.
Botha, W. 2005. Nuwe HAT modern, leier in klas. *Die Burger,* 31 Oktober 2005: 15.
Hambidge, J. 2015. Taal is nie nét taal, kepies en klanke nie. *By*, *Die Burger,* 22 Augustus 2015: 8.
Hugo, D. 2015. Daar is nog baie skop in dié 'antie'. *Rapport Weekliks*, 9 Augustus 2015: 13.
Pheiffer, F. 2015. *Sunday Argus,* 2 Augustus 2015: 17.

Willem Botha
Buro van die Woordeboek van die Afrikaanse Taal
Stellenbosch
Suid-Afrika
(wfb@sun.ac.za)

Michael Mann (Red.). *Digitale Lexikographie. Ein- und mehrsprachige elektronische Wörterbücher mit Deutsch: aktuelle Entwicklungen und Analysen*. Germanistische Linguistik 223–224. 2014, VI + 280 pp. ISBN 978-3-487-15191-5. Hildesheim/Zürich/New York: Georg Olms Verlag. Prys €54.

Hierdie uitgawe van die reeks fokus op digitale leksikografie en bevat ondersoeke na die ontwikkeling en analise van spesifiek een- en meertalige elektroniese Duitse woordeboeke. Dit bevat tien artikels plus 'n inleiding deur die redakteur.

Die bundel is geredigeer deur Michael Mann, 'n doktorale kandidaat aan die Universiteit van Erlangen, die skrywer van 'n blog oor leksikografie, woordeboeke en taal (*lexikographieblog*) en medewerker aan die *Wörterbücher zur Sprach- und Kommunikationswissenschaft* (*WSK*) deur Stephan J. Schierholz en Herbert Ernst Wiegand. Die redakteur het tien ander doktorale of postdoktorale navorsers genooi om van hulle navorsingsbevindinge in die veld van digitale leksikografie aan te bied, soos wat kortliks hier in Afrikaans opgesom sal word. Die doel van hierdie skrywe is dus nie om 'n kritiese resensie of bespreking van die bundel aan te bied nie, maar om dit ook vir lesers sonder 'n kennis van Duits moontlik te maak om kennis van die inhoud van die boek te neem.

Mann verduidelik in die inleiding tot die volume dat die term "digitale leksikografie", soos wat dit ook in die titel van die boek voorkom, meerduidig is en dat dit enersyds dui op die produksie- of vervaardigingsproses waar van 'n rekenaar gebruik gemaak word om tot 'n woordeboek te lei, en dat dit andersyds ook betrekking het op die produk van hierdie proses, die elektroniese woordeboek, met sy funksies en inhoud. In laasgenoemde geval kan verder onderskei word tussen die woordeboek as programmatuur wat op 'n gebruiker se rekenaar geïnstalleer word ("offline" of vanlyn) of wat deur 'n webblaaier van 'n internetbediener opgeroep en vertoon word ("online" of aanlyn), en die woordeboek as relatief onskeibare apparatuur-programmatuur-produk, soos 'n elektroniese sakwoordeboek of elektroniese sakvertaler. Daarby kan die produk 'n sogenaamde masjienwoordeboek wees wat nie as woordeboek deur *mense* gebruik word nie, maar waarvan die inhoud deur taalverwerkende sisteme met doelstellings soos taalherkenning, tekskorrigering, ens., opgeroep word.

Die boek behandel verskillende aspekte van die digitale leksikografiese proses en stel ook digitale woordeboeke as produkte voor en analiseer hulle. Al die woordeboeke wat spesifiek behandel word, behoort aan die kategorie aanlyn woordeboeke of internetwoordeboeke in die breë sin van die woord en sluit m.a.w. ook naslaanwerke met 'n ensiklopediese oriëntasie in. Tarp stel dit as volg: "The Internet seems to have consolidated itself as the dominant electronic platform for dictionaries" (Tarp 2012: 253). Digitale leksikografie vir menslike gebruikers is dus op die oomblik sinoniem met aanlyn leksikografie; naslaanwerke wat uitsluitlik op 'n CD of DVD, of as 'n aflaaibare produk aangebied word, is die uitsondering. Die sukses van aanlyn produkte is relatief maklik om te verduidelik: uit die oogpunt van die gebruiker kan aanlyn produkte direk by enige moderne rekenaar opgeroep word en is hulle meestal of

dikwels gratis beskikbaar; uit die oogpunt van die leksikograaf bied aanlyn produkte groot verspreidingsmoontlikhede, maklike bewerkbaarheid en ook verskeie tegniese moontlikhede, soos skakels na ander webblaaie, skakels na tekskorpora, multimedia, interaktiwiteit en maklike bywerking, gebruikeraanpasbaarheid en ook samewerkingsmoontlikhede.

Daar is uiteenlopende menings daaroor of aanlyn woordeboeke steeds oorwegend as digitale reproduksies van gedrukte woordeboeke geklassifiseer moet word en of hulle intussen direk vir die elektroniese medium gekonseptualiseer en bewerk word sodat die tegniese potensiaal beter benut word en hulle ontwerp duideliker van 'n gedrukte eksemplaar onderskei kan word. Enersyds is daar diegene, soos Tarp (2009) en Mann (2010) wat glo dat talle bekende en veelbenutte aanlyn woordeboeke ten tye van die ondersoek maar min van die toepaslike moontlikhede van die elektroniese medium benut; andersyds is daar diegene, soos Gouws, Schweickard en Wiegand (2013) wat aandui dat daar op die oomblik minder gedigitaliseerde internetwoordeboeke met 'n gedrukte woordeboek as basis is as wat daar internetwoordeboeke is wat gekonseptualiseer is vir aanlyn gebruik. Die uiteenlopende uitsprake dui volgens Mann daarop dat die digitale leksikografie in 'n fase van radikale verandering is en dat die "erfenis" van gedrukte woordeboeke stadig maar seker agtergelaat word (Mann 2014: 1-4).

Al die bydraes in die boek handel oor die onderwerp van digitale leksikografie met Duits as betrokke taal. Die bydraes bied 'n versameling insigte in verskillende benaderings tot die jongste onderwerpe, probleemvelde en uitdagings tydens verskillende projekfases van die onderskeidelik bespreekte aanlyn woordeboeke, en bied daarbenewens analises van woordeboeknavorsing vir digitale leksikografie. Al die skrywers is doktorale studente of postdoktorale genote wat werklike navorsingsbevindinge aanbied.

Die hoofstukke in die boek is in drie dele ingedeel: (i) *Planung und Vorbereitung von Wörterbüchern* [Ontwerp en voorbereiding van woordeboeke]; (ii) *Erstellung und (Weiter-)Entwicklung von digitalen Wörterbüchern* [Samestelling en (verdere) ontwikkeling van digitale woordeboeke]; en (iii) *Analyse solcher Wörterbücher, die bereits publiziert sind* [Analise en studie van bestaande digitale woordeboeke] (Mann 2014: 5).

Die onderskeie bydraes in elke afdeling is as volg:

(i) Ontwerp en voorbereiding van woordeboeke

Carolina Flinz, Università di Pisa: *Wörterbuchbenutzung: Ergebnisse einer Umfrage unter Studenten der Tourismuswirtschaft* [Woordeboekgebruik: bevindinge van 'n vraelys onder toerismestudente]. Die skrywer bespreek die bevindinge van 'n aanlyn vraelys onder 50 studente wat toerisme by die Universiteit van Pisa studeer. Die bevindinge word gebruik om 'n aanlyn Duits–Italiaans gespesialiseerde woordeboek oor toerisme te ontwikkel. Die skrywer vind dat die deelnemers algemene tweetalige woordeboeke bo eentalige en gespesialiseerde woordeboeke verkies, en dat hulle aanlyn woordeboeke sowel as gedrukte

woordeboeke gebruik. Slegs enkele buitetekste word as belangrik of nodig beskou, soos afkortingslyste. Deelnemers wil hê dat hulle woordeboek baie van hiperskakels gebruik moet maak en 'n groot verskeidenheid leksikografiese data moet aanbied, soos woordkeuse of sintagmatiese riglyne.

Christiane Zehrer, Universität Hildesheim: *Kriterien für die Terminus-Auswahl aus prozessbezogener Perspektive — ein Beitrag zur Meta-Lexikographie* [Kriteria vir die seleksie van terme vanuit 'n prosesverwante perspektief — 'n bydrae tot die metaleksikografie]. Die skrywer vergelyk die seleksie van terme as potensiële kandidate vir opname in 'n gespesialiseerde woordeboek deur studente met die outomatiese seleksie van sulke kandidate vir opname. Die studente het aansienlik minder termkandidate as die rekenaarlinguistiese algoritme geselekteer — 49 teenoor amper 250. Nie alle termkandidate wat met die hand uitgesoek is, is ook outomaties geselekteer nie en omgekeerd. Die menslike selekteerders het byvoorbeeld meerwoordige leksikale items as termkandidate geïdentifiseer maar die algoritme het nie. Die skrywer dui daarop dat verdere navorsing met terminoloë en vertalers as menslike termselekteerders nodig is om die voordele van gekombineerde seleksie van termkandidate vir 'n gespesialiseerde woordeboek verder te evalueer.

(ii) Samestelling en (verdere) ontwikkeling van digitale woordeboeke

Sandra Denzer en Franziska Horn, Technische Universität Darmstadt: *Die Arbeitsumgebung des Digitalen Familiennamenwörterbuch Deutschlands. Ein XML-basiertes Redaktionssystem* [Die werkomgewing van die digitale woordeboek van familiename in Duitsland. 'n XML-gebaseerde woordeboekskryfsisteem]. Die hantering van familiename in Duitsland in 'n omvattende aanlyn woordeboek wat 200 000 artikels beoog, vereis 'n woordeboekskryfprogram wat vir die spesifieke behoeftes van die projek aangepas is. Die projek beloop 'n tydperk van 24 jaar (2012–2036), drie projekvennote en 'n verskeidenheid data, wat bo en behalwe die familienaamartikels ook buitetekste met interessante feite, interaktiewe kaarte, 'n woordelys, ens. insluit — alles faktore wat die ontwikkeling van 'n spesifieke, aanpasbare en betroubare werkomgewing vereis. Dit behels 'n datamodel wat op TEI (*Text Encoding Initiative*) gebaseer is, 'n (aangepaste) XML-redigeerder, 'n databasis wat projekbestuur-nutsprogramme bevat en 'n inhoudsbestuursisteem vir aanlyn publisering. Die skrywers beskryf die besonderhede van alle komponente en beklemtoon die positiewe effek van die gedurige en intensiewe uitruiling en samewerking tussen leksikograwe en die ontwerpers, en bied daarmee aan lesers 'n kykie in die leksikografiese proses van 'n unieke Duitse woordeboekprojek.

Peter Meyer, Institut für Deutsche Sprache Mannheim: *Entlehnungsketten in einem Internetportal für Lehnwörterbücher. IT-Infrastruktur und computerlexikographischer Prozess in einem Projekt zu polnisch vermittelten Germanismen im Ostslavischen* [Ontleningsreekse in 'n internetportaal vir leenwoordwoordeboeke. IT-infrastruktur en die rekenaarleksikografiese proses in 'n projek oor Duitse leenwoorde in Oos-Slawies via Pools]. Die woordeboekportaal *Lehnwortportal*

Deutsch sal aangevul word met nuwe woordeboeke van Poolse leenwoorde uit Duits wat weer geleen word deur Russies, Belarussies en Oekraïens. Leksikografiese data oor hierdie leenwoorde word uit bestaande gedrukte woordeboeke van Oos-Slawies onttrek en sal later in die portaal se databasis geïntegreer word. Die skrywer beskryf die formele datamodel van hierdie databasis, en demonstreer hoe die model gebruik kan word om die verskillende "leenroetes" wat lei van spesifieke Duitse lekseme na Oos-Slawiese leenwoorde uit te beeld en te visualiseer, en om insig te bied in die uiters gekompliseerde leksikografiese proses van die projek (soos die feit dat een van die woordeboeke waaruit leenwoorde onttrek word nog nie voltooi is nie). Weens die spesifieke vereistes wat die projek stel, kon geen bestaande woordeboekskryfsisteem daarvoor aangepas word nie.

Mónica Mirazo Balsa, Universidade de Santiago de Compostela: *Überlegungen und Vorschläge zur Strukturierung und Darstellung lexikographischer Information in kontrastiven Onlinewörterbüchern* [Oorwegings en voorstelle vir die strukturering en ontwerp van leksikografiese data in kontrastiewe aanlyn woordeboeke]. Die artikel bespreek die moontlikhede om leksikografiese data volgens gebruikerbelangstelling en gebruikersituasie te modelleer. 'n Kontrastiewe Spaans-Duitse aanlyn woordeboek van naamwoordvalensie wat steeds saamgestel word, word as voorbeeld gebruik en stel 'n modulêre datamodel voor. Dit beteken dat 'n hele artikel of slegs sekere leksikografiese datatipes aan gebruikers getoon word op grond van hulle soektog. 'n Gebruiker wat na semantiese inligting op soek is, sal alle betekenisonderskeidings met sinonieme en ekwivalente saam met die semantiese kategorieë sien; 'n gebruiker wat sintagmatiese inligting soek, sal die moontlikhede van kombinering met die naamwoord in sintagmes sien. Die gebruikerkoppelvlak moet egter steeds ontwerp word en sal bepaal word deur resultate van 'n 2011 aanlyn vraelys oor die gebruik van Duitse woordeboeke en Duits–Spaanse woordeboeke deur Spaanse leerders van Duits.

Axel Herold, Berlin-Brandenburgische Akademie der Wissenschaften: *Das Wörterbuch als zentrale Komponente eines digitalen lexikografischen Systems* [Die woordeboek as die sentrale element in 'n digitale leksikografiese sisteem]. Die skrywer definieer digitale leksikografie as die elektroniese weergawes van linguistiese en veral leksikografiese data in een gebruikerkoppelvak. Sommige digitale leksikografiese sisteme kombineer verskillende woordeboeke, terwyl ander, soos die *Digitales Wörterbuch der Deutschen Sprache* (DWDS), ook leksikale bronne en/of korpusdata insluit. Die kernkomponent van die DWDS is 'n gedigitaliseerde eentalige woordeboek van Duits wat aangevul word deur outomaties gegenereerde inligting oor woordvorming. Dit is belangrik dat die status van elke inligtingstipe in hierdie digitale leksikografiese sisteem duidelik aangedui word sodat die gebruiker kan bepaal hoe nuut dit is al dan nie. Die DWDS-leksikograwe gebruik die korpusdata in die digitale leksikografiese sisteem wanneer hulle aan nuwe artikels werk of bestaande artikels herbewerk, sodat die twee sisteme begin om saam te vloei.

Antje Töpel, Institut für Deutsche Sprache Mannheim: *Die Beständigkeit*

von Wortartikeln im Onlinewörterbuch am Beispiel von elexiko [Die stabiliteit van artikels in 'n aanlyn woordeboek met *elexiko* as voorbeeld]. In hierdie ondersoek word die idee van artikels wat voortdurend in 'n aanlyn woordeboek bygewerk word, bekyk. Die skrywer gebruik as voorbeeld die eentalige Duitse internetwoordeboek *elexiko*, wat op 'n groot dinamiese korpus gebaseer is (*elexiko* is 'n afkorting vir *elektronisches, lexikalisch-lexikologisches und korpusbasiertes Informationssystem* — elektroniese leksikaal-leksikologiese en korpusgebaseerde inligtingstelsel). Die korpus groei en verander konstant en mag daarom vereis dat bestaande artikels en inskrywings reeds binne 'n paar jaar nadat hulle in die woordeboek opgeneem is, verander moet word, byvoorbeeld om nuwe kollokasies in te sluit. In vergelyking met die stand van sake selfs tien jaar gelede, bied die beskikbaarheid van meer korpusdata en ook meer gesofistikeerde korpusnutsprogramme *meer* data aan leksikograwe wat moontlik in artikels geïntegreer behoort te word. Taalverandering bring ook mee dat artikels bygewerk of herbewerk moet word. Daar is slegs 'n handjievol leksikograwe wat aan *elexiko* werk en die woordeboek kan as 'n woordeboek onder bewerking beskou word. Die projek vereis gevolglik 'n sekere mate van dissipline sodat nuwe artikels bygevoeg word en ook sodat ouer artikels terselfdertyd bygewerk word. Die skrywer beklemtoon dat die gebruikers se behoeftes in ag geneem moet word deur inligting oor die nuutste weergawe van elke artikel te verskaf.

(iii) Analise en studie van bestaande digitale woordeboeke

Julia Steube, Pädagogische Hochschule Luzern: *Wörterbücher und lexikografisches Wissen von Amateurlexikografen im Internet* [Woordeboeke en die leksikografiese kennis van amateurleksikograwe op die internet]. Die skrywer bespreek die bevindinge van haar empiriese studie van 75 gespesialiseerde aanlyn woordeboeke wat deur amateurleksikograwe saamgestel is. Hierdie woordeboeke word aan die hand van 'n aantal kriteria geëvalueer, soos toegangstruktuur, die gebruik van skakels en hipertekstualisering, multimediaelemente en ook web 2.0-elemente. Die meeste van die woordeboeke wat ondersoek word, is mediese of tegniese en rekenaarwoordeboeke, toon 'n eenvoudige ontwerp en vorm meestal deel van 'n woordeboekportaal. Die skrywer het ook amateurleksikograwe gekontak om uit te vind hoe en waar hulle van leksikografie geleer het en hoeveel hulle van leksikografie as 'n beroep weet. Sy vind dat amateurleksikograwe geneig is om aan hulle woordeboeke te begin werk sonder om inligting oor leksikografiese werk in die algemeen te soek, en dat hulle ook nie juis iets van leksikografie as 'n beroep weet nie.

Nathalie Mederake, Akademie der Wissenschaften zu Göttingen: *Artikel der Wikipedia aus lexikografischer und textlinguistischer Perspektive. Oder: Wenn das Leben dir Zitronen gibt …* [Wikipedia-artikels uit 'n leksikografiese en tekslinguistiese perspektief. Of: As die lewe jou suurlemoene gee …]. Die skrywer illustreer die ontwikkeling van Wikipedia-artikels aan die hand van die voorbeeld *Zitronenpresse* (suurlemoendrukker). Sy pas tekslinguistiese kriteria toe en fokus op koherensie in die proses van die konstruering van inhoud. Sy gebruik

'n sisteem van matriksraamwerke ("matrix frames") om die dinamika van veranderings in spesifieke inligtingstipes te beskryf, soos historiese inligting, inligting oor funksionaliteit, ens. Hierdie metode is slegs een manier om inhoudsontwikkeling in 'n verwysingsbron wat uit gemeenskaplike samewerking ontstaan, soos Wikipedia, te analiseer. Die metode sluit ongelukkig nie multimedia in die analise in nie, maar aan die hand van selfs 'n alledaagse voorbeeld soos *Zitronenpresse* kan gesien word tot watter hoë mate ensiklopediese vorm, die funksionaliteit van 'n wiki en die betrokkenheid van gebruikers in Wikipedia met mekaar verband hou, en die outeur stel gevolglik die verdere bespreking van hierdie moontlikhede vir digitale woordeboeke voor.

Susanne Dyka en Eva Scharf, Universität Bremen/Friedrich-Alexander-Universität Erlangen-Nürnberg: *Vergleichende Analyse von Artikelstrukturen deutsch-englisher Onlinewörterbücher* [Vergelykende analise van artikelstrukture in Duits–Engelse aanlyn woordeboeke]. Die outeurs vind uit gesprekke met studente dat gedrukte woordeboeke al hoe minder gebruik word en dat aanlyn woordeboeke verkies word, aangesien hulle maklik en dikwels ook gratis beskikbaar is.

Die outeurs bespreek, vergelyk en resenseer vyf gratis Duits–Engelse aanlyn woordeboeke (*LEO*, *dict.cc*, *Beolingus*, *Linguee* en *Pons*) met 'n analise van hulle toegangs- en mikrostrukture, veral wat betref kollokasies en aspekte van pragmatiek. Die sukses van die opsoek van frases word ook getoets, met betreklik swak resultate: een van die ooglopendste probleme is dat die soekenjin blykbaar nie die brontaal van die soekfrase kan identifiseer nie, sodat die soektog na 'n frase soos "ein Bad nehmen" (letterlik "'n bad neem") ook 'n lang lys van resultate vir die Engelse woord "bad" tot gevolg het. Frasesoektogte lewer oor die algemeen lang lyste met grotendeels irrelevante resultate waarin die toepaslike resultate ver onderaan die lys verskyn, wat dit veral vir niegevorderde gebruikers of taalleerders moeilik kan maak om die woordeboek suksesvol te gebruik.

Die outeurs noem dat geen van die woordeboeke wat hulle ondersoek het die potensiaal van die aanlyn medium ten volle benut nie, en dit lei tot probleme wat wel op die oog af anders is as die probleme in gedrukte woordeboeke, maar wat gedeeltelik vergelykbare resultate het en daarom ook swaar weeg — die woordeboeke bied byvoorbeeld wel leksikografiese data op 'n minder gekondenseerde wyse as gedrukte woordeboeke aan, maar steeds op 'n manier wat rigied of onaanpasbaar is. Ekwivalente word steeds in lang lyste aangebied, wat amper net so moeilik is om te verstaan as die gekondenseerde inskrywings in gedrukte woordeboeke. Die outeurs kom tot die gevolgtrekking dat die betrokke aanlyn woordeboeke hoofsaaklik in teksresepsie gebruik kan word, maar waarskynlik nie in teksproduksie nie.

Die bundel kan aanbeveel word aan enigiemand wat die vooruitgang in die maak van woordeboeke in die huidige tydperk van ontwikkeling en oorgang wil volg. Al die hoofstukke beskryf huidige ontwikkelinge in elektroniese leksikografie of lê aanlyn Duitse woordeboeke of ander verwysingsbronne en woordeboekportale voor. Die bydraes handel oor eentalige, tweetalige, alge-

mene of gespesialiseerde woordeboeke, en bevat verskeie voorbeelde oor die mate waarin leksikografiese werk verander het in die elektroniese omgewing. Verskillende vrae rondom die leksikografiese proses, woordeboekbeplanning, woordeboekvoorbereiding en woordeboeksamestelling word aan bod gestel.

Bronnelys

Gouws, R.H., W. Schweickard en H.E. Wiegand. 2013. Lexicography through the Ages: From the Early Beginnings to the Electronic age. Gouws, R.H., U. Heid, W. Schweickard en H.E. Wiegand (Reds.). 2013. *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography*: 1-24. Berlyn/Boston: Walter de Gruyter.

lexikographieblog. https://lexikographieblog.wordpress.com/.

Tarp, S. 2009. Beyond Lexicography: New Visions and Challenges in the Information Age. Bergenholtz, H., S. Nielsen en S. Tarp (Reds.). 2009. *Lexicography at a Crossroads. Dictionaries and Encyclopedias Today, Lexicographical Tools Tomorrow*: 17-32. Linguistic Insights. Studies in Language and Communication 90. Bern: Peter.

Tarp, S. 2012. Online Dictionaries: Today and Tomorrow. *Lexicographica* 28(1): 253-267.

Mann, M. 2010. Internet-Wörterbücher am Ende der "Nullerjahre": Der Stand der Dinge. Eine vergleichende Untersuchung beliebter Angebote hinsichtlich formaler Kriteria unter besonderer Berücksichtigung der Fachlexikographie. *Lexicographica* 26: 19-45.

Woordeboeke

Beolingus = Ihr Online-Wörterbuch. http://dict.tu-chemnitz.de/.

dict.cc = Online-Wörterbuch Englisch–Deutsch. http://www.dict.cc/.

Digitales Familiennamenwörterbuch Deutschlands. http://www.namenforschung.net/dfd/woerterbuch/liste/.

DWDS = Digitales Wörterbuch der deutschen Sprache. http://www.dwds.de/.

elexiko = Ein Onlinewörterbuch zur deutschen Gegenwartssprache. http://www.elexiko.de/.

Lehnwortportal Deutsch. http://lwp.ids-mannheim.de/.

LEO Dictionary. http://www.leo.org/ende/index_en.html.

Linguee = Wörterbuch Englisch-Deutsch. http://www.linguee.de/.

Pons Online-Wörterbuch. http://de.pons.com/.

Wikipedia. https://wikipedia.de/.

Wörterbucher zur Sprach- und Kommunikationswissenschaft (WSK) Online. http://www.degruyter.com/view/db/wsk en http://www.wsk.uni-erlangen.de/.

Hanelle Fourie Blair
Buro van die Woordeboek van die Afrikaanse Taal
Stellenbosch
Suid-Afrika
(hfourie@sun.ac.za)

Publikasieaankondigings / Publication Announcements

G.-M. de Schryver et al. (Editors). *Oxford Bilingual School Dictionary: IsiZulu and English*. Second edition. 2015, 672 pp. ISBN 978-0-19-907954-4 (Paperback). Cape Town: Oxford University Press Southern Africa. Price R129.95.

María José Domínguez Vázquez, Fabio Mollica and Martina Nied Curcio (Editors). *Zweisprachige Lexikographie zwischen Translation und Didaktik.* 2014, vi + 334 pp. ISBN 978-3-11-036973-1, e-ISBN 978-3-11-036663-1, ISSN 0175-9264. Lexicographica. Series Maior 147. Berlin/Boston: Walter de Gruyter. Price: €99.95.

Jana Luther, Fred Pheiffer en Rufus H. Gouws (Redakteurs). *HAT. Handwoordeboek van die Afrikaanse Taal.* Sesde uitgawe. 2015, xviii + 1618 pp. ISBN 978-1-77025-700-9. Kaapstad: Pearson. Prys: R490. (Resensie in hierdie nommer.)

Michael Mann (Red.). *Digitale Lexikographie. Ein- und mehrsprachige elektronische Wörterbücher mit Deutsch: aktuelle Entwicklungen und Analysen*. 2014, VI + 280 pp. ISBN 978-3-487-15191-5. Germanistische Linguistik 223–224. Hildesheim/Zürich/New York: Georg Olms Verlag. Prys €54. (Resensie in hierdie nommer.)

Oxford South African Pocket Dictionary. Fourth edition. 2015, 1128 pp. ISBN 978-0-19-904503-7 (Hardback). Cape Town: Oxford University Press Southern Africa. Price R264.95.

VivA Afrikaans. Webportaal vir Afrikaans met gratis inligting en betaalde intekening op o.a. Blokkiesraaiselwoordeboek, EWA, HAT Aanlyn, Longman Online, WAT. Navrae: navraag@viva-afrikaans.org Webblad: www.viva-afrikaans.org Prys: VivA-lede: gratis; VivA+-lede R 700 per jaar.

VOORSKRIFTE AAN SKRYWERS

(*Tree asseblief met ons in verbinding (lexikos@sun.ac.za) vir 'n uitvoeriger weergawe van hierdie instruksies of besoek ons webblad:* http://www.wat.co.za)

A. REDAKSIONELE BELEID

1. Aard en inhoud van artikels

Artikels kan handel oor die suiwer leksikografie of oor implikasies wat aanverwante terreine, bv. linguistiek, algemene taalwetenskap, terminologie, rekenaarwetenskap en bestuurskunde vir die leksikografie het.

Bydraes kan onder enigeen van die volgende rubrieke geklassifiseer word:

(1) **Artikels:** Grondige oorspronklike wetenskaplike navorsing wat gedoen en die resultate wat verkry is, of bestaande navorsingsresultate en ander feite wat op 'n oorspronklike wyse oorsigtelik, interpreterend, vergelykend of krities evaluerend aangebied word.
(2) **Resensieartikels:** Navorsingsartikels wat in die vorm van 'n kritiese resensie van een of meer gepubliseerde wetenskaplike bronne aangebied word.

Bydraes in kategorieë (1) en (2) word aan streng anonieme keuring deur onafhanklike akademiese vakgenote onderwerp ten einde die internasionale navorsingsgehalte daarvan te verseker.

(3) **Resensies:** 'n Ontleding en kritiese evaluering van gepubliseerde wetenskaplike bronne en produkte, soos boeke en rekenaarprogramme.
(4) **Projekte:** Besprekings van leksikografiese projekte.
(5) **Leksikonotas:** Enige artikel wat praktykgerigte inligting, voorstelle, probleme, vrae, kommentaar en oplossings betreffende die leksikografie bevat.
(6) **Leksikovaria:** Enigeen van 'n groot verskeidenheid artikels, aankondigings en nuusvrystellings van leksikografiese verenigings wat veral vir die praktiserende leksikograaf van waarde sal wees.
(7) **Ander:** Van tyd tot tyd kan ander rubrieke deur die redaksie ingevoeg word, soos Leksikoprogrammatuur, Leksiko-opname, Leksikobibliografie, Leksikonuus, Lexikofokus, Leksiko-eerbewys, Leksikohuldeblyk, Verslae van konferensies en werksessies.

Bydraes in kategorieë (3)-(7) moet almal aan die eise van akademiese geskrifte voldoen en word met die oog hierop deur die redaksie gekeur.

2. Wetenskaplike standaard en keuringsprosedure

Lexikos is deur die Departement van Hoër Onderwys van die Suid-Afrikaanse Regering as 'n gesubsidieerde, d.w.s. inkomstegenererende navorsingstydskrif goedgekeur. Dit verskyn ook op die *Institute of Science Index* (ISI).

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.

Manuskripte word vir publikasie oorweeg met dien verstande dat die redaksie die reg voorbehou om veranderinge aan te bring om die styl en aanbieding in ooreenstemming met die redaksionele beleid te bring. Outeurs moet toesien dat hulle bydraes taalkundig en stilisties geredigeer word voordat dit ingelewer word.

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 Direksie van die Woordeboek van die Afrikaanse Taal. Dit staan skrywers egter vry om hulle materiaal elders te gebruik mits *Lexikos* (AFRILEX-reeks) erken word as die oorspronklike publikasiebron.

5. Oorspronklikheid

Slegs oorspronklike werk sal vir opname oorweeg word. Skrywers dra die volle verantwoordelikheid vir die oorspronklikheid en feitelike inhoud van hulle publikasies. Indien van toepassing, moet besonderhede van die oorsprong van die artikel (byvoorbeeld 'n referaat by 'n kongres) verskaf word.

6. Gratis oordrukke en eksemplare

Skrywers ontvang vyf gratis oordrukke van elke artikel 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 Leksikonotas, Leksikovaria, ens. ontvang slegs vyf gratis oordrukke van hulle bydraes. In laasgenoemde 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 verkieslik in elektroniese formaat aan die volgende adres te stuur: lexikos@sun.ac.za, of 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

Manuskrip moet verkieslik in elektroniese formaat per e-pos of op rekenaarskyf voorgelê word in sagteware wat versoenbaar is met MS Word. Die lettersoort moet verkieslik 10-punt Palatino of Times Roman wees. Bydraes moet verkieslik nie **8 000 woorde** oorskry nie.

Elke artikel moet voorsien wees van 'n **opsomming** van ongeveer 200 woorde en ongeveer 10 **sleutelwoorde** in die taal waarin dit geskryf is, sowel as 'n opsomming en sleutelwoorde **in Engels**. Engelse artikels van Suid-Afrikaanse oorsprong moet 'n opsomming en sleutelwoorde in Afrikaans hê, terwyl Engelse artikels van buitelandse oorsprong 'n tweede opsomming en sleutelwoorde in enigeen van die aangeduide tale mag gee. As die outeur dit nie doen nie, sal die redaksie 'n Afrikaanse vertaling voorsien. Maak seker dat die opsomming in die tweede taal ook 'n **vertaling van die oorspronklike titel** bevat.

2. Grafika

Figure, soos tabelle, grafieke, diagramme en illustrasies, moet in 'n gepaste grootte wees dat dit versoen kan word met die bladspieël van *Lexikos*, naamlik 18 cm hoog by 12 cm breed. Die plasing van grafika binne die teks moet duidelik aangedui word. Indien skryftekens of grafika probleme oplewer, mag 'n uitdruk van die manuskrip of 'n e-pos in .pdf-formaat aangevra word.

3. Bibliografiese gegewens en verwysings binne die teks

Kyk na onlangse nommers van *Lexikos* vir meer inligting.

4. Aantekeninge/voetnote/eindnote

Aantekeninge moet deurlopend in die vorm van boskrifte genommer en aan die einde van die manuskrip onder die opskrif **Eindnote** gelys word.

INSTRUCTIONS TO AUTHORS

(*For a more detailed version of these instructions, please contact us (lexikos@sun.ac.za) or refer to our website:* http://www.wat.co.za)

A. EDITORIAL POLICY

1. Type and content of articles
Articles may treat pure lexicography or the implications that related fields such as linguistics, general linguistics, terminology, computer science and management have for lexicography.

Contributions may be classified in any one of the following categories:

(1) **Articles:** Fundamentally original scientific research done and the results obtained, or existing research results and other facts reflected in an original, synoptic, interpretative, comparative or critically evaluative manner.

(2) **Review articles:** Research articles presented in the form of a critical review of one or more published scientific sources.

Contributions in categories (1) and (2) are subjected to strict anonymous evaluation by independent academic peers in order to ensure the international research quality thereof.

(3) **Reviews:** An analysis and critical evaluation of published scientific sources and products, such as books and computer software.
(4) **Projects:** Discussions of lexicographical projects.
(5) **Lexiconotes:** Any article containing practice-oriented information, suggestions, problems, questions, commentary and solutions regarding lexicography.
(6) **Lexicovaria:** Any of a large variety of articles containing announcements and press releases by lexicographic societies which are of particular value to the practising lexicographer.
(7) **Other:** From time to time other categories may be inserted by the editors, such as Lexicosoftware, Lexicosurvey, Lexicobibliography, Lexiconews, Lexicofocus, Lexicohonour, Lexicotribute, Reports on conferences and workshops.

Contributions in categories (3)-(7) must all meet the requirements of academic writing and are evaluated by the editors with this in mind.

2. Academic standard and evaluation procedure
The Department of Higher Education of the South African Government has approved *Lexikos* as a subsidized, i.e. income-generating research journal. It is also included in the *Institute of Science Index* (ISI).

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.

Manuscripts are considered for publication on the understanding that the editors reserve the right to effect changes to the style and presentation in conformance with editorial policy. Authors are responsible for the linguistic and stylistic editing of their contributions prior their submission.

3. Language of contributions
Afrikaans, Dutch, English, French or German.

4. Copyright
Neither the Bureau of the WAT nor the African Association for Lexicography (AFRILEX) accepts any responsibility for claims which may arise from contributing authors' use of material from other sources.

Copyright of all material published in *Lexikos* will be vested in the Board of Directors of the Woordeboek van die Afrikaanse Taal. Authors are free, however, to use their material elsewhere provided that *Lexikos* (AFRILEX Series) is acknowledged as the original publication source.

5. Originality
Only original contributions will be considered for publication. Authors bear full responsibility for the originality and factual content of their contributions. If applicable, details about the origin of the article (e.g. paper read at a conference) should be supplied.

6. Free offprints and copies
Authors will receive five free offprints of each of their articles published, as well as one complimentary copy of the issue containing such article(s). Authors of purely evaluative reviews and of contributions to the categories Lexiconotes, Lexicovaria, etc., receive five free offprints of their contributions. In the case of the latter categories, the editors may, 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, preferably in electronic format, for publication in *Lexikos* to: lexikos@sun.ac.za, or

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. Format and length of articles
Manuscript should preferably be submitted in electronic format by email or on a disk, in software compatible with MS Word. The typeface used should preferably be 10-point Palatino or Times Roman. Contributions should not exceed **8 000 words**.

Each article must be accompanied by **abstracts** of approximately 200 words and approximately 10 **keywords** in the language in which it is written, as well as **in English**. English articles of South African origin should carry an abstract and keywords in Afrikaans, whilst English articles of foreign origin should carry a second abstract and keywords in any of the other languages mentioned. In cases where this is not done, the editors will provide an Afrikaans version. Ensure that the abstract in the second language also contains a **translation of the original title**.

2. Graphics
Figures such as tables, graphs, diagrams and illustrations should be in an appropriate size to be well accommodated within the page size of *Lexikos*, namely 18 cm high by 12 cm wide. The locations of figures within the text must be clearly indicated. If orthographic marks or graphics used in the text prove problematic, a printout of the manuscript or an email in .pdf format may be requested.

3. Bibliographical details and references in the text
Examine recent issues of *Lexikos* for details.

4. Notes/footnotes/endnotes
Notes must be numbered consecutively by superscript numbers and grouped together at the end of the manuscript under the heading **Endnotes**.